Random Dot Product Graphs as Dynamical Systems: Limitations and Opportunities

Connor Smith[1] and Giulio V. Dalla Riva[1]

[1] Baffelan.com

Abstract

Across ecology, history, economics, social behavior, cultural dynamics, and more, many phenomena can be described in terms of entities establishing and disestablishing interactions with each other. These scenarios are commonly represented mathematically as temporal networks, and the time evolution of these objects is studied as a time series with the goal of predicting future network states. Here, instead, we take a dynamical systems perspective: when the goal is to understand *why* networks evolve as they do, can we learn the differential equations that govern them? We investigate this question within the framework of Random Dot Product Graphs (RDPGs), where each network snapshot is generated from latent positions evolving under unknown dynamics. We identify three fundamental obstructions to recovering these dynamics: gauge freedom from the rotational ambiguity in latent positions, realizability constraints from the manifold structure of the probability matrix, and the trajectory recovery problem arising from spectral embedding artifacts. We develop a geometric framework based on principal fiber bundles that formalizes these obstructions and reveals their interplay. A sharp holonomy dichotomy emerges: polynomial dynamics have trivial holonomy, making gauge alignment purely statistical, while Laplacian dynamics generically produce non-trivial holonomy (proven to be full $\mathrm{SO}(2)$ for $d = 2$, conjectured $\mathrm{SO}(d)$ for $d \geq 3$), adding a topological obstruction. Cramér-Rao lower bounds show that the spectral gap controlling geometric difficulty simultaneously controls statistical difficulty, an inextricable duality. We establish an identifiability principle stating that symmetric dynamics cannot absorb skew-symmetric gauge contamination. Yet, we identify and present significant practical obstacles remain in finite samples. We frame the gap between identifiability and practical recovery as an open challenge and discuss directions for progress.

1 Introduction

Temporal networks are networks whose edges and nodes change over time. They appear throughout science, from ecological food webs that vary across space and time [1] to neural connectomes that rewire during development. RDPGs have proven particularly useful for such networks: Dalla Riva and Stouffer [2] first applied RDPG to ecological networks, showing that latent positions capture evolutionary signatures in food webs; subsequent work has used RDPG embeddings to predict trophic interactions [3], while the Athreya et al. survey [4] demonstrates applications to connectome

analysis and social network inference. A fundamental question is: **what differential equations govern the evolution of network structure?**

This paper investigates that question within the framework of Random Dot Product Graphs (RDPGs) [4], [5]. In an RDPG, each node i has a latent position $x_i \in \mathbb{R}^d$, and the probability of an edge between nodes i and j is $P_{ij} = x_i^\top x_j$. When these positions evolve according to some dynamics $\dot{X} = f(X)$, the network structure changes accordingly. Our goal is to recover f from observations.

This approach is appealing for several reasons. First, the latent space provides a continuous representation where dynamics can be modeled with standard tools from dynamical systems theory. Second, interpretable dynamics in X-space (e.g., attraction, repulsion, diffusion) translate to interpretable changes in network structure. Third, the RDPG framework has well-developed statistical theory connecting latent positions to spectral properties of observed adjacency matrices.

However, we identify **three fundamental obstructions** to learning dynamics from RDPG observations:

1. **Gauge freedom** (Section 3.1): The latent positions are determined only up to orthogonal transformation. Indeed, X and XQ produce identical networks for any $Q \in O(d)$. This means some dynamics are **invisible**: they change X without changing observable network structure.
2. **Realizability constraints** (Section 3.2): The probability matrix $P = XX^\top$ lives on a low-dimensional manifold. Not every symmetric perturbation $\dot{P}$ is achievable; those that would increase the rank of P are forbidden.
3. **Recovering trajectories from embeddings** (Section 4): Even when true dynamics are smooth, spectral embedding produces trajectories that jump erratically due to arbitrary gauge choices at each time step. This is not a statistical issue but a fundamental ambiguity in eigendecomposition.

We show that existing approaches fail to address these obstructions. Joint embedding methods like UASE [6] and Omnibus [7] assume generative models incompatible with ODE dynamics on latent positions. In fact, they model time-varying "activity" against fixed "identity," not genuinely evolving positions. Bayesian approaches with hierarchical smoothness priors [8] produce smooth trajectories but lack **dynamical consistency**: they interpolate well but don't enforce that velocity depends on state according to $\dot{X} = f(X)$. Pairwise Procrustes alignment is local and cannot ensure global consistency.

Our main contributions are theoretical. We develop a geometric framework based on principal fiber bundles that formalizes gauge freedom, realizability, and trajectory recovery as distinct but interrelated obstructions. We prove that random gauge artifacts introduce skew-symmetric contamination that cannot be absorbed by symmetric dynamics (Theorem 3), establishing an identifiability principle. We establish a holonomy dichotomy among horizontal dynamics families: polynomial dynamics have commuting generators and trivial holonomy, so gauge alignment is purely a statistical problem; Laplacian dynamics generically produce non-trivial holonomy (full $\mathrm{SO}(2)$ for $d = 2$; conjectured $\mathrm{SO}(d)$ for $d \geq 3$), adding a topological obstruction on top of the statistical one. We derive Cramér-Rao lower bounds showing that the spectral gap controlling curvature simultaneously controls Fisher information, so geometric and statistical difficulty are inextricable. However, we show that exploiting these structures in practice faces fundamental difficulties: the bias

induced by finite samples and the expressiveness of natural dynamics families make the constructive problem substantially harder than the identifiability theory suggests. We frame the gap between identifiability and practical recovery as an open problem and discuss directions for progress.

Paper outline. Section 2 reviews RDPG fundamentals. Section 3 develops the geometric framework: gauge freedom, realizability, and the fiber bundle perspective with connections, curvature, and holonomy. Section 4 addresses the practical problem of recovering trajectories from spectral embeddings, showing why existing methods fail and motivating the need for dynamics-aware approaches. Section 5 catalogs concrete dynamics families, analyzes their observable consequences through the Lyapunov equation, and establishes the holonomy dichotomy. Section 6 derives information-theoretic lower bounds revealing the statistical-geometric duality. Section 7 discusses the constructive problem: the identifiability principle, its algorithmic implications, the practical obstacles that remain, and a tractable special case using anchor nodes, illustrated with numerical experiments. Section 8 reflects on achievements and open problems.

Notation. $X \in \mathbb{R}^{n \times d}$ denotes the matrix of latent positions with rows $x_i^\top$. $\mathbb{R}_*^{n \times d} = \{X \in \mathbb{R}^{n \times d} : \text{rank}(X) = d\}$ denotes the set of full-rank $n \times d$ matrices. $P = XX^\top$ is the probability matrix. $O(d)$ is the orthogonal group; $\mathfrak{so}(d)$ is its Lie algebra of skew-symmetric matrices. $\hat{X}$ denotes an estimate (e.g., from spectral embedding). $\|\cdot\|_F$ is the Frobenius norm; $\|M\|_{2\to\infty} = \max_i \|e_i^\top M\|_2$ is the two-to-infinity norm, measuring the largest row norm of M. $O_{p(\cdot)}$ denotes stochastic order: $Y_n = O_{p(a_n)}$ means Y_n/a_n is bounded in probability. $\text{Sym}(d)$ denotes $d \times d$ symmetric matrices. Indices i, j refer to nodes; indices ι, γ refer to eigenvector directions when working in a spectral basis.

2 Random Dot Product Graphs

We begin with the RDPG framework for undirected networks. The extension to directed graphs appears in Section 15.

2.1 Latent positions and connection probabilities

In a Random Dot Product Graph, each node $i \in \{1, ..., n\}$ is associated with a latent position $x_i \in \mathbb{R}^d$. The probability of an edge between nodes i and j is:

$$P_{ij} = x_i^\top x_j \tag{1}$$

For this to be a valid probability, we need $P_{ij} \in [0, 1]$ for all pairs. A sufficient condition is that all positions lie in the positive orthant of the unit ball:

$$B_+^d = \{x \in \mathbb{R}^d : x_k \geq 0 \text{ for all } k, \quad \|x\| \leq 1\} \tag{2}$$

The non-negativity ensures $P_{ij} \geq 0$, and Cauchy-Schwarz gives $P_{ij} \leq \|x_i\| \, \|x_j\| \leq 1$.

Collecting positions into a matrix $X \in \mathbb{R}^{n \times d}$ with rows $x_i^\top$, the probability matrix is:

$$P = XX^\top \tag{3}$$

This is symmetric and positive semidefinite with rank at most d. Throughout this paper, we restrict attention to configurations where all edge probabilities lie in the open interval: $0 < P_{ij} < 1$ for all i, j. This excludes boundary cases (deterministic edges or guaranteed non-edges) that would create singularities in the Fisher information and complications in the differential geometry; see Section 3.3 for details.

Given latent positions, edges are drawn independently:

$$A_{ij} \sim \text{Bernoulli}\big(P_{ij}\big) \quad \text{for } i =< j \tag{4}$$

with $A_{ji} = A_{ij}$ (undirected) and $A_{ii} \sim \text{Bernoulli}(P_{ii})$, so $\mathbb{E}[A] = P$ exactly. In some applications it is common to disregard self-links by setting $A_{ii} = 0$, yielding $\mathbb{E}[A] = P - \text{diag}(P)$. This hollows the diagonal and introduces a systematic off-manifold bias that complicates the spectral theory (the perturbation $\text{diag}(P)$ has spectral norm at most 1, comparable to sampling noise for moderate n). We avoid this complication: throughout this paper we retain self-links, so that the data expectation lies exactly on the rank-d manifold $\{XX^\top\}$.

2.2 Adjacency spectral embedding

Given an observed adjacency matrix A, we estimate latent positions via **adjacency spectral embedding** (ASE) [4].

Since A is symmetric, it has an eigendecomposition $A = U\Lambda U^\top$ with real eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ and orthonormal eigenvectors. The rank-d ASE is:

$$\hat{X} = U_d \, |\Lambda_d|^{1/2} \tag{5}$$

where $U_d \in \mathbb{R}^{n \times d}$ contains the d leading eigenvectors (by eigenvalue magnitude) and $\Lambda_d = \text{diag}(\lambda_1, ..., \lambda_d)$. We take absolute values because A can have negative eigenvalues due to sampling noise, even though the true P is positive semidefinite. In the dense regime where $\lambda_d(P) = \Omega(\sqrt{n})$, Weyl's inequality guarantees that the top d eigenvalues of A are positive with high probability, so the absolute value has no effect asymptotically. We retain it for definiteness at finite n.

> *Remark.* For symmetric matrices, eigendecomposition and SVD are closely related: if $A = U\Lambda U^\top$ (eigen) and $A = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$ (SVD), then $\tilde{U} = \tilde{V} = U$ (up to signs) and $\sigma_i = |\lambda_i|$. We use eigendecomposition for ASE (the standard in RDPG literature) but SVD for Procrustes alignment (the standard for that problem).

Statistical properties. The statistical properties of RDPG are well studied, and here we remind some fundamental results from the literature. Under mild conditions, ASE is consistent: as $n \to \infty$, the rows of $\hat{X}$ converge to the true latent positions (up to orthogonal transformation) at rate $O(1/\sqrt{n})$ [4], [9]. Specifically, there exists $Q \in O(d)$ such that $\max_i \|\hat{x}_i - x_i Q\| = O_p\Big(\frac{1}{\sqrt{n}}\Big)$, and conditionally on each latent position, $\sqrt{n}(\hat{x}_i - Qx_i)$ converges to a Gaussian whose covariance depends on the edge variance profile [9], [10]. The perturbation expansion $\hat{X} = XW + (A - P)X(X^\top X)^{-1}W + R$, where $W \in O(d)$ and $\|R\|_{2\to\infty} = o\big(n^{-1/2}\big)$, shows that the leading error term is linear in $A - P$ and has conditional mean zero [11]. ASE is asymptotically unbiased but not fully efficient for individual latent positions; Xie and Xu's one-step procedure achieves what they term *local efficiency*: for each vertex, the estimator's asymptotic covariance matches that of the oracle MLE that knows all other latent positions, up to an orthogonal transformation [12]. For stochastic blockmodel parameters, the spectral estimator is itself asymptotically efficient [13]. The eigenvalues of A are also asymptotically normal about those of P, with an $O(1)$ bias from the Bernoulli variance [14]. This bias is negligible relative to the leading eigenvalue ($\lambda_1 \sim n$), but can be comparable to the spectral gap $\delta = \lambda_d$ when the gap is small, and is of the same order as the

eigenvalue fluctuations ($O(\sqrt{n})$); for dynamics estimation, what matters is the bias relative to the **changes** in eigenvalues across time steps, a point we revisit in Section 6.

The orthogonal matrix Q reflects the fundamental gauge ambiguity: eigenvectors are determined only up to sign, and rotations within repeated eigenspaces are arbitrary. This ambiguity is unavoidable and is the source of the trajectory estimation problem discussed in Section 4.

3 The Gauge Obstruction

We analyze the fundamental obstructions to learning dynamics from RDPG observations, characterizing what is and is not learnable from a theoretical perspective.

3.1 Gauge freedom and observability

The latent positions X are not uniquely determined by the probability matrix P. For any orthogonal matrix $Q \in O(d)$:

$$(XQ)(XQ)^\top = XQQ^\top X^\top = XX^\top = P \tag{6}$$

Thus X and XQ produce identical connection probabilities. This $O(d)$ symmetry is the **gauge freedom** of RDPG: the equivalence class $[X] = \{XQ : Q \in O(d)\}$ corresponds to a single observable network structure.

From a geometric point of view, a global rotation of all positions in latent space leaves all pairwise angles and magnitudes unchanged, hence all dot products are preserved.

This gauge freedom has profound implications for learning dynamics. If X and XQ are observationally indistinguishable, then any dynamics moving along the equivalence class, that is, rotating all positions by a common time-varying orthogonal transformation, produces **no observable change** in the network.

Consider dynamics $\dot{X} = f(X)$ on the latent positions. The induced dynamics on the probability matrix $P = XX^\top$ follow from the product rule:

$$\dot{P} = \dot{X}X^\top + X\dot{X}^\top = f(X)X^\top + Xf(X)^\top \tag{7}$$

> **Definition 1** (Observable and Invisible Dynamics). A vector field $f : \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$ produces **observable dynamics** if $\dot{P} \neq 0$. Otherwise, the dynamics are **invisible**: the latent positions change but the network structure remains static.

The definition calls for a natural question: which dynamics are invisible?

> **Theorem 1** (Characterization of Invisible Dynamics). For $X \in \mathbb{R}^{n\times d}_*$ (i.e., $\operatorname{rank}(X) = d$), a vector field f produces invisible dynamics ($\dot{P} = 0$) if and only if $f(X) = XA$ for some skew-symmetric matrix $A \in \mathfrak{so}(d)$.

Proof. ($\Leftarrow$) If $f(X) = XA$ with $A^\top = -A$, then:

$$\dot{P} = f(X)X^\top + Xf(X)^\top = XAX^\top + XA^\top X^\top = XAX^\top - XAX^\top = 0 \tag{8}$$

($\Rightarrow$) Suppose $\dot{P} = f(X)X^\top + Xf(X)^\top = 0$. Write $f(X) = XB + X_\perp C$ where $X_\perp$ spans the orthogonal complement of $\mathrm{col}(X)$. Then:

$$0 = XBX^\top + XB^\top X^\top + X_\perp CX^\top + XC^\top X_\perp^\top \tag{9}$$

The cross terms $X_\perp CX^\top$ and $XC^\top X_\perp^\top$ are in different subspaces from $XBX^\top$. For the equation to hold, we need $C = 0$ (so $f(X) = XB$) and $B + B^\top = 0$ (so $B \in \mathfrak{so}(d)$). □

We can readily interpret the result: invisible dynamics are exactly uniform rotations around the origin in latent space; all other dynamics (attraction, repulsion, non-uniform rotation, drift, …) produce observable changes in network structure.

This implies that the class of invisible dynamics is small (dimension $d(d-1)/2$, the dimension of $\mathfrak{so}(d)$), while observable dynamics span a much larger space.

3.2 Realizable dynamics

Beyond gauge freedom, RDPG dynamics face a geometric constraint: the probability matrix $P = XX^\top$ lives on a low-dimensional manifold, so most (symmetric) perturbations $\dot{P}$ are not achievable by dynamics of X.

Proposition 1 (Tangent Space Constraint). Let $V \in \mathbb{R}^{n\times d}$ be an orthonormal basis for $\mathrm{col}(P)$, and $V_\perp \in \mathbb{R}^{n\times(n-d)}$ span its orthogonal complement. Any realizable $\dot{P}$ (i.e., $\dot{P} = \dot{X}X^\top + X\dot{X}^\top$ for some $\dot{X}$) satisfies:

$$V_\perp^\top \dot{P} V_\perp = 0 \tag{10}$$

The realizable tangent space has dimension $nd - d(d-1)/2$.

Proof. Any realizable $\dot{P} = FX^\top + XF^\top$ for some $F \in \mathbb{R}^{n\times d}$. Since $\mathrm{col}(X) = \mathrm{col}(V)$, we have $X = VR$ for invertible R. Then:

$$V_\perp^\top \dot{P} V_\perp = V_\perp^\top FR^\top V^\top V_\perp + V_\perp^\top VRF^\top V_\perp = 0 \tag{11}$$

using $V^\top V_\perp = 0$. □

We can better understand the dynamics by considering a **block decomposition** of the generating matrices. Observe that any symmetric matrix M decomposes into blocks relative to $(V, V_\perp)$:

$$M = \underbrace{VAV^\top}_{\text{range-range}} + \underbrace{VBV_\perp^\top + V_\perp B^\top V^\top}_{\text{cross terms}} + \underbrace{V_\perp CV_\perp^\top}_{\text{null-null}} \tag{12}$$

For realizable $\dot{P}$: the A and B blocks can be arbitrary, but $C = 0$ always. The null-null block represents directions that would increase the rank of P. Movements in these directions are forbidden for **fixed** latent dimension d (but they might be achieavable for networks that start rank deficient and grow in dimension).

Corollary 1 (Dimension Count).

$$\dim(T_P\mathcal{B}) = nd - d(d-1)/2 \tag{13}$$

This equals the dimension of X-space (nd) minus the gauge freedom ($d(d-1)/2$).

The block decomposition suggest an immediate diagnostic for the model: if observed dynamics have nonzero structure in the null-null block $V_\perp^\top \dot{P} V_\perp$, this indicates either:

1. **Model misspecification**: The true dynamics don't preserve low-rank structure
2. **Dimensional emergence**: The latent dimension d is increasing, suggesting that new interaction dimensions are emerging

3.3 The fiber bundle perspective

The gauge freedom in RDPGs has a natural geometric structure that helps us understand what can and cannot be learned from network observations. The quotient geometry of $\mathbb{R}_*^{n\times d}/O(d)$ has been developed extensively in the optimization literature, particularly by Absil, Mahony, and Sepulchre [15], Journée, Bach, Absil, and Sepulchre [16], and Massart and Absil [17], [18]. We recall this geometry here, adapting it to the RDPG setting where the connections, curvature, and holonomy structures acquire concrete statistical meaning as obstructions to dynamics estimation.

3.3.1 Why fiber bundles?

A useful analogy is the relationship between the Earth's surface and its coordinates. Every point on the globe has a definite physical location, but many different coordinate systems (map projections) can describe it. Choosing coordinates is necessary for computation, but the choice is arbitrary and unreleated to the dynamic: it's a "gauge" choice. Different observers using different maps will describe the same journey differently, and must reconcile their descriptions when they meet.

The RDPG situation is similar. The observable quantity $P = XX^\top$ is like the physical location: it's what we can measure. The latent configuration X is like a choice of coordinates: useful for computation, but determined only up to an $O(d)$ rotation. Replacing X by XQ is like switching from one map projection to another.

A **fiber bundle** formalizes this structure. Think of the space of all possible observables P as the "ground level": this is the **base space** $\mathcal{B}$. Above each observable P, there is a "fiber" of equivalent latent configurations $\{XQ : Q \in O(d)\}$, all producing the same P. The full space of latent configurations X forms the **total space** $\mathcal{E}$, with a projection $\pi : X \mapsto XX^\top$ dropping from total space to base space.

The key concept is **lifting**: given a path $P(t)$ of evolving observables in the base space, we want to "lift" it to a path $X(t)$ in the total space. But there are infinitely many lifts, differing by time-varying gauge choices $Q(t)$. The connection and curvature of the bundle tell us which lifts are "natural" (no spurious rotation) and whether consistent lifting is possible at all. When it isn't, that is, when the bundle is curved, we get **holonomy**: gauge drift that accumulates even along the most careful trajectory.

We now make this precise.

3.3.2 The probability constraint

Before defining the bundle, we must address a constraint glossed over earlier: for P to represent connection probabilities, we need $P_{ij} \in [0,1]$ for all pairs i, j.

Not every $X \in \mathbb{R}^{n\times d}$ satisfies this. The constraint $\left(XX^\top\right)_{ij} = x_i^\top x_j \in [0,1]$ defines a semi-algebraic subset of $\mathbb{R}^{n\times d}$.

Definition 2 (Valid Latent Positions). The **valid configuration space** is:

$$\mathcal{E} = \left\{X \in \mathbb{R}^{n\times d} : \operatorname{rank}(X) = d, \quad 0 \le x_i^\top x_j \le 1 \text{ for all } i, j\right\} \tag{14}$$

The **valid probability space** is:

$$\mathcal{B} = \left\{P \in \mathbb{R}^{n\times n} : P = XX^\top \text{ for some } X \in \mathcal{E}\right\} \tag{15}$$

A sufficient condition for $X \in \mathcal{E}$ is that all rows lie in the **positive orthant of the unit ball**:

$$x_i \in B_+^d = \left\{x \in \mathbb{R}^d : x_k \ge 0 \text{ for all } k, \quad \|x\| \le 1\right\} \tag{16}$$

This ensures $x_i^\top x_j \ge 0$ (non-negative entries) and $x_i^\top x_j \le \|x_i\|\,\|x_j\| \le 1$ (Cauchy-Schwarz). However, $X \in B_+^d$ is sufficient but not necessary: many valid configurations exist with some entries outside the positive orthant, as long as all pairwise dot products remain in $[0,1]$.

Interior and boundary. The interior of $\mathcal{E}$ consists of configurations where all constraints are strict: $0 < x_i^\top x_j < 1$. The boundary includes configurations where some $P_{ij} = 0$ (nodes with orthogonal positions, hence no connection probability) or $P_{ij} = 1$ (nodes with aligned unit-length positions, hence certain connection).

The fiber bundle structure requires a smooth total space, which holds on the interior: there, $\mathcal{E}$ is an open subset of the smooth manifold $\left\{X \in \mathbb{R}^{n\times d} : \operatorname{rank}(X) = d\right\}$, and all the machinery of connections, curvature, and holonomy applies without caveat.

At the boundary, three issues arise. First, the feasible set $\left\{X : 0 \le x_i^\top x_j \le 1 \text{ for all } i, j\right\}$ is defined by polynomial inequalities, and has **corners** where multiple constraints are active simultaneously. Hence, it is a manifold with corners, not a smooth manifold. Second, at boundary points the tangent space is replaced by a **tangent cone**: not all directions are available, because some would push P_{ij} outside $[0,1]$. The horizontal-vertical decomposition of Proposition 2 may not respect feasibility: the horizontal component of a feasible velocity can point outside the constraint set. Third, if a trajectory reaches $P_{ij} = 0$ or $P_{ij} = 1$, the unconstrained ODE $\dot{X} = f(X)$ may attempt to leave the feasible region, requiring constraint handling (projection, reflection, or barrier terms as discussed in Section 7).

In practice, the boundary issue is rarely binding for temporal networks of interest: edge probabilities that are exactly 0 or 1 correspond to nodes that never interact, which are typically excluded from temporal network analysis (as they are "undetectable"), or always interact (which mean they don't have stochastic variability in the network structure). For the remainder of this paper, we work on a connected component of the interior of $\mathcal{E}$, where the geometry is smooth and the bundle structure is clean. This ensures that trajectories $X(t)$ do not cross rank-deficient strata where the fiber bundle structure degenerates.

3.3.3 Principal bundle structure

Definition 3 (RDPG Principal Bundle). On the interior of the valid spaces, $(\mathcal{E}, \mathcal{B}, \pi, O(d))$ forms a principal fiber bundle:

- **Total space** $\mathcal{E}$: valid latent configurations (interior)
- **Base space** $\mathcal{B}$: valid probability matrices (interior)
- **Projection** $\pi : \mathcal{E} \to \mathcal{B}$: sends $X \mapsto XX^\top$
- **Structure group** $O(d)$: acts on $\mathcal{E}$ by $X \cdot Q = XQ$

The **fiber** over P is $\pi^{-1}(P) = \{XQ : Q \in O(d)\} \cong O(d)$.

This is a **principal bundle** because $O(d)$ acts freely (no X is fixed by any non-identity Q) and transitively on each fiber (any two lifts of P differ by some Q).

3.3.4 Decomposing motion: vertical and horizontal

At each $X \in \mathcal{E}$, we can ask: which directions of motion change P, and which don't?

The **vertical subspace** $\mathcal{V}_X$ consists of directions along the fiber. Motion in these directions changes X but not $P = XX^\top$:

$$\mathcal{V}_X = \ker(d\pi_X) = \{X\Omega : \Omega \in \mathfrak{so}(d)\} \tag{17}$$

These are exactly the **invisible dynamics** from Theorem 1: infinitesimal rotations $\dot{X} = X\Omega$ with skew-symmetric Ω.

The **horizontal subspace** $\mathcal{H}_X$ consists of directions transverse to the fiber. Motion in these directions do change P (and X):

$$\mathcal{H}_X = \left\{Z \in T_X\mathcal{E} : X^\top Z \in \mathrm{Sym}(d)\right\} \tag{18}$$

Every tangent vector decomposes uniquely into vertical and horizontal parts:

$$T_X\mathcal{E} = \mathcal{V}_X \oplus \mathcal{H}_X \tag{19}$$

Proposition 2 (Horizontal Characterization). A tangent vector $\dot{X} \in T_X\mathcal{E}$ is horizontal if and only if $X^\top\dot{X}$ is symmetric.

Proof. Write $\dot{X} = X\Omega + H$ with $\Omega \in \mathfrak{so}(d)$ (vertical part) and $H \in \mathcal{H}_X$ (horizontal part). Then $X^\top\dot{X} = X^\top X\Omega + X^\top H$. The term $X^\top H$ is symmetric by definition of $\mathcal{H}_X$. The term $(X^\top X)\Omega$ is symmetric only if $\Omega = 0$, since the product of a positive definite symmetric matrix with a nonzero skew-symmetric matrix is never symmetric. Thus $X^\top\dot{X}$ is symmetric iff $\Omega = 0$ iff $\dot{X}$ is purely horizontal. □

The horizontal condition $X^\top\dot{X} \in \mathrm{Sym}(d)$ is a **gauge-fixing condition**: it picks out, among all ways to move in $\mathcal{E}$, the ones with no "wasted motion" along the fiber.

3.3.5 The connection 1-form: extracting gauge components

The choice of horizontal subspaces $\mathcal{H}_X$ varying smoothly over $\mathcal{E}$ is called an **Ehresmann connection**. It provides a consistent way to separate "observable" from "gauge" directions throughout the

bundle. Once we have a connection, we can define parallel transport (moving along the base while staying horizontal) and curvature (measuring how the connection twists).

The Ehresmann connection can be encoded by a **connection 1-form** ω that extracts the vertical (gauge) component of any motion:

Definition 4 (Connection 1-Form). The connection 1-form $\omega : T\mathcal{E} \to \mathfrak{so}(d)$ is defined by:

$$\omega_X(Z) = \Omega \quad \text{where} \quad (X^\top X)\Omega + \Omega(X^\top X) = X^\top Z - Z^\top X \tag{20}$$

That is, $\omega_{X(Z)}$ is the unique skew-symmetric matrix Ω solving this Lyapunov equation.

The horizontal space $\mathcal{H}_X = \{Z : X^\top Z \text{ symmetric}\}$ is not an arbitrary object: it is the **metric connection** induced by the Frobenius inner product. Equivalently, $\omega_X(Z)$ minimizes $\|Z - X\Omega\|_F^2$ over $\Omega \in \mathfrak{so}(d)$: it extracts the gauge component with minimal kinetic energy. This is the standard Riemannian connection on quotient manifolds [15], [17], ensuring that horizontal lifts are geodesics when projected appropriately.

To derive its expression, remember that any tangent vector decomposes as $Z = X\Omega + H$ with $\Omega \in \mathfrak{so}(d)$ and $X^\top H$ symmetric. Hence, computing $X^\top Z - Z^\top X$: since $\Omega^\top = -\Omega$, we have $Z^\top X = -\Omega(X^\top X) + H^\top X$. Thus $X^\top Z - Z^\top X = (X^\top X)\Omega + \Omega(X^\top X) + (X^\top H - H^\top X)$. Since $X^\top H$ is symmetric, $(X^\top H - H^\top X) = 0$, giving the Lyapunov equation. Uniqueness holds because the Lyapunov operator $\Omega \mapsto G\Omega + \Omega G$ is invertible when $G = X^\top X$ is positive definite: in the eigenbasis of G (with eigenvalues $\lambda_1, ..., \lambda_d$), this operator maps each entry $\Omega_{\iota\gamma}$ to $\big(\lambda_\iota + \lambda_\gamma\big)\Omega_{\iota\gamma}$ independently, so it acts as a diagonal operator on the vector space of skew-symmetric matrices, with all eigenvalues $\lambda_\iota + \lambda_\gamma > 0$. At rank-deficient points ($\text{rank}(X) < d$), G has a zero eigenvalue, the Lyapunov operator acquires a kernel, and the horizontal-vertical decomposition ceases to be unique. This is the algebraic counterpart of the geometric fact that the fiber bundle structure degenerates at the boundary of $\mathcal{E}$, and foreshadows the role of the spectral gap in Proposition 4.

The connection 1-form $\omega_X\big(\dot{X}\big)$ is the "instantaneous rotation rate" of the motion $\dot{X}$: if $\omega_X\big(\dot{X}\big) = 0$, the motion is purely horizontal (no gauge component); if $\omega_X\big(\dot{X}\big) = \Omega$, then $\dot{X}$ includes a rotation at rate Ω

The connection satisfies three key properties:
1. $\ker(\omega_X) = \mathcal{H}_X$: horizontal vectors have zero gauge component
2. $\omega_{X(X\Omega)} = \Omega$: vertical vectors are identified with their rotation rate
3. Equivariance: ω transforms appropriately under gauge changes

3.3.6 Horizontal lifts and parallel transport

Suppose we observe a trajectory $P(t)$ in the base space (observable dynamics). We want to "lift" this to a trajectory $X(t)$ in the total space. But there are infinitely many lifts, so which one should we choose?

The **horizontal lift** is the unique lift with no gauge drift:

$$\pi(X(t)) = P(t) \quad \text{and} \quad \omega_{X(t)}\big(\dot{X}(t)\big) = 0 \tag{21}$$

The second condition says the velocity $\dot{X}(t)$ has no vertical component at any time: the trajectory moves purely horizontally.

Existence and uniqueness: Given $P(t)$ and an initial lift $X(0)$ with $\pi(X(0)) = P(0)$, the horizontal lift exists and is unique [19]. The horizontal condition $\omega = 0$ defines an ODE on $\mathcal{E}$ whose solutions project to $P(t)$.

Why horizontal lifts matter: Among all trajectories $X(t)$ projecting to $P(t)$, the horizontal lift is the "gauge-canonical" one that tracks the observable dynamics without introducing spurious rotation. If we could compute horizontal lifts from data, we would solve the gauge problem. The difficulty is that in practice we observe noisy adjacency matrices, not $P(t)$ directly (see Section 5.2 for a detailed treatment of this estimation problem).

Computing horizontal lifts. The horizontal lift formula connects directly to the block decomposition of Proposition 1. Let $P = V\Lambda V^\top$ where $V \in \mathbb{R}^{n\times d}$ has orthonormal columns and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_d)$ with $\lambda_\iota > 0$. A natural lift is $X = V\Lambda^{1/2}$, giving $X^\top X = \Lambda$.

Given realizable $\dot{P}$, the block decomposition yields:

$$\dot{P} = VAV^\top + VBV_\perp^\top + V_\perp B^\top V^\top \tag{22}$$

where $A = V^\top \dot{P} V$ (symmetric, $d \times d$) and $B = V^\top \dot{P} V_\perp$ (the cross term).

We seek the horizontal lift $\dot{X} = VS + V_\perp T$ satisfying:
1. **Projection**: $\dot{P} = \dot{X}X^\top + X\dot{X}^\top$
2. **Horizontality**: $X^\top \dot{X} = \Lambda^{1/2} S$ is symmetric

From the projection condition, matching blocks:
- Range-range: $A = S\Lambda^{1/2} + \Lambda^{1/2} S^\top$
- Cross terms: $B = T^\top \Lambda^{1/2}$, giving $T = B^\top \Lambda^{-1/2}$

The horizontality condition requires $S = \Lambda^{-1/2}\Sigma$ for some symmetric Σ. Substituting into the range-range equation and setting $\tilde{\Sigma} = \Lambda^{-1/2}\Sigma\Lambda^{-1/2}$:

$$\Lambda\tilde{\Sigma} + \tilde{\Sigma}\Lambda = \Lambda^{-1/2} A \Lambda^{-1/2} \tag{23}$$

This is the same Lyapunov structure as in the connection 1-form above. And not coincidentally: both of them involve separating gauge from observable components. The following formula is essentially the horizontal projection of [15] and [17], specialized to the spectral decomposition of P.

Proposition 3 (Horizontal Lift Formula). Given $P = V\Lambda V^\top$ and realizable $\dot{P}$ with blocks $A = V^\top \dot{P} V$ and $B = V^\top \dot{P} V_\perp$, the horizontal lift at $X = V\Lambda^{1/2}$ is:

$$\dot{X} = V\tilde{\Sigma}\Lambda^{1/2} + V_\perp B^\top \Lambda^{-1/2} \tag{24}$$

where $\tilde{\Sigma}$ is the symmetric solution to Equation 23, given elementwise by:

$$\tilde{\Sigma}_{\iota\gamma} = \frac{\left(\Lambda^{-1/2} A \Lambda^{-1/2}\right)_{\iota\gamma}}{\lambda_\iota + \lambda_\gamma} \tag{25}$$

Proof. The derivation above (Equation 23 and the preceding block algebra) constructs $\dot{X}$ satisfying both the projection condition $\dot{P} = \dot{X}X^\top + X\dot{X}^\top$ and the horizontality condition $X^\top\dot{X}$ symmetric. The Lyapunov equation has a unique symmetric solution because $\lambda_\iota + \lambda_\gamma > 0$ for all ι, γ. □

The formula reveals two contributions to horizontal motion:

- The first term $V\tilde{\Sigma}\Lambda^{1/2}$ handles motion within the current column space of X, determined by the range-range block A via a Lyapunov equation
- The second term $V_\perp B^\top \Lambda^{-1/2}$ handles motion expanding into new directions, determined directly by the cross block B

In practice, computing horizontal lifts from discrete noisy observations requires interpolation and ODE integration. We discuss in Section 7 how discrete Procrustes alignment approximates horizontal transport.

3.3.7 Curvature and holonomy

A subtle issue: even horizontal lifts can accumulate gauge drift over closed loops.

The **curvature** of the connection measures how horizontal directions fail to commute:

Definition 5 (Curvature 2-Form). For horizontal vector fields H_1, H_2 on $\mathcal{E}$:

$$\Omega(H_1, H_2) = -\omega([H_1, H_2]) \tag{26}$$

where $[\cdot, \cdot]$ is the Lie bracket.

When curvature is nonzero, traveling horizontally in direction H_1 then H_2 doesn't end up at the same point as H_2 then H_1. So, there's a vertical (gauge) discrepancy.

This has a striking consequence for closed loops:

Definition 6 (Holonomy). Let $\gamma : [0, 1] \to \mathcal{B}$ be a closed curve with $\gamma(0) = \gamma(1) = P$. The horizontal lift starting at X ends at XQ_γ for some $Q_\gamma \in O(d)$. The **holonomy** of γ is this accumulated rotation Q_γ.

A classic result (see for example [19]) provides a precise characterization of the holonomy obstruction.

Theorem 2 (Holonomy Obstruction). If the bundle has nontrivial curvature, there exist closed paths in $\mathcal{B}$ such that no globally consistent gauge exists: any lift satisfies $X(1) = X(0)Q$ for some $Q \neq I$.

The results implies that, if the true dynamics $P(t)$ trace a closed loop (periodic network behavior), the underlying $X(t)$ may not close on itself. Instead, it returns rotated by the holonomy. This is a fundamental obstruction: even perfect local alignment accumulates global gauge drift over cycles.

It is interesting to notice that the holonomy obstructions provides a **connection to spectral properties** of the (observed) graphs: in fact, the curvature of $\mathcal{B}$ depends on the eigenvalues of $P = XX^\top$. Since the nonzero eigenvalues of P are exactly those of $X^\top X$ (the Gram matrix of the latent positions), small λ_d arises when, for example:

- **Latent positions cluster in a lower-dimensional subspace**: if nodes' positions are nearly coplanar in $\mathbb{R}^d$, the columns of X become nearly linearly dependent.
- **Stochastic block models with weak community structure**: for SBM with $X = ZB^{1/2}$ (membership Z, block matrix B), the spectral gap of P reflects that of B. When communities are hard to distinguish, λ_d is small.
- **Sparse networks**: if all entries of X scale as $\rho_n \to 0$ (sparsity parameter), then $\lambda_d \sim \rho_n^2 \to 0$.

This matters doubly: not only do geometric obstructions worsen as $\lambda_d \to 0$ (the injectivity radius vanishes as $\sqrt{\lambda_d}$, and curvature blows up when λ_{d-1} is also small), but ASE convergence rates also deteriorate (they scale as $O\big(\sqrt{\log n/\lambda_d}\big)$ [11]). Configurations with small spectral gap are thus problematic both statistically (harder to estimate) and geometrically (harder to track gauges).

Proposition 4 (Curvature and Spectral Gap). The quotient manifold $\mathcal{B} = \mathbb{R}_*^{n\times d}/O(d)$ with the Procrustes metric has **sectional curvature** K (a scalar measuring the Gaussian curvature of 2-dimensional sections, distinct from the Lie algebra-valued curvature 2-form Ω above) given by O'Neill's formula for Riemannian submersions [20]: for orthonormal horizontal vectors $\overline{\xi}, \overline{\eta} \in \mathcal{H}_X$,

$$K(\xi,\eta) = 3\,\|\mathcal{A}_{\overline{\xi}}\overline{\eta}\|^2 = \frac{3}{4}\,\|\left[\overline{\xi},\overline{\eta}\right]^{\mathcal{V}}\|^2 \tag{27}$$

where $\mathcal{A}_{\overline{\xi}}\overline{\eta} = \frac{1}{2}\left[\overline{\xi},\overline{\eta}\right]^{\mathcal{V}}$ is the O'Neill A-tensor and $\left[\overline{\xi},\overline{\eta}\right]^{\mathcal{V}}$ denotes the vertical component of the Lie bracket. Since the total space $\mathbb{R}_*^{n\times d}$ is flat, all base curvature arises from the A-tensor.

Massart, Hendrickx, and Absil [18] compute the sectional curvature explicitly for this quotient:

- If $d = 1$, the sectional curvature is identically zero (the base space is flat).
- If $d \geq 2$, the minimum sectional curvature is zero, achieved when the horizontal fields commute.
- The maximum sectional curvature at $[X]$ diverges when the two smallest eigenvalues λ_{d-1}, λ_d of $P = XX^\top$ approach zero simultaneously.

Two related but distinct geometric obstructions emerge from the spectral gap. The **curvature** depends on the A-tensor, which involves the connection coefficients $1/(\lambda_\iota + \lambda_\gamma)$ for $\iota \neq \gamma$. Since the connection is skew-symmetric, the worst case is $1/(\lambda_{d-1} + \lambda_d)$: curvature blows up only when **both** of the two smallest eigenvalues approach zero. If $\lambda_d \to 0$ but λ_{d-1} remains bounded, the curvature stays finite (the manifold looks locally like a cylinder in the collapsing direction). By contrast, the **injectivity radius** at $[X]$ equals $\sqrt{\lambda_d}$ [17] and vanishes whenever $\lambda_d \to 0$ alone: beyond this radius, geodesics cease to be length-minimizing and multiple paths can connect the same points.

For dynamics estimation, both phenomena matter. Large curvature means that even short paths produce substantial holonomy, making local Procrustes alignment inconsistent over moderate distances. Small injectivity radius means that interpolation between time steps is non-unique, even without curvature. The $d = 1$ case is degenerate in a useful way: $O(1) = \{\pm 1\}$ is discrete, so the only gauge ambiguity is a global sign flip, and the curvature vanishes because there is no continuous rotation to accumulate.

3.3.8 Riemannian structure

The quotient $\mathcal{B} \cong \mathcal{E}/O(d)$ inherits a natural metric:

Definition 7 (Quotient Metric). The Riemannian distance between $[X], [Y] \in \mathcal{B}$ is the **Procrustes distance**:

$$d_{\mathcal{B}}([X], [Y]) = \min_{Q \in O(d)} \|X - YQ\|_F \tag{28}$$

This connects the geometry to computation: Procrustes alignment computes geodesic distance on the base space [17]. The global injectivity radius of $\mathcal{B}$ is zero (since λ_d can be arbitrarily small), reflecting the nontrivial topology of the quotient.

4 Recovering trajectories from spectral embeddings

The obstructions above concern what dynamics are **theoretically** possible. We now turn to the **practical** problem: even when dynamics are observable and realizable, recovering trajectories from data is hard because spectral embedding introduces arbitrary gauge transformations at each time step.

4.1 ASE introduces arbitrary gauge transformations

At each time t, ASE computes the eigendecomposition of $A^{(t)}$. The eigenvectors are determined only up to sign (and rotation within repeated eigenspaces). This means:

$$\hat{X}^{(t)} = X^{(t)} R^{(t)} + E^{(t)} \tag{29}$$

where $E^{(t)}$ is statistical noise and $R^{(t)} \in O(d)$ is the gauge factor. Although $R^{(t)}$ is deterministic (fixed by the eigensolver's conventions, e.g. sign of the first nonzero entry), it is **discontinuous** in t: standard eigensolvers enforce conventions that jump whenever an entry crosses zero or eigenvalues change ordering. These jumps are uncorrelated with the dynamics, so $R^{(t)}$ behaves effectively like an arbitrary selection from the fiber $O(d)$ at each time step.

The key point: Even if the true positions $X^{(t)}$ evolve smoothly, the estimates $\hat{X}^{(t)}$ jump erratically because the $R^{(t)}$ are unrelated across time. In principle, a smooth choice of eigenvectors exists when eigenvalues are simple and vary analytically, but standard numerical eigensolvers do not enforce this choice.

4.2 Finite differences fail

Consider estimating velocity via finite differences:

$$\hat{\dot{X}}^{(t)} = \frac{\hat{X}^{(t+\delta t)} - \hat{X}^{(t)}}{\delta t} \tag{30}$$

Substituting the gauge-contaminated estimates:

$$\hat{\dot{X}}^{(t)} = \frac{X^{(t+\delta t)} R^{(t+\delta t)} - X^{(t)} R^{(t)}}{\delta t} + O(E) \tag{31}$$

Even ignoring noise E, this is dominated by the gauge jump $R^{(t+\delta t)} - R^{(t)}$, which is $O(1)$ regardless of δt. As the "velocity" diverges as $\delta t \to 0$, we're measuring gauge jumps, not dynamics.

4.3 Pairwise Procrustes is insufficient

One might try aligning consecutive embeddings via Procrustes:

$$Q^{(t)} = \arg\min_{Q \in O(d)} \|\hat{X}^{(t+1)} - \hat{X}^{(t)} Q\|_F \tag{32}$$

This finds the best rotation to match adjacent frames. However:

- The solution is **local**: it doesn't ensure **global** consistency across the full trajectory
- Errors accumulate: small misalignments at each step compound
- There's no guarantee the aligned trajectory corresponds to **any** consistent dynamics

4.4 Pairwise alignment and error accumulation

One might try aligning consecutive embeddings via Procrustes:

$$Q^{(t)} = \arg\min_{Q \in O(d)} \|\hat{X}^{(t+1)} - \hat{X}^{(t)} Q\|_F \tag{33}$$

This has a closed-form solution via SVD and finds the best rotation to match adjacent frames. However, sequential pairwise alignment suffers from fundamental limitations:

1. **Local, not global:** Each alignment minimizes error between adjacent frames but doesn't ensure consistency across the full trajectory.
2. **Error accumulation:** The ASE gauge $R^{(t)}$ is an arbitrary $O(1)$ rotation at each time step, completely uncorrelated across time. Pairwise Procrustes estimates the relative gauge $R^{((t))^{-1}} R^{(t+1)}$ between consecutive frames, with error determined by the ASE noise $E^{(t)}$. Since these noise-induced errors are approximately independent across steps, the accumulated rotation error after sequential alignment of T frames scales as $O\left(\sqrt{T}\sigma\right)$ where σ is the per-step noise level.
3. **Holonomy is a separate obstruction:** The $O\left(\sqrt{T}\sigma\right)$ diffusive error characterizes the **statistical** difficulty of alignment. The holonomy obstruction (Section 3.3) is **topological** and exists even with perfect data ($\sigma = 0$): for dynamics with nontrivial holonomy, no globally consistent gauge exists over closed or nearly-closed loops, regardless of how accurately each pairwise alignment is performed. The two obstructions interact but should not be conflated: the statistical problem makes alignment noisy; the topological problem makes exact global alignment impossible in principle.
4. **No dynamical constraint:** There's no guarantee the aligned trajectory corresponds to **any** consistent dynamics because the alignment is purely geometric.

Noise in spectral embeddings: RDPG spectral embeddings have estimation error $\|\hat{X} - XQ\|_F = O_{p(\sqrt{n})}$ for some $Q \in O(d)$, giving per-node error $O_{p(1/\sqrt{n})}$ [4], [11]. For pairwise Procrustes between consecutive frames, this translates to rotation estimation error that depends on both n and the separation between frames.

When the true trajectory moves slowly (small $\|X^{(t+1)} - X^{(t)}\|$), the signal-to-noise ratio for alignment degrades. In these cases, we're trying to detect small true rotations against a background of estimation noise.

4.5 Why existing joint embedding methods don't help

Joint embedding methods like UASE [6] embed all time points simultaneously, producing gauge-consistent estimates. However, they assume a different generative model.

UASE model (Multilayer RDPG):

$$P_{ij}^{(t)} = x_i \Lambda^{(t)} y_j^{(t)\top} \tag{34}$$

where x_i is a **fixed** "identity" and $y_j^{(t)}$ is a time-varying "activity."

Our model (temporal RDPG with ODE dynamics):

$$P_{ij}^{(t)} = x_i^{(t)\top} x_j^{(t)} \tag{35}$$

where the **same** position $x_i^{(t)}$ evolves according to $\dot{x}_i = f(x_i, ...)$.

The mismatch is fundamental and can be stated precisely in terms of the rank of the concatenated data. UASE assumes that the concatenated matrix $\left[A^{(1)}; A^{(2)}; ...; A^{(T)}\right]$ has a **common invariant column space** of dimension d: all time slices share the same left-singular vectors. Under ODE dynamics on latent positions, the column space of $P^{(t)} = X^{(t)} X^{((t))^\top}$ generically evolves (especially for Laplacian dynamics, where eigenvectors rotate), so the rank of the concatenated probability matrix grows with T. UASE fits a static average subspace to what is inherently an evolving one, compressing the dynamical signal into a deformation of "activity scores" within a fixed basis. Applying UASE to data from our model distorts the recovered trajectory, attributing genuine positional evolution to time-varying activities against fixed identities.

Similar issues affect Omnibus embedding [7] and COSIE [21]: all assume a shared subspace structure that ODE dynamics do not preserve.

4.6 Why Bayesian smoothing approaches are insufficient

Recent work [8] proposes Bayesian inference for dynamic RDPGs using hierarchical priors on latent positions. By placing priors on successive differences (or equivalently, using Gaussian processes with smooth kernels), the resulting trajectories are smooth: velocities and accelerations are well-defined and continuous.

However, smoothness is necessary but not sufficient for dynamical consistency. An ODE $\dot{X} = f(X)$ constrains the velocity to be a **function of the current state**; a smoothness prior constrains velocities to vary continuously, but does not require that $\dot{X}(t)$ is determined by $X(t)$.

> **Definition 8** (Dynamical Consistency). A trajectory $X(t)$ is **dynamically consistent** with respect to a function class $\mathcal{F}$ if there exists $f \in \mathcal{F}$ such that $\dot{X}(t) = f(X(t))$ for all t.

Hierarchical Bayesian priors ensure that $X(t)$, $\dot{X}(t)$, and higher derivatives are all smooth, but they do not enforce the state-dependence constraint $\dot{X}(t) = f(X(t))$ for any f. This distinction is the gap between kinematic regularity and dynamical structure.

Prior support and the ODE solution manifold. For a given initial condition $X(0) = X_0$ and dynamics $\dot{X} = f(X)$, the solution $X(t)$ traces a **unique** curve: there is no uncertainty in the trajectory given (f, X_0). The space of all ODE solutions (over all $f \in \mathcal{F}$ and X_0) forms a

finite-dimensional manifold in the infinite-dimensional space of smooth paths. A hierarchical smoothness prior assigns positive probability to all smooth paths, a much larger set than just those satisfying some ODE. So, the ODE solution manifold has measure zero under such priors.

Proposition 5 (Measure Zero). Let μ be any Gaussian measure on smooth paths $C^{k([0,T],\mathbb{R}^{n\times d})}$ with full support. Let $\mathcal{M}_{\mathcal{F}} = \left\{X(\cdot) : \dot{X} = f(X) \text{ for some } f \in \mathcal{F}\right\}$ be the manifold of ODE solutions. Then $\mu(\mathcal{M}_{\mathcal{F}}) = 0$.

Proof. See Section 12. □

The substantive implication concerns **posterior behavior**: Gaussian process posteriors with standard kernels (RBF, Matérn) cannot concentrate on $\mathcal{M}_{\mathcal{F}}$ as the number of observations grows, because the RKHS of these kernels spans the full path space and has no mechanism to enforce the ODE constraint. The posterior will converge to a smooth path that best fits the data in the RKHS norm, which is generically not an ODE solution.

The posterior under a smoothness prior therefore concentrates on trajectories that are smooth (from the prior) and explain observed edges well (from the likelihood), but do not satisfy $\dot{X} = f(X)$ for any reasonable f. This is the distinction between **interpolation** and **dynamics learning**: both produce smooth curves through the data, but only the latter respects the constraint that velocity is state-dependent.

Diagnostics for detecting dynamical inconsistency. Several tests can distinguish smooth-but-dynamically-inconsistent trajectories from genuine ODE solutions. First, a **state-dependence test**: for autonomous dynamics $\dot{X} = f(X)$, if the trajectory passes through similar positions at two different times ($X(t_1) \approx X(t_2)$), then the inferred velocities must also be similar ($\dot{X}(t_1) \approx \dot{X}(t_2)$); smoothing priors do not enforce this, and large velocity discrepancies at revisited states indicate dynamical inconsistency. Second, an **ODE residual test**: fit candidate dynamics $\hat{f}$ to the smoothed trajectory and compute residuals $r(t) = \dot{\hat{X}}(t) - \hat{f}\left(\hat{X}(t)\right)$; for dynamically consistent trajectories $\|r(t)\|$ should be small, while for smooth-but-wrong trajectories residuals will be systematically large. Third, **flow consistency**: integrate the fitted $\hat{f}$ forward from various initial conditions along the trajectory; dynamically consistent trajectories will track the integrated flow, while interpolated trajectories will diverge. Finally, **medium-to-long-term forecasting** discriminates the two approaches: kinematic extrapolation assumes velocity and acceleration persist, while dynamic extrapolation adapts velocity to the current state via $f(X)$, and only the latter remains accurate when the trajectory enters new regions of state space.

Alternative approaches with dynamical consistency.

Approach	**Smoothness**	**Dynamical Consistency**
Hierarchical GP prior	Yes (C^k)	No: interpolation only
SDE $dX = f(X)dt + \sigma dW$	No (Hölder $< 1/2$)	Yes, as $\sigma \to 0$
Neural ODE	Yes (C^k)	Yes (by construction)
GP-ODE (NPODE) [22]	Yes (C^1)	Yes: learns f as GP
Structure-constrained (ours)	Yes	Yes: family $\mathcal{F}$ enforced

Table 1: Comparison of approaches: smoothness vs dynamical consistency.

The SDE formulation $dX = f(X)dt + \sigma dW$ provides a principled bridge: Freidlin-Wentzell theory [23] shows that as $\sigma \to 0$, solutions concentrate around ODE solutions with rate function $J_{T(\varphi)} = \frac{1}{2}\int_0^T \|\dot{\varphi}(t) - f(\varphi(t))\|^2 \, dt$. This rate function is precisely the **dynamical consistency penalty**, measuring how far a path deviates from being an ODE solution.

4.7 Honest assessment

We must be candid: **aligning spectral embeddings to recover continuous-time trajectories is a hard open problem**. The methods described above address related but different problems. There is no existing method that provably recovers trajectories from ODE dynamics on RDPG latent positions.

Error accumulation in sequential alignment (Section 4.4) compounds over long trajectories. Dynamical consistency considerations (Section 4.6) distinguish interpolation from true dynamics learning. Holonomy (Section 3.3) implies that even perfect local alignment may accumulate global gauge drift.

This motivates investigating whether **structure of the dynamics themselves** can constrain the alignment problem, a question we return to in Section 7 after analyzing the dynamics families in Section 5.

5 Dynamics on RDPGs

The geometric framework of Section 3 characterizes the abstract structure of gauge freedom, curvature, and holonomy, and Section 4 demonstrates the practical difficulties of recovering trajectories from spectral embeddings. We now turn to concrete dynamics families, their observable consequences, and the interplay between geometric and statistical obstructions to estimation.

5.1 Families of RDPG dynamics

We now catalog concrete families of dynamics on RDPG latent positions. These families will later serve as inductive bias for the alignment problem (Section 7); the holonomy and statistical analyses that follow depend on their specific structure.

Linear dynamics. The simplest family is $\dot{X} = NX$ with $N \in \mathbb{R}^{n\times n}$, where N_{ij} determines how node j's position affects node i's velocity. By Proposition 2, this is horizontal if and only if $X^\top NX$ is symmetric; a sufficient condition is $N = N^\top$, which gives $X^\top NX$ symmetric for all X. The induced P-dynamics are $\dot{P} = NP + PN^\top$, which reduces to the Lyapunov equation $\dot{P} = NP + PN$ when N is symmetric. Symmetric linear dynamics have $\frac{n(n+1)}{2}$ free parameters.

Polynomial dynamics in P. A far more parsimonious family replaces the constant N with a polynomial in the probability matrix:

$$\dot{X} = N(P)X, \quad N(P) = \sum_{k=0}^{K} \alpha_k P^k \tag{36}$$

Since P is symmetric, $N(P)$ is automatically symmetric and these dynamics are **always horizontal**, regardless of the coefficients α_k. The terms have a clean interpretation: $P^0 = I$ gives self-dynamics (decay or growth), $P^1 = P$ encodes direct neighbor influence, and P^k captures k-hop effects through weighted path counts $\left(P^k\right)_{ij}$. Crucially, $P = XX^\top$ is gauge-invariant, so $N(P)$ can

be computed from observations without solving the alignment problem. The family has only $K+1$ parameters independent of n, a dramatic reduction from $\frac{n(n+1)}{2}$ for general symmetric N that is key for identifiability.

Graph Laplacian dynamics. Setting $\dot{X} = -LX$ with $L = D - P$ and $D = \mathrm{diag}(P\mathbf{1})$ gives diffusion on the graph: entry-wise, $\dot{x}_i = \sum_j P_{ij}(x_j - x_i)$, so each node moves toward the weighted average of its neighbors. Since L is symmetric, this is horizontal. Unlike polynomial dynamics, the Laplacian is **not** a polynomial in P: the degree matrix D depends on row sums, which mix eigenvector information that polynomials in P cannot access. This distinction will prove crucial for holonomy (Section 5.3).

Message-passing dynamics. More generally, $\dot{x}_i = \sum_j P_{ij} g(x_i, x_j)$ for some interaction function $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$. Special cases include neighbor attraction ($g(x_i, x_j) = x_j$, giving $\dot{X} = PX$) and Laplacian diffusion ($g(x_i, x_j) = x_j - x_i$). Horizontality depends on g; the locality constraint (velocity depends only on neighbors weighted by P) is itself a useful structural assumption.

Observable non-horizontal dynamics. Not all observable dynamics are horizontal. Centroid circulation $\dot{x}_i = (x_i - \bar{x})A$ with $A \in \mathfrak{so}(d)$ decomposes as $\dot{x}_i = x_i A - \bar{x}A$: the first term is pure gauge (Theorem 1), but the shared drift $-\bar{x}A$ shifts dot products, making $\dot{P} \neq 0$ generically. However, $X^\top \dot{X} = CA$ where C is the sample covariance of positions, and the product of a symmetric and skew-symmetric matrix is generically not symmetric, so the dynamics are not horizontal: the trajectory spirals through the fibers. Similarly, differential rotation $\dot{x}_i = x_i A_i$ with node-specific $A_i \in \mathfrak{so}(d)$ gives $\dot{P}_{ij} = x_i(A_i - A_j)x_j^\top$, which is observable whenever rotation rates differ.

Summary.

Family	Form	Horizontal?	Parameters
Linear symmetric	$\dot{X} = NX$, $N = N^\top$	Always	$n(n+1)/2$
Polynomial	$\dot{X} = \left(\sum_k \alpha_k P^k\right)X$	Always	$K+1$
Laplacian	$\dot{X} = -LX$	Always	0 (fixed)
Message-passing	$\dot{x}_i = \sum_j P_{ij} g(x_i, x_j)$	If g symmetric	$\lvert\theta\rvert$
Centroid circulation	$\dot{x}_i = (x_i - \bar{x})A$	No	$d(d-1)/2$
General linear	$\dot{X} = NX$	If $X^\top NX$ sym.	n^2

Table 2: Families of RDPG dynamics. The polynomial family is always horizontal, has few parameters ($K+1$ independent of n), and is gauge-invariant ($N(P)$ computable without alignment). Centroid circulation is observable but not horizontal.

5.2 Observable dynamics and the Lyapunov equation

Theorem 1 characterizes what is *lost* to gauge freedom. We now characterize what is *preserved*: the observable content of RDPG dynamics, projected to the base space $\mathcal{B}$ of probability matrices.

The key observation relies on two assumptions jointly. First, the RDPG model provides the factorization $P = XX^\top$, so that P inherits smoothness from X and the product rule applies. Second, considering for example the linear symmetric regime, the dynamical model $\dot{X} = NX$ specifies how positions evolve. Together, these give a closed equation for $\dot{P}$:

$$\dot{P} = \dot{X}X^\top + X\dot{X}^\top = NXX^\top + XX^\top N^\top = NP + PN^\top \tag{37}$$

When N is symmetric: $\dot{P} = NP + PN$, a Lyapunov equation.

This is *not* a general fact about temporal networks. Without the RDPG factorization, $P(t)$ is merely a time-varying matrix of edge probabilities whose entries could evolve independently, follow a completely different matrix ODE, or fail to be differentiable. Equation 37 is the **observable footprint** of RDPG dynamics: it is what the gauge-invariant content of $\dot{X} = NX$ looks like when projected to $\mathcal{B}$.

The Lyapunov operator $\mathcal{L}_P : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ defined by $\mathcal{L}_P(N) = NP + PN$ is a linear map on symmetric matrices. To analyze it, we pass to the eigenbasis of P. Let $P = U\Lambda U^\top$ be the eigendecomposition, and write $\tilde{N} = U^\top NU$ and $\tilde{\dot{P}} = U^\top \dot{P} U$ for the representations in this basis, with indices $\iota, \gamma \in \{1, ..., n\}$ labeling eigenvector directions (not nodes). In this basis, $\mathcal{L}_P$ acts diagonally: $\left(\mathcal{L}_P(N)\right)_{\iota\gamma} = \left(\lambda_\iota + \lambda_\gamma\right)\tilde{N}_{\iota\gamma}$.

For the RDPG setting, where $P = XX^\top$ with $X \in \mathbb{R}^{n\times d}$, the matrix P has rank d with eigenvalues $\lambda_1 \geq ... \geq \lambda_d > 0$ and $\lambda_{d+1} = ... = \lambda_n = 0$. The Lyapunov operator is therefore **not invertible** on all of $\mathrm{Sym}(n)$: when both $\iota, \gamma > d$, the equation $\left(\lambda_\iota + \lambda_\gamma\right)\tilde{N}_{\iota\gamma} = \tilde{\dot{P}}_{\iota\gamma}$ reduces to $0 \cdot \tilde{N}_{\iota\gamma} = \tilde{\dot{P}}_{\iota\gamma}$. The realizability constraint (Proposition 1) forces $\tilde{\dot{P}}_{\iota\gamma} = 0$ for this block, so the equation says nothing about $\tilde{N}_{\iota\gamma}$. For $n = 100$ and $d = 3$, this leaves $(n-d)(n-d+1)/2 = 4753$ out of $n(n+1)/2 = 5050$ entries of $\tilde{N}$ completely unconstrained by the data.

However, the resolution is clean: the undetermined part of N is exactly the part that does not affect the dynamics.

Proposition 6 (Partial Lyapunov Identifiability). Let $P = XX^\top$ with $\mathrm{rank}(P) = d < n$, and let $V \in \mathbb{R}^{n\times d}$ and $V_\perp \in \mathbb{R}^{n\times(n-d)}$ span the range and null space of P respectively. In the eigenbasis of P (indices ι, γ), the Lyapunov equation $\dot{P} = NP + PN$ uniquely determines the following blocks of $\tilde{N} = U^\top NU$:

- **Range-range block** ($\iota, \gamma \leq d$): $\tilde{N}_{\iota\gamma} = \tilde{\dot{P}}_{\iota\gamma} / \left(\lambda_\iota + \lambda_\gamma\right)$, giving $d(d+1)/2$ entries.
- **Cross block** ($\iota \leq d, \gamma > d$ or vice versa): $\tilde{N}_{\iota\gamma} = \tilde{\dot{P}}_{\iota\gamma}/\lambda_\iota$, giving $d(n-d)$ entries.

The **null-null block** ($\iota, \gamma > d$): $(n-d)(n-d+1)/2$ entries remain unconstrained.

The total number of determined entries is $nd - d(d-1)/2$, matching the dimension of the realizable tangent space.

Proof. In the eigenbasis, the Lyapunov equation gives $\left(\lambda_\iota + \lambda_\gamma\right)\tilde{N}_{\iota\gamma} = \tilde{\dot{P}}_{\iota\gamma}$ entry-wise. For $\iota, \gamma \leq d$: $\lambda_\iota + \lambda_\gamma > 0$, so $\tilde{N}_{\iota\gamma}$ is uniquely determined. For $\iota \leq d, \gamma > d$: $\lambda_\iota + \lambda_\gamma = \lambda_\iota > 0$, so $\tilde{N}_{\iota\gamma}$ is uniquely determined. For $\iota, \gamma > d$: $\lambda_\iota + \lambda_\gamma = 0$, and realizability forces $\tilde{\dot{P}}_{\iota\gamma} = 0$, leaving $\tilde{N}_{\iota\gamma}$ free. Counting: $d(d+1)/2 + d(n-d) = nd - d(d-1)/2$. □

The undetermined null-null block governs how N acts on vectors orthogonal to $\mathrm{col}(X)$. Since X has no component in this subspace, NX does not depend on this block: the dynamics $\dot{X} = NX$ are insensitive to it. The Lyapunov equation determines N on exactly the subspace that matters for the dynamics, and leaves unconstrained exactly the subspace that is dynamically irrelevant.

In summary symmetric linear dynamics are identifiable from $\left(P, \dot{P}\right)$ up to the dynamically irrelevant null-null block, without ever choosing a gauge. This extends to polynomial dynamics: $\dot{X} = N(P)X$ with $N(P) = \sum_k \alpha_k P^k$ gives $\dot{P} = N(P)P + PN(P)$, and the α_k can be recovered by projecting $\dot{P}$ onto the basis $\left\{P^k P + P P^k\right\}_{k=0}^{K}$ in the space of symmetric matrices. The polynomial structure is particularly natural here, as $N(P) = \sum_k \alpha_k P^k$ automatically has its null-null block determined by the α_k and the eigenvalues of P, eliminating the underdetermined component entirely.

However, two fundamental obstacles prevent direct application.

We do not observe $P(t)$. In practice, we observe adjacency matrices $A^{(t)}$ sampled as $A_{ij}^{(t)} \sim \text{Bernoulli}\big(P_{ij}(t)\big)$. Entry-wise, $A_{ij}^{(t)}$ is an unbiased estimate of $P_{ij}(t)$, with variance $P_{ij}\big(1 - P_{ij}\big)$. When m independent samples are available at each time t, averaging $\bar{A}^{(t)} = \frac{1}{m}\sum_{\ell=1}^{m} A_{\ell}^{(t)}$ reduces this per-entry variance by a factor of m, independently of the network size n.

However, the Lyapunov inversion in Proposition 6 requires the *spectral decomposition* of P, not merely its entries, and spectral accuracy depends on *both* n and m. Eigendecomposition is nonlinear, so the eigenvalues and eigenvectors of $\bar{A}$ are biased estimates of those of P. The bias is asymptotically negligible: the leading-order perturbation $\left(\bar{A} - P\right)X(X^\top X)^{-1}$ has conditional mean zero, and the residual bias is $o(1/\sqrt{nm})$ [9], [11]. But the approximation in Proposition 6 requires dividing by $\lambda_\iota + \lambda_\gamma$, and this division amplifies noise precisely where the spectral gap is small. This is the same $\frac{1}{\lambda_d}$ sensitivity that appears in the geometric analysis of Section 3.3, now manifesting as ill-conditioning of the Lyapunov inverse.

The parameter count of N on the relevant subspace. Even restricting to the dynamically relevant part of N, the number of free parameters can be large. The range-range and cross blocks together have $nd - d(d-1)/2$ entries. When N is constant, a single well-estimated $\left(P, \dot{P}\right)$ pair provides exactly this many scalar constraints, so the system is determined. But when N varies with time or state (for example in polynomial dynamics $N(P) = \sum_k \alpha_k P^k$, where the α_k are constant but $N(P(t))$ changes because P does) recovering the α_k requires sufficient variation in $P(t)$ across time points to separate the contributions of different powers P^k. The parsimonious polynomial family, with only $K+1$ parameters independent of n, is attractive precisely because it sidesteps the parameter-counting issue. For richer families, the estimation problem becomes underdetermined without strong structural assumptions or regularization.

> *Remark.* The P-dynamics perspective and the fiber bundle perspective address different aspects of the same problem. Proposition 6 says that the dynamics are identifiable *in principle* from the gauge-invariant observable P, up to the dynamically irrelevant null-null block. The fiber bundle theory (Section 3.3) addresses the harder question: what happens when we *must* work with X (e.g., because we need latent positions for interpretation or for the UDE pipeline), and how the geometry of the quotient space governs the difficulty of doing so. The two perspectives are complementary: the algebraic result tells us what information is available; the geometric framework tells us what it costs to extract it.

5.3 Holonomy for horizontal dynamics families

The holonomy obstruction (Section 3.3) raises a concrete question: for which horizontal dynamics families is holonomy trivial, and for which is it non-trivial? This distinction has direct practical consequences: trivial holonomy means global gauge consistency is achievable in principle; non-trivial holonomy means that even perfect local alignment must accumulate global drift.

We show that the answer depends sharply on whether the dynamics generator commutes with itself along the trajectory: a condition satisfied by polynomial dynamics but violated by Laplacian and message-passing dynamics (cf. Section 5.1).

Curvature from non-commuting generators. Consider horizontal dynamics $\dot{X} = M(X)X$ with $M(X)$ symmetric, so that $X^\top \dot{X} = X^\top M(X)X$ is symmetric and the dynamics are horizontal. At two distinct times t_1, t_2, the generators $M_1 = M(X(t_1))$ and $M_2 = M(X(t_2))$ are both symmetric. The commutator $[M_1, M_2] = M_1M_2 - M_2M_1$ is skew-symmetric (since transposing reverses the order).

The Lie bracket of the corresponding horizontal vector fields $\overline{\xi}_i(X) = M_iX$ (constant-coefficient extensions, horizontal since each M_i is symmetric) is:

$$\left[\overline{\xi}_1, \overline{\xi}_2\right](X) = (M_2M_1 - M_1M_2)X = -[M_1, M_2]X \tag{38}$$

Writing $S = -[M_1, M_2]$ (which is skew-symmetric), the bracket SX is *not* purely vertical: vertical vectors have the form $X\Omega$ for $\Omega \in \mathfrak{so}(d)$ (right multiplication), while SX is left multiplication by a skew-symmetric matrix, and these coincide only when S maps $\mathrm{col}(X)$ into itself, which generically fails for $n \gg d$.

However, the O'Neill formula requires only the vertical component $\left[\overline{\xi}_1, \overline{\xi}_2\right]^{\mathcal{V}}$. The vertical projection of $Z = SX$ at X onto the vertical space $\mathcal{V}_X = \{X\Omega : \Omega \in \mathfrak{so}(d)\}$ is $X\Omega^*$ where Ω^* solves the Lyapunov equation $G\Omega^* + \Omega^*G = 2\,\mathrm{skew}(X^\top Z) = 2X^\top SX$ with $G = X^\top X$ (see Section 13 for the derivation). Since S is skew-symmetric, $X^\top SX$ is skew-symmetric, and the Lyapunov equation has a unique solution because all eigenvalues of G are positive. In the eigenbasis of G (eigenvalues $\lambda_1, ..., \lambda_d$), the solution is:

$$\Omega^*_{\iota\gamma} = 2\frac{(X^\top SX)_{\iota\gamma}}{\lambda_\iota + \lambda_\gamma} \tag{39}$$

The vertical component thus vanishes if and only if $X^\top[M_1, M_2]X = 0$ (the commutator projected to the $d \times d$ fiber tangent space). This is weaker than $[M_1, M_2] = 0$: the full $n \times n$ commutator may be nonzero while its projection through X vanishes. However, for generic trajectories with $n \gg d$, if $[M_1, M_2] \neq 0$ then $X^\top[M_1, M_2]X \neq 0$ as well, since the "blind spot" $\mathrm{col}(X)^\perp$ is a high-codimension subspace.

Remembering Proposition 4, by O'Neill's formula [20] the sectional curvature of the 2-plane spanned by the projections of $\overline{\xi}_1, \overline{\xi}_2$ in $\mathcal{B}$ is: $K(\xi_1, \xi_2) = \frac{3}{4}\,\|\left[\overline{\xi}_1, \overline{\xi}_2\right]^{\mathcal{V}}\|^2 \,/\left(\|\overline{\xi}_1\|^2\,\|\overline{\xi}_2\|^2 - \langle\overline{\xi}_1, \overline{\xi}_2\rangle^2\right)$. The $1/\left(\lambda_\iota + \lambda_\gamma\right)$ factors in Ω^* connect the curvature directly to the connection coefficients of the fiber bundle: the same denominators that amplify gauge sensitivity also amplify the vertical bracket.

Since $\mathbb{R}^{n\times d}_*$ is flat (it is an open subset of Euclidean space), the base curvature arises **entirely** from the vertical bracket.

Proposition 7 (Curvature Criterion for Horizontal Dynamics). For horizontal dynamics $\dot{X} = M(X)X$ with M symmetric, the sectional curvature along the trajectory in $\mathcal{B}$ vanishes if and only if the projected commutator vanishes: $X^\top[M(X(t_1)), M(X(t_2))]X = 0$ for all t_1, t_2 along the trajectory. A sufficient condition is that the full generators commute: $[M(X(t_1)), M(X(t_2))] = 0$. When the projected commutator is nonzero, the curvature is strictly positive, with:

$$\|\left[\bar{\xi}_1, \bar{\xi}_2\right]^{\mathcal{V}}\|^2 = 4\sum_{\iota<\gamma} \frac{\lambda_\iota \lambda_\gamma}{\lambda_\iota + \lambda_\gamma}\left[\left(U^\top[M_1, M_2]U\right)_{\iota\gamma}\right]^2 \tag{40}$$

where $P = U\Lambda U^\top$ is the eigendecomposition. The $1/(\lambda_\iota + \lambda_\gamma)$ factors link the curvature directly to the connection coefficients of the fiber bundle.

Proof. The Lie bracket of the constant-coefficient horizontal extensions $\bar{\xi}_i(Y) = M_i Y$ is $\left[\bar{\xi}_1, \bar{\xi}_2\right](X) = -[M_1, M_2]X$ (standard commutator of linear vector fields). The vertical projection of SX (with $S = -[M_1, M_2]$ skew-symmetric) is $X\Omega^*$ where Ω^* solves $G\Omega^* + \Omega^* G = 2X^\top SX$ (Section 13). This vanishes iff $X^\top[M_1, M_2]X = 0$. The norm formula follows from $\|X\Omega^*\|^2 = \mathrm{tr}(\Omega^{*\top} G\Omega^*)$ evaluated in the eigenbasis of G (see Section 14 for the computation). □

Polynomial dynamics: trivial holonomy. For polynomial dynamics $\dot{X} = N(P)X$ with $N(P) = \sum_{k=0}^{K} \alpha_k P^k$, the induced P-dynamics are:

$$\dot{P} = N(P)P + PN(P) = 2\sum_{k=0}^{K} \alpha_k P^{k+1} \tag{41}$$

This is a polynomial in P alone. Writing $P = U\Lambda U^\top$ in its eigendecomposition, every power $P^k = U\Lambda^k U^\top$ shares the same eigenvectors U. Therefore $\dot{P} = U\left(2\sum \alpha_k \Lambda^{k+1}\right)U^\top$ also has eigenvectors U: the eigenvectors of P are stationary under polynomial dynamics, provided the spectrum of $P(0)$ is simple (all eigenvalues distinct).

The trajectory $P(t)$ lies on a d-dimensional submanifold of $\mathcal{B}$ parameterized by the eigenvalues $\left(\lambda_1(t), ..., \lambda_{d(t)}\right)$ alone, with each eigenvalue evolving independently:

$$\dot{\lambda}_\iota = 2\sum_{k=0}^{K} \alpha_k \lambda_\iota^{k+1} \tag{42}$$

Two consequences follow immediately.

Proposition 8 (Polynomial Dynamics: Trivial Holonomy). Polynomial dynamics $\dot{X} = N(P)X$ with $N(P) = \sum \alpha_k P^k$ and simple initial spectrum $(\lambda_1(0) > \lambda_2(0) > ... > \lambda_d(0) > 0)$ have trivial holonomy. Specifically:

1. The generators $N(P(t_1))$ and $N(P(t_2))$ commute at all times (they share eigenvectors), so the curvature along the trajectory vanishes by Proposition 7.
2. The eigenvalue ODEs are autonomous and one-dimensional, hence admit no periodic orbits. Closed loops in $\mathcal{B}$ are therefore impossible under polynomial dynamics.

Proof. For (1): since the eigenvectors of P are stationary, $N(P(t)) = U\big(\sum \alpha_k \Lambda(t)^k\big)U^\top$ for all t, with the same U. Any two diagonal matrices commute, so $[N(P(t_1)), N(P(t_2))] = U\Big[\sum \alpha_k \Lambda(t_1)^k, \sum \alpha_k \Lambda(t_2)^k\Big]U^\top = 0$.

For (2): each $\lambda_\iota(t)$ satisfies the autonomous ODE $\dot{\lambda}_\iota = f(\lambda_\iota)$ with $f(x) = 2\sum \alpha_k x^{k+1}$. By the uniqueness theorem for ODEs, if $\lambda_\iota(t_0) = \lambda_\iota(t_1)$ for $t_0 \neq t_1$, then λ_ι is constant. Non-constant orbits are monotone, so no closed loop in eigenvalue space, and hence in $\mathcal{B}$,is possible. □

What polynomial dynamics look like for nodes. The statement that eigenvectors of P are stationary deserves both rigorous justification and concrete interpretation.

Why eigenvectors are fixed. Write $P(t) = U(t)\Lambda(t)U(t)^\top$ and let $\Omega = U^\top \dot{U} \in \mathfrak{so}(d)$ be the eigenvector rotation rate. Differentiating $P = U\Lambda U^\top$ and projecting into the eigenbasis gives $\left(U^\top \dot{P} U\right)_{\iota\gamma} = \Omega_{\iota\gamma}\big(\lambda_\gamma - \lambda_\iota\big)$ for $\iota \neq \gamma$. Since $\dot{P} = 2\sum \alpha_k P^{k+1}$ is diagonal in the eigenbasis, the left side vanishes, so $\Omega_{\iota\gamma}\big(\lambda_\gamma - \lambda_\iota\big) = 0$ for all $\iota \neq \gamma$. It remains to show that eigenvalues remain distinct. All d eigenvalues satisfy the same autonomous ODE $\dot{\lambda} = f(\lambda)$ with $f(x) = 2\sum \alpha_k x^{k+1}$, differing only in initial conditions. If $\lambda_\iota(0) \neq \lambda_\gamma(0)$, then by uniqueness of ODE solutions $\lambda_\iota(t) \neq \lambda_\gamma(t)$ for all t: if they ever coincided, they would be the same solution, contradicting their distinct initial conditions. Therefore $\lambda_\gamma - \lambda_\iota \neq 0$, which forces $\Omega = 0$: the eigenvectors of P are genuinely stationary, not merely stationary to first order.

The simple spectrum assumption is essential: if $\lambda_{\iota(0)} = \lambda_{\gamma(0)}$ for some $\iota \neq \gamma$, the eigenbasis within the repeated eigenspace is not unique, and polynomial dynamics preserve this degeneracy (the repeated eigenvalues evolve identically, maintaining the degeneracy for all time). In this case, eigenvector "stationarity" is ill-defined because the basis itself is non-unique. For block models with equal block sizes, for instance, repeated eigenvalues arise from the model symmetry. This is a measure-zero condition in the space of initial conditions.

What this means for nodes. In the canonical gauge $X = U\Lambda^{1/2}$, node i has latent position $x_i(t) = \Big(U_{i1}\sqrt{\lambda_1(t)}, ..., U_{id}\sqrt{\lambda_{d(t)}}\Big)$. The loading $U_{i\iota}$ (which can be read as the "membership weight" of node i in eigenvector direction ι) is constant; what evolves is the scale $\sqrt{\lambda_\iota(t)}$ of each direction.

Since the eigenvalue ODE $\dot{\lambda} = f(\lambda)$ is nonlinear for $K \geq 1$, eigenvalues with different magnitudes evolve at different rates, and the ratios $\lambda_\iota(t)/\lambda_\gamma(t)$ change over time. The effect on the node positions is an **anisotropic rescaling** of the latent space along fixed axes: in $d = 2$, the point cloud deforms as if inscribed in an ellipse whose axes do not rotate but whose eccentricity changes. Node directions in $\mathbb{R}^d$ do change (they are not merely rescaled in magnitude), but the change is tightly constrained: all deformation is captured by the d scalar functions $\lambda_\iota(t)$.

The exception is linear dynamics ($K = 0$, $f(\lambda) = 2\alpha_0\lambda$), where $\lambda_\iota(t) = \lambda_\iota(0)e^{2\alpha_0 t}$ and all eigenvalues grow or decay at the same exponential rate. In this case the ratios are constant, the point cloud is rescaled isotropically, and node angular positions in $\mathbb{R}^d$ are truly fixed. For quadratic or higher dynamics ($K \geq 1$), the nonlinearity of f causes larger eigenvalues to evolve differently from smaller ones, producing genuine reshaping of the point cloud's geometry while preserving the eigenvector axes that define community alignment. Note, however, that preserving the axes does not guarantee preserving community **separability**: if eigenvalue ratios diverge (e.g., $\lambda_d(t) \to 0$ while λ_1 grows), the point cloud collapses onto a lower-dimensional subspace, and communities that were distinguishable along the λ_d direction may become indistinguishable.

Laplacian dynamics: non-trivial holonomy. Graph Laplacian dynamics $\dot{X} = -LX$ with $L = D - P$, where $D = \mathrm{diag}(P\mathbf{1})$, provide a natural example of horizontal dynamics with non-trivial holonomy. The generator $-L = P - D$ is symmetric, so the dynamics are horizontal. But L is **not** a polynomial in P: the degree matrix D depends on the row sums of P, which mix eigenvector information that a polynomial in P cannot access.

There is a notable exception: if D is a scalar multiple of the identity ($D = cI$, i.e., the graph is **regular** with all node degrees equal), then $L = cI - P$ is affine in P, and the Laplacian dynamics reduce to the polynomial family. Regular graphs (and more generally, vertex-transitive latent configurations where all rows of X have the same norm) thus have trivial holonomy under Laplacian dynamics. The non-trivial holonomy results below require that D is **not** scalar: structural heterogeneity in the graph is the source of holonomy.

The P-dynamics under Laplacian flow are:

$$\dot{P} = -LP - PL = -(D - P)P - P(D - P) = 2P^2 - DP - PD \tag{43}$$

The term $2P^2$ preserves eigenvectors of P (it is a polynomial in P), but $DP + PD$ does not: in the eigenbasis $P = U\Lambda U^\top$, the degree matrix $D = \mathrm{diag}(U\Lambda U^\top \mathbf{1})$ is diagonal in the **node** basis, not the eigenbasis. Thus $U^\top DU$ is generally a full matrix, and $\dot{P}$ has eigenvector components transverse to those of P.

Consequence: the eigenvectors of P rotate under Laplacian dynamics, the trajectory sweeps out area in $\mathcal{B}$ transverse to the eigenvalue-parameterized submanifold, and Proposition 7 guarantees positive curvature along any portion of the trajectory where $[L(P(t_1)), L(P(t_2))] \neq 0$.

What Laplacian dynamics look like for nodes. The contrast with polynomial dynamics is sharp at the node level. Under Laplacian diffusion, each node's velocity $\dot{x}_i = -\sum_j L_{ij} x_j$ is a weighted average of the positions of its neighbors (with weights from P), minus a restoring term from the node's own degree. This acts like heat diffusion on the latent positions: high-degree nodes experience stronger mean-reversion, while the diffusion couples all positions through the network structure. Because the coupling depends on the *row sums* of P, mixing contributions from all eigenvector directions, the eigenvectors of P rotate over time. In the ellipse analogy for $d = 2$: not only does the eccentricity change, but the axes themselves rotate. The community structure is no longer static; nodes' membership weights across latent dimensions genuinely evolve, creating the richer dynamics that produce non-trivial holonomy.

Proposition 9 (Laplacian Dynamics: Non-Trivial Holonomy). For $d \geq 2$ and initial conditions such that $D(P(0))$ is not a scalar multiple of the identity (i.e., the graph is not regular), Laplacian dynamics $\dot{X} = -L(P)X$ generically produce trajectories in $\mathcal{B}$ with strictly positive sectional curvature. If the trajectory visits states $P(t_1), P(t_2)$ such that $X^\top [L(P(t_1)), L(P(t_2))]X \neq 0$ (projected commutator nonzero), the holonomy around any loop enclosing the corresponding portion of the trajectory is non-trivial.

Proof. It suffices to show that $X^\top [L(P(t_1)), L(P(t_2))]X \neq 0$ generically. Write $L_i = D_i - P_i$ for $i = 1, 2$. Then $[L_1, L_2] = [D_1, P_2] - [D_2, P_1] + [P_1, P_2] - [D_1, D_2]$. Since D_1, D_2 are both diagonal, $[D_1, D_2] = 0$. And $[P_1, P_2] = 0$ when P_1, P_2 share eigenvectors, but under Laplacian dynamics they do

not, so generically $[P_1, P_2] \neq 0$. Moreover, $[D_1, P_2]$ is generically nonzero since the row sums encoded in D_1 depend on the full matrix P_1, not just its eigenvalues. For $n \gg d$, a nonzero $n \times n$ commutator $[L_1, L_2]$ generically projects to a nonzero $d \times d$ matrix $X^\top [L_1, L_2] X$, since vanishing would require $[L_1, L_2]$ to map $\mathrm{col}(X)$ entirely into $\mathrm{col}(X)^\perp$, a codimension-d condition. By Proposition 7, the curvature along the trajectory is positive, and by the Ambrose-Singer theorem, the holonomy group is generated by the curvature 2-form evaluated at such pairs. □

For $d = 2$, the result is particularly sharp.

Corollary 2 (Full Holonomy for Laplacian Dynamics in Dimension 2). For $d = 2$ and generic initial conditions, the restricted holonomy group of Laplacian dynamics is $\mathrm{SO}(2)$, the largest possible connected subgroup of $O(2)$. Concretely: for any target rotation angle $\varphi \in [0, 2\pi)$, there exists a loop in $\mathcal{B}$ traversable by Laplacian dynamics whose holonomy is rotation by φ.

Proof. The Lie algebra $\mathfrak{so}(2)$ is one-dimensional, spanned by the single generator $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. By Proposition 9, the curvature along a generic Laplacian trajectory is strictly positive, so the curvature 2-form produces a nonzero element of $\mathfrak{so}(2)$. A single nonzero element spans the one-dimensional Lie algebra, so by the Ambrose-Singer theorem, the holonomy algebra is all of $\mathfrak{so}(2)$, and the restricted holonomy group (holonomy of contractible loops) is $\mathrm{SO}(2)$. □

This means that for rank-2 latent spaces under Laplacian dynamics, the gauge ambiguity is not merely a discrete sign flip (as for polynomial dynamics) but a continuous rotation by an arbitrary angle: the worst possible case for alignment.

Quantitative estimate. For $d = 2$, the holonomy around a loop γ in $\mathcal{B}$ is a rotation by angle:

$$\varphi(\gamma) = \iint_\Sigma K \, dA \tag{44}$$

where Σ is a surface bounded by γ, K is the sectional curvature, and dA is the area element. By the curvature formula of [18], regions where the smallest eigenvalues of P are small contribute large curvature, so even short Laplacian trajectories that pass near rank-deficient states can accumulate substantial holonomy.

Higher dimensions ($d \geq 2$). For general d, the Lie algebra $\mathfrak{so}(d)$ has dimension $d(d-1)/2$. The curvature 2-form at each point produces holonomy algebra elements $\Omega^* \in \mathfrak{so}(d)$ with $\Omega^*_{\iota\gamma} = 2\big(X^\top [M_1, M_2] X\big)_{\iota\gamma} / \big(\lambda_\iota + \lambda_\gamma\big)$, and as the trajectory visits different states, these elements span an increasing subspace. For generic Laplacian dynamics with $n \gg d$, the degree matrices at different times are sufficiently varied that we expect the curvature elements to span all of $\mathfrak{so}(d)$, yielding holonomy group $\mathrm{SO}(d)$. We state this as a conjecture.

Conjecture 1 (Full Holonomy in Higher Dimensions). For $d \geq 2$, $n > d$, and initial conditions with $D(P(0))$ not scalar (non-regular graph), the restricted holonomy group of Laplacian dynamics $\dot{X} = -L(P)X$ is $\mathrm{SO}(d)$.

A proof would require showing that the holonomy algebra elements $\Omega^*_{\iota\gamma} = 2\big(X^\top [L(P(t_1)), L(P(t_2))] X\big)_{\iota\gamma} / \big(\lambda_\iota + \lambda_\gamma\big)$, as t_1, t_2 vary along the trajectory, span $\mathfrak{so}(d)$. Since the

Lyapunov weights $1/(\lambda_\iota + \lambda_\gamma)$ are positive scalars that do not change rank, this is equivalent to requiring that the projected commutators $X^\top[L(P(t_1)), L(P(t_2))]X$ span $\mathfrak{so}(d)$. For $d = 2$ this is automatic (one nonzero element spans a one-dimensional Lie algebra); for $d \geq 3$ it requires a transversality argument exploiting the fact that the degree matrix $D = \mathrm{diag}(P\mathbf{1})$ mixes all d eigenvector components. The potential obstruction to full holonomy would be an **invariant subspace** of the degree operator: if the map $P \mapsto D(P) = \mathrm{diag}(P\mathbf{1})$ always preserved some subspace of $\mathfrak{so}(d)$ aligned with the eigenvector structure, the holonomy group would be confined to a proper subgroup. Generically, the degree operator mixes all eigenvector components, precluding such invariant subspaces; the non-generic exceptions are configurations with symmetries that force D to commute with P (regular graphs being the extreme case identified above).

The same argument applies to message-passing dynamics $\dot{x}_i = \sum_j P_{ij} g(x_i, x_j)$ with symmetric generators that are not polynomials in P.

Summary: the holonomy obstruction creates a sharp dichotomy among horizontal dynamics families. Polynomial dynamics, which operate on the spectral structure of P through commuting generators, preserve eigenvectors and have trivial holonomy: global gauge consistency is achievable without topological obstruction. Laplacian and message-passing dynamics, which mix spectral and spatial structure through non-commuting generators, rotate eigenvectors and produce non-trivial holonomy: gauge drift accumulates even along horizontal trajectories, and no local alignment procedure can achieve global consistency over cycles.

This distinction has practical implications: for polynomial dynamics, the constructive alignment problem (Section 7) is purely a statistical challenge (overcoming Bernoulli noise); for Laplacian dynamics, it is simultaneously a statistical *and* topological challenge.

Remark. **Ambrose-Singer perspective.** The relationship between local curvature and global holonomy is formalized by the Ambrose-Singer theorem [19]: the Lie algebra $\mathfrak{hol}_p$ of the restricted holonomy group at p equals the subalgebra of $\mathfrak{so}(d)$ spanned by all curvature endomorphisms $\Omega_q(u, v)$, where q varies over points reachable by horizontal curves from p and the endomorphisms are parallel-transported back to p.

For $d = 2$, $\mathfrak{so}(2)$ has no proper nonzero subalgebras, so a single nonzero curvature element generates the full algebra and forces $\mathrm{Hol}_p^0 = \mathrm{SO}(2)$ (as used in the $d = 2$ corollary above).

For $d \geq 3$, the algebra $\mathfrak{so}(d)$ has dimension $d(d-1)/2$, and a single curvature value spans only a one-dimensional subspace. Full holonomy requires a bracket-generating condition (analogous to Hörmander's condition in sub-Riemannian geometry): the Lie brackets of the curvature tensors at different trajectory points, $[\Omega(t_1), \Omega(t_2)]$ and their iterated brackets, must span $\mathfrak{so}(d)$. Since the degree matrix $D(X)$ depends nonlinearly on the latent positions, the orientation of Ω^* within $\mathfrak{so}(d)$ rotates as the trajectory evolves, preventing confinement to a sub-variety where curvature values commute. This is the mechanism behind Conjecture 1.

As shown in Proposition 7, the curvature magnitude is weighted by $1/(\lambda_\iota + \lambda_\gamma)$; a small spectral gap therefore not only amplifies statistical noise (through ASE error bounds) but also enlarges the holonomy effect, since the curvature-generated rotations are larger in the ill-conditioned directions.

6 Information-theoretic limits

The geometric obstructions of Section 3 (gauge freedom, curvature, holonomy), the trajectory recovery difficulties of Section 4, and the dynamics analysis of Section 5 constrain what is learnable in principle. We now complement these with **statistical** obstructions: given T noisy adjacency matrices, how accurately can we estimate the parameters governing the dynamics, regardless of the estimation method?

The answer reveals a duality: the same spectral gap λ_d that controls geometric difficulty also controls statistical difficulty, so the two obstructions reinforce each other.

6.1 Fisher information for dynamics parameters

Consider a parametric dynamics family with parameter $\theta \in \mathbb{R}^k$ generating a trajectory $P(t;\theta)$. We observe T adjacency matrices $A^{(t)} \sim \text{Bernoulli}(P(t;\theta))$ at times $t_1, ..., t_T$, with all entries conditionally independent given $P(t;\theta)$. The log-likelihood factorizes:

$$\ell(\theta) = \sum_{t=1}^{T} \sum_{i<j} \Big[A_{ij}^{(t)} \log P_{ij}(t;\theta) + \Big(1 - A_{ij}^{(t)}\Big) \log\big(1 - P_{ij}(t;\theta)\big) \Big] \tag{45}$$

The Fisher information matrix $\mathcal{I}(\theta) \in \mathbb{R}^{k \times k}$ has entries:

$$\mathcal{I}(\theta)_{ab} = \sum_{t=1}^{T} \sum_{i<j} \frac{\partial P_{ij}}{\partial \theta_a} \frac{\partial P_{ij}}{\partial \theta_b} \cdot \frac{1}{P_{ij}(1 - P_{ij})} \tag{46}$$

where $P_{ij} = P_{ij}(t;\theta)$ and the derivatives $\partial P_{ij}/\partial \theta_a$ capture how perturbations in the dynamics parameters propagate to the observations.

6.2 Sensitivity propagation through the Lyapunov equation

The key quantity in Equation 46 is the sensitivity $\partial P_{ij}/\partial \theta_a$. For polynomial dynamics $N(\theta) = \sum_{k=0}^{K} \theta_k P^k$, the P-dynamics are:

$$\dot{P} = 2 \sum_{k=0}^{K} \theta_k P^{k+1} \tag{47}$$

A crucial simplification comes from Proposition 8: the eigenvectors of P are stationary under polynomial dynamics. Writing $P = U \Lambda U^\top$ with $\Lambda = \text{diag}(\lambda_1, ..., \lambda_d)$ and $U \in \mathbb{R}^{n \times d}$ fixed, the sensitivity $\partial P/\partial \theta_a = U(\partial \Lambda/\partial \theta_a)U^\top$ is diagonal in the eigenbasis.

Each eigenvalue satisfies $\dot{\lambda}_\iota = 2 \sum_{k=0}^{K} \theta_k \lambda_\iota^{k+1}$, so:

$$\frac{\partial}{\partial t} \frac{\partial \lambda_\iota}{\partial \theta_a} = 2\lambda_\iota^{a+1} + \frac{\partial \dot{\lambda}_\iota}{\partial \lambda_\iota} \cdot \frac{\partial \lambda_\iota}{\partial \theta_a} \tag{48}$$

where $\partial \dot{\lambda}_\iota / \partial \lambda_\iota = 2 \sum_{k=0}^{K} \theta_k (k+1) \lambda_\iota^k$ is the linearized eigenvalue dynamics. This is a scalar linear ODE for each pair (ι, a), with forcing $2\lambda_\iota^{a+1}$ and feedback through the linearized dynamics.

The resulting sensitivity of the observed probabilities is:

$$\frac{\partial P_{ij}}{\partial \theta_a} = \sum_{\iota=1}^{d} U_{i\iota} U_{j\iota} \frac{\partial \lambda_\iota}{\partial \theta_a} \tag{49}$$

Each node pair (i, j) receives signal from all d eigenvalue sensitivities, weighted by the eigenvector loadings $U_{i\iota}U_{j\iota}$.

Example: linear dynamics ($K = 0$). For $\dot{X} = \alpha_0 X$ (single parameter), the eigenvalue dynamics are $\dot{\lambda}_\iota = 2\alpha_0\lambda_\iota$, giving $\lambda_\iota(t) = \lambda_\iota(0)e^{2\alpha_0 t}$. The sensitivity is $\partial\lambda_\iota/\partial\alpha_0 = 2t\lambda_\iota(t)$, and therefore:

$$\frac{\partial P_{ij}}{\partial \alpha_0} = 2t \sum_\iota U_{i\iota}\lambda_\iota(t)U_{j\iota} = 2tP_{ij}(t) \tag{50}$$

The Fisher information at a single snapshot t is:

$$\mathcal{I}_t(\alpha_0) = 4t^2 \sum_{i<j} P_{ij} \frac{(t)^2}{P_{ij}(t)\big(1 - P_{ij}(t)\big)} = 4t^2 \sum_{i<j} P_{ij} \frac{t}{1 - P_{ij}(t)} \tag{51}$$

Summing over T equally-spaced snapshots gives $\mathcal{I}(\alpha_0) = 4\sum_{t=1}^{T} t^2 \sum_{i<j} P_{ij}(t)/\big(1 - P_{ij}(t)\big)$, which scales as $O(T^3 \cdot n^2)$ when the edge probabilities are bounded away from 0 and 1.

Higher-degree terms and the spectral gap. For $a \geq 1$, the forcing $2\lambda_\iota^{a+1}$ is small for eigenvalues near the spectral gap: $2\delta^{a+1}$ versus $2\lambda_1^{a+1}$. This means the sensitivity $\partial\lambda_d/\partial\theta_a$ is small relative to $\partial\lambda_1/\partial\theta_a$, so the Fisher information for higher-degree parameters θ_a receives proportionally less signal from the small-eigenvalue directions. Concretely, the contribution of the λ_d direction to $\mathcal{I}(\theta)_{ab}$ scales as δ^{a+b+2}, while the λ_1 direction contributes λ_1^{a+b+2}.

6.3 General dynamics and the Lyapunov amplification

The polynomial case is clean because eigenvectors are fixed, confining all sensitivity to the diagonal of the eigenbasis. For general symmetric dynamics $\dot{X} = M(P)X$ with M symmetric but not a polynomial in P, eigenvectors rotate and the sensitivity $\partial P/\partial\theta$ has both diagonal and off-diagonal components in the eigenbasis.

In this setting, the Lyapunov equation $\dot{P} = MP + PM$ must be inverted to recover M from $\dot{P}$. In the eigenbasis of P, the inversion acts componentwise: $M_{\iota\gamma} = \dot{P}_{\iota\gamma}/\big(\lambda_\iota + \lambda_\gamma\big)$. The factor $1/\big(\lambda_\iota + \lambda_\gamma\big)$ amplifies noise in the off-diagonal components, and is the same factor that appears in the connection 1-form (Section 3.3) and governs curvature. This yields a direct correspondence: directions in parameter space that are hard geometrically (small $\lambda_\iota + \lambda_\gamma$ produces large curvature and large connection coefficients) are also hard statistically (small $\lambda_\iota + \lambda_\gamma$ amplifies estimation noise).

For Laplacian dynamics, the situation is compounded by holonomy. The eigenvector rotation contributes additional degrees of freedom to the sensitivity, but this information is entangled with gauge degrees of freedom that the holonomy obstruction (Proposition 9) prevents from being resolved locally.

6.4 The statistical-geometric duality

Combining the geometric and statistical pictures yields a unified obstruction. We first state the result for polynomial dynamics, then explain the broader duality.

Proposition 10 (Cramér-Rao Bound for Polynomial Dynamics). For polynomial dynamics of degree K with parameter $\theta = (\theta_0, ..., \theta_K) \in \mathbb{R}^k$ ($k = K + 1$) on an RDPG with n nodes, d latent dimensions, and spectral gap $\delta = \lambda_d$, any unbiased estimator $\hat{\theta}$ satisfies:

$$\mathbb{E}\Big[\|\hat{\theta} - \theta\|^2\Big] \geq \operatorname{tr}(\mathcal{I}(\theta)^{-1}) \tag{52}$$

The Fisher information matrix $\mathcal{I}(\theta)$ has entries:

$$\mathcal{I}(\theta)_{ab} = \sum_{t=1}^{T} \sum_{\iota,\gamma=1}^{d} \frac{\partial \lambda_\iota}{\partial \theta_a} \frac{\partial \lambda_\gamma}{\partial \theta_b} \cdot C_{\iota\gamma}(t) \tag{53}$$

where $C_{\iota\gamma}(t) = \sum_{i<j} U_{i\iota} U_{j\iota} U_{i\gamma} U_{j\gamma} / \big[P_{ij}(t)\big(1 - P_{ij}(t)\big)\big]$. The diagonal entries $C_{\iota\iota}(t) = \mathcal{W}_\iota(t)$ measure the information content of eigenvalue direction ι at time t; the off-diagonal entries $C_{\iota\gamma}(t)$ for $\iota \neq \gamma$ capture cross-eigenvalue correlations. A baseline lower bound on each parameter's variance follows from the matrix inequality $\left(\mathcal{I}^{-1}\right)_{aa} \geq 1/\mathcal{I}_{aa}$:

$$\operatorname{Var}\big(\hat{\theta}_a\big) \geq \frac{1}{\mathcal{I}_{aa}} = \frac{1}{\sum_t \sum_\iota \left(\frac{\partial \lambda_\iota}{\partial \theta_a}\right)^2 \mathcal{W}_\iota(t) + \text{cross terms}} \tag{54}$$

Non-zero cross-eigenvalue terms arising from structured graphs (where Fisher weights correlate with eigenvector components) can only *increase* the true Cramér-Rao bound above this baseline.

Proof. By Equation 49, the sensitivity of P_{ij} decomposes as $\partial P_{ij}/\partial \theta_a = \sum_\iota U_{i\iota} U_{j\iota} (\partial \lambda_\iota / \partial \theta_a)$. Substituting into Equation 46:

$$\mathcal{I}(\theta)_{ab} = \sum_t \sum_{i<j} \left[\sum_\iota U_{i\iota} U_{j\iota} \frac{\partial \lambda_\iota}{\partial \theta_a}\right] \left[\sum_\gamma U_{i\gamma} U_{j\gamma} \frac{\partial \lambda_\gamma}{\partial \theta_b}\right] \cdot \frac{1}{P_{ij}\big(1 - P_{ij}\big)} \tag{55}$$

Expanding the product over ι, γ yields:

$$\mathcal{I}(\theta)_{ab} = \sum_t \sum_{\iota,\gamma} \frac{\partial \lambda_\iota}{\partial \theta_a} \frac{\partial \lambda_\gamma}{\partial \theta_b} \underbrace{\sum_{i<j} \frac{U_{i\iota} U_{j\iota} U_{i\gamma} U_{j\gamma}}{P_{ij}\big(1 - P_{ij}\big)}}_{C_{\iota\gamma}(t)} \tag{56}$$

On the cross-eigenvalue terms. Column orthogonality of U gives $\sum_i U_{i\iota} U_{i\gamma} = \delta_{\iota\gamma}$, but the Fisher weights $1/\big[P_{ij}\big(1 - P_{ij}\big)\big]$ are not uniform. In structured RDPGs (e.g., stochastic block models with equal-sized blocks), these weights are constant within blocks and the blocks are defined by the eigenvectors themselves, so the weights correlate with eigenvector components and the cross terms $C_{\iota\gamma}$ for $\iota \neq \gamma$ are generically nonzero. The Fisher information matrix is therefore **not** diagonal in the eigenvalue basis for structured graphs.

However, the spectral gap scaling is unaffected: the eigenvalue sensitivity $\partial \lambda_\iota / \partial \theta_a$ is determined by the scalar ODE $\dot{\lambda}_\iota = 2 \sum_k \theta_k \lambda_\iota^{k+1}$. The forcing term $2\lambda_\iota^{a+1}$ scales as δ^{a+1} for $\iota = d$, so the λ_d direction contributes $O\big(\delta^{a+b+2}\big)$ to $\mathcal{I}_{ab}$ regardless of whether this contribution arises from diagonal or cross terms. When $\delta \to 0$, the Fisher information matrix becomes ill-conditioned: the rows and columns involving high powers of δ shrink, making $\operatorname{tr}\big(\mathcal{I}^{-1}\big)$ diverge.

The baseline bound $\left(\mathcal{I}^{-1}\right)_{aa} \geq 1/\mathcal{I}_{aa}$ follows from the Schur complement: for any positive definite matrix, the diagonal of the inverse is at least the reciprocal of the diagonal. This bound is tight when $\mathcal{I}$ is diagonal and loose when off-diagonal terms are large (i.e., when parameter correlations from structured graphs make estimation strictly harder than the diagonal analysis suggests). □

The bound reveals three scaling regimes. Each snapshot provides $O(n^2)$ conditionally independent Bernoulli observations, so the Fisher information grows as n^2. Information accumulates linearly in T, assuming the dynamics produce sufficient variation in P across time. The spectral gap δ controls estimation difficulty through the eigenvalue sensitivities: the parameter θ_a receives signal proportional to δ^{a+1} from the smallest eigenvalue direction. For the constant term θ_0 (linear dynamics), the δ-direction contributes $O(\delta^2)$ to the Fisher information; for the highest-degree term θ_K, the contribution degrades to $O(\delta^{2K+2})$. These scalings hold for both the diagonal entries and the full matrix; the cross-eigenvalue terms share the same δ-dependence.

For linear dynamics (single scalar parameter α_0), the Fisher information is exact with no cross-term ambiguity:

Corollary 3 (Linear Dynamics: Explicit Fisher Information). For linear dynamics $\dot{X} = \alpha_0 X$ with T snapshots at times $t_1, ..., t_T$, the Fisher information is:

$$\mathcal{I}(\alpha_0) = 4 \sum_{t=1}^{T} t^2 \sum_{i<j} P_{ij} \frac{t}{1 - P_{ij}(t)} \tag{57}$$

This scales as $\Theta(T^3 \cdot n^2)$ when edge probabilities remain bounded away from 0 and 1 throughout the trajectory, giving a Cramér-Rao lower bound of $\Omega(1/(T^3 n^2))$.

Proof. See Section 12. □

The cubic dependence on T (rather than linear) reflects the fact that the sensitivity $\partial P_{ij}/\partial \alpha_0 = 2tP_{ij}$ grows with time: later snapshots are more informative because the perturbation has had longer to propagate. This is a general feature of dynamics estimation, in contrast to i.i.d. settings where information accumulates linearly.

The scaling $\Theta(T^3 \cdot n^2)$ holds only while the trajectory remains in the interior of the probability space. Linear dynamics produce exponential growth or decay of eigenvalues: $\lambda_\iota(t) = \lambda_\iota(0) e^{2\alpha_0 t}$. For $\alpha_0 > 0$, the probabilities P_{ij} grow toward 1, and the Fisher weights $P_{ij}/\left(1 - P_{ij}\right)$ diverge; but the model becomes invalid as probabilities saturate, so the growth of $\mathcal{I}$ eventually halts. For $\alpha_0 < 0$, probabilities decay toward 0, the sensitivities $\partial P_{ij}/\partial \alpha_0 = 2tP_{ij}(t)$ shrink exponentially, and the Fisher information per snapshot saturates. In both cases, the T^3 scaling describes a **short-time regime** before boundary effects dominate; the effective time horizon is $T_{\text{eff}} \sim 1/|\alpha_0|$.

The duality. The statistical-geometric correspondence is sharpest for general (non-polynomial) symmetric dynamics, where the full Lyapunov structure appears. The connection 1-form involves factors $1/\left(\lambda_\iota + \lambda_\gamma\right)$ controlling gauge sensitivity (Section 3.3); the vertical bracket norm in the sectional curvature involves the same factors, with $\left\| \left[\overline{\xi}_1, \overline{\xi}_2\right]^{\mathcal{V}} \right\|^2 \sim \sum_{\iota<\gamma} \lambda_\iota \lambda_\gamma / \left(\lambda_\iota + \lambda_\gamma\right) \cdot \left[\left(U^\top [M_1, M_2] U\right)_{\iota\gamma}\right]^2$ (Proposition 7), which diverges as $\lambda_d \to 0$ for fixed commutator; and the

Fisher information for estimating $M_{\iota\gamma}$ from the Lyapunov equation $\dot{P}_{\iota\gamma} = (\lambda_\iota + \lambda_\gamma) M_{\iota\gamma}$ involves the same factor $(\lambda_\iota + \lambda_\gamma)^2$ that appears in the curvature denominator.

An important qualification: this duality applies to **full operator recovery** (estimating the action of M on all eigenvector directions, including the weak subspace near λ_d). For **parsimonious parameterizations** (e.g., a single scalar parameter α_0 shared across all eigenvalue directions), estimation can proceed primarily from the dominant modes: the contribution of λ_1 alone may suffice to identify α_0, bypassing the ill-conditioned λ_d direction entirely. The geometric-statistical coupling bites when the parameter of interest specifically governs the weak subspace, or when the full operator M must be recovered.

With this qualification, the correspondence remains tight: for any component of the dynamics that depends on the spectral gap, the same δ controls both curvature and Fisher information. Networks near rank-deficiency ($\delta \to 0$, with λ_{d-1} also small) are simultaneously harder to align (curvature $\sim 1/(\lambda_{d-1} + \delta)$), harder to interpolate (injectivity radius $\sim \sqrt{\delta}$), and harder to estimate statistically (Fisher information for the weakest direction $\sim \delta^2$).

> *Remark.* For non-polynomial dynamics (e.g., Laplacian), the sensitivity propagation is more complex because the eigenvectors of P also evolve. The additional eigenvector sensitivity contributes positively to the Fisher information (there are more directions in which observations carry signal), but also introduces the holonomy obstruction: the extra information is entangled with gauge degrees of freedom that Proposition 9 shows cannot be resolved locally. A complete minimax theory for dynamics estimation in the presence of holonomy remains an important open problem.

> *Remark.* **Relationship to ASE estimation theory.** Proposition 10 is derived from the Bernoulli likelihood of the observed adjacency matrices and applies to *any* estimator of θ, not to spectral methods specifically. The bound is therefore conceptually distinct from the CLT for adjacency spectral embedding [4], [9], the two-to-infinity perturbation theory [11], and the efficiency results for spectral estimators of static network parameters [12], [13]. Those results characterize the accuracy of estimating *latent positions* X or *block probability matrices* B from a *single* graph (or a fixed collection of graphs sharing the same P). ASE is asymptotically unbiased with per-vertex error $O(1/\sqrt{n})$ [9], and the spectral estimator of SBM block probabilities achieves asymptotic efficiency [13]; for individual latent positions, a one-step procedure achieves local efficiency (matching the oracle MLE covariance, up to orthogonal transformation) [12]. The minimax rate for latent position estimation under Frobenius loss is $\Theta(d/n)$ per vertex [24], with two-to-infinity rates depending on the spectral gap [25].
>
> By contrast, Proposition 10 concerns estimation of the *dynamics parameters* θ from a *time series* of graphs, a setting for which no prior efficiency theory exists. The estimation target is qualitatively different: not the $O(nd)$ latent position coordinates, but the $O(1)$ parameters governing their temporal evolution. The eigenvalue-direction decomposition of the Fisher information in Proposition 10, with weights $\mathcal{W}_\iota(t)$ involving eigenvector loadings, has no counterpart in the static theory. Extending the existing minimax framework to parametric temporal models is an open problem that our Fisher information structure could help resolve.

The "any unbiased estimator" qualification deserves comment. The CRB is a bound on the Bernoulli likelihood, which is well-defined for any n, so the bound itself does not require asymptotic arguments. However, whether useful estimators of θ are unbiased at finite n is a separate question. Any estimator that first recovers P spectrally and then fits θ inherits the finite-sample bias of eigendecomposition: the nonlinearity of the spectral map introduces bias, but this is $o(1/\sqrt{n})$ [11] and does not affect the asymptotic bound. Near rank-deficiency ($\delta \to 0$), the situation is more severe: the ASE collapses the weak eigenvalue direction, and the resulting bias in $\hat{\lambda}_d$ can dominate the signal λ_d itself. In this regime, the **mean squared error** for parameters that depend on the λ_d direction is likely dominated by bias rather than variance, and the CRB (which bounds variance alone) understates the true difficulty. For finite-sample inference, the biased Cramér-Rao variant $\mathrm{Var}\left(\hat{\theta}\right) \geq (1 + b'(\theta))^2 / \mathcal{I}(\theta)$ provides the appropriate generalization, but characterizing the bias $b(\theta)$ as a function of δ remains open.

7 The Constructive Problem: Identifiability and Its Limits

The obstructions in the previous sections characterize what is and isn't observable in principle. We now turn to the constructive question: given that observable dynamics exist, can we *recover* them from spectral embeddings? We establish a theoretical identifiability result showing that dynamics structure can, in principle, resolve gauge ambiguity. We then discuss why this identifiability does not straightforwardly translate into a practical algorithm.

7.1 Structure-constrained alignment

The alignment problem is underdetermined: without additional information, many gauge choices produce plausible trajectories. A natural idea is to use **inductive bias about dynamics structure** to constrain which trajectories are admissible.

Suppose dynamics belong to a family $\mathcal{F}$ from Section 5.1. For instance, polynomial dynamics $\dot{X} = N(P)X$ with $N(P) = \sum_k \alpha_k P^k$. These families have two properties that suggest they could regularize alignment: they are **horizontal** (symmetric N ensures no gauge drift), and they have **restricted structure** that random gauge errors would violate.

If the true dynamics belong to $\mathcal{F}$, then correct gauges produce trajectories fittable by some $f \in \mathcal{F}$, while wrong gauges produce trajectories requiring dynamics outside $\mathcal{F}$. The dynamics family acts as regularization on the gauge choice.

This leads to a joint optimization formulation:

Definition 9 (Joint Alignment-Learning Problem). Given ASE embeddings $\left\{\hat{X}^{(t)}\right\}_{t=0}^{T}$ and dynamics family $\mathcal{F}$, find gauge corrections $\{Q_t \in O(d)\}$ and $f \in \mathcal{F}$ minimizing:

$$\mathcal{L}(\{Q_t\}, f) = \sum_{t=0}^{T-1} \|\hat{X}^{(t+1)} Q_{t+1} - \hat{X}^{(t)} Q_t - \delta t \cdot f\left(\hat{X}^{(t)} Q_t\right)\|_F^2 \tag{58}$$

The objective measures how well the aligned trajectory $\left\{\hat{X}^{(t)} Q_t\right\}$ is explained by dynamics f: if the gauges are correct and f captures the true dynamics, each step's displacement should match the predicted velocity.

7.2 The identifiability principle

The theoretical case for structure-constrained alignment rests on a clean separation between gauge artifacts and horizontal dynamics.

Theorem 3 (Gauge Velocity Contamination). Let $X(t)$ follow true dynamics $\dot{X} = NX$ with $N = N^\top$. Let $\tilde{X} = XS$ for time-varying gauge error $S(t) \in O(d)$. Then the apparent dynamics are:

$$\dot{\tilde{X}} = N\tilde{X} + \tilde{X}\Omega \tag{59}$$

where $\Omega = S^{-1}\dot{S} \in \mathfrak{so}(d)$ is skew-symmetric.

Moreover, $\dot{\tilde{X}} = \tilde{N}\tilde{X}$ for some symmetric $\tilde{N}$ if and only if $\Omega = 0$.

Proof. By the product rule: $\dot{\tilde{X}} = \dot{X}S + X\dot{S} = NXS + X\dot{S}$. Since $X = \tilde{X}S^{-1}$: $\dot{\tilde{X}} = N\tilde{X} + \tilde{X}S^{-1}\dot{S} = N\tilde{X} + \tilde{X}\Omega$.

For this to equal $\tilde{N}\tilde{X}$ with $\tilde{N}$ symmetric, we need $\tilde{X}^\top\dot{\tilde{X}}$ symmetric. But $\tilde{X}^\top\dot{\tilde{X}} = \tilde{X}^\top N\tilde{X} + \tilde{X}^\top\tilde{X}\Omega$. The first term is symmetric; the second is $\left(\tilde{X}^\top\tilde{X}\right)\Omega$. For positive definite $\tilde{X}^\top\tilde{X}$ and skew Ω, this is symmetric only if $\Omega = 0$. □

The mechanism is elegant: random ASE gauge errors $R^{(t)}$ introduce skew-symmetric contamination that **cannot be absorbed** by symmetric dynamics. Requiring symmetric dynamics implicitly selects gauges where the contamination vanishes.

The reader may notice an affinity with Proposition 2, which states that $\dot{X}$ is horizontal if and only if $X^\top\dot{X}$ is symmetric. The two results use the same algebraic fact (a positive definite matrix times a nonzero skew-symmetric matrix is never symmetric) but answer different questions. Proposition 2 characterizes directions at a single point: given a velocity $\dot{X}$, does it have a gauge component? Theorem 3 characterizes trajectories over time: given an entire trajectory $\tilde{X}(t)$ observed in the wrong gauge, can any symmetric dynamics explain it? The former is a pointwise decomposition; the latter is an identifiability statement about the dynamics-gauge coupling along a path.

For specific families, this yields concrete algorithms. For linear symmetric dynamics, one obtains an alternating scheme: fix gauges and solve a least-squares problem with symmetry constraint for the dynamics matrix M; fix M and solve orthogonal Procrustes for each gauge. For polynomial dynamics $\dot{X} = \left(\sum_k \alpha_k P^k\right)X$, the gauge invariance of $P = \hat{X}\hat{X}^\top$ simplifies the dynamics step to linear regression in the α_k coefficients. Both have closed-form updates per step.

7.3 The downstream pipeline: from trajectories to equations

Suppose, optimistically, that the trajectory recovery problem were solved, that is, some method (structure-constrained alignment, a future P-level estimator, or domain-specific prior information) produced gauge-consistent estimates $\tilde{X}^{(t)}$ of the latent positions, up to noise. When multiple independent network samples $A_1^{(t)}, ..., A_m^{(t)}$ are available at each time t, averaging $\bar{A}^{(t)} = \frac{1}{m}\sum_i A_i^{(t)}$ before embedding reduces per-entry variance by a factor of m and per-vertex ASE error by $\sqrt{m}$, potentially making the trajectory recovery feasible even when individual snapshots are noisy.

Given such a recovered trajectory, the remaining steps are well-established. **Universal Differential Equations** (UDEs) [26] provide a framework for learning dynamics that combine known mechanistic structure with neural network flexibility:

$$\dot{X} = g(f_{\text{known}}(X, \varphi), f_{\text{NN}}(X, \theta)) \tag{60}$$

For RDPG dynamics, the known structure comes from the families in Section 5.1. For example, the polynomial architecture $\dot{X} = \left(\sum_k \alpha_k(\theta, X) P^k\right) X$ ensures horizontality by construction while allowing the coefficients to be learned. Gradients flow through the ODE solver via adjoint sensitivity methods, and the gauge-consistent architecture constrains the hypothesis space.

Once a UDE is trained, **symbolic regression** [27] can extract interpretable closed-form equations. By sampling state-velocity pairs $\left(X, f_{\theta(X)}\right)$ from the trained model and searching for symbolic expressions of the form $N(P)X$, one can recover explicit dynamics equations completing the path from observed adjacency matrices to differential equations. A subtlety: symbolic regression on latent variables X is gauge-dependent. Unless the UDE output is pre-aligned to a canonical frame (e.g., the eigenbasis of P), the regression may fail to find sparse coefficients because X in the UDE gauge is a rotated version of the "natural" coordinates. Working with **gauge-invariant features** (eigenvalues of P, or the P-dynamics $\dot{P} = NP + PN$ directly) avoids this issue, and is natural for polynomial families where the dynamics separate by eigenvalue.

The bottleneck is step 2: obtaining $\tilde{X}^{(t)}$. The UDE and symbolic regression machinery is mature; the trajectory recovery problem is not.

7.4 Why the constructive problem remains open

Despite the clean identifiability theory, translating Theorem 3 into a reliable recovery method faces three interacting difficulties.

Finite-sample bias. Each spectral embedding $\hat{X}^{(t)}$ estimates the true positions with error $O_{p(1/\sqrt{n})}$. When the true dynamics are slow (small $\|X^{(t+1)} - X^{(t)}\|$), the signal-to-noise ratio for alignment degrades: we are trying to distinguish small true displacements from estimation noise of comparable magnitude. The alternating optimization inherits this: the dynamics step fits a combination of true signal and noise, while the gauge step aligns to a noisy target. The two sources of error reinforce rather than cancel.

Expressiveness of dynamics families. The identifiability argument requires that $\mathcal{F}$ be restrictive enough that wrong gauges produce trajectories outside $\mathcal{F}$. But in practice, even low-degree polynomial families can be surprisingly expressive. With $K + 1$ free parameters and T time steps, the polynomial family can fit trajectories contaminated by moderate gauge errors, absorbing the skew-symmetric artifact into the coefficients rather than rejecting it. The problem is sharpest for small d: when $d = 2$, the skew-symmetric space $\mathfrak{so}(2)$ is one-dimensional, and a single-parameter gauge drift can be partially absorbed by adjusting the dynamics coefficients. A compounding numerical difficulty is **multicollinearity** of the polynomial basis: the matrices $I, P, P^2, \ldots$ tend to align with the dominant eigenvector of P as the degree increases (by the power iteration phenomenon), so the regression problem for the coefficients α_k involves a Vandermonde-like matrix that is increasingly ill-conditioned for $K \geq 2$. The "gap" between true dynamics and gauge artifacts can be swallowed by this conditioning: small gauge drifts may be perfectly fitted by large, canceling

coefficients in higher-order terms. Making $\mathcal{F}$ more restrictive risks excluding the true dynamics; making it more expressive loses the regularization.

Holonomy and global consistency. Even if local alignment succeeds (each consecutive pair of frames is well-aligned), holonomy (Section 3.3) implies that gauge drift can accumulate over the trajectory. For dynamics tracing closed or nearly-closed loops in the base space $\mathcal{B}$, no sequence of local corrections can achieve global consistency: the horizontal lift of a closed curve need not close. This is not a statistical issue but a topological one, and no amount of data resolves it.

The fundamental tradeoff. The core difficulty is a tension between two requirements. The dynamics family must be restrictive enough to reject gauge-contaminated trajectories (identifiability), but expressive enough to capture the true dynamics (model adequacy). For the problem to be well-posed, the "gap" between what $\mathcal{F}$ can fit and what gauge contamination produces must exceed the noise level. Characterizing when this gap is sufficient (as a function of n, d, T, the spectral gap of P, and the dynamics family) is the central open problem.

> *Remark.* The difficulty is not that the identifiability principle fails theoretically. Theorem 3 is sharp: in the continuous-time, infinite-data limit, symmetric dynamics *do* identify gauges. The difficulty is that the finite-sample, discrete-time setting introduces errors that interact with the gauge-dynamics coupling in ways that the asymptotic theory does not capture. Progress likely requires either stronger structural assumptions (e.g., working directly on P-dynamics to bypass the gauge problem entirely) or fundamentally different estimation strategies (e.g., information-theoretic approaches that characterize the minimax rate for dynamics recovery).

7.5 Anchor-based alignment: a tractable special case

The difficulties above are genuine in the general case. However, there is a natural special case in which the gauge problem admits a clean solution: when a subset of nodes is known (or assumed) to be stationary.

The anchor principle. Suppose a subset $S \subset \{1, ..., n\}$ of nodes has $\dot{x}_i = 0$ for $i \in S$. These are "anchor" points whose latent positions do not move. At each time t, the ASE produces $\hat{X}^{(t)} = X^{(t)}R^{(t)} + E^{(t)}$ with random gauge $R^{(t)} \in O(d)$. For anchor nodes $i \in S$, the true positions satisfy $x_i^{(t)} = x_i^{(0)}$ for all t, so the anchor rows of the embedding are:

$$\hat{X}_S^{(t)} = X_S R^{(t)} + E_S^{(t)} \tag{61}$$

where $X_S \in \mathbb{R}^{|S| \times d}$ is constant. Aligning the anchor rows to a reference frame (say $t = 0$) via Procrustes gives:

$$\hat{Q}^{(t)} = \arg\min_{Q \in O(d)} \|\hat{X}_S^{(t)} Q - \hat{X}_S^{(0)}\|_F \tag{62}$$

This estimates $\left(R^{(t)}\right)^{-1} R^{(0)}$ directly, without any dynamics model or holonomy considerations. Applying $\hat{Q}^{(t)}$ to the full embedding $\hat{X}^{(t)}$ then aligns all nodes to the $t = 0$ gauge.

Why this works. The anchor nodes provide a fixed reference frame that "pins" the gauge at each time step. The alignment is **global** (all frames aligned to $t = 0$, not sequentially) so there is no error accumulation. The method requires no knowledge of the dynamics family $\mathcal{F}$: it works

for polynomial, Laplacian, or any other dynamics, as long as the anchors are truly stationary. Holonomy is irrelevant because we never attempt to propagate a gauge along the trajectory; each frame is independently aligned to the fixed reference.

Conditions and limitations. The anchor principle requires:

1. **Well-conditioned anchors:** the anchor position matrix $X_S \in \mathbb{R}^{|S| \times d}$ must have full column rank for the Procrustes problem to be determined, and must be well-conditioned ($\sigma_d(X_S) \gg 0$) for robustness to noise. Counting anchors alone is insufficient: $|S| \gg d$ anchors that are nearly collinear or clustered around a single point leave the rotation underdetermined in the orthogonal directions. The Procrustes rotation error scales as $O_p\left(\sigma_d(X_S)^{-1}/\sqrt{n}\right)$, so anchor nodes must robustly span $\mathbb{R}^d$ in their latent positions.
2. **Known anchor identity:** we must know which nodes are stationary. In practice this could come from domain knowledge (e.g., established species in an ecological network, institutional nodes in a social network) or from a preliminary analysis identifying nodes with low temporal variance.
3. **Approximately stationary anchors:** if anchor nodes drift slowly (with velocity $\|\dot{x}_i\| = \varepsilon$ for $i \in S$), the alignment inherits a bias of order εT over the trajectory. This bias is acceptable when $\varepsilon T \ll 1/\sqrt{n}$, i.e., when the anchor drift is small relative to the ASE noise. Since the bias grows linearly in T while the stochastic alignment error remains $O(1/\sqrt{n})$, this imposes a **short-time validity regime**: the method works for $T \ll 1/(\varepsilon\sqrt{n})$, beyond which the systematic bias from anchor drift dominates the noise floor.

The third condition explains a phenomenon we observed in preliminary experiments: networks with a large block of slowly-moving nodes could be aligned successfully by naive Procrustes, while networks where all nodes moved at comparable rates could not. The slow nodes were acting as *de facto* anchors, stabilizing the gauge without our explicitly recognizing it.

When anchors are realistic. Several application domains naturally feature nodes with heterogeneous dynamics rates. In ecological food webs, basal species (primary producers) often have stable trophic positions while higher-level consumers undergo rapid changes. In social networks, institutional actors (organizations, permanent positions) may persist while individuals fluctuate. In neural connectomes, structural hub regions may be stable on timescales over which peripheral connections rewire. More generally, any system with a separation of timescales — a slowly-evolving "backbone" and rapidly-evolving periphery — is a natural candidate.

The anchor approach does not solve the general gauge problem. It replaces a hard geometric-statistical problem with a domain knowledge requirement: identifying stationary nodes. But when such knowledge is available, it provides a clean and computationally trivial path to gauge-consistent trajectories, enabling the downstream UDE pipeline of Section 7.

7.6 Numerical illustration

We illustrate the theory with two controlled experiments on synthetic RDPG data. The first demonstrates anchor-based alignment on gauge-equivariant polynomial dynamics, where alignment quality can be assessed independently of dynamics recovery. The second demonstrates the full downstream pipeline: UDE training and symbolic regression; to do this more visibily, we move to a non-gauge-equivariant dynamics, where alignment quality directly impacts dynamics recovery.

7.6.1 Experiment 1: Anchor-based alignment

Setup. We generate an RDPG with $n = 200$ nodes in $d = 2$ latent dimensions. The initial positions $X(0)$ are drawn uniformly from $B_+^2 = \{x \in \mathbb{R}^2 : x_1, x_2 \geq 0, \|x\| \leq 1\}$, with a designated anchor set S of n_a nodes. The non-anchor nodes evolve under polynomial dynamics $\dot{X} = (\alpha_0 I + \alpha_1 P)X$ with $(\alpha_0, \alpha_1) = (-0.3, 0.003)$, while anchor nodes remain fixed: $\dot{x}_i = 0$ for $i \in S$. We integrate the ODE to produce a trajectory $X(t)$ at T equally-spaced times, generate $K = 3$ independent adjacency matrices $A_k^{(t)} \sim \text{Bernoulli}(X(t)X(t)^\top)$ at each time (averaging to reduce noise), compute ASE embeddings $\hat{X}^{(t)}$, and compare anchor-based alignment to sequential Procrustes.

Metrics. We measure alignment quality by the mean Procrustes error:

$$\text{err}(t) = \frac{1}{n} \|\hat{X}^{(t)}\hat{Q}^{(t)} - X^{(t)}Q^*\|_F \tag{63}$$

where Q^* is the best global alignment to the true trajectory (accounting for the residual gauge at $t = 0$). We sweep over four experimental conditions: anchor count n_a, trajectory length T, anchor drift rate ε, and initial embedding norm scale (controlling signal-to-noise ratio).

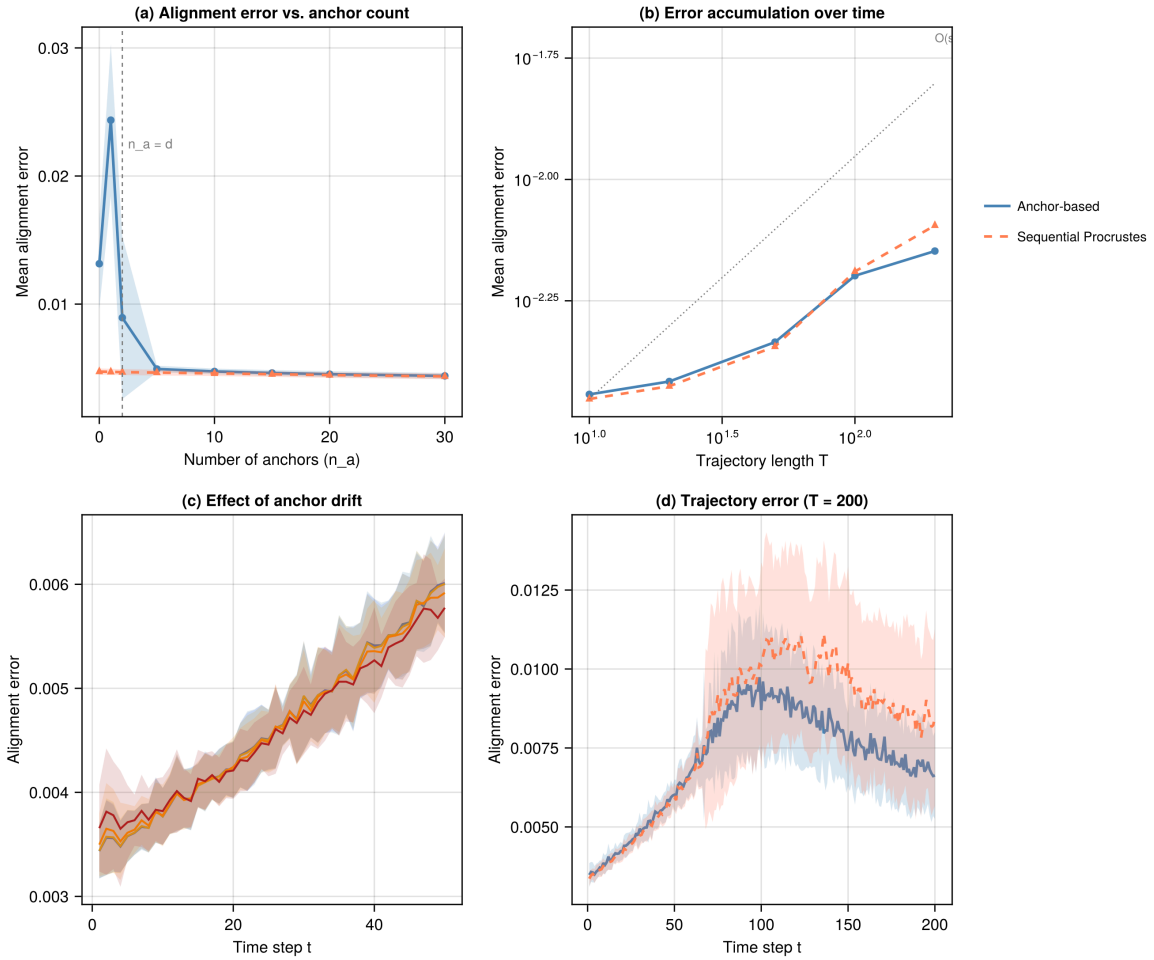

Figure 1: Anchor-based alignment experiment ($n = 200$, $d = 2$, polynomial dynamics $\dot{X} = (\alpha_0 I + \alpha_1 P)X$ with $K = 3$ Bernoulli samples per time step). **(a)** Alignment error vs. number of anchor nodes: below $n_a = d = 2$ (dashed vertical line) Procrustes is underdetermined and alignment fails; with sufficient anchors, error stabilizes at the ASE noise floor. **(b)** Error accumulation over trajectory length: anchor-based alignment (blue) remains bounded while sequential Procrustes (coral, dashed) grows with T, consistent with $O\left(\sqrt{T}\right)$ drift accumulation. **(c)** Effect of anchor drift rate ε: with larger ε, systematic bias from drifting anchors becomes visible. **(d)** Per-timestep alignment error for the $T = 200$ trajectory: the anchor-based error (blue) remains approximately flat, while sequential Procrustes (coral) accumulates drift, with the gap widening toward later time steps. Shaded bands show ± 1 standard deviation across 20 Monte Carlo repetitions.

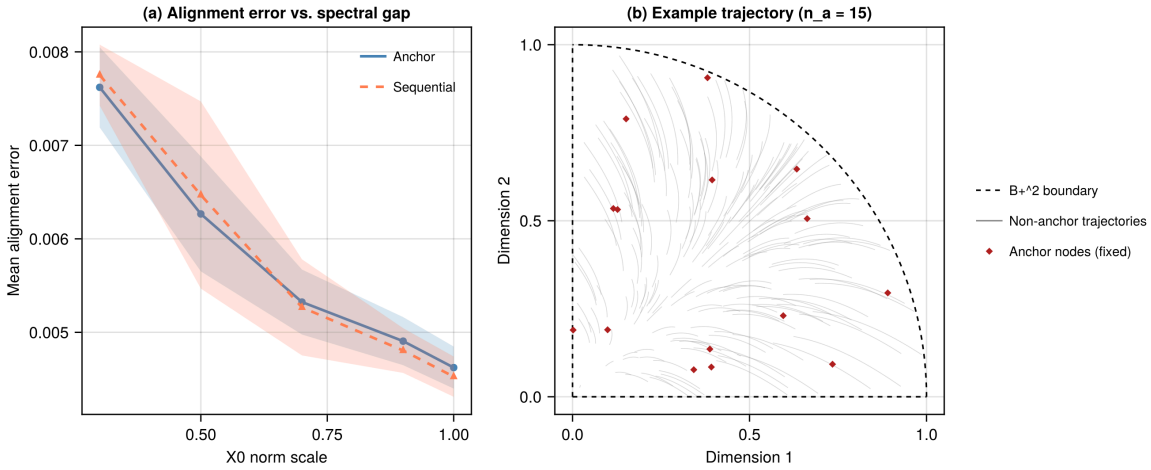

Figure 2: **(a)** Alignment error as a function of the initial embedding norm scale (proxy for signal strength): smaller norms yield weaker signals and less informative Bernoulli observations, increasing ASE noise and degrading alignment for both methods. Note that uniform scaling preserves the condition number σ_1/σ_2 (constant at ≈ 2.2); this experiment varies signal *magnitude* rather than the spectral gap ratio. **(b)** Phase portrait with $n_a = 15$ anchor nodes (red diamonds, stationary) and non-anchor nodes (gray trajectories evolving under polynomial dynamics) in B_+^2.

Results. The experiment confirms the theoretical predictions. Alignment error stabilizes at the ASE noise floor once $n_a \geq d$ (Figure 1a); below $n_a = d = 2$, Procrustes is underdetermined and fails catastrophically, consistent with the well-conditioned anchor requirement of Section 7. At short trajectory lengths ($T \leq 50$), sequential Procrustes is marginally better than anchor-based alignment (by $\approx 2\%$), because it uses all $n = 200$ nodes for each Procrustes step rather than only the $n_a = 15$ anchors. The crossover occurs around $T \approx 100$, after which the sequential method's accumulated drift dominates: at $T = 200$, sequential Procrustes is approximately 13% worse (Figure 1b, d), consistent with $O\left(\sqrt{T}\right)$ drift accumulation. A systematic bias from drifting anchors becomes visible for $\varepsilon \geq 0.05$ in alignment error, and for $\varepsilon \geq 0.1$ in the variability of the recovered dynamics parameters (standard deviation of $\hat{\alpha}_0$ roughly doubles); at $\varepsilon = 0.01$ and below, the effect is negligible (Figure 1c).

Dynamics recovery is gauge-free. A notable observation: for polynomial dynamics, the coefficients (α_0, α_1) can be estimated from the P-space trajectory via the gauge-invariant estimator, since $P = XX^\top$ is rotation-invariant. With $n_a = 15$, we recover $(\hat{\alpha}_0, \hat{\alpha}_1) = (-0.250 \pm 0.008, 0.0024 \pm 0.0001)$ against true values $(-0.3, 0.003)$; with $n_a = 0$ the estimates are $(-0.282 \pm 0.008, 0.0028 \pm 0.0001)$. The comparable precision (standard errors ≈ 0.008) regardless of anchor count confirms that the P-based estimator does not require alignment. The systematic shift in the point estimate (from -0.282 at $n_a = 0$ to -0.250 at $n_a = 15$) reflects the fact that more frozen anchor nodes change the effective trajectory of P, not a gauge effect. This motivates a natural question: when the dynamics *cannot* be recovered from P alone, does alignment quality directly impact dynamics recovery? The second experiment addresses this.

7.6.2 Experiment 2: UDE pipeline with non-gauge-equivariant dynamics

Motivation. The polynomial dynamics in Experiment 1 are gauge-equivariant: $P = XX^\top$ is rotation-invariant, so the dynamics parameters live in P-space and can be recovered without alignment. To demonstrate the full UDE pipeline and the necessity of anchor-based alignment for dynamics recovery, we design dynamics that depend on the coordinates of X directly, so that alignment quality is a bottleneck.

Setup. We generate an RDPG with $n = 200$ nodes in $d = 3$ latent dimensions, with $n_a = 100$ anchor nodes organized into 3 communities near the vertices of B_+^3 (at positions $(0.7, 0.2, 0.2)$, $(0.2, 0.7, 0.2)$, $(0.2, 0.2, 0.7)$ plus Gaussian noise). The 100 non-anchor nodes evolve under damped spiral dynamics around their community centroids μ_k:

$$\dot{x}_i = (-\gamma + \beta \, \|x_i - \mu_k\|^2)(x_i - \mu_k) + \omega J(x_i - \mu_k) \tag{64}$$

where $J = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}$ is the rotation generator around the $\frac{1,1,1}{\sqrt{3}}$ axis, and $(\gamma, \beta, \omega) = (0.3, -0.5, 1.0)$. These dynamics are *not* gauge-equivariant: the centroids μ_k and rotation axis live in X-space coordinates, so a misaligned trajectory $\tilde{X}(t)$ produces different offsets $\tilde{x}_i - \mu_k \neq x_i - \mu_k$, corrupting the dynamics structure.

We observe $K = 10$ independent adjacency matrices per time step over $T = 50$ steps at $\delta t = 0.1$, compute ASE embeddings, and apply three alignment conditions: anchor-based Procrustes, sequential Procrustes, and no alignment (raw ASE with random per-frame rotations).

UDE architecture. Following the framework of Section 7, we decompose the dynamics into known and unknown components:

$$\dot{x}_i = \underbrace{-\hat{\gamma}(x_i - \mu_k)}_{f_{\text{known}}} + \underbrace{f_\theta(x_i - \mu_k)}_{f_{\text{NN}}} \tag{65}$$

where $\hat{\gamma}$ is a learnable scalar parameter and $f_\theta : \mathbb{R}^3 \to \mathbb{R}^3$ is a small neural network (architecture: $3 \to 16 \to 16 \to 3$ with tanh activations, ≈ 390 parameters total). The known part encodes the structural assumption that non-anchor nodes are attracted toward their community centroids; the unknown part captures whatever additional dynamics exist. L2 regularization on the network weights ($\lambda = 10^{-3}$) discourages the network from absorbing the linear damping term, improving the identifiability of the known-unknown decomposition. The UDE is trained by solving the neural ODE forward from the initial condition and minimizing the trajectory MSE via adjoint sensitivity methods.

Primary metric: total dynamics MSE. Since the neural network component can absorb part of the linear damping, the additive UDE decomposition has an inherent identifiability issue. This produces a shifting of $\hat{\gamma}$ values away from the generative truth, but without changing the goodness of fit of the learned dynamics (indeed, this is essentially an algebraic sum zero movement). For this reason, we evaluate the *total* learned dynamics function $f_{\text{learned}(\delta)} = -\hat{\gamma}\delta + f_\theta(\delta)$ against the true dynamics $f_{\text{true}(\delta)} = (-\gamma + \beta \, \|\delta\|^2)\delta + \omega J\delta$, evaluated on random test inputs $\delta \in [-0.3, 0.3]^3$. This metric is invariant to the gamma–network split.

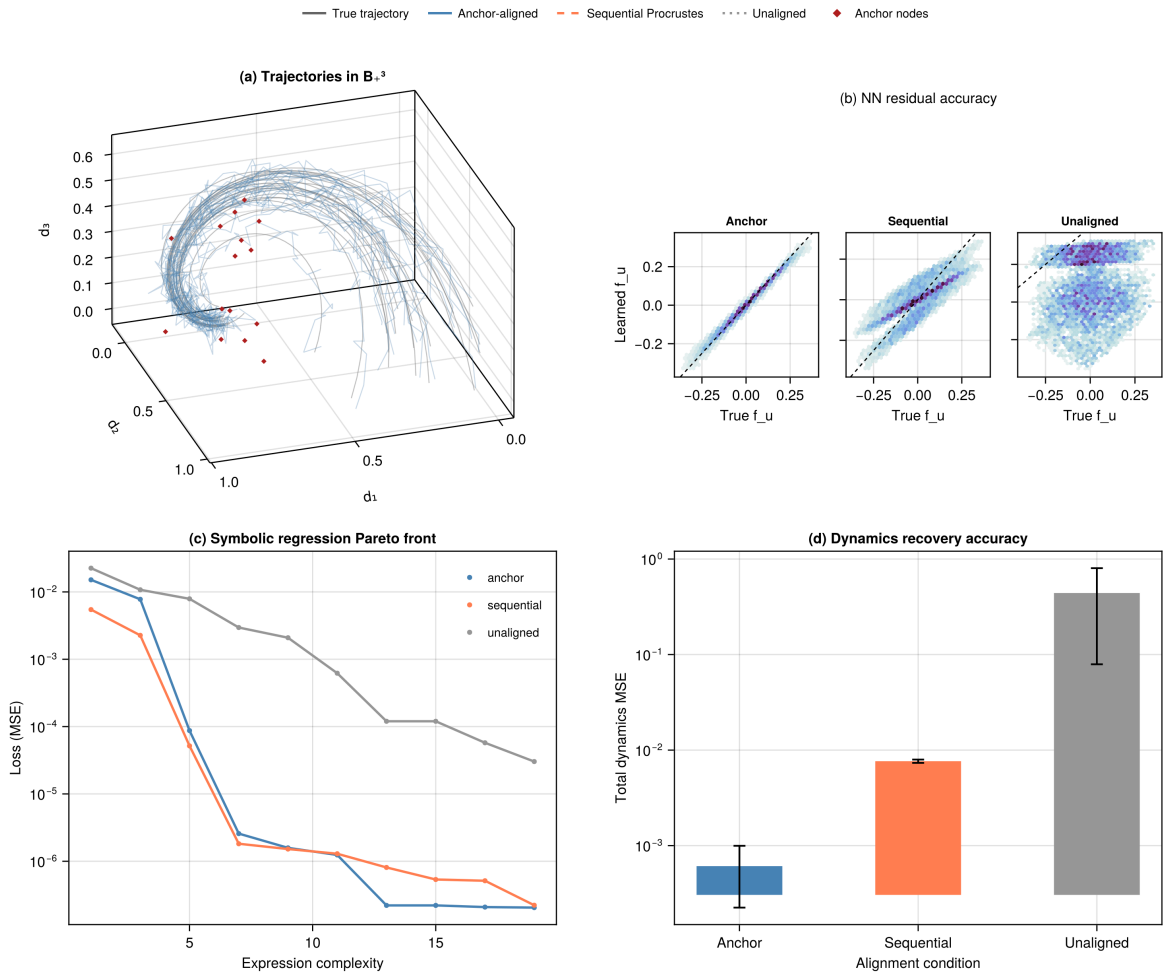

Figure 3: UDE pipeline experiment ($n = 200$, $d = 3$, damped spiral dynamics with rotation around $\frac{1,1,1}{\sqrt{3}}$, $K = 10$ samples). **(a)** True trajectories (gray) and anchor-aligned ASE (blue) in B_+^3; anchor nodes shown as red diamonds. **(b)** Learned NN residual $f_{\theta(\delta)}$ vs. true residual $f_{u(\delta)}$: anchor-aligned data (left) produces tight agreement along the diagonal; sequential Procrustes (center) and unaligned data (right) degrade progressively. **(c)** Symbolic regression Pareto fronts (loss vs. expression complexity): anchor-aligned data achieves 1–2 orders of magnitude lower loss at each complexity level. **(d)** Total dynamics MSE (log scale) by alignment condition: anchor alignment achieves MSE $\approx 6 \times 10^{-4}$, sequential Procrustes $\approx 8 \times 10^{-3}$ ($13\times$ worse), and unaligned ≈ 0.44 ($\approx 700\times$ worse). Error bars show ± 1 standard deviation across 5 repetitions.

Results. The total dynamics MSE cleanly separates the three alignment conditions (Figure 3d):

- **Anchor-aligned:** MSE $= 6 \times 10^{-4} \pm 4 \times 10^{-4}$: the UDE accurately recovers the complete dynamics function.
- **Sequential Procrustes:** MSE $= 7.6 \times 10^{-3} \pm 3 \times 10^{-4}$: approximately $13\times$ worse, reflecting systematic gauge drift that corrupts the coordinate-dependent dynamics.
- **Unaligned:** MSE $= 0.44 \pm 0.35$: approximately $700\times$ worse, with high variance; the network cannot learn coherent dynamics from incoherently-rotated frames.

The NN residual accuracy (Figure 3b) shows the mechanism: with anchor alignment, the learned f_θ closely matches the true nonlinear residual $\beta \|\delta\|^2 \delta + \omega J\delta$; without alignment, the network fits noise rather than structure.

Symbolic regression. Symbolic regression on the trained network's input-output pairs (Figure 3c) achieves Pareto-optimal expressions with losses 1–2 orders of magnitude lower under anchor alignment than under the other conditions at each complexity level. The anchor-aligned expressions contain the expected structural terms (products of coordinates and cross-terms consistent with the rotation generator J). As noted in Section 7, symbolic regression requires working with gauge-invariant features or a canonically aligned frame; the anchor alignment provides the latter.

The identifiability caveat. The learned damping parameter $\hat{\gamma}$ shows high variance across repetitions (0.26 ± 0.08 for anchor-aligned, vs.
true $\gamma = 0.3$), reflecting the fundamental identifiability issue in additive UDE decompositions: the network can absorb part of the linear term. This is a known limitation, and the total dynamics MSE confirms it is benign: the complete dynamics function $f_{\text{known}} + f_{\text{NN}}$ is well-recovered regardless of the $\hat{\gamma}$-network split.

Takeaway. When dynamics are gauge-equivariant ($\dot{X} = N(P)X$), alignment quality is irrelevant for parameter recovery (Experiment 1). When dynamics depend on X-space coordinates, alignment quality directly controls the fidelity of the learned dynamics (Experiment 2). Anchor-based alignment provides the gauge-consistent trajectories that the downstream UDE pipeline requires.

8 Discussion

What we achieved. We provided a rigorous geometric framework for understanding dynamics on RDPGs. The fiber bundle perspective (Section 3.3) formalizes gauge freedom via principal bundles, with explicit formulas for the connection 1-form, curvature, and holonomy. We established a sharp holonomy dichotomy (Section 5.3): polynomial dynamics have trivial holonomy through commuting generators, while Laplacian dynamics generically produce non-trivial holonomy (full SO(2) for $d = 2$), which constitutes the worst case for alignment. We derived Cramér-Rao lower bounds (Section 6) revealing that geometric and statistical difficulty are controlled by the same spectral gap, an inextricable duality. We cataloged concrete families of dynamics on RDPG latent positions, characterizing their horizontality, observability, and parameter complexity. We showed that existing methods, such as joint embeddings assuming the wrong generative model (Section 4.5), Bayesian smoothing approaches that interpolate rather than learn dynamics (Section 4.6), fail to address the fundamental obstructions. We established the identifiability principle (Theorem 3): symmetric dynamics cannot absorb skew-symmetric gauge contamination, providing a theoretical foundation for structure-constrained alignment. We identified the anchor-based alignment strategy (Section 7.5) as a tractable special case when stationary nodes are available, and demonstrated its effectiveness numerically.

What remains open. The gap between identifiability and practical recovery is the central challenge. Theorem 3 guarantees that, in the continuous-time limit with perfect data, dynamics structure resolves gauge ambiguity. But the discrete, finite-sample setting introduces errors that interact with the gauge-dynamics coupling in ways not captured by the asymptotic theory. The expressiveness-restrictiveness tradeoff for dynamics families, the finite-sample signal-to-noise ratio for alignment, and holonomy over long trajectories all remain unresolved.

Directions for progress. Several avenues seem promising. First, the P-dynamics approach (Section 5.2) bypasses the gauge problem in principle, but the estimation theory for recovering N from noisy Bernoulli samples $A^{(t)}$ remains undeveloped. The Lyapunov inverse amplifies noise at rate $1/(\lambda_\iota + \lambda_\gamma)$, suggesting that the minimax rate for estimating N from T adjacency matrices will depend critically on the spectral gap of P. Characterizing this rate, and understanding how it interacts with the choice of dynamics family (polynomial vs. general symmetric N), would clarify whether the P-level approach is practically viable or faces the same fundamental barriers as X-level methods.

Second, the holonomy characterization (Section 5.3) establishes that polynomial dynamics have trivial holonomy while Laplacian dynamics generically have full $\mathrm{SO}(d)$ holonomy, but several questions remain. Proving the full holonomy conjecture for $d \geq 3$ requires a transversality argument that the commutators $X^\top[L_1, L_2]X$ span $\mathfrak{so}(d)$ along generic trajectories. More practically, quantifying the **rate** at which holonomy accumulates along specific trajectories, and not just its generic nontriviality, would connect the abstract topological obstruction to concrete alignment error budgets.

Third, the Cramér-Rao bound of Proposition 10 establishes the Fisher information structure and its spectral gap dependence, but leaves open whether the implied rates are achievable. For polynomial dynamics, where holonomy is trivial and the P-dynamics are explicit, a matching upper bound via maximum likelihood on the P-trajectory seems within reach. For Laplacian dynamics, the entanglement of Fisher information with gauge degrees of freedom (Section 6) suggests that achievability may require estimators that account for holonomy, potentially through the P-level approach of Section 5.2.

Fourth, the geometry changes qualitatively in the sparse regime $P = \rho_n XX^\top$ with $\rho_n \to 0$: the injectivity radius shrinks as $\sqrt{\rho_n \lambda_d}$, curvature blows up, and ASE convergence rates deteriorate. The interplay between sparsity, curvature, and dynamics recovery in this regime is largely unexplored.

Broader impact. Learning interpretable dynamics from network data could enable mechanistic understanding in domains from neuroscience to ecology. However, the RDPG assumption, that connection probabilities arise from latent position dot products, is strong. Real networks may violate this assumption, and any learned dynamics should be validated against domain knowledge.

9 Conclusion

We investigated the problem of learning differential equations governing time-evolving Random Dot Product Graphs, developing a rigorous geometric framework based on principal fiber bundles.

We identified three fundamental obstructions: gauge freedom (formalized via the connection 1-form and characterized by Theorem 1), realizability constraints (the tangent space of $\mathcal{B}$), and recovering trajectories from spectral embeddings (complicated by holonomy and information-theoretic limits). We established a sharp dichotomy: polynomial dynamics have trivial holonomy and commuting generators (Proposition 8), making alignment a purely statistical problem, while Laplacian dynamics generically produce non-trivial holonomy (Proposition 9, with full $\mathrm{SO}(2)$ proved for $d = 2$), making alignment simultaneously a statistical and topological challenge. We derived Cramér-Rao lower bounds (Proposition 10) revealing a statistical-geometric duality: the

same spectral gap that controls curvature also controls Fisher information, so geometric and statistical difficulty are inextricable. We proved that existing joint embedding methods assume generative models incompatible with ODE dynamics, and that Bayesian smoothing approaches lack the dynamical consistency required for learning state-dependent velocity fields.

We established the identifiability principle that dynamics structure can resolve gauge ambiguity: symmetric dynamics cannot absorb the skew-symmetric contamination from wrong gauges (Theorem 3). However, we showed that this theoretical identifiability faces significant practical obstacles in the finite-sample, discrete-time setting. We identified anchor-based alignment as a tractable special case when stationary nodes are available, providing a practical entry point for the downstream dynamics-learning pipeline, and demonstrated its effectiveness and limitations numerically.

This work establishes mathematical foundations for the study of dynamics on temporal network data. The characterization of invisible dynamics, the horizontal lift formula, the holonomy dichotomy between polynomial and Laplacian families, and the statistical-geometric duality, provides both a geometrical language and the concrete results needed to formulate and attack the open problems that remain. Progress on these problems, whether through P-level dynamics, tighter information-theoretic bounds, or estimation strategies that account for holonomy, would open the door to mechanistic understanding of how and why networks evolve.

10 Acknowledgments

11 Data and Code Availability

Bibliography

[1] T. Poisot, D. B. Stouffer, and D. Gravel, "Beyond species: why ecological interaction networks vary through space and time," *Oikos*, vol. 124, no. 3, pp. 243–251, 2015, doi: 10.1111/oik.01719.

[2] G. V. Dalla Riva and D. B. Stouffer, "Exploring the evolutionary signature of food webs' backbones using functional traits," *Oikos*, vol. 125, no. 4, pp. 446–456, 2016, doi: 10.1111/oik.02305.

[3] T. Strydom *et al.*, "Food web reconstruction through phylogenetic transfer of low-rank network representation," *Methods in Ecology and Evolution*, vol. 13, no. 12, pp. 2838–2849, 2022, doi: 10.1111/2041-210X.13835.

[4] A. Athreya *et al.*, "Statistical inference on random dot product graphs: a survey," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 8393–8484, 2018.

[5] S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," in *Algorithms and Models for the Web-Graph: 5th International Workshop, WAW 2007*, 2007, pp. 138–149.

[6] I. Gallagher, A. Jones, and P. Rubin-Delanchy, "Spectral embedding for dynamic networks with stability guarantees," in *Advances in Neural Information Processing Systems*, 2021, pp. 10158–10170.

[7] K. Levin, A. Athreya, M. Tang, V. Lyzinski, Y. Park, and C. E. Priebe, "A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference," *arXiv preprint arXiv:1705.09355*, 2017.

[8] J. D. Loyal, "Bayesian Inference for Generalized Random Dot Product Graphs with Scaled Mixtures of Normals," *arXiv preprint arXiv:2509.19748*, 2025.

[9] A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman, "A limit theorem for scaled eigenvectors of random dot product graphs," *Sankhya A*, vol. 78, no. 1, pp. 1–18, 2016.

[10] P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe, "A statistical interpretation of spectral embedding: the generalised random dot product graph," *Journal of the Royal Statistical Society: Series B*, vol. 84, no. 4, pp. 1446–1473, 2022.

[11] J. Cape, M. Tang, and C. E. Priebe, "The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics," *The Annals of Statistics*, vol. 47, no. 5, pp. 2405–2439, 2019.

[12] F. Xie and Y. Xu, "Efficient estimation for random dot product graphs via a one-step procedure," *Journal of the American Statistical Association*, vol. 118, no. 541, pp. 651–664, 2023.

[13] M. Tang, J. Cape, and C. E. Priebe, "Asymptotically efficient estimators for stochastic blockmodels: the naive MLE, the rank-constrained MLE, and the spectral estimator," *Bernoulli*, vol. 28, no. 2, pp. 1049–1073, 2022.

[14] M. Tang, J. Cape, and C. E. Priebe, "The eigenvalues of stochastic blockmodel graphs," *arXiv preprint arXiv:1803.11551*, 2018.

[15] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. doi: 10.1515/9781400830244.

[16] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, "Low-rank optimization on the cone of positive semidefinite matrices," SIAM, 2010, pp. 2327–2351. doi: 10.1137/080731359.

[17] E. Massart and P.-A. Absil, "Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 41, no. 1, pp. 171–198, 2020, doi: 10.1137/18M1231389.

[18] E. Massart, J. M. Hendrickx, and P.-A. Absil, "Curvature of the manifold of fixed-rank positive-semidefinite matrices endowed with the Bures–Wasserstein metric," in *Geometric Science of Information (GSI 2019)*, in Lecture Notes in Computer Science, vol. 11712. Springer, 2019, pp. 739–748.

[19] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vol. 1. Wiley, 1963.

[20] B. O'Neill, "The fundamental equations of a submersion," *Michigan Mathematical Journal*, vol. 13, no. 4, pp. 459–469, 1966, doi: 10.1307/mmj/1028999604.

[21] J. Arroyo, A. Athreya, J. Cape, G. Chen, C. E. Priebe, and J. T. Vogelstein, "Inference for multiple heterogeneous networks with a common invariant subspace," *Journal of Machine Learning Research*, vol. 22, no. 142, pp. 1–49, 2021.

[22] M. Heinonen, C. Yildiz, H. Mannerström, J. Intosalmi, and H. Lähdesmäki, "Learning unknown ODE models with Gaussian processes," *International Conference on Machine Learning*, pp. 1959–1968, 2018.

[23] M. I. Freidlin and A. D. Wentzell, "Random perturbations of dynamical systems," *Grundlehren der mathematischen Wissenschaften*, vol. 260, 1998.

[24] F. Xie and Y. Xu, "Optimal Bayesian estimation for random dot product graphs," *Biometrika*, vol. 107, no. 4, pp. 875–889, 2020.

[25] J. Agterberg, Q. Yan, and K. Levin, "Minimax rates for latent position estimation in the generalized random dot product graph," *arXiv preprint arXiv:2307.01942*, 2023.

[26] C. Rackauckas *et al.*, "Universal differential equations for scientific machine learning," *arXiv preprint arXiv:2001.04385*, 2020.

[27] M. Cranmer, "SymbolicRegression.jl: Distributed High-Performance Symbolic Regression in Julia." [Online]. Available: https://astroautomata.com/SymbolicRegression.jl/dev/

12 Deferred Proofs

Proof of Proposition 5 (Measure Zero). The function class $\mathcal{F}$ is assumed to be a finite-dimensional family with p parameters (e.g., polynomial dynamics with $p = K + 1$ coefficients). For each $(f, X_0) \in \mathcal{F} \times \mathbb{R}^{n\times d}$, the Picard-Lindelöf theorem yields a unique solution $X(\cdot; f, X_0) \in C^{k([0,T],\mathbb{R}^{n\times d})}$. The map $\Phi : (f, X_0) \mapsto X(\cdot; f, X_0)$ is smooth (by smooth dependence on parameters and initial conditions), so $\mathcal{M}_{\mathcal{F}} = \mathrm{im}(\Phi)$ is the image of a smooth map from an m-dimensional manifold (with $m = p + nd$) into the infinite-dimensional path space $\mathcal{X} = C^{k([0,T],\mathbb{R}^{n\times d})}$. Choose $m + 1$ continuous linear functionals $\ell_1, ..., \ell_{m+1}$ on $\mathcal{X}$ (e.g., evaluations of specific coordinates at distinct times) and let $L = (\ell_1, ..., \ell_{m+1}) : \mathcal{X} \to \mathbb{R}^{m+1}$. The composition $L\circ\Phi : \mathbb{R}^m \to \mathbb{R}^{m+1}$ is a smooth map between finite-dimensional spaces, and by Sard's theorem its image has Lebesgue measure zero in $\mathbb{R}^{m+1}$. Since $L(\mathcal{M}_{\mathcal{F}}) \subseteq \mathrm{im}(L\circ\Phi)$, the set $L(\mathcal{M}_{\mathcal{F}})$ also has Lebesgue measure zero. The pushforward $L_*\mu$ is an $(m + 1)$-dimensional Gaussian (absolutely continuous with respect to Lebesgue measure), so $\mu(\mathcal{M}_{\mathcal{F}}) \le (L_*\mu)(L(\mathcal{M}_{\mathcal{F}})) = 0$. □

Proof of Corollary 3 (Linear Dynamics: Explicit Fisher Information). For linear dynamics $\dot{X} = \alpha_0 X$, the solution is $X(t) = e^{\alpha_0 t}X(0)$, giving $P(t) = e^{2\alpha_0 t}P(0)$. The sensitivity is $\partial P_{ij}(t)/\partial\alpha_0 = 2te^{2\alpha_0 t}P_{ij}(0) = 2tP_{ij}(t)$. Since there is a single scalar parameter, the Fisher information is:

$$\mathcal{I}(\alpha_0) = \sum_{t=1}^{T}\sum_{i<j}\left(\frac{\partial P_{ij}(t)}{\partial\alpha_0}\right)^2 \cdot \frac{1}{P_{ij}(t)\big(1 - P_{ij}(t)\big)} = 4\sum_{t=1}^{T} t^2 \sum_{i<j} P_{ij}\frac{t}{1 - P_{ij}(t)} \tag{66}$$

For the scaling claim: when $P_{ij}(t) \in [\varepsilon, 1 - \varepsilon]$ for some $\varepsilon > 0$, the inner sum $\sum_{i<j} P_{ij}/\big(1 - P_{ij}\big)$ is $\Theta\big(n^2\big)$ (there are $\binom{n}{2}$ terms, each bounded away from zero and infinity). The outer sum $\sum_{t=1}^{T} t^2 = T(T + 1)(2T + 1)/6 = \Theta\big(T^3\big)$. Therefore $\mathcal{I}(\alpha_0) = \Theta\big(T^3 \cdot n^2\big)$, giving a Cramér-Rao lower bound $\mathrm{Var}(\hat{\alpha}_0) \ge \mathcal{I}^{-1} = \Omega(1/(T^3n^2))$. □

13 Vertical Projection onto the Fiber

We derive the formula for projecting a tangent vector $Z \in T_X\mathbb{R}^{n\times d}_*$ onto the vertical subspace $\mathcal{V}_X = \{X\Omega : \Omega \in \mathfrak{so}(d)\}$.

The vertical component $Z^{\mathcal{V}} = X\Omega^*$ is defined by the orthogonality condition $Z - X\Omega^* \in \mathcal{H}_X$ (the remainder is horizontal). By Proposition 2, a vector W at X is horizontal if and only if $X^\top W$ is symmetric. Applying this to $W = Z - X\Omega^*$:

$$X^\top(Z - X\Omega^*) = X^\top Z - G\Omega^* \quad \text{must be symmetric} \tag{67}$$

where $G = X^\top X$ is positive definite. Decomposing $X^\top Z$ into symmetric and skew-symmetric parts, $X^\top Z = \mathrm{sym}\big(X^\top Z\big) + \mathrm{skew}\big(X^\top Z\big)$, the condition becomes:

$$\mathrm{skew}(G\Omega^*) = \mathrm{skew}\big(X^\top Z\big) \tag{68}$$

Since Ω^* is skew-symmetric and G is symmetric, the skew-symmetric part of $G\Omega^*$ is:

$$\mathrm{skew}(G\Omega^*) = \frac{1}{2}\Big(G\Omega^* - (G\Omega^*)^\top\Big) = \frac{1}{2}(G\Omega^* + \Omega^* G) \tag{69}$$

where the last step uses $\Omega^{*\top} = -\Omega^*$ and $G^\top = G$. Equating:

$$G\Omega^* + \Omega^* G = 2\ \mathrm{skew}(X^\top Z) \tag{70}$$

This is a Lyapunov equation in Ω^*. Since G is positive definite, all sums $\lambda_\iota + \lambda_\gamma > 0$, so the equation has a unique solution. In the eigenbasis of $G = \mathrm{diag}(\lambda_1, ..., \lambda_d)$, the solution is elementwise:

$$\Omega^*_{\iota\gamma} = \frac{2\mathrm{skew}(X^\top Z)_{\iota\gamma}}{\lambda_\iota + \lambda_\gamma} \tag{71}$$

For the application in Proposition 7, $Z = SX$ with S skew-symmetric. Then $X^\top Z = X^\top SX$, which is already skew-symmetric (since $(X^\top SX)^\top = X^\top S^\top X = -X^\top SX$), so $\mathrm{skew}(X^\top SX) = X^\top SX$ and the Lyapunov equation becomes $G\Omega^* + \Omega^* G = 2X^\top SX$.

14 Norm of the Vertical Bracket Component

We derive the explicit formula for $\|\left[\bar{\xi}_1, \bar{\xi}_2\right]^{\mathcal{V}}\|^2$ stated in Proposition 7.

The vertical component is $X\Omega^*$ with $\Omega^*_{\iota\gamma} = 2(X^\top SX)_{\iota\gamma}/(\lambda_\iota + \lambda_\gamma)$ where $S = -[M_1, M_2]$ (Section 13). Its squared norm in the ambient Euclidean metric is:

$$\|X\Omega^*\|^2 = \mathrm{tr}\Big((X\Omega^*)^\top X\Omega^*\Big) = \mathrm{tr}(\Omega^{*\top} G\Omega^*) \tag{72}$$

Working in the eigenbasis of $G = \mathrm{diag}(\lambda_1, ..., \lambda_d)$:

$$\mathrm{tr}(\Omega^{*\top} G\Omega^*) = \sum_{\iota,\gamma} \lambda_\gamma \big(\Omega^*_{\iota\gamma}\big)^2 \tag{73}$$

(using $(\mathrm{tr}(A^\top BA))$ in coordinates with diagonal B; diagonal terms vanish since $\Omega^*_{\iota\iota} = 0$). Collecting the pair (ι, γ) and (γ, ι) for $\iota < \gamma$ and using skew-symmetry $\big(\Omega^*_{\gamma\iota}\big)^2 = \big(\Omega^*_{\iota\gamma}\big)^2$:

$$= \sum_{\iota<\gamma} (\lambda_\iota + \lambda_\gamma)\big(\Omega^*_{\iota\gamma}\big)^2 \tag{74}$$

Substituting $\Omega^*_{\iota\gamma} = 2(X^\top SX)_{\iota\gamma}/(\lambda_\iota + \lambda_\gamma)$:

$$= \sum_{\iota<\gamma} (\lambda_\iota + \lambda_\gamma) \cdot \frac{4(X^\top SX)^2_{\iota\gamma}}{(\lambda_\iota + \lambda_\gamma)^2} = 4\sum_{\iota<\gamma} \frac{(X^\top SX)^2_{\iota\gamma}}{\lambda_\iota + \lambda_\gamma} \tag{75}$$

It remains to express $(X^\top SX)_{\iota\gamma}$ in terms of $(U^\top [M_1, M_2]U)_{\iota\gamma}$. Writing $X = U\Lambda^{1/2}$ in the canonical gauge (where U has orthonormal columns spanning $\mathrm{col}(X)$ and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_d)$):

$$X^\top SX = \Lambda^{1/2} U^\top SU \Lambda^{1/2} \tag{76}$$

so $(X^\top SX)_{\iota\gamma} = \sqrt{\lambda_\iota}(U^\top SU)_{\iota\gamma}\sqrt{\lambda_\gamma}$. Since $S = -[M_1, M_2]$:

$$(X^\top SX)^2_{\iota\gamma} = \lambda_\iota \lambda_\gamma (U^\top [M_1, M_2]U)^2_{\iota\gamma} \tag{77}$$

Substituting:

$$\|X\Omega^*\|^2 = 4\sum_{\iota<\gamma} \frac{\lambda_\iota \lambda_\gamma}{\lambda_\iota + \lambda_\gamma} \Big[(U^\top [M_1, M_2]U)_{\iota\gamma}\Big]^2 \tag{78}$$

This is the formula stated in Proposition 7. The $1/\big(\lambda_\iota + \lambda_\gamma\big)$ weighting shows that curvature is amplified when both eigenvalues in a pair are small, matching the connection coefficient structure of the fiber bundle.

15 Extension to Directed Graphs

For directed graphs, each node has source position g_i and target position r_i. The probability matrix is $P = GR^\top$ (not symmetric).

The gauge group becomes $(G, R) \sim (GQ, RQ)$ for $Q \in O(d)$, and the invisible dynamics are $\dot{G} = GA$, $\dot{R} = RA$ with $A \in \mathfrak{so}(d)$.

The gauge-consistent architecture generalizes to:

$$\dot{G} = N_G(P)G, \quad \dot{R} = N_R(P)R \tag{79}$$

with appropriate symmetry constraints on (N_G, N_R) ensuring that the induced dynamics on $P = GR^\top$ are well-defined. The fiber bundle framework carries over: the total space is pairs $(G, R) \in \mathbb{R}_*^{n\times d} \times \mathbb{R}_*^{n\times d}$, the structure group is $O(d)$ acting diagonally, and the connection, curvature, and holonomy theory applies with the same Lyapunov-type structures, though the asymmetry of P introduces additional subtleties in the spectral decomposition.