Decentralized Coupled Representation Learning

Zilin Li* **Weiwei Xu** **Xuchun Tong**
Xuanbo Lu **Xuanqi Zhao** **Ge Zhang†**
School of Information and Intelligent Science, Donghua University, Shanghai
{tzulamlee,yumuzaijie}@gmail.com
{brucetongxc,Luethris2023,zxq15091779116}@outlook.com
gezhang@dhu.edu.cn
Lixeeone/Decentralized-Coupled-Representation-Learning

Abstract

The learning dynamics of decentralized coupled representation learning can be formulated as a non-autonomous stochastic dynamical system, yet the macroscopic behavior of the resulting system on continuous manifolds remains poorly understood. We establish three theoretical results for such coupled slow-fast dynamics. First, under a standard random geometric graph scaling regime, we prove that the infinitesimal generator of the discrete Markov process converges to the generator of an overdamped Langevin diffusion for every $f \in C^3(\mathcal{M})$. Second, under timescale separation, the stochastic parameter dynamics converge to a deterministic averaged flow that admits a row-orthogonality Lyapunov function, controlling row-space degeneracy and ruling out parameter divergence. Third, under a spectral gap condition, the row space of the averaged parameter trajectory converges to the principal eigenspace of the induced covariance matrix. Together, these results give a stochastic-approximation limit theory for the coupled regime in which Markovian sampling and local subspace learning co-evolve, showing that memoryless local interactions can induce a stable and structured macroscopic limit.

1 Introduction

A central objective in unsupervised representation learning is to discover low-dimensional structure in data supported on continuous, non-linear manifolds [5, 6]. Standard approaches often rely on centralized objectives and synchronous gradient coordination [62, 90]. In contrast, decentralized learning procedures update local states or parameters using only neighborhood-level interactions and local observations [12, 43]. Such procedures are attractive when observations and updates are inherently local, making centralized aggregation or synchronous gradient coordination impractical. Their analysis is substantially harder because the sampling process and the parameter dynamics are no longer independent.

This locality makes the data stream endogenous, so the macroscopic convergence of localized swarm dynamics on data manifolds remains poorly understood. Existing analyses of local representation learning, such as Oja's rule [76, 77], typically assume stationary Euclidean inputs [10, 14]. When autonomous agents explore a Riemannian manifold through randomized local transitions while their representation parameters are updated concurrently, the resulting process becomes a coupled, non-autonomous stochastic dynamical system. The central question is: *Can such a memoryless decentralized system converge to a stable low-dimensional representation without parameter divergence or unstable oscillation?*

*First author.
†Corresponding author.

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

We model decentralized representation learning as a coupled slow-fast system on $\mathcal{M} \times \mathbb{R}^{m \times n}$ [4, 6, 56, 80]. The system consists of two interacting streams: the fast kinematics of the single-agent spatial state X_t, which explores the data manifold, and the slow plasticity of the representation matrix $\mathbf{W}_t$, which learns the induced principal subspace. We reserve $\mathbf{X}_t$ for the finite-population configuration matrix used later in the coupled comparison skeleton. Throughout the paper, microscopic events are tracked in continuous time t, while the parameter dynamics are analyzed on the macroscopic timescale $\tau = \epsilon t$. Classical Oja-type analyses give rigorous convergence theory for stationary, streaming, or externally supplied data streams [8, 65, 76, 77, 91]; here the stream is generated by Markovian spatial exploration whose states feed the local updates, so the object of analysis is the coupled slow-fast system rather than Oja dynamics alone. The conceptual chain is local graph substrate $\rightarrow$ continuous exploration $\rightarrow$ correlated observations $\rightarrow$ principal-subspace learning: Section 2 establishes the graph-to-diffusion limit, while Section 3 analyzes the resulting correlated observations through stochastic averaging, stability, and principal-subspace selection.

To our knowledge, this is the first stochastic-approximation limit theory for decentralized representation learning that treats Markovian spatial exploration and Oja-type plasticity as a single coupled system. We isolate three mathematical questions and address them through separate convergence and stability results:

- **Discrete-to-continuous limit.** We prove that, under the random geometric graph scaling in Eq. 1, memoryless transitions on random geometric graphs [82] converge at the generator level to an overdamped Langevin diffusion on the manifold [28, 45, 48].
- **Averaged stability of the parameter flow.** Under timescale separation, we show that the stochastic Oja-type parameter dynamics converge to an averaged deterministic flow. By applying stochastic averaging to filter fast Markovian fluctuations [10, 55, 56], we obtain a Lyapunov stability mechanism that rules out divergent parameter trajectories.
- **Limit-set characterization.** Under a spectral gap condition for the induced covariance matrix [44], we characterize the invariant set of the averaged flow [55, 70] and show that the row space of the representation matrix aligns with the principal eigenspace. This yields a decorrelated stationary representation without requiring centralized gradient coordination.

2 Problem Formulation: From Discrete Swarms to Continuous Kinematics

We first formalize the agents' memoryless spatial exploration as a random walk on a geometric graph and identify its continuous diffusion limit on a Riemannian manifold [28, 45, 82].

2.1 Geometric Substrate and Graph Scaling

Let $(\mathcal{M}, g)$ be a compact, connected, boundaryless d-dimensional Riemannian manifold [63, 83]. We use the uniform probability measure $d\mu(x) = dV_g(x)/\mathrm{Vol}_g(\mathcal{M})$ as the reference geometric measure [49, 83]; non-uniform exploration is introduced through the potential-modulated transition kernel below.

We model the static spatial substrate as a sequence of random geometric graphs [19, 38, 82]. Let $\mathcal{X}_N = \{z_i\}_{i=1}^N$ be i.i.d. substrate vertices sampled from μ, with empirical measure $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{z_i}$. Two substrate vertices z_i, z_j are connected whenever $0 < d_{\mathcal{M}}(z_i, z_j) \leq \varepsilon_N$.

Agents are Markovian walkers moving on this fixed substrate graph; the vertices z_i are spatial substrate locations rather than agents themselves. To obtain a non-degenerate continuum limit, the connectivity radius satisfies

$$\lim_{N\to\infty} \frac{N\varepsilon_N^{d+2}}{\log N} = \infty. \tag{1}$$

This scaling balances geometric locality with empirical concentration and matches classical graph Laplacian convergence regimes [41, 93, 98].

2.2 Localized Swarm Dynamics and the Langevin Limit

Given an agent currently located at a substrate vertex $x \in \mathcal{X}_N$, let P_N denote its one-step local Markov transition kernel on the graph [52, 64]. For the uniform-in-space convergence statement

below, we extend the empirical neighborhood to arbitrary $x \in \mathcal{M}$ by $\mathcal{N}_N(x) = \{y \in \mathcal{X}_N : 0 < d_{\mathcal{M}}(x, y) \leq \varepsilon_N\}$. When $x \in \mathcal{X}_N$, this is the finite graph neighborhood used by the agents; for general $x \in \mathcal{M}$, it defines the empirical graph operator appearing in the supremum over $\mathcal{M}$.

Given a smooth potential $U \in C^3(\mathcal{M}, \mathbb{R})$, the local Gibbs weights are

$$P_N(x, y) = \frac{1}{Z_N(x)} \exp\left(-\frac{\beta}{2}[U(y) - U(x)]\right), \qquad y \in \mathcal{N}_N(x), \tag{2}$$

where $\beta > 0$ controls the potential bias and $Z_N(x)$ normalizes the local kernel [72, 87]. We analyze the associated continuous-time jump process obtained by running this local kernel at rate $\tau_N = 2D(d+2)/\varepsilon_N^2$. Its scaled graph generator is

$$\mathcal{L}_N f(x) = \tau_N \sum_y P_N(x, y)\big[f(y) - f(x)\big]. \tag{3}$$

Since $\mathcal{M}$ is compact and U is smooth, the Gibbs weights are uniformly bounded above and below [89]; hence the potential tilt does not change the local sample-complexity order required by Eq. 1. This is stated as a generator convergence result for Markov processes [28, 80].

Theorem 1 (Continuous Limit of the Localized Graph Flow) *For every fixed $f \in C^3(\mathcal{M})$, under the scaling regime in Eq. 1, as $N \to \infty$ and $\varepsilon_N \to 0$, the scaled graph generator $\mathcal{L}_N f$ converges uniformly in probability to the continuous generator $\mathcal{L}f$:*

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{x\in\mathcal{M}} |\mathcal{L}_N f(x) - \mathcal{L}f(x)| > \delta\right) = 0, \quad \forall \delta > 0, \tag{4}$$

where

$$\mathcal{L}f(x) = -D\beta\langle \nabla U(x), \nabla f(x)\rangle_g + D\Delta_{\mathcal{M}} f(x). \tag{5}$$

Proof Outline (Full proof in Appendix A.5). We expand the potential-tilted graph transition kernel in $T_x\mathcal{M}$ via $\exp_x^{-1}$ [36, 63, 83]. The first-order tilt yields the drift term, the second-order isotropic component yields $\Delta_{\mathcal{M}}$, and compactness controls curvature remainders uniformly [18, 36, 41]. Eq. 1 supplies the empirical concentration required for uniform convergence in probability [11, 93, 96, 98].

Implication for Fast Kinematics. Theorem 1 identifies the limiting generator associated with the overdamped Langevin SDE [33, 45, 78, 80]

$$dX_t = -D\beta\nabla U(X_t)dt + \sqrt{2D}\, dB_t^{\mathcal{M}}. \tag{6}$$

This diffusion supplies the fast spatial process used in the representation dynamics. Figure 1 provides numerical diagnostics for the graph-to-diffusion limit, including local operator bias–variance behavior, the dependence on the scaling statistic in Eq. 1, representative Langevin sample paths, and generator-error decay under satisfied versus violated scaling regimes.

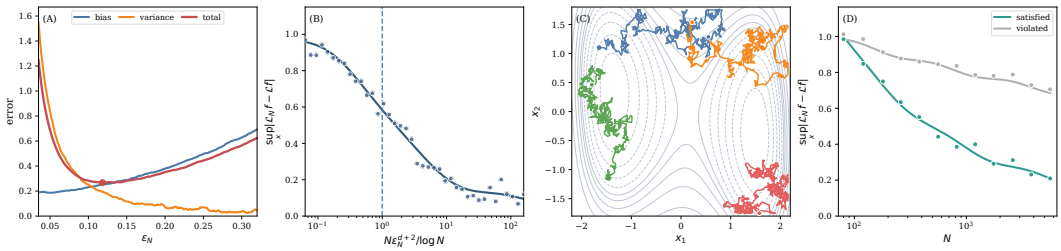

Figure 1: **Numerical Diagnostics for the Graph-to-Diffusion Limit.** Subfigures **(A)–(D)** illustrate the finite-sample behavior underlying Theorem 1. **(A)** Bias–variance tradeoff for local graph operators as a function of the neighborhood radius. **(B)** Operator error as a function of the scaling statistic $N\varepsilon_N^{d+2}/\log N$. **(C)** Representative sample paths of the limiting Langevin diffusion on a potential landscape. **(D)** Generator error versus population size under satisfied and violated scaling regimes.

3 Representation Dynamics: Stability and Principal Subspace Alignment

While Section 2 establishes the continuous kinematics of spatial exploration, representation learning is driven by the interaction between the agent's spatial state and its local representation parameters [14, 10, 56]. In a decentralized swarm, the local representation matrix $\mathbf{W}_t \in \mathbb{R}^{m\times n}$ updates from instantaneous observations generated by the fast spatial process.

To construct a stability proof, we formalize the mapping between the Riemannian manifold $\mathcal{M}$ and the Euclidean feature space. Let $\Phi : \mathcal{M} \hookrightarrow \mathbb{R}^n$ be a smooth isometric embedding [73]. For an agent at $X_t \in \mathcal{M}$, the local observation is $\mathbf{x}_t = \Phi(X_t) \in \mathbb{R}^n$, and the linear latent projection is $\mathbf{y}_t = \mathbf{W}_t\mathbf{x}_t \in \mathbb{R}^m$. By lifting a subspace Oja-type rule with lateral inhibition [76, 77, 81, 91] into the continuous domain with a slow timescale parameter $\epsilon > 0$, the joint system on $\mathcal{M} \times \mathbb{R}^{m\times n}$ is governed by:

$$\text{Fast Kinematics:} \quad dX_t = -D\beta\nabla U(X_t)dt + \sqrt{2D}\, dB_t^{\mathcal{M}}, \tag{7}$$

$$\text{Slow Plasticity:} \quad d\mathbf{W}_t = \epsilon\gamma \left(\mathbf{y}_t\mathbf{x}_t^T - \mathbf{y}_t\mathbf{y}_t^T\mathbf{W}_t\right) dt, \tag{8}$$

where $\gamma > 0$ is the plasticity rate. The inhibitory term $\mathbf{y}_t\mathbf{y}_t^T\mathbf{W}_t$ prevents row-space collapse and is the continuous subspace analogue of lateral competition among latent coordinates [81, 91]. Since Eq. 8 is driven by observations generated from the Markovian spatial process X_t, the analytical objective is to extract the deterministic macroscopic limit governing the asymptotic representation geometry [56, 59].

3.1 Macroscopic Limit and Lyapunov Dissipativity

Exploiting timescale separation ($\epsilon \ll 1$), the fast spatial process equilibrates to its invariant measure π on the timescale relevant to the slow parameter dynamics [56, 80]. The embedding Φ induces a pushforward measure $\Phi_{\#}\pi$ on $\mathbb{R}^n$ [97], and we define the induced spatial covariance matrix [4, 44]

$$\Sigma_\pi \equiv \int_{\mathcal{M}} \Phi(x)\Phi(x)^T d\pi(x). \tag{9}$$

To control row-space degeneracy and characterize the subspace normalization induced by lateral inhibition, we use the orthogonality Lyapunov functional [2, 27, 54]

$$V_{\text{orth}}(\mathbf{W}) = \frac{1}{4}\left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F^2. \tag{10}$$

We assume that the fast Langevin process on the compact manifold is exponentially mixing, for example through a spectral gap on the centered observables used below, and that the matrix drift $H_\gamma(\mathbf{W}, x)$ is locally Lipschitz in $\mathbf{W}$ and smooth in x on the compact sublevel sets considered. These standard conditions ensure solvability of the Poisson equation and justify the finite-horizon stochastic averaging limit.

Theorem 2 (Global Dissipativity of the Averaged Subspace Flow) *For any finite macroscopic horizon $T > 0$, define $\mathbf{W}^\epsilon(\tau) = \mathbf{W}_{\tau/\epsilon}$. As $\epsilon \to 0$, $\mathbf{W}^\epsilon$ converges uniformly in probability on $[0, T]$ to the limit trajectory $\bar{\mathbf{W}}$, which solves the averaged ODE [10]:*

$$\frac{d\bar{\mathbf{W}}(\tau)}{d\tau} = \gamma\left(\bar{\mathbf{W}}\Sigma_\pi - \bar{\mathbf{W}}\Sigma_\pi\bar{\mathbf{W}}^T\bar{\mathbf{W}}\right). \tag{11}$$

Along this averaged vector field, the orthogonality Lyapunov functional is non-increasing [54]:

$$\frac{d}{d\tau}V_{\text{orth}}(\bar{\mathbf{W}}(\tau)) = -\gamma\,\text{Tr}\left[\left(\bar{\mathbf{W}}(\tau)\bar{\mathbf{W}}(\tau)^T - \mathbf{I}_m\right)^2\bar{\mathbf{W}}(\tau)\Sigma_\pi\bar{\mathbf{W}}(\tau)^T\right] \le 0. \tag{12}$$

3.2 Principal Subspace Projection via LaSalle Invariance of the Averaged Flow

The non-increasing V_{orth} confines the trajectory to a precompact sublevel set, enabling a LaSalle-type stationary-set characterization [55, 60]. We assume Σ_π is positive on reached full-rank row spaces, $\mathbf{W}\Sigma_\pi\mathbf{W}^T \succ 0$ [44]. On the row-orthonormal component $\mathbf{W}\mathbf{W}^T = \mathbf{I}_m$, the condition $\text{Row}(\mathbf{W}) = \text{Col}(\mathbf{Q}_m)$ is equivalently written as $\mathbf{W} = \mathbf{R}\mathbf{Q}_m^T$ for some $\mathbf{R} \in O(m)$.

Theorem 3 (Principal Eigenspace Projection) *Let $\mathcal{I}$ be the maximal invariant set of the averaged flow satisfying $\dot{V}_{\mathrm{orth}}(\mathbf{W}) = 0$. Assume that Σ_π has a spectral gap $\lambda_m > \lambda_{m+1}$ with $\lambda_m > 0$, that $\mathbf{W}_0$ has full row rank, and that its projection onto the top-m eigenspace is nondegenerate $\mathrm{rank}(\mathbf{W}_0\mathbf{Q}_m) = m$, where $\mathbf{Q}_m \in \mathbb{R}^{n\times m}$ denotes the matrix of top-m eigenvectors of Σ_π. Then the averaged subspace Oja flow approaches the row-orthonormal invariant component $\mathbf{W}\mathbf{W}^T = \mathbf{I}_m$, and the row space of $\bar{\mathbf{W}}(\tau)$ converges to the m-dimensional principal eigenspace of Σ_π, up to an orthogonal rotation of the latent coordinates [77, 91].*

Proof Outline (Full proof in Appendix B). The transition from Markovian spatial transport to principal subspace alignment proceeds through two steps.

1. Time-Scale Separation via Poisson Resolvents (Appendix B.2): The fast spatial process continuously perturbs the slow parameter update through $\mathbf{x}_t = \Phi(X_t)$. Define the local drift $H_\gamma(\mathbf{W}, x) = \gamma\left[(\mathbf{W}\Phi(x))\Phi(x)^T - (\mathbf{W}\Phi(x))(\mathbf{W}\Phi(x))^T\mathbf{W}\right]$, and its averaged counterpart $\bar{H}_\gamma(\mathbf{W}) = \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right)$. The centered fluctuation is controlled by a matrix-valued Poisson corrector [79, 80] solving $\mathcal{L}\nu(\mathbf{W}, x) = -\left(H_\gamma(\mathbf{W}, x) - \bar{H}_\gamma(\mathbf{W})\right)$, where $\mathcal{L}$ is the generator of the fast spatial process [79]. Applying Itô's formula [78] converts the centered Markovian fluctuation into remainder terms that vanish over the macroscopic timescale $\tau = \epsilon t$ [10, 56].

2. Subspace Alignment via LaSalle Invariance of the Averaged Flow (Appendices B.3 & B.4): The orthogonality Lyapunov functional in Eq. 10 controls row-space collapse and identifies the Stiefel-type invariant set [2, 27]. Projecting the averaged subspace Oja flow onto the eigenbasis of Σ_π [44, 77] shows that, under the spectral gap $\lambda_m > \lambda_{m+1}$, components outside the principal eigenspace decay relative to the top-m components and vanish asymptotically. Thus the limiting row space aligns with the principal eigenspace rather than a lower-dimensional collapsed direction [76, 77, 91].

Implication for Representation Stability: Theorem 3 serves as the final invariant-set step of the slow-fast limit: the covariance driving the subspace flow is induced by Markovian geometric exploration, rather than supplied as a fixed stationary stream. In the averaged limit, the subspace Oja dynamics control row-space degeneracy through V_{orth} and characterize the principal spectral structure selected by the flow. The numerical consequences of this result, including Lyapunov decay, row-orthogonality, and subspace-angle convergence, are evaluated in Section 5.

4 The Coupled Flow and Feature Decorrelation

The preceding analyses isolated the spatial kinematics (Section 2) and representation dynamics (Section 3). Decentralized representation learning, however, is coupled: the spatial distribution determines the covariance seen by the parameters, while the representation changes the exploration landscape. We therefore lift the dynamics to $\mathcal{M}^N \times \mathbb{R}^{m\times n}$ [1, 2] and characterize the decorrelated stationary latent structure (Appendix C). Here, the fixed-π guarantees above provide the frozen- or exogenous-U limit, while $U_{\mathbf{W}}(x) = -\frac{1}{2}\|\mathbf{W}x\|_2^2$ defines a quasi-static comparison skeleton with $\mathbf{W}$ frozen during fast spatial relaxation.

4.1 The Macroscopic Unsupervised Loss

To study the coupled system at the macroscopic level, we introduce a Lyapunov-type energy functional $\mathcal{E}(\mathbf{X}, \mathbf{W})$ (Definition 10, Appendix C). Following PCA-style energy formulations [4, 61], for a spatial configuration $\mathbf{X}_t \in \mathbb{R}^{N\times n}$ and representation matrix $\mathbf{W}_t \in \mathbb{R}^{m\times n}$, define

$$\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) = \underbrace{-\frac{1}{2N}\mathrm{Tr}\left(\mathbf{X}_t\mathbf{W}_t^T\mathbf{W}_t\mathbf{X}_t^T\right)}_{\text{Feature variance maximization}} + \underbrace{\frac{\lambda}{4}\left\|\mathbf{W}_t\mathbf{W}_t^T - \mathbf{I}_m\right\|_F^2}_{\text{Row-orthogonality penalty}}. \tag{13}$$

In the drift-only coupled skeleton, U is instantiated by the representation-induced part of $\mathcal{E}$, so the spatial drift follows the tangent-projected negative $\mathbf{X}$-gradient of the macroscopic energy. The variance term is PCA-like for $\mathbf{Y}_t = \mathbf{X}_t\mathbf{W}_t^T$ [4, 44], while the row-orthogonality penalty matches the subspace Oja lateral-inhibition dynamics in Section 3 [91] and prevents latent-coordinate collapse.

Accordingly, Section 4 provides a macroscopic comparison principle for the coupled skeleton: $\mathcal{E}$ supplies a Lyapunov surrogate for the deterministic quasi-static dynamics, while the stochastic graph

limit and averaged Oja limit are the formal guarantees established in Sections 2 and 3. Writing $\widehat{\Sigma}_{\mathbf{X}} = \frac{1}{N}\mathbf{X}^T\mathbf{X}$, the row-space component of the skeleton takes the subspace Oja form

$$\dot{\mathbf{W}} = \eta_W \left(\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} - \mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\mathbf{W} \right), \tag{14}$$

which corresponds to a projected variance-maximization flow on the row-orthonormal manifold. Eq. 14 represents the projected Lyapunov dynamics of the comparison skeleton. The spatial component follows the drift associated with the macroscopic energy landscape, while the stochastic Brownian term is treated separately in the Langevin formulation of Section 2.

4.2 Global Stabilization of the Deterministic Skeleton

In decentralized networks, concurrent updates over non-stationary data streams may cause unstable oscillations. We therefore analyze a deterministic macroscopic skeleton that removes the diffusion martingale term while retaining the coupled drift. The Lyapunov identity applies to this skeleton [54]; the stochastic flow requires Itô corrections [78]. Thus, Proposition 4 characterizes the macroscopic comparison flow, with stochastic corrections handled by the Langevin and averaging results above.

For the row-Stiefel manifold $\mathcal{S}_{m,n} = \{\mathbf{W} \in \mathbb{R}^{m\times n} : \mathbf{W}\mathbf{W}^T = \mathbf{I}_m\}$ [2, 27], we denote by $\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\,\mathcal{E}$ the Euclidean gradient projected onto the tangent space of $\mathcal{S}_{m,n}$ [2]. This projection removes the normal component so that the parameter flow remains tangent to the row-orthonormal constraint.

Proposition 4 (Lyapunov Dissipation of the Deterministic Coupled Skeleton) *Along the deterministic macroscopic skeleton associated with $\mathcal{E}$, the spatial component follows the tangent-projected negative energy drift and the parameter component follows the negative row-Stiefel projected gradient direction. Then the energy is non-increasing:*

$$\frac{d}{dt}\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) = -\eta_X \left\|\Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\|_F^2 - \eta_W \left\|\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\,\mathcal{E}\right\|_F^2 \le 0. \tag{15}$$

Implication. Proposition 4 shows that the deterministic macroscopic skeleton admits a non-increasing Lyapunov description. Thus, the interaction between spatial drift and local parameter plasticity can be analyzed through an energy-decreasing comparison flow, even though the original decentralized dynamics are stochastic and locally implemented.

4.3 Orthogonal Decorrelation at Equilibrium

We next characterize the latent representation at a stationary point of the macroscopic flow [4, 44]. Let $(\mathbf{X}^*, \mathbf{W}^*)$ be a row-orthonormal principal-subspace stationary point, and let $\mathbf{Y}^* = \mathbf{X}^*(\mathbf{W}^*)^T \in \mathbb{R}^{N\times m}$. Define the empirical spatial covariance $\mathbf{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^*$.

Theorem 5 (Orthogonal Decorrelation) *Assume that $\mathbf{\Sigma}^* = \mathbf{Q}^*\mathbf{\Lambda}^*(\mathbf{Q}^*)^T$ with $\lambda_m^* > \lambda_{m+1}^*$, and let $\mathbf{Q}_m^*$ be the top-m columns of $\mathbf{Q}^*$. If $\mathrm{Row}(\mathbf{W}^*) = \mathrm{Col}(\mathbf{Q}_m^*)$, then for some $\mathbf{R} \in O(m)$,*

$$\mathbf{W}^* = \mathbf{R}(\mathbf{Q}_m^*)^T. \tag{16}$$

Moreover, the latent covariance satisfies

$$\mathbf{\Sigma}_{\mathbf{Y}^*} = \frac{1}{N}(\mathbf{Y}^*)^T\mathbf{Y}^* = \mathbf{R}\mathbf{\Lambda}_m^*\mathbf{R}^T. \tag{17}$$

Equivalently, in the rotated latent coordinates,

$$\mathbf{R}^T\mathbf{\Sigma}_{\mathbf{Y}^*}\mathbf{R} = \mathbf{\Lambda}_m^* = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_m^*). \tag{18}$$

Proof Outline (Full proof in Appendix C). The stationary representation is characterized by projecting the row space of $\mathbf{W}^*$ onto the spectral basis of the empirical spatial covariance.

1. Macroscopic Eigenvalue Resolution (Appendix C.3): At a row-orthonormal stationary point of the subspace flow, the rows of $\mathbf{W}^*$ span an invariant subspace of $\mathbf{\Sigma}^*$. Under the spectral gap $\lambda_m^* > \lambda_{m+1}^*$, the stable invariant subspace is the top-m principal eigenspace of $\mathbf{\Sigma}^*$.

2. Spectral Projection and Decorrelation (Appendices C.3 & C.4): Since the row space of $\mathbf{W}^*$ equals the principal subspace $\mathrm{Col}(\mathbf{Q}_m^*)$, there exists an orthogonal matrix $\mathbf{R}$ such that $\mathbf{W}^* =$

$\mathbf{R}(\mathbf{Q}_m^*)^T$. Substituting this into the latent covariance gives $\frac{1}{N}(\mathbf{Y}^*)^T\mathbf{Y}^* = \mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T = \mathbf{R}\mathbf{\Lambda}_m^*\mathbf{R}^T$. Thus the covariance is orthogonally diagonalizable [44], and it becomes diagonal after choosing an appropriate orthogonal basis for the latent coordinates.

Implication. The diagonalization of the latent covariance establishes feature decorrelation in principal-subspace coordinates [46]. This explains how local memoryless updates can produce stable low-dimensional representations. Downstream linear separability depends on additional data assumptions such as class alignment, margin structure, and probe optimization [3, 6], and can be studied on top of the representation-level guarantees.

4.4 Algorithmic Realization of the Coupled Flow

To translate dynamics into a procedure, we use Euler–Maruyama [57] for the spatial step under the quasi-static frozen-$\mathbf{W}$ potential $U_{\mathbf{W}}(x) = -\frac{1}{2}\|\mathbf{W}x\|_2^2$, and a finite-step subspace Oja update for the representation, yielding Decentralized Coupled Representation Learning (DCRL, **Algorithm 1**).

Algorithm 1 Decentralized Coupled Representation Learning (DCRL)

Require: Manifold $\mathcal{M}$; population N; rank m; parameters D, β; steps η_x, η_w with $\eta_w \ll \eta_x$.

Init: Configurations $\mathbf{X}^{(0)} = \{\mathbf{x}_i^{(0)}\}_{i=1}^N \sim \mu_0$ on $\mathcal{M}$, base weights $\mathbf{W}^{(0)} \in \mathbb{R}^{m\times n}$.
for discrete step $k = 0, \dots, K-1$ **do**
  {**Phase 1: Fast Kinematics**}
  for each agent $i \in \{1, \dots, N\}$ **in parallel do**
    $\mathbf{g}_i^{(k)} \leftarrow \Pi_{\mathbf{x}_i^{(k)}}\left((\mathbf{W}^{(k)})^T\mathbf{W}^{(k)}\mathbf{x}_i^{(k)}\right)$; $\widetilde{\boldsymbol{\xi}}_i^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$; $\boldsymbol{\xi}_i^{(k)} \leftarrow \Pi_{\mathbf{x}_i^{(k)}}\widetilde{\boldsymbol{\xi}}_i^{(k)}$
    $\mathbf{x}_i^{(k+1)} \leftarrow \Pi_{\mathcal{X}}\left(\mathbf{x}_i^{(k)} + \eta_x D\beta\mathbf{g}_i^{(k)} + \sqrt{2D\eta_x}\boldsymbol{\xi}_i^{(k)}\right)$
  end for
  {**Phase 2: Slow Plasticity with Lateral Inhibition**}
  $\mathbf{y}_i^{(k+1)} \leftarrow \mathbf{W}^{(k)}\mathbf{x}_i^{(k+1)}$ for all i
  $\Delta\mathbf{W}^{(k)} \leftarrow \frac{1}{N}\sum_{i=1}^N\left[\mathbf{y}_i^{(k+1)}(\mathbf{x}_i^{(k+1)})^T - \mathbf{y}_i^{(k+1)}(\mathbf{y}_i^{(k+1)})^T\mathbf{W}^{(k)}\right]$
  $\mathbf{W}^{(k+1)} \leftarrow \mathbf{W}^{(k)} + \eta_w\Delta\mathbf{W}^{(k)}$
end for
return Stabilized matrix $\mathbf{W}^{(K)}$, latent features $\mathbf{Y}^{(K)} = \mathbf{X}^{(K)}(\mathbf{W}^{(K)})^T$.
Note: Phase 2 is synchronous plasticity. A-DCRL uses pairwise asynchronous gossip with $\mathcal{O}(nm)$ updates and $\mathcal{O}(nm)$ payload per active edge; see Appendix D.2. $\Pi_{\mathcal{X}}$ denotes substrate-domain retraction; $\Pi_{\mathbf{x}}$ denotes tangent projection [2]; $\eta_w \ll \eta_x$ matches the continuous-time regime up to finite-step error; see Appendix D.3.

5 Empirical Evaluations: Numerical Verification of Macroscopic Invariants

The experiments examine macroscopic invariants of the theory, with emphasis on stability, subspace recovery, and latent decorrelation. We evaluate DCRL on low-dimensional manifolds, high-dimensional controlled distributions, and feature spaces extracted from pretrained models. Experiments are organized around three predictions: generator convergence of the spatial process (Theorem 1), Lyapunov stability of the averaged subspace flow (Theorem 2), and principal-subspace alignment with latent covariance decorrelation (Results 3–5). Comparisons use theory-aligned eigenspace references—exact PCA [47], streaming/subspace Oja–Sanger [76, 91], EigenGame [34], and decentralized spectral/distributed PCA [53, 66]. These references calibrate fixed-stream or fixed-covariance eigenspace recovery, while the object studied here is the endogenous Markovian sampling process coupled to local Oja-type plasticity. The fixed-stream invariant calibration is reported in Appendix D.5, Table D.4. Modern deep self-supervised learning pipelines such as DINO, BYOL, or SimCLR [13, 16, 37] target encoder-level contrastive or teacher-student objectives; accordingly, baselines focus on theory-aligned eigenspaces.

5.1 Experimental Protocol

We evaluate substrates ranging from low-dimensional manifolds to high-dimensional neural features. Unless stated otherwise, results are reported as mean $\pm$ standard deviation over random seeds, with $\eta_w \ll \eta_x$ fixed to match the slow-fast regime. We monitor generator error, Stiefel orthogonality, subspace angle, and canonical-gauge covariance off-diagonal mass [2, 27, 44]:

$$E_{\mathcal{L}} = \|\mathcal{L}_N - \mathcal{L}\|_\infty, \qquad E_{\text{orth}} = \|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\|_F,$$
$$E_{\text{sub}} = \sin\Theta(\text{Row}(\mathbf{W}), \mathbf{Q}_m), \qquad E_{\text{off}}^{\text{can}} = \frac{\left\|\mathbf{R}^T\mathbf{\Sigma_Y}\mathbf{R} - \text{diag}\left(\mathbf{R}^T\mathbf{\Sigma_Y}\mathbf{R}\right)\right\|_F}{\|\mathbf{\Sigma_Y}\|_F}.$$

Here $E_{\mathcal{L}}$ measures generator approximation error, E_{orth} row-orthogonality, E_{sub} principal-subspace alignment, and $E_{\text{off}}^{\text{can}}$ latent covariance decorrelation after fixing an orthogonal latent gauge. The gauge $\mathbf{R}$ is fixed by Procrustes alignment to the target eigenspace, independently of the measured covariance: if $\mathbf{W}\mathbf{Q}_m = \mathbf{A}_R\mathbf{S}_R\mathbf{B}_R^T$ is the SVD, then $\mathbf{R} = \mathbf{A}_R\mathbf{B}_R^T$.

5.2 Verification of the Macroscopic Invariants

Figure 2 and Table 1 report trajectory-level and endpoint evidence for the diagnostics: generator convergence, Lyapunov decay, row-orthogonality, subspace alignment, and covariance decorrelation.

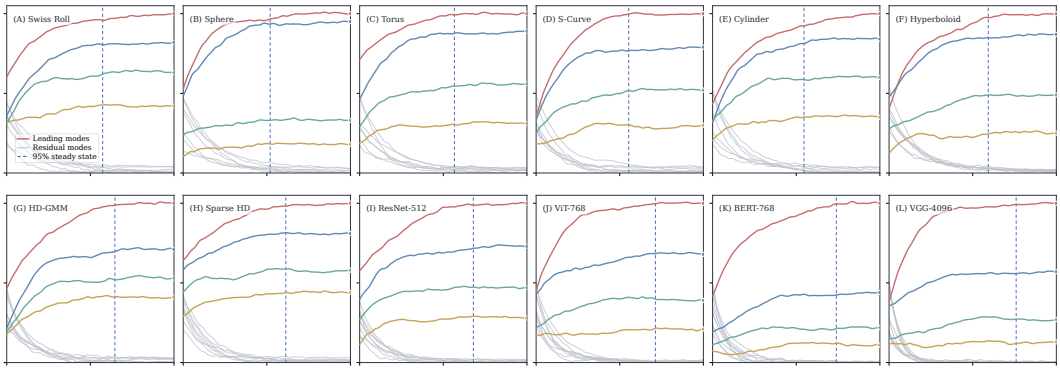

Figure 2: **Macroscopic Spectral Evolution Across Substrates.** Panels **(A)–(L)** show normalized latent eigenspectrum trajectories over 1.5×10^4 steps, with dotted blue lines marking the 95% steady-state transition. Stabilization of leading modes and contraction of residual modes are consistent with principal-subspace selection and covariance decorrelation.

Table 1: **Numerical Diagnostics of Macroscopic Invariants.** Mean $\pm$ standard deviation over seeds. Error columns are scaled by the headers; $E_{\mathcal{L}}$ is reported only where graph-to-diffusion diagnostics apply. Entries shown as 0.000 are below printed precision after column scaling.

Substrate	N	$n \to m$	$10^2 E_{\mathcal{L}}$	$10^5 E_{\text{orth}}$	$10^4 E_{\text{sub}}$	$10^2 E_{\text{off}}^{\text{can}}$
Swiss Roll	2048	$3 \to 2$	1.910 ± 0.024	0.041 ± 0.049	0.000 ± 0.000	0.005 ± 0.008
Sphere	2048	$3 \to 2$	1.690 ± 0.043	0.013 ± 0.007	0.000 ± 0.000	0.143 ± 0.121
Torus	2048	$3 \to 2$	1.599 ± 0.035	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
HD-GMM	1024	$500 \to 20$	7.608 ± 0.003	3.075 ± 1.379	4.051 ± 0.371	1.639 ± 1.208
ResNet-512	60k	$512 \to 64$	N/A	4.553 ± 0.051	0.000 ± 0.000	7.183 ± 0.256
ViT-768	60k	$768 \to 64$	N/A	0.214 ± 0.047	0.000 ± 0.000	0.159 ± 0.045
BERT-768	127.6k	$768 \to 32$	N/A	5.123 ± 0.113	0.000 ± 0.000	0.396 ± 0.029

5.3 Ablations: Lateral Inhibition and Timescale Separation

We next test whether the stability mechanisms used in the analysis are necessary for stable behavior. Table 2 compares the full DCRL dynamics with variants that remove lateral inhibition, replace the subspace Oja update by scalar inhibition, or violate the slow-fast hierarchy. The relevant failure modes are Lyapunov violations, row-space collapse, subspace misalignment, and parameter divergence.

Table 2: **Ablation of Lateral Inhibition and Timescale Separation.** Each row removes or violates one stability mechanism. All results use the controlled HD-GMM substrate with $N = 1024$, $n = 500$, $m = 20$, and a fixed step-size schedule. Lower is better.

Variant	Timescale	Viol. (%)	E_{orth}	E_{sub}
Full DCRL	$\eta_w \ll \eta_x$	0.21	2.935e−05 ± 1.090e−07	4.328e−03 ± 7.697e−05
No lateral inhibition	$\eta_w \ll \eta_x$	Diverged	Diverged	Diverged
Scalar Oja inhibition	$\eta_w \ll \eta_x$	53.17	9.487e+00 ± 7.959e−07	9.996e−01 ± 2.050e−04
No timescale separation	$\eta_w \approx \eta_x$	48.69	3.435e−03 ± 8.507e−05	4.302e−02 ± 1.311e−03
Large-step regime	large steps	50.22	1.349e+00 ± 2.729e−01	7.065e−01 ± 3.084e−01

5.4 Scalability and Communication Footprint

Having tested the stability and alignment diagnostics, we report the computational footprint in decentralized settings. Here the relevant question is storage and communication under local agent interactions, especially when forming a centralized covariance is undesirable. As shown in Table 3, DCRL stores representation-level quantities, while A-DCRL replaces the synchronous accumulator with pairwise gossip over the communication graph [7, 74]. Runtime numbers are implementation-dependent reference timings; accordingly, the table separates one-shot batch costs from per-update iterative costs.

Table 3: **Complexity and Communication Footprint.** Time entries report one-shot cost for batch eigensolvers and per-update cost for iterative methods. Thus K DCRL iterations cost $\mathcal{O}(KNnm)$, while K pairwise A-DCRL rounds cost $\mathcal{O}(Knm)$. Measured time is reported in milliseconds for a reference NumPy implementation with $N = 2048$, $n = 512$, and $m = 64$.

Method	Per-step / One-shot Time	Memory / Node	Payload / Node	Centralized Matrix	Ref. Time (ms)
Exact PCA	$\mathcal{O}(Nn^2 + n^3)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n)$	Yes	22.93 ± 1.46
Streaming Oja	$\mathcal{O}(Nnm)$	$\mathcal{O}(nm)$	$\mathcal{O}(m)$	No	61.70 ± 2.51
DCRL	$\mathcal{O}(Nnm)$	$\mathcal{O}(nm)$	$\mathcal{O}(nm)$	No	2.58 ± 0.09
A-DCRL	$\mathcal{O}(nm)$ per active edge	$\mathcal{O}(nm)$	$\mathcal{O}(nm)$	No	0.19 ± 0.01

Summary. The experiments test graph-to-diffusion convergence, Lyapunov decay, row-space collapse avoidance, principal-eigenspace alignment, and covariance decorrelation. Extended protocols, ablations, scalability, and substrate results appear in Appendix D; supervised linear probing is treated as a downstream evaluation rather than an invariant diagnostic.

6 Discussion and Limitations

This work builds on local Hebbian and Oja-type representation learning [67, 76, 81], continuous-time learning dynamics [15, 71, 94], and stochastic analysis on manifolds [45, 21]. Unlike fixed-stream subspace analyses, our setting couples Markovian spatial exploration with local subspace learning. The proof combines graph-to-diffusion convergence, stochastic averaging through Poisson equations [79], and invariant-set analysis of the averaged subspace flow.

We establish three macroscopic results: localized graph transitions converge at the generator level to Langevin diffusion; slow subspace Oja dynamics admit an averaged Lyapunov stability mechanism that rules out divergence under the stated assumptions; and, under a spectral gap, the learned row space aligns with the principal eigenspace, yielding decorrelated latent coordinates up to rotation.

The results are proved under structural assumptions. The geometric limit assumes a compact, connected, resolved manifold and the graph scaling in Eq. 1 [31, 75]. The representation model is linear, $\mathbf{y}_t = \mathbf{W}_t \mathbf{x}_t$, so extending the guarantees to nonlinear encoders remains open [15, 71, 94]. The stability result relies on timescale separation and Markovian spatial dynamics; asynchronous delays, sparse communication, and non-Markovian effects remain natural extensions [7, 84]. Similarly, downstream linear separability depends on label geometry and builds on the representation-level guarantees. Overall, the results provide a tractable limit theory for studying how local sampling and local plasticity can induce global subspace structure in decentralized representation learning.

Acknowledgements

This work was supported by the Shanghai Pujiang Program (No. 24PJA004). The authors gratefully acknowledge the support provided by this program, which contributed to the development and completion of this research.

References

[1] Ralph Abraham, Jerrold E Marsden, and Tudor Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75. Springer Science & Business Media, 2012.

[2] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[3] Sanjeev Arora, Hrishikesh Khandeparkar, Mikhail Khodak, Orestis Plevrakis, and Nikunj Saunshi. A theoretical analysis of contrastive unsupervised representation learning. In *International Conference on Machine Learning*, pages 5628–5637. PMLR, 2019.

[4] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53–58, 1989.

[5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, pages 585–591, 2001.

[6] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1798–1828, 2013.

[7] Dimitri P Bertsekas and John N Tsitsiklis. *Parallel and distributed computation: numerical methods*. Prentice hall, 1989.

[8] Kush Bhatia, Aldo Pacchiano, Nicolas Flammarion, Peter L. Bartlett, and Michael I. Jordan. Gen-oja: Simple & efficient algorithm for streaming generalized eigenvector computation. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[9] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 3rd edition, 1995.

[10] Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Springer, 2009.

[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[12] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[13] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[14] Han-Fu Chen. *Stochastic Approximation and Its Applications*, volume 64. Springer Science & Business Media, 2002.

[15] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[16] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[17] Fan RK Chung. *Spectral Graph Theory*, volume 92. American Mathematical Society, 1997.

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

[19] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66(1): 016121, 2002.

[20] Chandler Davis and William M Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[21] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*, volume 35, pages 2406–2422, 2022.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186, 2019.

[23] Alexandros G Dimakis, Soummya Kar, José MF Moura, Michael G Rabbat, and Anna Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98 (11):1847–1864, 2010.

[24] Manfredo Perdigao Do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.

[25] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

[26] Richard M Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.

[27] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[28] Stewart N Ethier and Thomas G Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2009.

[29] Lawrence C Evans. *Partial Differential Equations*, volume 19. American Mathematical Society, 2nd edition, 2010.

[30] Herbert Federer. *Geometric Measure Theory*. Springer, 1969.

[31] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[32] Gerald B Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2nd edition, 1999.

[33] Crispin W Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, volume 13. Springer Berlin, 3rd edition, 2004.

[34] Ian Gemp, Brian McWilliams, Claire Vernade, and Thore Graepel. EigenGame: PCA as a Nash equilibrium. In *International Conference on Learning Representations*, 2021.

[35] Gene H Golub and Charles F Van Loan. *Matrix Computations*. JHU press, 2013.

[36] Alfred Gray. *Tubes*, volume 221. Springer Science & Business Media, 2nd edition, 2004.

[37] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, 2020.

[38] Piyush Gupta and P. R. Kumar. Critical power for asymptotic connectivity in wireless networks. In William M. McEneaney, G. George Yin, and Qing Zhang, editors, *Stochastic Analysis, Control, Optimization and Applications*, pages 547–566. Birkhäuser, Boston, MA, 1999. doi: 10.1007/978-1-4612-1784-8_33.

[39] Brian C Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, volume 222. Springer, 2015.

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[41] Matthias Hein, Jean-Yves Audibert, and Ulrike Von Luxburg. From graphs to manifolds–weak and strong pointwise consistency of graph laplacians. In *Learning Theory: 18th Annual Conference on Learning Theory, COLT 2005*, pages 470–485. Springer, 2005.

[42] Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a definition of disentangled representations, 2018.

[43] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[44] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012.

[45] Elton P Hsu. *Stochastic Analysis on Manifolds*, volume 38. American Mathematical Society, 2002.

[46] Aapo Hyvärinen and Erkki Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[47] Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2002.

[48] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[49] Jürgen Jost. *Riemannian Geometry and Geometric Analysis*. Springer Science & Business Media, 5th edition, 2008.

[50] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer Science & Business Media, 2nd edition, 1991.

[51] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, 2013.

[52] Frank P Kelly. *Reversibility and Stochastic Networks*. John Wiley & Sons, 1979.

[53] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. *Journal of Computer and System Sciences*, 74(1):70–83, 2008.

[54] Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

[55] Rafail Khasminskii. *Stochastic stability of differential equations*, volume 66. Springer Science & Business Media, 2011.

[56] Rafail Z Khasminskii. On the principle of averaging for itô's stochastic differential equations. *Kibernetika*, 4(3):260–279, 1968.

[57] Peter E Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.

[58] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[59] Harold J Kushner. *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*. MIT Press, 1984.

[60] Joseph P LaSalle. Some extensions of liapunov's second method. *IRE Transactions on Circuit Theory*, 7(4):520–527, 1960.

[61] Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu-Jie Huang. A tutorial on energy-based learning. In Gökhan H. Bakır, Thomas Hofmann, Bernhard Schölkopf, Alexander J. Smola, and Ben Taskar, editors, *Predicting Structured Data*, pages 191–246. MIT Press, 2006.

[62] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015.

[63] John M Lee. *Introduction to Riemannian Manifolds*. Springer, 2018.

[64] David A Levin and Yuval Peres. *Markov Chains and Mixing Times*, volume 107. American Mathematical Society, 2017.

[65] Chris Junchi Li, Mengdi Wang, Han Liu, and Tong Zhang. Diffusion approximations for online principal component estimation and global convergence. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[66] Fangshuo Liao, Wenyi Su, and Anastasios Kyrillidis. Provable model-parallel distributed principal component analysis with parallel deflation. *arXiv preprint arXiv:2502.17615*, 2025.

[67] Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

[68] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *International conference on machine learning*, pages 4114–4124. PMLR, 2019.

[69] Jan R Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 3rd edition, 2019.

[70] Xuerong Mao. Stochastic versions of the lasalle theorem. *Journal of Differential Equations*, 153(1):175–195, 1999.

[71] Song Mei, Andrea Montanari, and Phan-Minh Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33): E7665–E7671, 2018.

[72] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[73] John Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, pages 20–63, 1956.

[74] Angelia Nedić and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[75] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3): 419–441, 2008.

[76] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

[77] Erkki Oja and Juha Karhunen. Stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 106(1):69–84, 1985.

[78] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, 6th edition, 2003.

[79] Etienne Pardoux and Alexander Yu Veretennikov. On the Poisson equation and diffusion approximation. I. *The Annals of Probability*, 29(3):1061–1085, 2001.

[80] Grigorios A Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker–Planck and Langevin Equations*, volume 60. Springer, 2014.

[81] Cengiz Pehlevan and Dmitri B Chklovskii. A normative theory of adaptive dimensionality reduction in neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

[82] Mathew Penrose. *Random Geometric Graphs*, volume 5. Oxford University Press, 2003.

[83] Peter Petersen. *Riemannian Geometry*, volume 171 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 2006.

[84] Benjamin Recht, Christopher Re, Stephen Wright, and Feng Niu. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

[85] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293. Springer Science & Business Media, 3rd edition, 2013.

[86] Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Berlin Heidelberg, 1996.

[87] Christian P Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.

[88] Steven Rosenberg. *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds*. Number 31 in London Mathematical Society Student Texts. Cambridge University Press, 1997.

[89] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3rd edition, 1976.

[90] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[91] Terence D Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2(6):459–473, 1989.

[92] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015.

[93] Amit Singer. From graph to manifold laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.

[94] Justin Sirignano and Konstantinos Spiliopoulos. Mean field analysis of neural networks: A law of large numbers. *SIAM Journal on Applied Mathematics*, 80(2):725–752, 2020.

[95] Aad W Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000.

[96] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge University Press, 2018.

[97] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[98] Ulrike Von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[99] Kôsaku Yosida. *Functional Analysis*. Springer Science & Business Media, 6th edition, 1995.

[100] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Table of Notation

Group	Notation	Meaning
Geometry	$(\mathcal{M}, g)$	Compact, connected, boundaryless d-dimensional Riemannian manifold and metric.
	$d_{\mathcal{M}}, dV_g, \mu$	Geodesic distance, Riemannian volume measure, and normalized reference probability measure.
	$T_x\mathcal{M}, \exp_x$	Tangent space at x and Riemannian exponential map.
	$\nabla, \Delta_{\mathcal{M}}$	Riemannian gradient and Laplace–Beltrami operator.
Graph process	$\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N)$	Random geometric graph built from N sampled substrate vertices on $\mathcal{M}$.
	$\varepsilon_N, \mathcal{N}_N(x)$	Connectivity radius and empirical local neighborhood of x.
	P_N, Z_N, π_N	Local Gibbs transition kernel, partition function, and discrete invariant measure.
	$\mathcal{L}_N, \mathcal{L}$	Discrete infinitesimal generator and limiting Langevin generator.
	D, β, τ_N	Diffusion coefficient, inverse temperature, and temporal jump-rate scaling.
	$X_t, B_t^{\mathcal{M}}$	Single-agent fast state and Brownian motion on $\mathcal{M}$.
Representation	$\Phi : \mathcal{M} \hookrightarrow \mathbb{R}^n$	Smooth embedding from the manifold to the ambient feature space.
	n, m	Ambient feature dimension and latent dimension.
	$\mathbf{x}_t, \mathbf{y}_t$	Euclidean observation $\Phi(X_t)$ and latent coordinate $\mathbf{W}_t\mathbf{x}_t$.
	$\mathbf{W}_t \in \mathbb{R}^{m\times n}$	Representation matrix learned by the subspace Oja dynamics.
	$\pi, \Phi_{\#}\pi, \Sigma_\pi$	Fast invariant measure, its pushforward, and induced spatial covariance.
	$\mathbf{Q}, \Lambda, \mathbf{Q}_m$	Eigenbasis, eigenvalues, and $n \times m$ top-eigenvector matrix of Σ_π.
	$\lambda_m > \lambda_{m+1}$	Spectral gap condition for principal-subspace identification.
Averaged dynamics	$H_\gamma(\mathbf{W}, x), \bar{H}_\gamma(\mathbf{W})$	Instantaneous Oja drift and its invariant-measure average.
	$\epsilon, \gamma, t, \tau$	Timescale parameter, plasticity rate, microscopic time, and slow time $\tau = \epsilon t$.
	$\mathbf{W}^\epsilon(\tau), \bar{\mathbf{W}}(\tau)$	Slow-time trajectory $\mathbf{W}_{\tau/\epsilon}$ and its averaged ODE limit.
	$V_{\mathrm{orth}}(\mathbf{W})$	Row-orthogonality Lyapunov functional $\frac{1}{4}\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\|_F^2$.
	$\nu(\mathbf{W}, x)$	Matrix-valued Poisson corrector for centered Markovian fluctuations.
	$\mathrm{Row}(\mathbf{W}) = \mathrm{Col}(\mathbf{Q}_m)$	Row-space alignment convention; on $\mathbf{W}\mathbf{W}^T = \mathbf{I}_m$, equivalently $\mathbf{W} = \mathbf{R}\mathbf{Q}_m^T$ for some $\mathbf{R} \in O(m)$.
Coupled skeleton	$\mathbf{X}_t \in \mathbb{R}^{N\times n}$	Finite-population configuration matrix used in the comparison skeleton.
	$U_{\mathbf{W}}(x)$	Representation-induced quasi-static potential $-\frac{1}{2}\|\mathbf{W}x\|_2^2$.
	$\mathcal{E}(\mathbf{X}, \mathbf{W})$	Macroscopic Lyapunov energy for the deterministic comparison skeleton.
	$\widehat{\Sigma}_{\mathbf{X}}$	Empirical covariance $\frac{1}{N}\mathbf{X}^T\mathbf{X}$ of a finite configuration.
	$\mathcal{S}_{m,n}$	Row-Stiefel manifold $\{\mathbf{W} : \mathbf{W}\mathbf{W}^T = \mathbf{I}_m\}$.
	$\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\mathcal{E}$	Euclidean $\mathbf{W}$-gradient projected onto the tangent space of $\mathcal{S}_{m,n}$.
	$\Pi_{\mathbf{x}}, \Pi_{\mathcal{X}}$	Tangent projection at $\mathbf{x}$ and substrate-domain retraction/projection.
Stationary covariance	$(\mathbf{X}^*, \mathbf{W}^*)$	Row-orthonormal principal-subspace stationary point of the comparison flow.
	$\mathbf{Y}^* = \mathbf{X}^*(\mathbf{W}^*)^T$	Latent representation at stationarity.
	$\mathbf{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^*$	Empirical spatial covariance at stationarity.
	$\mathbf{Q}^*, \mathbf{\Lambda}^*, \mathbf{Q}^*_m, \mathbf{\Lambda}^*_m$	Eigenbasis, eigenvalues, top-m eigenvector matrix, and top eigenvalue block of the stationary empirical covariance $\mathbf{\Sigma}^*$.
	$\mathbf{\Sigma}_{\mathbf{Y}^*}$	Latent covariance $\frac{1}{N}(\mathbf{Y}^*)^T\mathbf{Y}^*$.
	$\mathbf{R} \in O(m)$	Orthogonal latent rotation used to express covariance decorrelation.
Algorithm and diagnostics	η_x, η_w	Spatial and representation step sizes, with $\eta_w \ll \eta_x$ in the slow-fast regime.
	$\mathcal{G}_c, \alpha$	Communication graph and pairwise gossip weight in A-DCRL.
	$E_{\mathcal{L}}$	Generator approximation error.
	E_{orth}	Row-orthogonality error $\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\|_F$.
	E_{sub}	Principal-subspace alignment error based on principal angles.
	$E_{\mathrm{off}}^{\mathrm{can}}$	Canonical-gauge off-diagonal mass of the latent covariance.

Table of Contents for Appendices

A Asymptotic Limits of Localized Graph Flows on Riemannian Manifolds 18
- A.1 Measure-Theoretic Formulation of the Geometric Substrate 18
- A.2 Discrete Transition Kernel and Detailed Balance 20
- A.3 Riemannian Taylor Expansion and Covariant Remainder Bounds 22
- A.4 Asymptotic Expansion of the Partition Function 24
- A.5 Convergence of the Discrete Infinitesimal Generator 27
- A.6 Summary of Part: From Local Graph Flow to Langevin Kinematics 29

B Averaged Stability and Principal Subspace Alignment 30
- B.1 State-Parameter Embedding and the Spatial Covariance Operator 30
- B.2 Two-Time-Scale Separation via the Poisson Resolvent 32
- B.3 Row-Orthogonality Lyapunov Functional and Dissipativity 35
- B.4 LaSalle's Invariance and Eigenspace Alignment 37

C Coupled Flow and Feature Decorrelation 42
- C.1 The Coupled Langevin–Subspace Oja Flow on the Product Space 42
- C.2 The Joint Lyapunov Functional and Lyapunov Dissipation 46
- C.3 Stationary State and Principal Component Alignment 50
- C.4 Latent Feature Decorrelation 53
- C.5 Degenerate Spectra and Subspace Gauge Symmetry 55

D Algorithmic Implementation and Extended Empirical Protocols 59
- D.1 Atomic Discretization of the Coupled System 59
- D.2 Asynchronous Decentralized Realization 60
- D.3 Finite-Step Lyapunov Diagnostics 61
- D.4 Resource Complexity and Scalability 62
- D.5 Relation to Static PCA and Fixed-Stream Baselines 63
- D.6 Benchmark Suite and Substrate Diagnostics 64
- D.7 Operator Ablations and Timescale Robustness 66
- D.8 Extended Invariant Validation and Downstream Sanity Checks 67
- D.9 Summary of Diagnostic Coverage 68

A Asymptotic Limits of Localized Graph Flows on Riemannian Manifolds

In this section, we formalize the continuum limit of the localized agent dynamics within the framework of measure theory and stochastic analysis on Riemannian manifolds. Rather than relying on heuristic physical expansions such as the Kramers–Moyal expansion [33, 86], our objective is aligned with Theorem 1: for every fixed $f \in C^3(\mathcal{M})$, we prove that the discrete infinitesimal generator $\mathcal{L}_N f$ converges uniformly in probability to the Kolmogorov backward generator of an overdamped Langevin diffusion [45, 80]. This generator-level result justifies the fast spatial process used by the slow–fast representation dynamics in the main text.

A.1 Measure-Theoretic Formulation of the Geometric Substrate

We first construct the discrete spatial substrate as a random geometric discretization of a continuous probability space. This step is necessary because the local graph operator used by the agents is only meaningful asymptotically after specifying how sampled neighborhoods approximate geodesic neighborhoods on the underlying manifold.

Let $(\mathcal{M}, g)$ be a compact, connected, boundaryless d-dimensional smooth Riemannian manifold [63].[3] The assumptions play the following roles:

- **Compactness:** Compactness implies bounded curvature and a positive global injectivity radius $r_{\text{inj}}(\mathcal{M}) > 0$, so local exponential coordinates are uniformly valid for sufficiently small neighborhoods.
- **No boundary:** The absence of a boundary avoids additional reflection or boundary conditions for the limiting diffusion.
- **Connectedness:** Connectedness prevents the limiting spatial process from decomposing into disconnected invariant components.

Let dV_g denote the Riemannian volume measure induced by the metric tensor g [49]. The geodesic distance between $\mathbf{x}, \mathbf{y} \in \mathcal{M}$ is denoted by $d_{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ [24]. We equip $\mathcal{M}$ with the normalized uniform probability measure μ [97],

$$d\mu(\mathbf{x}) = \frac{1}{\text{Vol}_g(\mathcal{M})}\, dV_g(\mathbf{x}), \qquad \text{Vol}_g(\mathcal{M}) = \int_{\mathcal{M}} dV_g < \infty.$$

Definition 1 (Random Geometric Graph Sequence) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a background probability space, and let*

$$\mathcal{X}_N = \{z_1, z_2, \ldots, z_N\}$$

be i.i.d. substrate vertices sampled from $\mathcal{M}$ according to the uniform measure μ. The empirical measure is

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i},$$

which converges weakly to μ almost surely by Varadarajan's theorem [26, 31, 59]. The random geometric graph is $\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N)$ with vertex set $\mathcal{V}_N = \mathcal{X}_N$ [19, 82]. For a deterministic radius $\varepsilon_N > 0$, two distinct vertices are connected when

$$(z_i, z_j) \in \mathcal{E}_N \quad \Longleftrightarrow \quad 0 < d_{\mathcal{M}}(z_i, z_j) \leq \varepsilon_N.$$

Figure A.1 illustrates the geometric setting underlying this construction: local neighborhoods are sampled from a compact manifold whose curvature may vary, while compactness provides the uniform coordinate control required by the subsequent expansion.

Although the agent graph is defined on the sampled vertices $\mathcal{X}_N$, the uniform convergence statement in Theorem 1 is taken over $x \in \mathcal{M}$. We therefore use the continuum extension of the empirical neighborhood:

$$\mathcal{N}_N(x) = \{z_i \in \mathcal{X}_N : 0 < d_{\mathcal{M}}(x, z_i) \leq \varepsilon_N\}, \qquad x \in \mathcal{M}.$$

[3] **Remark on the manifold assumption.** Compact smoothness is used only to ensure bounded differential operators, positive injectivity radius, and uniform control of curvature remainders. Empirical data supports may have singularities or boundaries; our result should be read as a regular limiting model for the bulk geometry of the sampled distribution [31, 75].

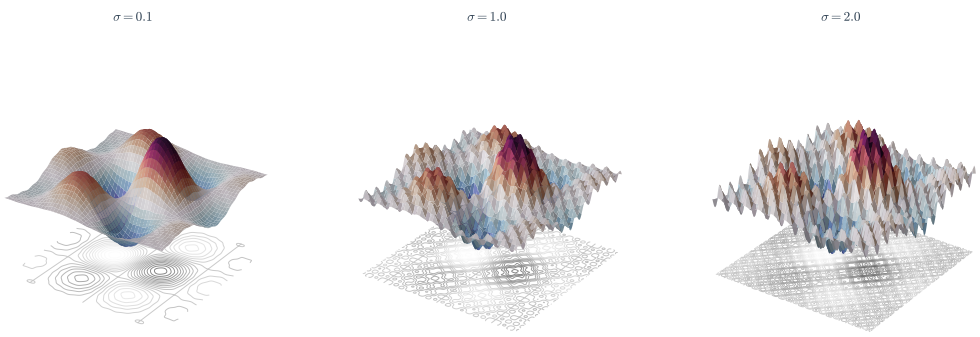

Figure A.1: **Geometric Substrate for the Local Expansion.** The localized graph flow is formulated on a smooth compact manifold $\mathcal{M}$ with controlled local coordinate neighborhoods. The figure illustrates how local geometry can vary with curvature while compactness controls curvature-dependent remainders uniformly.

When $x \in \mathcal{X}_N$, this agrees with the graph neighborhood used by the discrete agents. This convention makes $\mathcal{L}_N f(\mathbf{x})$ well-defined for all $\mathbf{x} \in \mathcal{M}$ in the uniform generator estimate.

To obtain a non-degenerate local differential limit, the connectivity radius must shrink to zero while retaining enough samples in each local neighborhood. We impose the following sufficient scaling regime.

Lemma 1 (Sufficient Scaling Regime for Uniform Generator Consistency [41, 93]) *Let $\varepsilon_N \to 0$ be a deterministic sequence. A sufficient scaling regime for graph connectivity and uniform concentration of the empirical second-order graph operator is*

$$\lim_{N\to\infty} \frac{N\varepsilon_N^{d+2}}{\log N} = \infty.$$

Proof: The scaling combines two requirements: connectivity of the local communication graph and uniform concentration of second-order empirical averages.

Step 1: Connectivity. For $\varepsilon_N < r_{\text{inj}}(\mathcal{M})$, the volume of a geodesic ball satisfies

$$\mu(B_g(\mathbf{x}, \varepsilon_N)) = \Theta(\varepsilon_N^d)$$

uniformly over $\mathbf{x} \in \mathcal{M}$ [83]. Hence the expected degree of a vertex is of order $N\varepsilon_N^d$. Standard connectivity bounds for random geometric graphs require

$$\frac{N\varepsilon_N^d}{\log N} \to \infty, \tag{A.1}$$

up to manifold-dependent constants [38, 82]. Since

$$\frac{N\varepsilon_N^{d+2}}{\log N} \to \infty \quad \text{and} \quad \varepsilon_N \to 0$$

imply

$$\frac{N\varepsilon_N^d}{\log N} = \frac{N\varepsilon_N^{d+2}}{\log N}\varepsilon_N^{-2} \to \infty,$$

the adopted scaling is stronger than the connectivity threshold.

Step 2: Uniform second-order concentration. The infinitesimal generator approximates a second-order differential operator and therefore involves a local empirical average scaled at order ε_N^{-2}. The deterministic Taylor bias of such a local operator is of order $\mathcal{O}(\varepsilon_N^2)$ [18]. The stochastic fluctuation is

controlled by the effective local sample size $N\varepsilon_N^d$ together with the second-order scaling. Uniformly over an ε_N-net of $\mathcal{M}$ [96], Bernstein-type concentration gives a bound of the form

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{M}} |\mathcal{L}_N f(\mathbf{x}) - \mathbb{E}\mathcal{L}_N f(\mathbf{x})| > \delta\right) \leq C_1 \varepsilon_N^{-d} \exp\left(-C_2 N \varepsilon_N^{d+2} \delta^2\right),$$

for constants $C_1, C_2 > 0$ depending on the manifold geometry and uniform bounds on f [11, 41, 93, 98]. The condition

$$N\varepsilon_N^{d+2} \gg \log N$$

ensures that this probability tends to zero after the covering union bound. Thus the stated scaling is sufficient for the uniform concentration needed in the generator convergence proof. ■

A.2 Discrete Transition Kernel and Detailed Balance

Building on the measure-theoretic construction of the geometric substrate $\mathcal{G}_N = (\mathcal{V}_N, \mathcal{E}_N)$, we now define the localized Markov transition rule used by the agents. The purpose of this subsection is not to establish a separate discrete-time mixing theorem, but to specify a reversible local kernel whose continuous-time generator

$$\mathcal{L}_N = \tau_N(P_N - I)$$

is analyzed in the generator convergence proof.

Let $U \in C^3(\mathcal{M}, \mathbb{R})$ be a smooth scalar potential. Since $\mathcal{M}$ is compact, the Extreme Value Theorem implies that U and its derivatives up to order three are uniformly bounded [89]. This ensures that the Gibbs weights below are uniformly controlled on local neighborhoods.

Definition 2 (Energy-Modulated Transition Kernel) *For any $x \in \mathcal{M}$, let the empirical local neighborhood be*

$$\mathcal{N}_N(x) = \{y \in \mathcal{X}_N : 0 < d_{\mathcal{M}}(x, y) \leq \varepsilon_N\}.$$

We define the local Gibbs transition kernel by

$$P_N(x, y) = \begin{cases} \dfrac{1}{Z_N(x)} \exp\left(-\dfrac{\beta}{2}\left[U(y) - U(x)\right]\right), & y \in \mathcal{N}_N(x), \\ 0, & y \notin \mathcal{N}_N(x), \end{cases} \tag{A.2}$$

where $\beta > 0$ is the inverse-temperature parameter [72], and

$$Z_N(x) = \sum_{z\in\mathcal{X}_N} \exp\left(-\frac{\beta}{2}\left[U(z) - U(x)\right]\right) \mathbb{I}_{\{0<d_{\mathcal{M}}(x,z)\leq\varepsilon_N\}} \tag{A.3}$$

is the local partition function [61, 87]. When $x \in \mathcal{X}_N$, this kernel coincides with the finite graph transition rule used by the agents. For general $x \in \mathcal{M}$, it gives the continuum extension needed to interpret $\sup_{x\in\mathcal{M}} |\mathcal{L}_N f(x) - \mathcal{L}f(x)|$.

Well-posedness of the kernel. Under the scaling regime in Lemma 1, and on the corresponding uniform concentration event, local neighborhoods are nonempty for all sampled vertices, while the continuum extension is well-defined on the same uniform-covering event used in the generator estimate. Since U is bounded on compact $\mathcal{M}$, the weights in Eq. A.2 are uniformly bounded above and below on each local neighborhood. Hence $Z_N(\mathbf{x}) > 0$ whenever $\mathcal{N}_N(\mathbf{x}) \neq \emptyset$, and

$$\sum_{y\in\mathcal{X}_N} P_N(x, y) = 1.$$

Thus P_N is a row-stochastic local transition kernel on the graph vertices and a consistent empirical extension on $\mathcal{M}$.

We next record the reversible measure of the finite graph chain. This is used to identify the stationary spatial weighting and to connect the discrete construction with the invariant measure π appearing in the averaged representation dynamics.

Lemma 2 (Discrete Detailed Balance and Invariant Measure) *Following the classical construction for reversible Markov chains on graphs [52, 64], define*

$$\pi_N(\mathbf{x}) = \frac{1}{\mathcal{Z}_{\text{global}}} Z_N(\mathbf{x}) \exp(-\beta U(\mathbf{x})), \qquad \mathbf{x} \in \mathcal{V}_N, \tag{A.4}$$

where

$$\mathcal{Z}_{\text{global}} = \sum_{\mathbf{v} \in \mathcal{V}_N} Z_N(\mathbf{v}) \exp(-\beta U(\mathbf{v})). \tag{A.5}$$

Then P_N satisfies detailed balance with respect to π_N:

$$\pi_N(\mathbf{x}) P_N(\mathbf{x}, \mathbf{y}) = \pi_N(\mathbf{y}) P_N(\mathbf{y}, \mathbf{x}), \qquad \forall \mathbf{x}, \mathbf{y} \in \mathcal{V}_N. \tag{A.6}$$

Consequently, π_N is an invariant distribution of the finite graph transition kernel.

Proof: We verify the detailed balance identity by considering two cases.

Case 1: Disconnected states. If $(\mathbf{x}, \mathbf{y}) \notin \mathcal{E}_N$, then $P_N(\mathbf{x}, \mathbf{y}) = 0$. Since the graph is undirected, $(\mathbf{x}, \mathbf{y}) \notin \mathcal{E}_N$ implies $(\mathbf{y}, \mathbf{x}) \notin \mathcal{E}_N$, and hence $P_N(\mathbf{y}, \mathbf{x}) = 0$. The identity is therefore trivial.

Case 2: Connected states. If $(\mathbf{x}, \mathbf{y}) \in \mathcal{E}_N$, then $\mathbf{y} \in \mathcal{N}_N(\mathbf{x})$ and $\mathbf{x} \in \mathcal{N}_N(\mathbf{y})$. Substituting Eq. A.4 and the transition kernel gives [72, 87]

$$\begin{aligned} \pi_N(\mathbf{x}) P_N(\mathbf{x}, \mathbf{y}) &= \left[\frac{1}{\mathcal{Z}_{\text{global}}} Z_N(\mathbf{x}) \exp(-\beta U(\mathbf{x})) \right] \left[\frac{1}{Z_N(\mathbf{x})} \exp\left(-\frac{\beta}{2} [U(\mathbf{y}) - U(\mathbf{x})] \right) \right] \\ &= \frac{1}{\mathcal{Z}_{\text{global}}} \exp\left(-\frac{\beta}{2} U(\mathbf{x}) - \frac{\beta}{2} U(\mathbf{y}) \right). \end{aligned} \tag{A.7}$$

The final expression is symmetric in $\mathbf{x}$ and $\mathbf{y}$, so the same calculation gives $\pi_N(\mathbf{y}) P_N(\mathbf{y}, \mathbf{x})$ [52, 64]. Therefore detailed balance holds, and summing over $\mathbf{x}$ shows that π_N is invariant [10]. ■

The continuous-time generator $\mathcal{L}_N = \tau_N(P_N - I)$ is the object used in the asymptotic analysis. Therefore, discrete-time aperiodicity is not needed for the generator convergence argument. What is needed is irreducibility on the connected graph and reversibility of the associated generator.

Theorem 6 (Finite-Graph Reversibility and Stationarity) *On the connectivity event guaranteed by Lemma 1, the continuous-time Markov chain generated by*

$$\mathcal{L}_N = \tau_N(P_N - I)$$

is irreducible and reversible with respect to π_N. Consequently, π_N is the unique stationary distribution of the finite graph dynamics.

Proof: Since $\mathcal{G}_N$ is connected on the event under consideration, for any two vertices $\mathbf{x}, \mathbf{y} \in \mathcal{V}_N$ there exists a path

$$\mathbf{x} = \mathbf{x}_{i_0}, \mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k} = \mathbf{y}$$

with $(\mathbf{x}_{i_\ell}, \mathbf{x}_{i_{\ell+1}}) \in \mathcal{E}_N$ for every ℓ. Along each edge, the Gibbs weight in Eq. A.2 is strictly positive, hence $P_N(\mathbf{x}_{i_\ell}, \mathbf{x}_{i_{\ell+1}}) > 0$. Thus the finite graph chain is irreducible.

By Lemma 2, P_N is reversible with respect to π_N. The same detailed balance identity implies that the continuous-time generator $\tau_N(P_N - I)$ is self-adjoint in $L^2(\mathcal{V}_N, \pi_N)$ [17]. Since the state space is finite and the chain is irreducible, the invariant distribution is unique [44, 64]. This proves the claim. ■

Finally, we connect the discrete invariant measure to the limiting measure used in the averaged representation dynamics. By Theorem 7, the local partition factor satisfies

$$Z_N(\mathbf{x}) = \mathcal{K}_N \left(1 + \mathcal{O}(\varepsilon_N^2) + \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log N}{N \varepsilon_N^d}} \right) \right)$$

uniformly on the high-probability concentration event. Hence the leading non-uniform weighting in Eq. A.4 is the Gibbs factor $\exp(-\beta U(\mathbf{x}))$. In the continuum limit, the corresponding invariant measure is

$$d\pi(\mathbf{x}) = \frac{\exp(-\beta U(\mathbf{x}))}{\int_{\mathcal{M}} \exp(-\beta U(\mathbf{z}))\, dV_g(\mathbf{z})} dV_g(\mathbf{x}).$$

This is the invariant spatial measure used in Section 3.1 to define the induced covariance Σ_π in Eq. 9.

A.3 Riemannian Taylor Expansion and Covariant Remainder Bounds

Roadmap of the Proof Strategy: Sections A.3–A.5 derive the local asymptotic expansion needed to prove convergence of the empirical graph generator to the Kolmogorov backward generator in Theorem 1. The key step is to express neighboring sample points through Riemannian normal coordinates and to control the Taylor remainders uniformly on the compact manifold. Readers primarily interested in the final continuum limit may proceed directly to Section A.6.

Having defined the reversible local transition kernel, we now evaluate the asymptotic action of the empirical infinitesimal generator $\mathcal{L}_N$ on a fixed test function $f \in C^3(\mathcal{M})$. This requires expanding both $f(\mathbf{y})$ and the potential $U(\mathbf{y})$ for local neighbors $\mathbf{y} \in \mathcal{N}_N(\mathbf{x})$.

Unlike Euclidean spaces, where local displacements can be written as vector differences, transitions on $\mathcal{M}$ must be represented along geodesics. We therefore pull local neighborhoods back to the tangent space through the exponential map.

Definition 3 (Injectivity Radius and Riemannian Normal Coordinates) *Let $T_{\mathbf{x}}\mathcal{M}$ denote the tangent space at $\mathbf{x} \in \mathcal{M}$. The exponential map*

$$\exp_{\mathbf{x}} : T_{\mathbf{x}}\mathcal{M} \to \mathcal{M}$$

maps a tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}$ to $\gamma(1)$, where $\gamma : [0,1] \to \mathcal{M}$ is the unique constant-speed geodesic satisfying

$$\gamma(0) = \mathbf{x}, \qquad \dot{\gamma}(0) = \mathbf{v}$$

[24].

By the Hopf–Rinow theorem [63], compactness and connectedness imply geodesic completeness. Moreover, compactness gives a positive global injectivity radius [83],

$$r_{\mathrm{inj}}(\mathcal{M}) = \inf_{\mathbf{x}\in\mathcal{M}} \sup\{r > 0 : \exp_{\mathbf{x}} \text{ is a diffeomorphism on } B(0,r) \subset T_{\mathbf{x}}\mathcal{M}\} > 0.$$

Since $\varepsilon_N \to 0$, for all sufficiently large N we have $\varepsilon_N < r_{\mathrm{inj}}(\mathcal{M})$. Thus, for every neighbor $\mathbf{y} \in \mathcal{N}_N(\mathbf{x})$, there exists a unique tangent vector $\mathbf{v}_{\mathbf{xy}} \in T_{\mathbf{x}}\mathcal{M}$ such that

$$\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v}_{\mathbf{xy}}), \qquad \|\mathbf{v}_{\mathbf{xy}}\|_g = d_{\mathcal{M}}(\mathbf{x},\mathbf{y}) \le \varepsilon_N.$$

This defines valid Riemannian normal coordinates for all local neighborhoods used in the empirical generator [63].

To state the Taylor expansion, we recall the differential operators acting on scalar fields $f \in C^3(\mathcal{M})$.

Definition 4 (Covariant Derivatives and Tensors) *Let ∇ denote the Levi–Civita connection associated with the metric g [49].*

- ***Riemannian Gradient:*** *The gradient $\nabla f(\mathbf{x}) \in T_{\mathbf{x}}\mathcal{M}$ is the unique vector satisfying*

  $$g(\nabla f, \mathbf{A}) = df(\mathbf{A})$$

  for every tangent vector $\mathbf{A} \in T_{\mathbf{x}}\mathcal{M}$ [99].

- ***Riemannian Hessian:*** *The Hessian $\nabla^2 f(\mathbf{x})$ is the symmetric $(0,2)$-tensor*

  $$\nabla^2 f(\mathbf{x})(\mathbf{A},\mathbf{B}) = (\nabla_{\mathbf{A}} df)(\mathbf{B}) = \mathbf{A}(\mathbf{B}f) - (\nabla_{\mathbf{A}}\mathbf{B})f = \langle \nabla_{\mathbf{A}}\nabla f, \mathbf{B}\rangle_g$$

  for $\mathbf{A},\mathbf{B} \in T_{\mathbf{x}}\mathcal{M}$ [63].

- ***Third Covariant Derivative:*** *The third derivative $\nabla^3 f(\mathbf{x})$ is the $(0,3)$-tensor obtained by covariantly differentiating $\nabla^2 f$.*

Figure A.2 visualizes the local pullback $\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v})$ used below: graph neighbors on $\mathcal{M}$ are represented as tangent vectors in $T_{\mathbf{x}}\mathcal{M}$ before applying the Taylor expansion in Lemma 3.

Using these coordinates, we obtain the local expansion with a uniform third-order remainder.

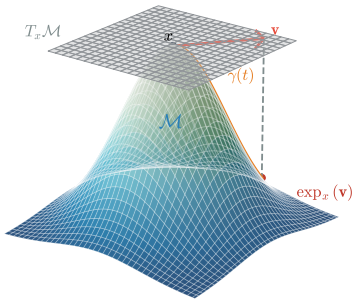

Figure A.2: **Riemannian Normal Coordinates.** The exponential map $\exp_{\mathbf{x}}$ identifies a sufficiently small geodesic neighborhood of $\mathbf{x}$ with a ball in $T_{\mathbf{x}}\mathcal{M}$, enabling the local Taylor expansion used in the generator convergence proof.

Lemma 3 (Third-Order Riemannian Taylor Expansion) *Let $f \in C^3(\mathcal{M}, \mathbb{R})$. For any $\mathbf{x} \in \mathcal{M}$ and any tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}$ satisfying $\|\mathbf{v}\|_g \leq \varepsilon_N < r_{\text{inj}}(\mathcal{M})$, define $\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v})$. Then [83, 49]*

$$f(\mathbf{y}) = f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle_g + \frac{1}{2}\nabla^2 f(\mathbf{x})(\mathbf{v}, \mathbf{v}) + R_f(\mathbf{x}, \mathbf{v}), \tag{A.8}$$

where the remainder satisfies the uniform bound [1]

$$|R_f(\mathbf{x}, \mathbf{v})| \leq \frac{1}{6}\left(\sup_{\mathbf{z}\in\mathcal{M}} \|\nabla^3 f(\mathbf{z})\|_{\text{op}}\right) \|\mathbf{v}\|_g^3 = \mathcal{O}(\varepsilon_N^3). \tag{A.9}$$

Proof: Let

$$\gamma(t) = \exp_{\mathbf{x}}(t\mathbf{v}), \qquad t \in [0, 1],$$

be the unique geodesic from $\mathbf{x}$ to $\mathbf{y}$, and define $F(t) = f(\gamma(t))$. Applying the one-dimensional Taylor formula to $F(1)$ gives [89]

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(0) + \frac{1}{6}F'''(\xi), \qquad \xi \in (0, 1). \tag{A.10}$$

We compute these derivatives using covariant differentiation along γ [1].

First derivative.

$$F'(t) = df(\dot\gamma(t)) = \langle \nabla f(\gamma(t)), \dot\gamma(t) \rangle_g. \tag{A.11}$$

At $t = 0$, since $\gamma(0) = \mathbf{x}$ and $\dot\gamma(0) = \mathbf{v}$,

$$F'(0) = \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle_g.$$

Second derivative. Using metric compatibility of the Levi–Civita connection [49],

$$\begin{aligned} F''(t) &= \frac{d}{dt}\langle \nabla f(\gamma(t)), \dot\gamma(t) \rangle_g \\ &= \langle \nabla_{\dot\gamma} \nabla f(\gamma(t)), \dot\gamma(t) \rangle_g + \langle \nabla f(\gamma(t)), \nabla_{\dot\gamma}\dot\gamma(t) \rangle_g. \end{aligned} \tag{A.12}$$

Because γ is an affinely parameterized geodesic,

$$\nabla_{\dot\gamma}\dot\gamma(t) = 0$$

for all t [83]. Hence

$$F''(t) = \nabla^2 f(\gamma(t))\big(\dot\gamma(t), \dot\gamma(t)\big),$$

and therefore

$$F''(0) = \nabla^2 f(\mathbf{x})(\mathbf{v}, \mathbf{v})$$

[24].

Third derivative and remainder. Differentiating the Hessian term along the geodesic and using the Leibniz rule for tensor fields gives [1]

$$\begin{aligned} F'''(t) &= \frac{d}{dt}\left[\nabla^2 f(\gamma(t))\big(\dot\gamma(t),\dot\gamma(t)\big)\right] \\ &= (\nabla_{\dot\gamma}\nabla^2 f)\big(\dot\gamma(t),\dot\gamma(t)\big) + 2\nabla^2 f\big(\nabla_{\dot\gamma}\dot\gamma(t),\dot\gamma(t)\big). \end{aligned} \tag{A.13}$$

The geodesic equation removes the second term. Thus

$$F'''(t) = \nabla^3 f(\gamma(t))\big(\dot\gamma(t),\dot\gamma(t),\dot\gamma(t)\big) \quad [49]. \tag{A.14}$$

Let

$$\|\nabla^3 f(\mathbf{z})\|_{\mathrm{op}} = \sup_{\|\mathbf{u}\|_g=1} \left|\nabla^3 f(\mathbf{z})(\mathbf{u},\mathbf{u},\mathbf{u})\right|$$

be the induced tensor operator norm [44]. Geodesics have constant speed, so

$$\|\dot\gamma(\xi)\|_g = \|\mathbf{v}\|_g.$$

Since $f \in C^3(\mathcal{M})$ and $\mathcal{M}$ is compact,

$$C_f = \sup_{\mathbf{z}\in\mathcal{M}} \|\nabla^3 f(\mathbf{z})\|_{\mathrm{op}} < \infty$$

[63, 89]. Therefore

$$|R_f(\mathbf{x},\mathbf{v})| = \frac{1}{6}|F'''(\xi)| \le \frac{C_f}{6}\|\mathbf{v}\|_g^3 \le \frac{C_f}{6}\varepsilon_N^3.$$

The bound is uniform over $\mathbf{x} \in \mathcal{M}$, completing the proof. ■

A.4 Asymptotic Expansion of the Partition Function

Before computing the limiting action of the infinitesimal generator $\mathcal{L}_N$, we control the local normalization factor $Z_N(\mathbf{x})$ in Definition 2. This step is needed because the transition probabilities are locally normalized, and the expansion of $Z_N(\mathbf{x})$ determines which terms survive in the normalized generator.

We work in Riemannian normal coordinates at a pole $\mathbf{x} \in \mathcal{M}$. For a local neighbor $\mathbf{y} \in \mathcal{N}_N(\mathbf{x})$, write

$$\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v}), \qquad \mathbf{v} \in T_{\mathbf{x}}\mathcal{M}, \qquad \|\mathbf{v}\|_g \le \varepsilon_N.$$

In this chart, $T_{\mathbf{x}}\mathcal{M}$ is identified locally with $\mathbb{R}^d$, the metric at the pole satisfies $g_{ij}(\mathbf{x}) = \delta_{ij}$, and the Christoffel symbols vanish at the pole [63]. The Riemannian volume element pulled back to $T_{\mathbf{x}}\mathcal{M}$ is

$$dV_g(\exp_{\mathbf{x}}(\mathbf{v})) = \theta(\mathbf{x},\mathbf{v})\, d\mathbf{v},$$

where the volume density satisfies the classical expansion [36, 83]

$$\theta(\mathbf{x},\mathbf{v}) = 1 - \frac{1}{6}\mathrm{Ric}_{\mathbf{x}}(\mathbf{v},\mathbf{v}) + \mathcal{O}(\|\mathbf{v}\|_g^3). \tag{A.15}$$

We first record the local moment identities used repeatedly in the generator expansion.

Lemma 4 (Moments on the Tangent Space Ball) *Let $d\mathbf{v}$ denote the Lebesgue measure on $T_{\mathbf{x}}\mathcal{M} \cong \mathbb{R}^d$, and let $V_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}$ be the volume of the Euclidean unit ball [30]. By spherical symmetry, all odd-degree monomials integrate to zero over $B_g(0,\varepsilon_N)$. Moreover [36, 83],*

$$\int_{B_g(0,\varepsilon_N)} \theta(\mathbf{x},\mathbf{v})\, d\mathbf{v} = V_d \varepsilon_N^d \left(1 - \frac{R(\mathbf{x})}{6(d+2)}\varepsilon_N^2\right) + \mathcal{O}(\varepsilon_N^{d+4}), \quad [49] \tag{A.16}$$

$$\int_{B_g(0,\varepsilon_N)} v^i v^j \theta(\mathbf{x},\mathbf{v})\, d\mathbf{v} = \frac{V_d}{d+2}\varepsilon_N^{d+2} g^{ij}(\mathbf{x}) + \mathcal{O}(\varepsilon_N^{d+4}), \quad [88] \tag{A.17}$$

where $R(\mathbf{x}) = g^{ij}\mathrm{Ric}_{ij}(\mathbf{x})$ is the scalar curvature.

Proof: Equation A.16 is the standard small-ball volume expansion obtained by substituting Eq. A.15 and integrating in polar coordinates [32]. For Eq. A.17, first consider the unweighted integral

$$I^{ij} = \int_{B_g(0,\varepsilon_N)} v^i v^j \, d\mathbf{v}.$$

By $O(d)$ symmetry of the Euclidean tangent ball, $I^{ij} = 0$ for $i \neq j$, and $I^{ij} = C\delta^{ij}$ on the diagonal [30]. Taking traces gives

$$Cd = \int_{B_g(0,\varepsilon_N)} \|\mathbf{v}\|^2 \, d\mathbf{v} = \int_0^{\varepsilon_N} r^2 (dV_d r^{d-1}) \, dr = \frac{dV_d}{d+2} \varepsilon_N^{d+2},$$

so

$$I^{ij} = \frac{V_d}{d+2} \varepsilon_N^{d+2} \delta^{ij} = \frac{V_d}{d+2} \varepsilon_N^{d+2} g^{ij}(\mathbf{x})$$

in Riemannian normal coordinates [32]. The curvature correction from $-\frac{1}{6}\mathrm{Ric}_{kl} v^k v^l$ contributes fourth-order monomials and is therefore $\mathcal{O}(\varepsilon_N^{d+4})$. ■

Figure A.3 gives a geometric interpretation of Lemma 4: the Euclidean tangent-ball moments provide the leading term, while curvature enters through the volume-density correction in Eq. A.15.

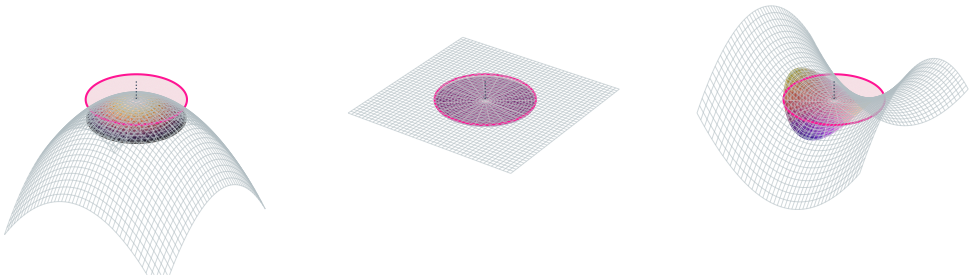

Figure A.3: **Geometric Distortion of the Local Volume Measure.** The tangent-space expansion accounts for the curvature-dependent distortion between a geodesic ball and its Euclidean tangent approximation. Lemma 4 quantifies the leading curvature correction through the scalar curvature $R(\mathbf{x})$.

We next connect empirical sums over $\mathcal{N}_N(\mathbf{x})$ to their tangent-space integrals.

Lemma 5 (Empirical Measure Concentration on Local Neighborhoods) *Let $\psi : \mathcal{M} \to \mathbb{R}$ be a bounded C^2 function appearing in the local partition or generator expansion [28]. Since the vertices are sampled i.i.d. from $d\mu = dV_g/\mathrm{Vol}_g(\mathcal{M})$ [26], the neighborhood sum satisfies*

$$\sum_{\mathbf{y} \in \mathcal{N}_N(\mathbf{x})} \psi(\mathbf{y}) = \frac{N}{\mathrm{Vol}_g(\mathcal{M})} \int_{B_g(0,\varepsilon_N)} \psi(\exp_{\mathbf{x}}(\mathbf{v})) \theta(\mathbf{x}, \mathbf{v}) \, d\mathbf{v} + \mathcal{E}_N(\mathbf{x}), \tag{A.18}$$

where the stochastic fluctuation satisfies, for each fixed $\mathbf{x}$,

$$\mathcal{E}_N(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}\left(\sqrt{N\varepsilon_N^d}\right).$$

The corresponding uniform version over $\mathbf{x} \in \mathcal{M}$ satisfies

$$\sup_{\mathbf{x} \in \mathcal{M}} |\mathcal{E}_N(\mathbf{x})| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{N\varepsilon_N^d \log N}\right),$$

under the covering argument used in Lemma 1 [11, 95].

Proof: The sum can be written as

$$\sum_{i=1}^{N} \psi(\mathbf{x}_i) \mathbb{I}_{\{\mathbf{x}_i \in B_{\mathcal{M}}(\mathbf{x}, \varepsilon_N)\}}.$$

Its expectation equals

$$N \int_{B_{\mathcal{M}}(\mathbf{x},\varepsilon_N)} \psi(\mathbf{y})\, d\mu(\mathbf{y})$$

[87]. Pulling this integral back through $\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v})$ gives the deterministic integral term [97]. Since ψ is bounded,

$$\mathrm{Var}\left[\sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \psi(\mathbf{y})\right] \le N\|\psi\|_\infty^2 \mu(B_{\mathcal{M}}(\mathbf{x},\varepsilon_N)) = \mathcal{O}(N\varepsilon_N^d).$$

Chebyshev's inequality yields the stated fixed-$\mathbf{x}$ fluctuation scale [9]. The logarithmic factor in the uniform bound comes from taking a union bound over an ε_N-net of the compact manifold and extending the estimate to all $\mathbf{x} \in \mathcal{M}$ by the uniform Lipschitz control of the local kernel, as in Lemma 1. ■

We now expand the local partition function. This expansion is used only to normalize the transition probabilities; its curvature and potential corrections are lower-order in the final generator limit.

Theorem 7 (Asymptotic Expansion of the Partition Function [41, 93]) *Under the scaling regime of Lemma 1, the partition function satisfies, uniformly on the high-probability concentration event,*

$$Z_N(\mathbf{x}) = \mathcal{K}_N\left[1 + \frac{\varepsilon_N^2}{8(d+2)}\left(\beta^2\|\nabla U(\mathbf{x})\|_g^2 - 2\beta\Delta_{\mathcal{M}}U(\mathbf{x}) - \frac{4}{3}R(\mathbf{x})\right) + \mathcal{O}(\varepsilon_N^3)\right] + \mathcal{O}_{\mathbb{P}}\left(\sqrt{N\varepsilon_N^d \log N}\right), \tag{A.19}$$

where

$$\mathcal{K}_N = \frac{NV_d\varepsilon_N^d}{\mathrm{Vol}_g(\mathcal{M})}.$$

Equivalently, after division by $\mathcal{K}_N$, the uniform relative stochastic fluctuation is of order

$$\mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N\varepsilon_N^d}}\right).$$

Proof: By Definition 2,

$$Z_N(\mathbf{x}) = \sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \exp\left(-\frac{\beta}{2}[U(\mathbf{y}) - U(\mathbf{x})]\right).$$

Write

$$\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v}), \qquad \Delta U(\mathbf{v}) = U(\exp_{\mathbf{x}}(\mathbf{v})) - U(\mathbf{x}).$$

By Lemma 3,

$$\Delta U(\mathbf{v}) = \langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g + \frac{1}{2}\nabla^2 U(\mathbf{x})(\mathbf{v},\mathbf{v}) + \mathcal{O}(\|\mathbf{v}\|_g^3).$$

Expanding the Gibbs weight gives [89]

$$\begin{aligned}\exp\left(-\frac{\beta}{2}\Delta U(\mathbf{v})\right) = 1 - \frac{\beta}{2}\langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g - \frac{\beta}{4}\nabla^2 U(\mathbf{x})(\mathbf{v},\mathbf{v}) \\ + \frac{\beta^2}{8}\langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g^2 + \mathcal{O}(\|\mathbf{v}\|_g^3).\end{aligned} \tag{A.20}$$

Applying Lemma 5, the expectation of $Z_N(\mathbf{x})$ is

$$\frac{N}{\mathrm{Vol}_g(\mathcal{M})}\int_{B_g(0,\varepsilon_N)} \exp\left(-\frac{\beta}{2}\Delta U(\mathbf{v})\right)\theta(\mathbf{x},\mathbf{v})\, d\mathbf{v}.$$

We integrate Eq. A.20 term by term using Lemma 4.

First, the zeroth-order term gives

$$V_d\varepsilon_N^d\left(1 - \frac{R(\mathbf{x})}{6(d+2)}\varepsilon_N^2\right) + \mathcal{O}(\varepsilon_N^{d+4}).$$

Second, the first-order term integrates to zero by odd symmetry:

$$\int \left(-\frac{\beta}{2}\langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g\right) \theta(\mathbf{x}, \mathbf{v})\, d\mathbf{v} = 0.$$

Third, the Hessian term gives [88]

$$-\frac{\beta}{4}(\nabla^2 U)_{ij} \int v^i v^j\, d\mathbf{v} = -\frac{\beta V_d}{4(d+2)} \varepsilon_N^{d+2} \Delta_{\mathcal{M}} U(\mathbf{x}).$$

Finally, the squared-gradient term gives [36]

$$\frac{\beta^2}{8}(\partial_i U)(\partial_j U) \int v^i v^j\, d\mathbf{v} = \frac{\beta^2 V_d}{8(d+2)} \varepsilon_N^{d+2} \|\nabla U(\mathbf{x})\|_g^2.$$

Combining these contributions,

$$\mathbb{E}[Z_N(\mathbf{x})] = \frac{N V_d \varepsilon_N^d}{\mathrm{Vol}_g(\mathcal{M})} \left[1 - \frac{R(\mathbf{x})}{6(d+2)}\varepsilon_N^2 - \frac{\beta \Delta_{\mathcal{M}} U(\mathbf{x})}{4(d+2)}\varepsilon_N^2 + \frac{\beta^2 \|\nabla U(\mathbf{x})\|_g^2}{8(d+2)}\varepsilon_N^2\right] + \mathcal{O}\left(N\varepsilon_N^{d+3}\right) \quad [59, 88, 36]. \tag{A.21}$$

Factoring out $\mathcal{K}_N = N V_d \varepsilon_N^d / \mathrm{Vol}_g(\mathcal{M})$ gives Eq. A.19. The stochastic fluctuation term follows from the uniform version of Lemma 5. The logarithmic factor comes from taking a union bound over an ε_N-net of the compact manifold and extending the estimate to all $\mathbf{x} \in \mathcal{M}$ by the uniform Lipschitz control of the local kernel. ■

Remark on Normalization. The leading term of $Z_N(\mathbf{x})$ is the isotropic local neighborhood volume $\mathcal{K}_N$. The potential and curvature corrections enter at order ε_N^2, while the relative empirical fluctuation is controlled by the scaling law. In the generator expansion, these normalization corrections multiply second-order local increments and therefore contribute only lower-order remainders to the limiting Kolmogorov backward generator.

A.5 Convergence of the Discrete Infinitesimal Generator

We now combine the local Taylor expansion, the partition-function estimate, and the empirical concentration bounds to prove the generator-level convergence stated in Theorem 1. The object analyzed here is the continuous-time generator obtained from the localized transition kernel, not a separate discrete-time mixing limit.

Definition 5 (Time-Scaled Infinitesimal Generator) *To recover a macroscopic diffusion process with diffusion coefficient $D > 0$ [50], we choose the jump rate*

$$\tau_N = \frac{2D(d+2)}{\varepsilon_N^2}. \tag{A.22}$$

The factor $(d+2)$ calibrates the second moment of the uniform tangent ball [18], while the factor 2 matches the standard quadratic-variation normalization of Brownian motion [50]. For a fixed test function $f \in C^3(\mathcal{M})$, the empirical infinitesimal generator is

$$\mathcal{L}_N f(x) = \tau_N \sum_{y \in \mathcal{X}_N} P_N(x, y)\,[f(y) - f(x)], \qquad x \in \mathcal{M}, \tag{A.23}$$

where P_N is the continuum extension of the local Gibbs kernel in Definition 2 [28].

This scaling depends on the intrinsic dimension d and the local radius ε_N, and the resulting theorem supplies the discrete-to-continuous kinematic limit used as the fast process in Section 3.

Theorem 8 (Convergence to the Langevin Generator, Restating Theorem 1) *Let $\mathcal{M}$ satisfy the assumptions in Section A.1. For every fixed $f \in C^3(\mathcal{M})$, as $N \to \infty$ and $\varepsilon_N \to 0$ under the scaling regime $N\varepsilon_N^{d+2} \gg \log N$, the empirical generator $\mathcal{L}_N f$ converges uniformly in probability to the Kolmogorov backward generator of the limiting Langevin diffusion [45, 80]:*

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{M}} |\mathcal{L}_N f(\mathbf{x}) - \mathcal{L} f(\mathbf{x})| > \delta\right) = 0, \qquad \forall \delta > 0, \tag{A.24}$$

where

$$\mathcal{L} f(\mathbf{x}) = -D\beta \langle \nabla U(\mathbf{x}), \nabla f(\mathbf{x})\rangle_g + D\Delta_{\mathcal{M}} f(\mathbf{x}). \tag{A.25}$$

Proof: Fix $\mathbf{x} \in \mathcal{M}$ and write each local neighbor as

$$\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v}), \qquad \mathbf{v} \in T_{\mathbf{x}}\mathcal{M}, \qquad \|\mathbf{v}\|_g \le \varepsilon_N.$$

Using Definition 2, the generator can be written as

$$\mathcal{L}_N f(\mathbf{x}) = \frac{\tau_N}{Z_N(\mathbf{x})} \sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \exp\left(-\frac{\beta}{2}\left[U(\mathbf{y}) - U(\mathbf{x})\right]\right)\left[f(\mathbf{y}) - f(\mathbf{x})\right]. \tag{A.26}$$

Equivalently, under the pullback $\mathbf{y} = \exp_{\mathbf{x}}(\mathbf{v})$, the summation is over tangent vectors associated with the local empirical neighborhood [83, 93].

Step 1: Local cross-expansion. By Lemma 3,

$$f(\exp_{\mathbf{x}}(\mathbf{v})) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{v}\rangle_g + \frac{1}{2}\nabla^2 f(\mathbf{x})(\mathbf{v},\mathbf{v}) + \mathcal{O}(\|\mathbf{v}\|_g^3). \tag{A.27}$$

Similarly, Eq. A.20 gives

$$\exp\left(-\frac{\beta}{2}[U(\exp_{\mathbf{x}}(\mathbf{v})) - U(\mathbf{x})]\right) = 1 - \frac{\beta}{2}\langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g + \mathcal{O}(\|\mathbf{v}\|_g^2). \tag{A.28}$$

Multiplying these two expansions and retaining all terms contributing at order ε_N^2 after normalization yields [1]

$$\begin{aligned}
&\exp\left(-\frac{\beta}{2}[U(\exp_{\mathbf{x}}(\mathbf{v})) - U(\mathbf{x})]\right)[f(\exp_{\mathbf{x}}(\mathbf{v})) - f(\mathbf{x})] \\
&\quad = \langle \nabla f(\mathbf{x}), \mathbf{v}\rangle_g - \frac{\beta}{2}\langle \nabla U(\mathbf{x}), \mathbf{v}\rangle_g \langle \nabla f(\mathbf{x}), \mathbf{v}\rangle_g + \frac{1}{2}\nabla^2 f(\mathbf{x})(\mathbf{v},\mathbf{v}) + \mathcal{R}(\mathbf{x},\mathbf{v}),
\end{aligned} \tag{A.29}$$

where the remainder satisfies

$$|\mathcal{R}(\mathbf{x},\mathbf{v})| \le C_f \|\mathbf{v}\|_g^3$$

uniformly over $\mathbf{x} \in \mathcal{M}$, for a constant C_f depending on uniform C^3 bounds of f and U.

Step 2: Moment evaluation. Applying Lemma 5, the local empirical sums are replaced by tangent-space integrals, up to the uniform stochastic fluctuations controlled by Lemma 1. The first-order term integrates to zero by odd symmetry of the tangent ball [36, 32]:

$$\sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \langle \nabla f(\mathbf{x}), \mathbf{v}_{\mathbf{xy}}\rangle_g = \text{empirical fluctuation}.$$

After multiplication by the prefactor $\tau_N/Z_N(\mathbf{x})$, this empirical first-moment fluctuation is controlled by the same uniform graph-operator concentration rate used below. The two quadratic terms give, using Lemma 4,

$$\sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \langle \nabla U(\mathbf{x}), \mathbf{v}_{\mathbf{xy}}\rangle_g \langle \nabla f(\mathbf{x}), \mathbf{v}_{\mathbf{xy}}\rangle_g = \mathcal{K}_N \frac{\varepsilon_N^2}{d+2}\langle \nabla U(\mathbf{x}), \nabla f(\mathbf{x})\rangle_g + \mathcal{E}_N^{(1)}(\mathbf{x}), \quad [49] \tag{A.30}$$

$$\sum_{\mathbf{y}\in\mathcal{N}_N(\mathbf{x})} \nabla^2 f(\mathbf{x})(\mathbf{v}_{\mathbf{xy}}, \mathbf{v}_{\mathbf{xy}}) = \mathcal{K}_N \frac{\varepsilon_N^2}{d+2}\Delta_{\mathcal{M}} f(\mathbf{x}) + \mathcal{E}_N^{(2)}(\mathbf{x}), \quad [88] \tag{A.31}$$

where

$$\mathcal{K}_N = \frac{N V_d \varepsilon_N^d}{\mathrm{Vol}_g(\mathcal{M})}$$

is the deterministic leading local mass [93]. The error processes $\mathcal{E}_N^{(1)}$ and $\mathcal{E}_N^{(2)}$ satisfy the corresponding uniform concentration bounds under $N\varepsilon_N^{d+2} \gg \log N$.

Step 3: Normalization and limiting drift-diffusion terms. By Theorem 7,

$$Z_N(\mathbf{x}) = \mathcal{K}_N\left(1 + \mathcal{O}(\varepsilon_N^2)\right) + \mathcal{O}_{\mathbb{P}}\left(\sqrt{N\varepsilon_N^d \log N}\right),$$

and equivalently, on the uniform high-probability concentration event,

$$\frac{1}{Z_N(\mathbf{x})} = \frac{1}{\mathcal{K}_N}\left(1 + \mathcal{O}(\varepsilon_N^2) + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N\varepsilon_N^d}}\right)\right). \tag{A.32}$$

Combining this with

$$\tau_N = \frac{2D(d+2)}{\varepsilon_N^2},$$

we obtain

$$\frac{\tau_N}{Z_N(\mathbf{x})} = \frac{2D(d+2)}{\varepsilon_N^2\mathcal{K}_N}\left(1 + \mathcal{O}(\varepsilon_N^2) + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N\varepsilon_N^d}}\right)\right). \tag{A.33}$$

Substituting Eqs. A.30–A.31 into Eq. A.29, and then applying the prefactor in Eq. A.33, gives

$$\begin{aligned}\mathcal{L}_N f(\mathbf{x}) &= \frac{2D(d+2)}{\varepsilon_N^2\mathcal{K}_N}\left[\mathcal{K}_N\frac{\varepsilon_N^2}{d+2}\left(-\frac{\beta}{2}\left\langle\nabla U(\mathbf{x}), \nabla f(\mathbf{x})\right\rangle_g + \frac{1}{2}\Delta_{\mathcal{M}} f(\mathbf{x})\right)\right] \\ &\quad + \mathcal{O}(\varepsilon_N) + \text{uniform stochastic fluctuation} \\ &= -D\beta\left\langle\nabla U(\mathbf{x}), \nabla f(\mathbf{x})\right\rangle_g + D\Delta_{\mathcal{M}} f(\mathbf{x}) + \mathcal{O}(\varepsilon_N) + \text{uniform stochastic fluctuation}.\end{aligned} \tag{A.34}$$

Step 4: Vanishing of the remainders. The deterministic remainder $\mathcal{O}(\varepsilon_N)$ vanishes uniformly because $f, U \in C^3(\mathcal{M})$ and $\mathcal{M}$ is compact [83]. The stochastic fluctuation is the empirical second-order graph-operator fluctuation; standard graph-Laplacian consistency estimates on point clouds yield the rate

$$\mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log N}{N\varepsilon_N^{d+2}}}\right)$$

after the uniform covering argument [93, 98]. Under Lemma 1, this term converges to zero in probability. Therefore

$$\sup_{\mathbf{x}\in\mathcal{M}} |\mathcal{L}_N f(\mathbf{x}) - (-D\beta\langle\nabla U(\mathbf{x}), \nabla f(\mathbf{x})\rangle_g + D\Delta_{\mathcal{M}} f(\mathbf{x}))| \xrightarrow{\mathbb{P}} 0.$$

This proves uniform convergence in probability of the empirical generator to the Kolmogorov backward generator of the limiting Langevin diffusion [28, 80]. ■

A.6 Summary of Part: From Local Graph Flow to Langevin Kinematics

The preceding sections establish the spatial continuum limit used in the main text. Section A.1 constructs the random geometric graph sequence as an empirical discretization of the compact Riemannian manifold $(\mathcal{M}, g)$. Section A.2 defines the local Gibbs transition kernel and its reversible invariant measure. Sections A.3–A.4 control the local Riemannian Taylor expansion, curvature-dependent volume distortion, empirical concentration, and partition-function normalization. Combining these ingredients in Section A.5 proves Theorem 8, which is the detailed proof of Theorem 1 in Section 2.

The result is a generator-level convergence statement. For every fixed $f \in C^3(\mathcal{M})$, under the scaling regime

$$N\varepsilon_N^{d+2} \gg \log N,$$

the empirical generator $\mathcal{L}_N f$ converges uniformly in probability to the Kolmogorov backward generator

$$\mathcal{L} f(\mathbf{x}) = -D\beta\langle\nabla U(\mathbf{x}), \nabla f(\mathbf{x})\rangle_g + D\Delta_{\mathcal{M}} f(\mathbf{x}).$$

Equivalently, this is the backward generator associated with the overdamped Langevin diffusion used in the main text:

$$dX_t = -D\beta\nabla U(X_t)\, dt + \sqrt{2D}\, dB_t^{\mathcal{M}}, \tag{A.35}$$

where $B_t^{\mathcal{M}}$ denotes Brownian motion on the Riemannian manifold [45, 78].

Role in the main analysis. This continuum limit justifies replacing the localized graph exploration by the Langevin fast process in the slow–fast representation dynamics. Appendix B then takes Eq. A.35 as the Markovian spatial component and studies the averaged subspace Oja dynamics driven by the induced invariant measure π. Thus Appendix A supplies the fast kinematic substrate, while the subsequent appendices analyze stability, principal-subspace alignment, and decorrelation of the representation parameters.

B Averaged Stability and Principal Subspace Alignment

Appendix A establishes the discrete-to-continuous kinematic limit: the localized graph exploration converges at the generator level to the overdamped Langevin process on the Riemannian manifold $\mathcal{M}$ [45, 80]. This appendix uses that Langevin process as the fast Markovian component driving the local observations. Building on that kinematic limit, this appendix analyzes the slow representation dynamics induced by the Markovian observation stream.

We proceed in two steps. First, stochastic averaging is used to replace the Markovian observation stream by its invariant spatial average, yielding the deterministic averaged subspace Oja flow [10, 56, 80]. Second, a row-orthogonality Lyapunov functional and a LaSalle-type invariant-set argument characterize the limiting row-space geometry of the deterministic averaged flow [54, 60, 55]. This appendix therefore supports Theorem 2 and Theorem 3 in the main text.

B.1 State-Parameter Embedding and the Spatial Covariance Operator

Roadmap of the proof strategy. Sections B.1–B.2 formalize how the fast Langevin state generates the observation stream entering the slow subspace Oja dynamics. Section B.3 then proves row-orthogonality dissipation for the averaged flow, and Section B.4 uses the resulting invariant-set structure to identify the principal eigenspace selected by the dynamics.

We first specify the map from the Riemannian state space to the Euclidean feature space used by the representation update. Let

$$\Phi : \mathcal{M} \hookrightarrow \mathbb{R}^n$$

be the smooth isometric embedding used in the main text [63, 73]. For an agent located at $X_t \in \mathcal{M}$, the local Euclidean observation is

$$\mathbf{x}_t = \Phi(X_t) \in \mathbb{R}^n.$$

The representation matrix

$$\mathbf{W}_t \in \mathbb{R}^{m \times n}$$

maps this observation to the latent coordinate

$$\mathbf{y}_t = \mathbf{W}_t \mathbf{x}_t \in \mathbb{R}^m.$$

Thus the slow learning rule analyzed in this appendix is the subspace Oja/Sanger-type update with lateral inhibition,

$$d\mathbf{W}_t = \epsilon\gamma \left(\mathbf{y}_t \mathbf{x}_t^T - \mathbf{y}_t \mathbf{y}_t^T \mathbf{W}_t \right) dt,$$

which matches Eq. 8 in the main text [76, 77, 81, 91]. The term $\mathbf{y}_t \mathbf{y}_t^T \mathbf{W}_t$ is the lateral-inhibition component responsible for controlling row-space degeneracy.

Since the slow update depends on the Euclidean observation $\mathbf{x}_t = \Phi(X_t)$ rather than on X_t directly, averaging the Markovian input stream requires pushing the invariant spatial measure π through the embedding Φ.

Definition 6 (Pushforward Measure and the Spatial Covariance Operator) *Let π be the invariant probability measure of the limiting Langevin fast process established in Appendix A. Equivalently,*

$$d\pi(\mathbf{z}) = \frac{\exp(-\beta U(\mathbf{z}))}{\int_{\mathcal{M}} \exp(-\beta U(\mathbf{u}))\, dV_g(\mathbf{u})} dV_g(\mathbf{z}), \qquad \mathbf{z} \in \mathcal{M}, \tag{B.1}$$

where dV_g is the Riemannian volume measure [45, 80, 72].

The embedding Φ induces the pushforward measure $\Phi_{\#}\pi$ on $\mathbb{R}^n$, defined by

$$(\Phi_{\#}\pi)(A) = \pi(\Phi^{-1}(A))$$

for every Borel set $A \subseteq \mathbb{R}^n$ [97]. The induced spatial covariance matrix is

$$\Sigma_\pi = \mathbb{E}_{\mathbf{x} \sim \Phi_{\#}\pi} \left[\mathbf{x}\mathbf{x}^T \right] = \int_{\mathcal{M}} \Phi(\mathbf{z}) \Phi(\mathbf{z})^T d\pi(\mathbf{z}) \in \mathbb{R}^{n \times n}. \tag{B.2}$$

This is the same covariance operator used in Eq. 9 of the main text.

Lemma 6 (Boundedness and Positive Semi-Definiteness of Σ_π) *The spatial covariance matrix Σ_π is finite, symmetric, and positive semi-definite:*

$$\Sigma_\pi = \Sigma_\pi^T \succeq 0.$$

Proof: We verify the three properties directly from Definition 6.

Finiteness. Since $\mathcal{M}$ is compact and Φ is smooth, Φ is bounded on $\mathcal{M}$ by the extreme value theorem [89]. Hence there exists $C_\Phi < \infty$ such that

$$\sup_{\mathbf{z}\in\mathcal{M}} \|\Phi(\mathbf{z})\|_2 \le C_\Phi.$$

Using the Frobenius norm and the triangle inequality for matrix-valued integrals,

$$\begin{aligned}
\|\Sigma_\pi\|_F &= \left\| \int_{\mathcal{M}} \Phi(\mathbf{z})\Phi(\mathbf{z})^T d\pi(\mathbf{z}) \right\|_F \\
&\le \int_{\mathcal{M}} \left\| \Phi(\mathbf{z})\Phi(\mathbf{z})^T \right\|_F d\pi(\mathbf{z}) \\
&= \int_{\mathcal{M}} \|\Phi(\mathbf{z})\|_2^2 d\pi(\mathbf{z}) \\
&\le C_\Phi^2 \int_{\mathcal{M}} d\pi(\mathbf{z}) = C_\Phi^2 < \infty.
\end{aligned} \tag{B.3}$$

Thus Σ_π is well-defined and finite [44].

Symmetry. For every $\mathbf{z} \in \mathcal{M}$,

$$\left(\Phi(\mathbf{z})\Phi(\mathbf{z})^T\right)^T = \Phi(\mathbf{z})\Phi(\mathbf{z})^T.$$

Therefore,

$$\Sigma_\pi^T = \left(\int_{\mathcal{M}} \Phi(\mathbf{z})\Phi(\mathbf{z})^T d\pi(\mathbf{z}) \right)^T = \int_{\mathcal{M}} \Phi(\mathbf{z})\Phi(\mathbf{z})^T d\pi(\mathbf{z}) = \Sigma_\pi.$$

Positive semi-definiteness. For any vector $\mathbf{v} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{v}^T\Sigma_\pi\mathbf{v} &= \mathbf{v}^T \left(\int_{\mathcal{M}} \Phi(\mathbf{z})\Phi(\mathbf{z})^T d\pi(\mathbf{z}) \right) \mathbf{v} \\
&= \int_{\mathcal{M}} \mathbf{v}^T\Phi(\mathbf{z})\Phi(\mathbf{z})^T\mathbf{v} d\pi(\mathbf{z}) \\
&= \int_{\mathcal{M}} \left(\mathbf{v}^T\Phi(\mathbf{z})\right)^2 d\pi(\mathbf{z}) \\
&= \|\langle \mathbf{v}, \Phi(\cdot)\rangle_{\mathbb{R}^n}\|^2_{L^2(\mathcal{M},\pi)} \ge 0.
\end{aligned} \tag{B.4}$$

Hence $\Sigma_\pi \succeq 0$ [44]. ■

Since Σ_π is real symmetric and positive semi-definite, the spectral theorem gives an orthogonal eigendecomposition

$$\Sigma_\pi = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n), \qquad \lambda_1 \ge \cdots \ge \lambda_n \ge 0. \tag{B.5}$$

Let $\mathbf{Q}_m \in \mathbb{R}^{n\times m}$ denote the matrix of the top m eigenvectors. The principal-subspace theorem in the main text assumes the spectral gap

$$\lambda_m > \lambda_{m+1}, \qquad \lambda_m > 0,$$

and studies the convergence of $\mathrm{Row}(\mathbf{W})$ to $\mathrm{Col}(\mathbf{Q}_m)$ up to an orthogonal rotation of the latent coordinates [44, 77, 91]. We do not require the covariance to be full rank in all ambient directions; possible null directions of Σ_π are simply directions with zero second moment under $\Phi_\#\pi$, and the alignment result concerns the nondegenerate top-m spectral subspace selected by the gap condition.

B.2 Two-Time-Scale Separation via the Poisson Resolvent

We now derive the deterministic averaged flow corresponding to the coupled slow–fast system in Eq. 7–8. The fast component is the Langevin process obtained in Appendix A, while the slow component is the subspace Oja/Sanger-type lateral-inhibition dynamics used in the main text [10, 56, 80].

Averaging assumptions. Throughout this subsection, we assume that the fast Langevin diffusion on the compact manifold admits a unique invariant measure π and a spectral gap on centered $L^2(\pi)$. For each $\mathbf{W}$ in the compact sublevel set reached by the slow dynamics, the centered drift $H(\mathbf{W},\cdot) - \bar{H}(\mathbf{W})$ has a Poisson solution $\nu(\mathbf{W},\cdot)$ with bounded first derivatives in $\mathbf{W}$ and bounded derivatives in x. Since $\mathcal{M}$ is compact and Φ is smooth, these requirements reduce to uniform boundedness and Lipschitz continuity of the finite-dimensional matrix drift on the considered set.

Let $\epsilon > 0$ denote the slow learning-rate parameter and let $\gamma > 0$ be the plasticity rate. The coupled system on $\mathcal{M} \times \mathbb{R}^{m\times n}$ is

$$\text{Fast Spatial SDE:} \qquad dX_t = -D\beta\nabla U(X_t)\,dt + \sqrt{2D}\,dB_t^{\mathcal{M}}, \tag{B.6}$$

$$\text{Slow Subspace Oja Dynamics:} \qquad d\mathbf{W}_t = \epsilon\gamma\left(\mathbf{y}_t\mathbf{x}_t^T - \mathbf{y}_t\mathbf{y}_t^T\mathbf{W}_t\right)dt,$$
$$\mathbf{x}_t = \Phi(X_t), \quad \mathbf{y}_t = \mathbf{W}_t\mathbf{x}_t. \tag{B.7}$$

Equivalently, define the instantaneous matrix-valued drift. For compact notation in this appendix, write $H = H_\gamma$ and $\bar{H} = \bar{H}_\gamma$:

$$H(\mathbf{W},z) = \gamma\left[(\mathbf{W}\Phi(z))\Phi(z)^T - (\mathbf{W}\Phi(z))(\mathbf{W}\Phi(z))^T\mathbf{W}\right], \qquad z\in\mathcal{M}. \tag{B.8}$$

Using

$$(\mathbf{W}\Phi(z))(\mathbf{W}\Phi(z))^T\mathbf{W} = \mathbf{W}\Phi(z)\Phi(z)^T\mathbf{W}^T\mathbf{W},$$

we can also write

$$H(\mathbf{W},z) = \gamma\left[\mathbf{W}\Phi(z)\Phi(z)^T - \mathbf{W}\Phi(z)\Phi(z)^T\mathbf{W}^T\mathbf{W}\right]. \tag{B.9}$$

Thus Eq. B.7 becomes

$$d\mathbf{W}_t = \epsilon H(\mathbf{W}_t, X_t)\,dt.$$

The goal of this subsection is to show that, on the macroscopic timescale

$$\tau = \epsilon t,$$

the Markovian fluctuations in $H(\mathbf{W}_t, X_t)$ average out and the slow trajectory follows an autonomous deterministic subspace flow.

Lemma 7 (Deterministic Limit via Itô–Poisson Decomposition) *Let*

$$\mathcal{L} = -D\beta\langle\nabla U,\nabla\cdot\rangle_g + D\Delta_{\mathcal{M}}$$

be the Kolmogorov backward generator of the fast Langevin process in Eq. B.6. Let

$$\mathbf{W}^\epsilon(\tau) = \mathbf{W}_{\tau/\epsilon}$$

be the slow parameter trajectory viewed on the macroscopic timescale. For every finite horizon $\mathcal{T} > 0$, as $\epsilon \to 0$, the locally stopped slow trajectory, with the stopping time introduced in the proof, converges uniformly in probability on $[0,\mathcal{T}]$ to the solution of the averaged ODE [10, 14, 59]

$$\frac{d\bar{\mathbf{W}}(\tau)}{d\tau} = \bar{H}(\bar{\mathbf{W}}(\tau)), \tag{B.10}$$

where

$$\bar{H}(\mathbf{W}) = \int_{\mathcal{M}} H(\mathbf{W},z)\,d\pi(z)$$
$$= \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right). \tag{B.11}$$

Here Σ_π is the induced spatial covariance from Definition 6. The averaged ODE in Eq. B.10 is the same macroscopic flow stated in Eq. 11 of the main text.

Proof: We prove the averaging statement by the Poisson-equation method for Markovian perturbations [10, 56, 79, 80]. The proof is first localized on bounded subsets of parameter space; global boundedness of the averaged trajectory is established by the row-orthogonality Lyapunov functional in Section B.3.

Step 1: Averaged drift and centered fluctuation.

For fixed $\mathbf{W}$, define

$$\bar{H}(\mathbf{W}) = \int_{\mathcal{M}} H(\mathbf{W}, z)\, d\pi(z).$$

Substituting Eq. B.9 and using the definition

$$\Sigma_\pi = \int_{\mathcal{M}} \Phi(z)\Phi(z)^T\, d\pi(z),$$

we obtain

$$\begin{aligned}\bar{H}(\mathbf{W}) &= \gamma \int_{\mathcal{M}} \left[\mathbf{W}\Phi(z)\Phi(z)^T - \mathbf{W}\Phi(z)\Phi(z)^T\mathbf{W}^T\mathbf{W}\right] d\pi(z) \\ &= \gamma \left[\mathbf{W}\left(\int_{\mathcal{M}} \Phi(z)\Phi(z)^T\, d\pi(z)\right) - \mathbf{W}\left(\int_{\mathcal{M}} \Phi(z)\Phi(z)^T\, d\pi(z)\right)\mathbf{W}^T\mathbf{W}\right] \\ &= \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right). \end{aligned} \tag{B.12}$$

This is the subspace Oja/Sanger averaged flow, not the scalar norm-normalized Oja flow [76, 77, 91].

Define the centered Markovian fluctuation

$$\widetilde{H}(\mathbf{W}, z) = H(\mathbf{W}, z) - \bar{H}(\mathbf{W}). \tag{B.13}$$

By construction,

$$\int_{\mathcal{M}} \widetilde{H}(\mathbf{W}, z)\, d\pi(z) = \mathbf{0}_{m\times n}. \tag{B.14}$$

Step 2: Poisson resolvent for the fast generator.

For each fixed $\mathbf{W}$, we introduce a matrix-valued corrector

$$\nu(\mathbf{W}, z) \in \mathbb{R}^{m\times n}$$

solving the Poisson equation

$$\mathcal{L}\nu(\mathbf{W}, z) = -\widetilde{H}(\mathbf{W}, z) = -\left(H(\mathbf{W}, z) - \bar{H}(\mathbf{W})\right). \tag{B.15}$$

The solvability condition is exactly Eq. B.14. Since the Langevin process has invariant measure $d\pi \propto e^{-\beta U} dV_g$, the generator $\mathcal{L}$ is symmetric in $L^2(\mathcal{M}, \pi)$:

$$\int_{\mathcal{M}} f(\mathcal{L}h)\, d\pi = -D\int_{\mathcal{M}} \langle \nabla f, \nabla h\rangle_g\, d\pi = \int_{\mathcal{M}} (\mathcal{L}f)h\, d\pi. \tag{B.16}$$

On a compact boundaryless manifold, ellipticity of $\mathcal{L}$ and the zero-mean condition yield a smooth corrector, unique after imposing

$$\int_{\mathcal{M}} \nu(\mathbf{W}, z)\, d\pi(z) = \mathbf{0}_{m\times n},$$

by the Fredholm alternative and standard Poisson-equation regularity [29, 45, 79]. On every bounded parameter set

$$\mathcal{B}_R = \{\mathbf{W} \in \mathbb{R}^{m\times n} : \|\mathbf{W}\|_F \le R\},$$

the compactness of $\mathcal{B}_R \times \mathcal{M}$ and smoothness of H imply uniform bounds on ν, $\partial_{\mathbf{W}}\nu$, and $d_z\nu$.

Step 3: Itô expansion of the corrector.

Apply Itô's formula to the matrix-valued function

$$\nu(\mathbf{W}_t, X_t) \quad [45,\ 78].$$

The parameter process $\mathbf{W}_t$ has finite variation, so no second-order term in $\mathbf{W}$ appears. Writing $\mathcal{L}_z$ for the fast generator acting on the manifold variable and $\partial_{\mathbf{W}}\nu[\cdot]$ for the Fréchet derivative in the matrix variable, we obtain

$$\begin{aligned} d\nu(\mathbf{W}_t, X_t) &= \mathcal{L}_z\nu(\mathbf{W}_t, X_t)\,dt + \partial_{\mathbf{W}}\nu(\mathbf{W}_t, X_t)[d\mathbf{W}_t] + \sqrt{2D}\,d_z\nu(\mathbf{W}_t, X_t)[dB_t^{\mathcal{M}}] \\ &= \mathcal{L}_z\nu(\mathbf{W}_t, X_t)\,dt + \epsilon\,\partial_{\mathbf{W}}\nu(\mathbf{W}_t, X_t)[H(\mathbf{W}_t, X_t)]\,dt + \sqrt{2D}\,d_z\nu(\mathbf{W}_t, X_t)[dB_t^{\mathcal{M}}] \\ &= -\widetilde{H}(\mathbf{W}_t, X_t)\,dt + \epsilon\,\partial_{\mathbf{W}}\nu(\mathbf{W}_t, X_t)[H(\mathbf{W}_t, X_t)]\,dt + \sqrt{2D}\,d_z\nu(\mathbf{W}_t, X_t)[dB_t^{\mathcal{M}}] \\ &= \big(\bar{H}(\mathbf{W}_t) - H(\mathbf{W}_t, X_t)\big)\,dt \\ &\quad + \epsilon\,\partial_{\mathbf{W}}\nu(\mathbf{W}_t, X_t)[H(\mathbf{W}_t, X_t)]\,dt + \sqrt{2D}\,d_z\nu(\mathbf{W}_t, X_t)[dB_t^{\mathcal{M}}]. \end{aligned} \tag{B.17}$$

Rearranging Eq. B.17 gives the exact identity

$$\begin{aligned} H(\mathbf{W}_t, X_t)\,dt &= \bar{H}(\mathbf{W}_t)\,dt - d\nu(\mathbf{W}_t, X_t) \\ &\quad + \epsilon\,\partial_{\mathbf{W}}\nu(\mathbf{W}_t, X_t)[H(\mathbf{W}_t, X_t)]\,dt + \sqrt{2D}\,d_z\nu(\mathbf{W}_t, X_t)[dB_t^{\mathcal{M}}]. \end{aligned} \tag{B.18}$$

Step 4: Integral rescaling on the macroscopic time axis.

Fix $R > 0$ and define the stopping time

$$\tau_R = \inf\{t \geq 0 : \|\mathbf{W}_t\|_F > R\}.$$

For $t \leq \mathcal{T}/\epsilon$, integrate Eq. B.18 over $[0, t \wedge \tau_R]$ and multiply by ϵ. Since

$$\mathbf{W}_{t\wedge\tau_R} - \mathbf{W}_0 = \epsilon \int_0^{t\wedge\tau_R} H(\mathbf{W}_s, X_s)\,ds,$$

we obtain

$$\begin{aligned} \mathbf{W}_{t\wedge\tau_R} - \mathbf{W}_0 &= \epsilon \int_0^{t\wedge\tau_R} \bar{H}(\mathbf{W}_s)\,ds - \epsilon \int_0^{t\wedge\tau_R} d\nu(\mathbf{W}_s, X_s) \\ &\quad + \epsilon^2 \int_0^{t\wedge\tau_R} \partial_{\mathbf{W}}\nu(\mathbf{W}_s, X_s)[H(\mathbf{W}_s, X_s)]\,ds \\ &\quad + \epsilon\sqrt{2D} \int_0^{t\wedge\tau_R} d_z\nu(\mathbf{W}_s, X_s)[dB_s^{\mathcal{M}}]. \end{aligned} \tag{B.19}$$

Equivalently, setting $\mathbf{W}^\epsilon(\tau) = \mathbf{W}_{\tau/\epsilon}$ and taking $t = \tau/\epsilon$, Eq. B.19 becomes, for $0 \leq \tau \leq \mathcal{T}$,

$$\begin{aligned} \mathbf{W}^\epsilon(\tau \wedge \epsilon\tau_R) &= \mathbf{W}_0 + \int_0^{\tau\wedge\epsilon\tau_R} \bar{H}(\mathbf{W}^\epsilon(r))\,dr \\ &\quad - \epsilon\left[\nu(\mathbf{W}_{(\tau/\epsilon)\wedge\tau_R}, X_{(\tau/\epsilon)\wedge\tau_R}) - \nu(\mathbf{W}_0, X_0)\right] \\ &\quad + \epsilon \int_0^{\tau\wedge\epsilon\tau_R} \partial_{\mathbf{W}}\nu(\mathbf{W}^\epsilon(r), X_{r/\epsilon})[H(\mathbf{W}^\epsilon(r), X_{r/\epsilon})]\,dr \\ &\quad + \epsilon\sqrt{2D}\,\mathcal{M}_{(\tau/\epsilon)\wedge\tau_R}, \end{aligned} \tag{B.20}$$

where

$$\mathcal{M}_t = \int_0^t d_z\nu(\mathbf{W}_s, X_s)[dB_s^{\mathcal{M}}]$$

is a matrix-valued continuous local martingale.

Step 5: Uniform control of the remainder terms.

On the stopped domain $\|\mathbf{W}_t\|_F \leq R$, the functions

$$\nu, \qquad \partial_{\mathbf{W}}\nu, \qquad d_z\nu, \qquad H$$

are uniformly bounded on $\mathcal{B}_R \times \mathcal{M}$. Let $C_R < \infty$ denote a constant controlling these bounds.

The boundary corrector in Eq. B.20 satisfies

$$\sup_{0\leq\tau\leq\mathcal{T}} \epsilon \left\|\nu(\mathbf{W}_{(\tau/\epsilon)\wedge\tau_R}, X_{(\tau/\epsilon)\wedge\tau_R}) - \nu(\mathbf{W}_0, X_0)\right\|_F \leq 2\epsilon C_R = \mathcal{O}(\epsilon). \tag{B.21}$$

The higher-order variation term satisfies

$$\sup_{0\le\tau\le\mathcal{T}} \epsilon \left\| \int_0^{\tau\wedge\epsilon\tau_R} \partial_{\mathbf{W}}\nu[H]\, dr \right\|_F \le \epsilon\mathcal{T}C_R = \mathcal{O}(\epsilon). \tag{B.22}$$

For the martingale term, the Burkholder–Davis–Gundy inequality gives [50, 85]

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le\tau\le\mathcal{T}} \left\|\epsilon\sqrt{2D}\,\mathcal{M}_{(\tau/\epsilon)\wedge\tau_R}\right\|_F^2\right] &\le 2D\epsilon^2 C_{\mathrm{BDG}}\mathbb{E}\left[\langle\mathcal{M}\rangle_{\mathcal{T}/\epsilon\wedge\tau_R}\right] \\ &\le 2D\epsilon^2 C_{\mathrm{BDG}}C_R^2\left(\frac{\mathcal{T}}{\epsilon}\right) \\ &= \mathcal{O}(\epsilon\mathcal{T}).\end{aligned} \tag{B.23}$$

Therefore the martingale remainder converges to zero in $L^2(\mathbb{P})$, hence also in probability, uniformly on $[0,\mathcal{T}]$.

Combining Eqs. B.21–B.23, we have

$$\sup_{0\le\tau\le\mathcal{T}} \left\| \mathbf{W}^\epsilon(\tau\wedge\epsilon\tau_R) - \mathbf{W}_0 - \int_0^{\tau\wedge\epsilon\tau_R} \bar{H}(\mathbf{W}^\epsilon(r))\, dr \right\|_F \xrightarrow{\mathbb{P}} 0. \tag{B.24}$$

Since $\bar{H}$ is a polynomial vector field in $\mathbf{W}$ with coefficients determined by the bounded matrix Σ_π, it is locally Lipschitz on $\mathbb{R}^{m\times n}$ [54]. Standard perturbation estimates for ODEs then imply that the stopped process $\mathbf{W}^\epsilon(\cdot\wedge\epsilon\tau_R)$ converges uniformly in probability on $[0,\mathcal{T}]$ to the solution of

$$\dot{\bar{\mathbf{W}}} = \bar{H}(\bar{\mathbf{W}}), \qquad \bar{\mathbf{W}}(0) = \mathbf{W}_0.$$

Finally, the Lyapunov analysis in Section B.3 shows that the averaged solution remains in a bounded row-orthogonality sublevel set on every finite horizon. Choose R larger than this deterministic bound by a fixed margin. The stopped convergence above then implies

$$\mathbb{P}(\tau_R \le \mathcal{T}/\epsilon) \to 0,$$

so the stopped and unstopped trajectories coincide with probability tending to one on $[0,\mathcal{T}]$. This removes the localization and proves the finite-horizon averaging statement in Eq. B.10. ■

B.3 Row-Orthogonality Lyapunov Functional and Dissipativity

Having obtained the averaged subspace Oja flow in Eq. B.10, we now establish the Lyapunov property used in Theorem 2. The purpose of this subsection is not to constrain the entire matrix $\mathbf{W}$ to a scalar Frobenius sphere. Instead, the lateral-inhibition term in the subspace Oja dynamics induces a row-orthogonality geometry. Accordingly, the correct Lyapunov functional measures deviation from the row-Stiefel condition

$$\mathbf{W}\mathbf{W}^T = \mathbf{I}_m.$$

For notational clarity, we drop the bar notation and write the averaged macroscopic state as $\mathbf{W}(\tau)$.

Definition 7 (Row-Orthogonality Lyapunov Functional) *We define the Lyapunov functional, corresponding to Eq. 10 in the main text, by*

$$V_{\mathrm{orth}}(\mathbf{W}) = \frac{1}{4}\left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F^2 = \frac{1}{4}\mathrm{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2\right]. \tag{B.25}$$

This functional is smooth and non-negative on $\mathbb{R}^{m\times n}$ [54, 44]. Its zero set is precisely the row-orthonormal manifold

$$\mathcal{S}_{m,n} = \left\{\mathbf{W}\in\mathbb{R}^{m\times n} : \mathbf{W}\mathbf{W}^T = \mathbf{I}_m\right\},$$

which is the row-Stiefel geometry associated with subspace Oja/Sanger lateral inhibition [2, 27, 91].

Theorem 9 (Row-Orthogonality Dissipation) *This formalizes Theorem 2 in the main text. Consider the averaged subspace Oja flow*

$$\dot{\mathbf{W}} = \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right), \qquad \gamma > 0, \tag{B.26}$$

where $\Sigma_\pi \succeq 0$ is the spatial covariance from Definition 6. Then the orbital derivative of V_{orth} along Eq. B.26 satisfies

$$\frac{d}{d\tau}V_{\text{orth}}(\mathbf{W}(\tau)) = -\gamma \operatorname{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2 \mathbf{W}\Sigma_\pi \mathbf{W}^T\right] \leq 0. \tag{B.27}$$

Consequently, the row-orthogonality Lyapunov functional is non-increasing along the averaged flow.

Proof: Let

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T, \qquad \mathbf{A} = \mathbf{S} - \mathbf{I}_m.$$

Then

$$V_{\text{orth}}(\mathbf{W}) = \frac{1}{4}\operatorname{Tr}(\mathbf{A}^2).$$

We also define the projected covariance

$$\mathbf{B} = \mathbf{W}\Sigma_\pi \mathbf{W}^T.$$

Since $\Sigma_\pi \succeq 0$, the matrix $\mathbf{B}$ is symmetric positive semi-definite.

Step 1: Fréchet differential of the functional. Since $\dot{\mathbf{A}} = \dot{\mathbf{S}}$, the trace product rule gives

$$\begin{aligned}\frac{d}{d\tau}V_{\text{orth}}(\mathbf{W}(\tau)) &= \frac{1}{4}\operatorname{Tr}\left(\dot{\mathbf{A}}\mathbf{A} + \mathbf{A}\dot{\mathbf{A}}\right) \\ &= \frac{1}{2}\operatorname{Tr}\left(\mathbf{A}\dot{\mathbf{A}}\right) = \frac{1}{2}\operatorname{Tr}\left(\mathbf{A}\dot{\mathbf{S}}\right),\end{aligned} \tag{B.28}$$

where the second equality uses cyclic invariance of the trace.

Step 2: Vector-field substitution. The averaged vector field in Eq. B.26 can be written as

$$\dot{\mathbf{W}} = \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{B}\mathbf{W}\right).$$

Therefore

$$\begin{aligned}\dot{\mathbf{S}} &= \frac{d}{d\tau}\left(\mathbf{W}\mathbf{W}^T\right) = \dot{\mathbf{W}}\mathbf{W}^T + \mathbf{W}\dot{\mathbf{W}}^T \\ &= \gamma\left(\mathbf{W}\Sigma_\pi\mathbf{W}^T - \mathbf{B}\mathbf{W}\mathbf{W}^T\right) + \gamma\left(\mathbf{W}\Sigma_\pi\mathbf{W}^T - \mathbf{W}\mathbf{W}^T\mathbf{B}\right) \\ &= \gamma\left(2\mathbf{B} - \mathbf{B}\mathbf{S} - \mathbf{S}\mathbf{B}\right),\end{aligned} \tag{B.29}$$

where we used the symmetry of Σ_π and $\mathbf{B}$.

Step 3: Cyclic permutation and row-orthogonality factorization. Since $\mathbf{S} = \mathbf{A} + \mathbf{I}_m$, Eq. B.29 becomes

$$\begin{aligned}\dot{\mathbf{S}} &= \gamma\left[2\mathbf{B} - \mathbf{B}(\mathbf{A} + \mathbf{I}_m) - (\mathbf{A} + \mathbf{I}_m)\mathbf{B}\right] \\ &= -\gamma\left(\mathbf{B}\mathbf{A} + \mathbf{A}\mathbf{B}\right).\end{aligned} \tag{B.30}$$

Substituting Eq. B.30 into Eq. B.28 yields

$$\begin{aligned}\frac{d}{d\tau}V_{\text{orth}}(\mathbf{W}(\tau)) &= -\frac{\gamma}{2}\operatorname{Tr}\left[\mathbf{A}\left(\mathbf{B}\mathbf{A} + \mathbf{A}\mathbf{B}\right)\right] \\ &= -\frac{\gamma}{2}\left(\operatorname{Tr}(\mathbf{A}\mathbf{B}\mathbf{A}) + \operatorname{Tr}(\mathbf{A}^2\mathbf{B})\right).\end{aligned} \tag{B.31}$$

By cyclic invariance of trace,

$$\operatorname{Tr}(\mathbf{A}\mathbf{B}\mathbf{A}) = \operatorname{Tr}(\mathbf{B}\mathbf{A}^2) = \operatorname{Tr}(\mathbf{A}^2\mathbf{B}).$$

Therefore,

$$\frac{d}{d\tau}V_{\text{orth}}(\mathbf{W}(\tau)) = -\gamma\operatorname{Tr}\left(\mathbf{A}^2\mathbf{B}\right). \tag{B.32}$$

Returning to the definitions of $\mathbf{A}$ and $\mathbf{B}$ gives exactly

$$\frac{d}{d\tau}V_{\text{orth}}(\mathbf{W}(\tau)) = -\gamma\operatorname{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2 \mathbf{W}\Sigma_\pi\mathbf{W}^T\right]. \tag{B.33}$$

Step 4: Positivity and sign of the derivative. It remains to show that the trace term is non-negative. Since $\Sigma_\pi \succeq 0$, for every $\mathbf{u} \in \mathbb{R}^m$,

$$\mathbf{u}^T \mathbf{B} \mathbf{u} = \mathbf{u}^T \mathbf{W} \Sigma_\pi \mathbf{W}^T \mathbf{u} = (\mathbf{W}^T \mathbf{u})^T \Sigma_\pi (\mathbf{W}^T \mathbf{u}) \geq 0.$$

Thus $\mathbf{B} \succeq 0$. Also, $\mathbf{A}$ is symmetric, so $\mathbf{A}^2 \succeq 0$. Let $\mathbf{B}^{1/2}$ be the positive semi-definite square root of $\mathbf{B}$. Then

$$\begin{aligned} \mathrm{Tr}(\mathbf{A}^2 \mathbf{B}) &= \mathrm{Tr}\left(\mathbf{B}^{1/2} \mathbf{A}^2 \mathbf{B}^{1/2}\right) \\ &= \mathrm{Tr}\left[(\mathbf{A}\mathbf{B}^{1/2})^T (\mathbf{A}\mathbf{B}^{1/2})\right] \\ &= \left\|\mathbf{A}\mathbf{B}^{1/2}\right\|_F^2 \geq 0. \end{aligned} \tag{B.34}$$

Therefore,

$$\frac{d}{d\tau} V_{\mathrm{orth}}(\mathbf{W}(\tau)) \leq 0,$$

which proves Lyapunov dissipativity.

Moreover, Eq. B.34 identifies the zero-derivative set:

$$\dot{V}_{\mathrm{orth}}(\mathbf{W}) = 0 \quad \Longleftrightarrow \quad \left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right) \left(\mathbf{W} \Sigma_\pi \mathbf{W}^T\right)^{1/2} = \mathbf{0}.$$

In particular, whenever $\mathbf{W}\Sigma_\pi \mathbf{W}^T \succ 0$ along a full-row-rank trajectory, the equality condition reduces to

$$\mathbf{W}\mathbf{W}^T = \mathbf{I}_m.$$

This is the row-orthonormal invariant geometry used in the LaSalle argument of Section B.4. ■

This establishes the Lyapunov dissipation identity used in Theorem 2 of the main text.

Boundedness of row-orthogonality sublevel sets. The preceding theorem also supplies the deterministic boundedness needed to remove the stopping-time localization in Lemma 7. Let

$$\Omega_c = \left\{\mathbf{W} \in \mathbb{R}^{m \times n} : V_{\mathrm{orth}}(\mathbf{W}) \leq c\right\}.$$

If $\mathbf{W} \in \Omega_c$, then

$$\left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F \leq 2\sqrt{c}.$$

Let $\sigma_1(\mathbf{W}), \ldots, \sigma_m(\mathbf{W})$ denote the singular values of $\mathbf{W}$. Since the eigenvalues of $\mathbf{W}\mathbf{W}^T$ are $\sigma_i(\mathbf{W})^2$ [44], we have

$$\sum_{i=1}^m \left(\sigma_i(\mathbf{W})^2 - 1\right)^2 = \left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F^2 \leq 4c.$$

Hence, for each i,

$$\sigma_i(\mathbf{W})^2 \leq 1 + 2\sqrt{c},$$

and therefore

$$\|\mathbf{W}\|_F^2 = \sum_{i=1}^m \sigma_i(\mathbf{W})^2 \leq m(1 + 2\sqrt{c}).$$

Thus every sublevel set Ω_c is bounded and closed, hence compact in finite-dimensional parameter space. Since V_{orth} is non-increasing along the averaged flow, the trajectory remains in the compact sublevel set determined by its initial value. This gives the precompactness required for the LaSalle-type invariant-set characterization in the next subsection [54, 55].

B.4 LaSalle's Invariance and Eigenspace Alignment

The previous subsection proves that the averaged subspace Oja flow decreases the row-orthogonality Lyapunov functional

$$V_{\mathrm{orth}}(\mathbf{W}) = \frac{1}{4}\left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F^2.$$

Moreover, its sublevel sets are compact in $\mathbb{R}^{m\times n}$. Hence every averaged trajectory has a nonempty compact ω-limit set, and LaSalle's invariance principle can be applied to characterize the limiting row-space geometry [54, 60, 55]. This subsection first identifies the row-orthonormal invariant component selected by V_{orth}, and then uses the spectral gap of Σ_π to prove principal-subspace alignment. Here "nondegenerate" refers to full-row-rank trajectories for which the projected covariance $\mathbf{W}\Sigma_\pi\mathbf{W}^T$ remains positive definite.

Theorem 10 (Invariant Row-Orthonormal Limit Set) *Let*

$$\mathcal{E}_0 = \left\{\mathbf{W}\in\mathbb{R}^{m\times n} : \dot{V}_{\mathrm{orth}}(\mathbf{W}) = 0\right\}$$

be the zero-dissipation set of the averaged flow

$$\dot{\mathbf{W}} = \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right).$$

Let $\mathcal{I}$ denote the maximal invariant subset of $\mathcal{E}_0$. Then every averaged trajectory approaches $\mathcal{I}$. Furthermore, on any full-row-rank component satisfying

$$\mathbf{W}\Sigma_\pi\mathbf{W}^T \succ 0,$$

the zero-dissipation condition reduces to the row-orthonormal constraint

$$\mathbf{W}\mathbf{W}^T = \mathbf{I}_m.$$

Consequently, the relevant nondegenerate invariant component is contained in the row-Stiefel manifold

$$\mathcal{S}_{m,n} = \left\{\mathbf{W}\in\mathbb{R}^{m\times n} : \mathbf{W}\mathbf{W}^T = \mathbf{I}_m\right\}.$$

Proof: By Theorem 9,

$$\dot{V}_{\mathrm{orth}}(\mathbf{W}) = -\gamma\mathrm{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2\mathbf{W}\Sigma_\pi\mathbf{W}^T\right] \le 0. \tag{B.35}$$

The compactness of the sublevel set containing the trajectory implies precompactness. Therefore LaSalle's invariance principle implies convergence to the maximal invariant subset of the zero-dissipation set

$$\mathcal{E}_0 = \{\mathbf{W} : \dot{V}_{\mathrm{orth}}(\mathbf{W}) = 0\}.$$

We now identify the nondegenerate part of this set. Define

$$\mathbf{A} = \mathbf{W}\mathbf{W}^T - \mathbf{I}_m, \qquad \mathbf{B} = \mathbf{W}\Sigma_\pi\mathbf{W}^T.$$

Since $\Sigma_\pi \succeq 0$, we have $\mathbf{B}\succeq 0$. As shown in Eq. B.34,

$$\begin{aligned}\mathrm{Tr}(\mathbf{A}^2\mathbf{B}) &= \mathrm{Tr}\left(\mathbf{B}^{1/2}\mathbf{A}^2\mathbf{B}^{1/2}\right)\\ &= \left\|\mathbf{A}\mathbf{B}^{1/2}\right\|_F^2.\end{aligned} \tag{B.36}$$

Thus

$$\dot{V}_{\mathrm{orth}}(\mathbf{W}) = 0 \quad\Longleftrightarrow\quad \mathbf{A}\mathbf{B}^{1/2} = \mathbf{0}.$$

If $\mathbf{B}\succ 0$, then $\mathbf{B}^{1/2}$ is invertible, so

$$\mathbf{A} = \mathbf{0}.$$

Equivalently,

$$\mathbf{W}\mathbf{W}^T = \mathbf{I}_m.$$

This proves that, on reached full-rank row spaces satisfying $\mathbf{W}\Sigma_\pi\mathbf{W}^T\succ 0$, the zero-dissipation invariant component is row-orthonormal.

It remains to note that $\mathcal{S}_{m,n}$ is invariant under the averaged subspace flow. If

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T$$

and

$$\mathbf{B} = \mathbf{W}\Sigma_\pi\mathbf{W}^T,$$

then the computation in Eq. B.29 gives

$$\dot{\mathbf{S}} = \gamma\left(2\mathbf{B} - \mathbf{B}\mathbf{S} - \mathbf{S}\mathbf{B}\right).$$

When $\mathbf{S} = \mathbf{I}_m$, this becomes

$$\dot{\mathbf{S}} = \gamma\left(2\mathbf{B} - \mathbf{B} - \mathbf{B}\right) = \mathbf{0}.$$

Hence a row-orthonormal initialization remains row-orthonormal, and every nondegenerate ω-limit trajectory identified by LaSalle evolves on the row-Stiefel manifold. ■

Figure B.1 illustrates the same Lyapunov mechanism geometrically: singular values are driven toward the row-orthonormal manifold while V_{orth} and $\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\|_F$ decay along representative trajectories.

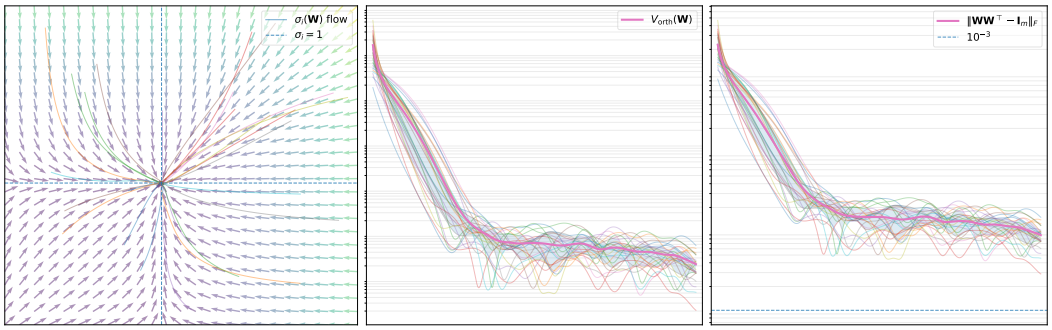

Figure B.1: **Row-Orthogonality Stabilization.** Numerical illustration of the row-orthogonality invariant geometry for the averaged subspace Oja flow. The left panel visualizes the singular-value vector field driving $\sigma_i(\mathbf{W})$ toward 1; the middle panel shows the decay of the averaged Lyapunov trend $V_{\mathrm{orth}}(\mathbf{W})$ across trajectories; the right panel shows convergence of $\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\|_F$ toward zero, consistent with approach to the row-Stiefel manifold.

The invariant row-Stiefel geometry prevents row-space collapse, but it does not by itself determine which m-dimensional row space is selected. To identify the selected subspace, we now diagonalize the covariance operator and analyze the induced flow on the Grassmannian row space [2, 27, 44].

Theorem 11 (Principal Eigenspace Alignment, formalizing Theorem 3) *Let*

$$\Sigma_\pi = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \qquad \lambda_1 \geq \cdots \geq \lambda_n \geq 0,$$

and assume the spectral gap

$$\lambda_m > \lambda_{m+1}, \qquad \lambda_m > 0.$$

Let $\mathbf{Q}_m \in \mathbb{R}^{n\times m}$ be the top-m eigenvector matrix. Suppose that the initial condition has full row rank and nondegenerate projection onto the top-m eigenspace:

$$\mathrm{rank}(\mathbf{W}_0\mathbf{Q}_m) = m.$$

Then the row space of the averaged subspace Oja flow converges to the principal eigenspace:

$$\mathrm{Row}(\mathbf{W}(\tau)) \longrightarrow \mathrm{Col}(\mathbf{Q}_m), \qquad \tau \to \infty, \tag{B.37}$$

up to an orthogonal rotation of the latent coordinates [77, 91].

Proof: By Theorem 10, every nondegenerate ω-limit trajectory lies on the row-Stiefel manifold. For row-space selection, the spectral-ratio argument can be applied after the LaSalle reduction to the nondegenerate row-orthonormal component; the nondegenerate projection condition ensures that the relative coordinate used below is well-defined along the flow. Indeed, the averaged subspace Oja flow is

$$\dot{\mathbf{W}} = \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right), \tag{B.38}$$

and V_{orth} keeps the trajectory in a compact sublevel set. On the invariant row-orthonormal component, this same vector field coincides with the row-Stiefel projected gradient flow of the variance objective

$$\frac{1}{2}\mathrm{Tr}(\mathbf{W}\Sigma_\pi\mathbf{W}^T)$$

under the constraint $\mathbf{W}\mathbf{W}^T = \mathbf{I}_m$.

Step 1: Spectral change of basis.

Define the spectral-coordinate representation

$$\mathbf{C}(\tau) = \mathbf{W}(\tau)\mathbf{Q} \in \mathbb{R}^{m\times n}.$$

Since $\mathbf{Q}$ is orthogonal,

$$\mathbf{C}\mathbf{C}^T = \mathbf{W}\mathbf{Q}\mathbf{Q}^T\mathbf{W}^T = \mathbf{W}\mathbf{W}^T = \mathbf{I}_m.$$

Differentiating $\mathbf{C} = \mathbf{W}\mathbf{Q}$ and substituting Eq. B.38 gives

$$\begin{aligned}\dot{\mathbf{C}} &= \dot{\mathbf{W}}\mathbf{Q}\\ &= \gamma\left(\mathbf{W}\Sigma_\pi - \mathbf{W}\Sigma_\pi\mathbf{W}^T\mathbf{W}\right)\mathbf{Q}\\ &= \gamma\left(\mathbf{W}\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{Q} - \mathbf{W}\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{W}^T\mathbf{W}\mathbf{Q}\right)\\ &= \gamma\left(\mathbf{C}\mathbf{\Lambda} - \mathbf{C}\mathbf{\Lambda}\mathbf{C}^T\mathbf{C}\right).\end{aligned} \tag{B.39}$$

This is the spectral form of the subspace Oja/Sanger flow. Unlike the scalar Oja case, the columns are coupled through the matrix factor

$$\mathbf{C}\mathbf{\Lambda}\mathbf{C}^T.$$

Step 2: Block decomposition into principal and residual coordinates.

Partition

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_1 & \mathbf{C}_2\end{bmatrix}, \qquad \mathbf{\Lambda} = \begin{bmatrix}\mathbf{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_2\end{bmatrix},$$

where

$$\mathbf{C}_1 \in \mathbb{R}^{m\times m}, \qquad \mathbf{C}_2 \in \mathbb{R}^{m\times(n-m)},$$

and

$$\mathbf{\Lambda}_1 = \mathrm{diag}(\lambda_1,\dots,\lambda_m), \qquad \mathbf{\Lambda}_2 = \mathrm{diag}(\lambda_{m+1},\dots,\lambda_n).$$

Let

$$\mathbf{M} = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^T = \mathbf{C}_1\mathbf{\Lambda}_1\mathbf{C}_1^T + \mathbf{C}_2\mathbf{\Lambda}_2\mathbf{C}_2^T.$$

Then Eq. B.39 yields the block system

$$\dot{\mathbf{C}}_1 = \gamma\left(\mathbf{C}_1\mathbf{\Lambda}_1 - \mathbf{M}\mathbf{C}_1\right), \tag{B.40}$$

$$\dot{\mathbf{C}}_2 = \gamma\left(\mathbf{C}_2\mathbf{\Lambda}_2 - \mathbf{M}\mathbf{C}_2\right). \tag{B.41}$$

The assumption

$$\mathrm{rank}(\mathbf{W}_0\mathbf{Q}_m) = m$$

is exactly the condition that $\mathbf{C}_1(0)$ is invertible. We now verify that this nondegeneracy is preserved along the averaged flow. Whenever $\mathbf{C}_1$ is invertible, Eq. B.40 gives

$$\dot{\mathbf{C}}_1 = \gamma\left(\mathbf{C}_1\mathbf{\Lambda}_1 - \mathbf{M}\mathbf{C}_1\right).$$

Using Jacobi's formula for the determinant [44],

$$\frac{d}{d\tau}\log|\det\mathbf{C}_1| = \mathrm{Tr}\left(\mathbf{C}_1^{-1}\dot{\mathbf{C}}_1\right) = \gamma\mathrm{Tr}(\mathbf{\Lambda}_1) - \gamma\mathrm{Tr}(\mathbf{M}).$$

Since the trajectory remains in a compact Lyapunov sublevel set, $\mathbf{M}(\tau)$ is bounded on every finite interval. Hence the right-hand side is integrable on every finite interval, and

$$\det\mathbf{C}_1(\tau) = \det\mathbf{C}_1(0)\exp\left(\gamma\int_0^\tau\left[\mathrm{Tr}(\mathbf{\Lambda}_1) - \mathrm{Tr}(\mathbf{M}(s))\right]ds\right) \neq 0$$

for every finite τ. Thus $\mathbf{C}_1(\tau)$ remains invertible, and the relative coordinate

$$\mathbf{G}(\tau) = \mathbf{C}_1(\tau)^{-1}\mathbf{C}_2(\tau)$$

is well-defined along the nondegenerate trajectory.

Step 3: Relative residual coordinate and exact cancellation of the nonlinear inhibition.

Define the relative residual coordinate

$$\mathbf{G}(\tau) = \mathbf{C}_1(\tau)^{-1}\mathbf{C}_2(\tau) \in \mathbb{R}^{m\times(n-m)}.$$

Differentiating gives

$$\dot{\mathbf{G}} = -\mathbf{C}_1^{-1}\dot{\mathbf{C}}_1\mathbf{C}_1^{-1}\mathbf{C}_2 + \mathbf{C}_1^{-1}\dot{\mathbf{C}}_2. \tag{B.42}$$

Substituting Eqs. B.40–B.41,

$$\begin{aligned}\dot{\mathbf{G}} &= -\gamma\mathbf{C}_1^{-1}\left(\mathbf{C}_1\mathbf{\Lambda}_1 - \mathbf{M}\mathbf{C}_1\right)\mathbf{C}_1^{-1}\mathbf{C}_2 + \gamma\mathbf{C}_1^{-1}\left(\mathbf{C}_2\mathbf{\Lambda}_2 - \mathbf{M}\mathbf{C}_2\right)\\ &= -\gamma\left(\mathbf{\Lambda}_1 - \mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_1\right)\mathbf{G} + \gamma\mathbf{G}\mathbf{\Lambda}_2 - \gamma\mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_2\\ &= -\gamma\mathbf{\Lambda}_1\mathbf{G} + \gamma\mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_1\mathbf{G} + \gamma\mathbf{G}\mathbf{\Lambda}_2 - \gamma\mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_2.\end{aligned} \tag{B.43}$$

Since $\mathbf{C}_2 = \mathbf{C}_1\mathbf{G}$, the two nonlinear inhibition terms cancel:

$$\mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_1\mathbf{G} = \mathbf{C}_1^{-1}\mathbf{M}\mathbf{C}_2.$$

Therefore,

$$\dot{\mathbf{G}} = \gamma\left(\mathbf{G}\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1\mathbf{G}\right). \tag{B.44}$$

This identity is the key spectral-decoupling step: the nonlinear lateral-inhibition matrix $\mathbf{M}$ disappears from the relative coordinate dynamics.

Step 4: Exponential decay from the spectral gap.

Let G_{ij} denote the (i,j) entry of $\mathbf{G}$, where $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,n-m\}$. From Eq. B.44,

$$\dot{G}_{ij} = \gamma\left(\lambda_{m+j} - \lambda_i\right)G_{ij}. \tag{B.45}$$

Hence

$$G_{ij}(\tau) = G_{ij}(0)\exp\left[-\gamma\left(\lambda_i - \lambda_{m+j}\right)\tau\right]. \tag{B.46}$$

By the spectral gap,

$$\lambda_i - \lambda_{m+j} \geq \lambda_m - \lambda_{m+1} > 0.$$

Therefore,

$$|G_{ij}(\tau)| \leq |G_{ij}(0)|\exp\left[-\gamma\left(\lambda_m - \lambda_{m+1}\right)\tau\right], \tag{B.47}$$

and consequently

$$\|\mathbf{G}(\tau)\|_F \leq \|\mathbf{G}(0)\|_F\exp\left[-\gamma\left(\lambda_m - \lambda_{m+1}\right)\tau\right]. \tag{B.48}$$

Since the row-orthonormal invariant component satisfies

$$\mathbf{C}\mathbf{C}^T = \mathbf{I}_m,$$

both $\mathbf{C}_1$ and $\mathbf{C}_2$ are bounded. Because

$$\mathbf{C}_2(\tau) = \mathbf{C}_1(\tau)\mathbf{G}(\tau),$$

Eq. B.48 implies

$$\|\mathbf{C}_2(\tau)\|_F \longrightarrow 0.$$

But

$$\mathbf{C}_2 = \mathbf{W}\mathbf{Q}_\perp,$$

where $\mathbf{Q}_\perp$ contains the eigenvectors outside the top-m subspace. Thus

$$\|\mathbf{W}(\tau)\mathbf{Q}_\perp\|_F \longrightarrow 0,$$

which is equivalent to convergence of the row space of $\mathbf{W}(\tau)$ to $\mathrm{Col}(\mathbf{Q}_m)$ in the Grassmannian principal-angle sense [2, 27].

Finally, because $\mathbf{W}\mathbf{W}^T \to \mathbf{I}_m$ on the nondegenerate invariant component, the limiting row-orthonormal basis may differ from $\mathbf{Q}_m^T$ only by an orthogonal transformation in the latent coordinates. Hence there exists an orthogonal matrix

$$\mathbf{R} \in \mathbb{R}^{m\times m}$$

such that the limiting representation has the form

$$\mathbf{W}_\infty = \mathbf{R}\mathbf{Q}_m^T.$$

This proves principal eigenspace alignment up to orthogonal rotation, completing the appendix proof of Theorem 3. ■

Remark on spectral degeneracy. The exponential decay in Eq. B.48 relies on the strict gap $\lambda_m > \lambda_{m+1}$. If $\lambda_m = \lambda_{m+1}$, the top-$m$ subspace is not uniquely identifiable. In that case the dynamics can still converge to a row-orthonormal invariant subspace within the degenerate eigenspace, but the limiting basis is defined only up to a Grassmann equivalence class [2, 27].

Figure B.2 visualizes this spectral-gap mechanism by tracking the relative residual coordinate $\mathbf{G}$, the residual projection outside the top eigenspace, and the resulting principal-angle decay.

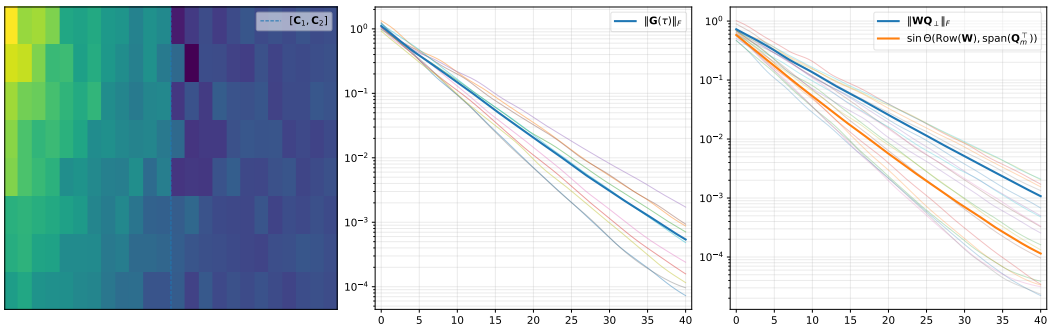

Figure B.2: **Spectral Gap Mechanism for Principal Subspace Alignment.** Numerical illustration of the block spectral mechanism used in Theorem 11. In the eigenbasis of Σ_π, the row-space error is represented by the relative residual coordinate $\mathbf{G} = \mathbf{C}_1^{-1}\mathbf{C}_2$. The spectral gap $\lambda_m - \lambda_{m+1} > 0$ induces exponential decay of $\|\mathbf{G}(\tau)\|_F$, forcing the residual projection $\|\mathbf{W}\mathbf{Q}_\perp\|_F$ and the principal-angle error to vanish.

C Coupled Flow and Feature Decorrelation

Appendices A and B analyze the two limiting components of the model separately: the discrete-to-continuous spatial kinematics and the averaged subspace Oja dynamics. The present appendix connects these components to the coupled macroscopic skeleton used in Section 4. Building on those two limits, this appendix shows how the representation-induced spatial potential, the finite-sample covariance, and the row-orthogonality Lyapunov structure combine into the deterministic coupled comparison flow.

The coupled analysis is organized as follows. Section C.1 formulates the product-space Langevin–subspace Oja system and derives the finite-sample row-space update. Section C.2 introduces the macroscopic Lyapunov surrogate used in Eq. 13 and proves dissipation for the deterministic drift-only skeleton. Sections C.3–C.4 then use the principal-subspace characterization from Appendix B to describe the decorrelated latent covariance at stationarity, and Section C.5 records the corresponding gauge non-identifiability under degenerate spectra.

C.1 The Coupled Langevin–Subspace Oja Flow on the Product Space

We now lift the fast spatial process and the slow representation dynamics to the product space

$$\mathcal{M}^N \times \mathbb{R}^{m\times n}.$$

This is a constrained product geometry: the spatial coordinates lie on the embedded manifold, while $\mathbf{W}_t$ is represented in an ambient matrix space before the row-orthonormal component is selected [1, 63]. The N agents evolve on the geometric substrate while sharing a common representation matrix $\mathbf{W}_t$. This section provides the finite-population counterpart of the averaged flow in Eq. 11 and the coupled skeleton used in Eq. 14 of the main text.

Definition 8 (Representation-Dependent Interaction Potential) *Let $\mathbf{W}_t \in \mathbb{R}^{m\times n}$ be the representation matrix at time t. For an agent located at $x \in \mathcal{M}$, let*

$$\mathbf{x} = \Phi(x) \in \mathbb{R}^n$$

be its Euclidean observation, consistent with the embedding and pushforward construction in Definition 6. The corresponding latent coordinate is

$$\mathbf{y}_t(\mathbf{x}) = \mathbf{W}_t\mathbf{x} \in \mathbb{R}^m.$$

We define the representation-induced spatial potential by

$$U_{\mathbf{W}_t}(\mathbf{x}) = -\frac{1}{2}\|\mathbf{W}_t\mathbf{x}\|_2^2 = -\frac{1}{2}\mathbf{x}^T\mathbf{W}_t^T\mathbf{W}_t\mathbf{x}. \tag{C.1}$$

This is the finite-dimensional version of the representation-induced part of the macroscopic energy in Eq. 13, following the standard energy-based view of unsupervised representation learning [61, 6].

We compute the embedded Euclidean gradient of U with respect to $\mathbf{x}$. Let

$$\mathbf{A}_t = \mathbf{W}_t^T\mathbf{W}_t \in \mathbb{R}^{n\times n}.$$

Since $\mathbf{A}_t$ is symmetric positive semi-definite, for an infinitesimal perturbation $\mathbf{v} \in \mathbb{R}^n$,

$$\begin{aligned} U(\mathbf{x}+\mathbf{v}, \mathbf{W}_t) &= -\frac{1}{2}(\mathbf{x}+\mathbf{v})^T\mathbf{A}_t(\mathbf{x}+\mathbf{v}) \\ &= -\frac{1}{2}\left[\mathbf{x}^T\mathbf{A}_t\mathbf{x} + \mathbf{x}^T\mathbf{A}_t\mathbf{v} + \mathbf{v}^T\mathbf{A}_t\mathbf{x} + \mathbf{v}^T\mathbf{A}_t\mathbf{v}\right] \\ &= U_{\mathbf{W}_t}(\mathbf{x}) - \frac{1}{2}\left[\mathbf{x}^T\mathbf{A}_t\mathbf{v} + \mathbf{v}^T\mathbf{A}_t\mathbf{x}\right] - \frac{1}{2}\mathbf{v}^T\mathbf{A}_t\mathbf{v} \\ &= U_{\mathbf{W}_t}(\mathbf{x}) - \mathbf{v}^T\mathbf{A}_t\mathbf{x} - \frac{1}{2}\mathbf{v}^T\mathbf{A}_t\mathbf{v} \\ &= U_{\mathbf{W}_t}(\mathbf{x}) + \left\langle -\mathbf{W}_t^T\mathbf{W}_t\mathbf{x}, \mathbf{v}\right\rangle_{\mathbb{R}^n} + \mathcal{O}(\|\mathbf{v}\|_2^2). \end{aligned} \tag{C.2}$$

Therefore,

$$\nabla_{\mathbf{x}} U_{\mathbf{W}_t}(\mathbf{x}) = -\mathbf{W}_t^T\mathbf{W}_t\mathbf{x}. \tag{C.3}$$

The corresponding ambient negative-gradient direction is

$$-\nabla_{\mathbf{x}} U_{\mathbf{W}_t}(\mathbf{x}) = \mathbf{W}_t^T\mathbf{W}_t\mathbf{x}.$$

Since the spatial state is constrained to the embedded manifold $\Phi(\mathcal{M}) \subset \mathbb{R}^n$, only the tangent component of this ambient direction is admissible. For $\mathbf{x} = \Phi(x)$, let

$$\Pi_{\mathbf{x}} : \mathbb{R}^n \to T_{\mathbf{x}}\Phi(\mathcal{M})$$

denote the orthogonal projection onto the embedded tangent space. The manifold gradient of the restricted potential is therefore represented in ambient coordinates as

$$\nabla_{\mathcal{M}} U_{\mathbf{W}_t}(\mathbf{x}) = \Pi_{\mathbf{x}} \nabla_{\mathbf{x}} U_{\mathbf{W}_t}(\mathbf{x}),$$

and the admissible negative-gradient drift is

$$-\nabla_{\mathcal{M}} U_{\mathbf{W}_t}(\mathbf{x}) = \Pi_{\mathbf{x}}\left(\mathbf{W}_t^T\mathbf{W}_t\mathbf{x}\right).$$

For a configuration matrix $\mathbf{X}$, we use the row-wise projection

$$\Pi_{\mathbf{X}}\mathbf{A} = \begin{bmatrix} --- & \left(\Pi_{\mathbf{x}^{(1)}}\mathbf{a}^{(1)}\right)^T & --- \\ --- & \left(\Pi_{\mathbf{x}^{(2)}}\mathbf{a}^{(2)}\right)^T & --- \\ & \vdots & \\ --- & \left(\Pi_{\mathbf{x}^{(N)}}\mathbf{a}^{(N)}\right)^T & --- \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} --- & (\mathbf{a}^{(1)})^T & --- \\ --- & (\mathbf{a}^{(2)})^T & --- \\ & \vdots & \\ --- & (\mathbf{a}^{(N)})^T & --- \end{bmatrix}.$$

This notation keeps the finite-dimensional row-vector representation while preserving the manifold constraint imposed by the embedding [1, 63, 44, 69].

Definition 9 (Finite-Population Configuration Matrix) *Let N be the number of agents. We write the embedded observations as row vectors and collect them into*

$$\mathbf{X}_t = \begin{bmatrix} - - - & (\mathbf{x}_t^{(1)})^T & - - - \\ - - - & (\mathbf{x}_t^{(2)})^T & - - - \\ & \vdots & \\ - - - & (\mathbf{x}_t^{(N)})^T & - - - \end{bmatrix} \in \mathbb{R}^{N \times n}. \tag{C.4}$$

The corresponding Brownian increments are collected as

$$d\mathbf{B}_t^{\mathcal{M}} = \begin{bmatrix} - - - & (d\mathbf{b}_t^{(1)})^T & - - - \\ - - - & (d\mathbf{b}_t^{(2)})^T & - - - \\ & \vdots & \\ - - - & (d\mathbf{b}_t^{(N)})^T & - - - \end{bmatrix}. \tag{C.5}$$

This notation denotes the embedded coordinate representation of Brownian increments along the manifold-valued spatial process. The increments are independent across agents and are tangent to the embedded manifold. Thus, in ambient coordinates, their conditional covariance is

$$\mathbb{E}\left[d\mathbf{b}_t^{(i)} (d\mathbf{b}_t^{(j)})^T \,\middle|\, \mathbf{x}_t^{(i)} \right] = \delta_{ij} \Pi_{\mathbf{x}_t^{(i)}}\, dt.$$

Equivalently, in any local orthonormal frame of $T_{\mathbf{x}_t^{(i)}}\Phi(\mathcal{M})$, this is the usual d-dimensional Brownian covariance. The ambient matrix $\Pi_{\mathbf{x}_t^{(i)}}$ has rank d, not n [50, 78, 63].

Theorem 12 (Coupled Langevin–Subspace Oja System) *The finite-population coupled system consists of the representation-induced Langevin spatial dynamics and the finite-sample subspace Oja parameter dynamics:*

$$\text{Fast spatial dynamics:} \qquad d\mathbf{X}_t = D\beta\, \Pi_{\mathbf{X}_t} \left(\mathbf{X}_t \mathbf{W}_t^T \mathbf{W}_t \right) dt + \sqrt{2D}\, d\mathbf{B}_t^{\mathcal{M}}, \tag{C.6}$$

$$\text{Slow row-space dynamics:} \qquad \dot{\mathbf{W}}_t = \eta_W \left(\mathbf{W}_t \widehat{\Sigma}_{\mathbf{X}_t} - \mathbf{W}_t \widehat{\Sigma}_{\mathbf{X}_t} \mathbf{W}_t^T \mathbf{W}_t \right), \qquad \widehat{\Sigma}_{\mathbf{X}_t} = \frac{1}{N} \mathbf{X}_t^T \mathbf{X}_t. \tag{C.7}$$

Equivalently, the slow dynamics can be written in expanded finite-sample form as

$$\dot{\mathbf{W}}_t = \frac{\eta_W}{N} \left(\mathbf{W}_t \mathbf{X}_t^T \mathbf{X}_t - \mathbf{W}_t \mathbf{X}_t^T \mathbf{X}_t \mathbf{W}_t^T \mathbf{W}_t \right). \tag{C.8}$$

Proof: We derive the two components separately and then express the slow update in a form that matches the subspace Oja dynamics of Section 3.

Step 1: Fast spatial dynamics. For the i-th agent, the overdamped Langevin equation with the representation-induced potential is written on the embedded manifold. In ambient coordinates, this means that the Euclidean drift is projected onto the tangent space:

$$d\mathbf{x}_t^{(i)} = -D\beta\, \Pi_{\mathbf{x}_t^{(i)}} \nabla_{\mathbf{x}} U(\mathbf{x}_t^{(i)}, \mathbf{W}_t)\, dt + \sqrt{2D}\, d\mathbf{b}_t^{(i)}. \tag{C.9}$$

Using Eq. C.3,

$$\begin{aligned} d\mathbf{x}_t^{(i)} &= -D\beta\, \Pi_{\mathbf{x}_t^{(i)}} \left(-\mathbf{W}_t^T \mathbf{W}_t \mathbf{x}_t^{(i)} \right) dt + \sqrt{2D}\, d\mathbf{b}_t^{(i)} \\ &= D\beta\, \Pi_{\mathbf{x}_t^{(i)}} \left(\mathbf{W}_t^T \mathbf{W}_t \mathbf{x}_t^{(i)} \right) dt + \sqrt{2D}\, d\mathbf{b}_t^{(i)}. \end{aligned} \tag{C.10}$$

The Brownian increment $d\mathbf{b}_t^{(i)}$ is the embedded-coordinate representation of a manifold Brownian increment, so it is tangent-projected and has conditional ambient covariance

$$\mathbb{E}\left[d\mathbf{b}_t^{(i)} (d\mathbf{b}_t^{(j)})^T \,\middle|\, \mathbf{x}_t^{(i)} \right] = \delta_{ij} \Pi_{\mathbf{x}_t^{(i)}}\, dt.$$

Since $\mathbf{X}_t$ stores observations as rows, we transpose Eq. C.10:

$$(d\mathbf{x}_t^{(i)})^T = D\beta \left[\Pi_{\mathbf{x}_t^{(i)}} \left(\mathbf{W}_t^T \mathbf{W}_t \mathbf{x}_t^{(i)} \right) \right]^T dt + \sqrt{2D}\, (d\mathbf{b}_t^{(i)})^T. \tag{C.11}$$

Stacking Eq. C.11 over $i = 1, \dots, N$ yields the row-wise projected matrix equation

$$d\mathbf{X}_t = D\beta\, \Pi_{\mathbf{X}_t}\left(\mathbf{X}_t \mathbf{W}_t^T \mathbf{W}_t\right) dt + \sqrt{2D} d\mathbf{B}_t^{\mathcal{M}}. \tag{C.12}$$

This proves Eq. C.6 while preserving the constraint $\mathbf{x}_t^{(i)} \in \Phi(\mathcal{M})$ for each agent [50, 80, 63].

Step 2: Finite-sample covariance induced by the swarm. At time t, the finite agent population defines the empirical covariance

$$\widehat{\Sigma}_{\mathbf{X}_t} = \frac{1}{N}\mathbf{X}_t^T \mathbf{X}_t = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_t^{(i)} (\mathbf{x}_t^{(i)})^T \in \mathbb{R}^{n\times n}. \tag{C.13}$$

This is the finite-population counterpart of the invariant covariance Σ_π in Eq. 9; unlike Σ_π, it changes with the evolving spatial configuration $\mathbf{X}_t$ [96].

Step 3: Finite-sample subspace Oja update. For each observation $\mathbf{x}_t^{(i)}$, define

$$\mathbf{y}_t^{(i)} = \mathbf{W}_t \mathbf{x}_t^{(i)} \in \mathbb{R}^m.$$

The finite-population subspace Oja/Sanger update is the sample average of the local lateral-inhibition rule [76, 77, 91, 81]:

$$\dot{\mathbf{W}}_t = \frac{\eta_W}{N}\sum_{i=1}^{N}\left[\mathbf{y}_t^{(i)}(\mathbf{x}_t^{(i)})^T - \mathbf{y}_t^{(i)}(\mathbf{y}_t^{(i)})^T \mathbf{W}_t\right]. \tag{C.14}$$

We now expand both terms.

For the Hebbian term,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_t^{(i)}(\mathbf{x}_t^{(i)})^T &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_t\mathbf{x}_t^{(i)}(\mathbf{x}_t^{(i)})^T \\ &= \mathbf{W}_t\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_t^{(i)}(\mathbf{x}_t^{(i)})^T\right) \\ &= \mathbf{W}_t\widehat{\Sigma}_{\mathbf{X}_t}.\end{aligned} \tag{C.15}$$

For the lateral-inhibition term,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_t^{(i)}(\mathbf{y}_t^{(i)})^T\mathbf{W}_t &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{W}_t\mathbf{x}_t^{(i)}\right)\left(\mathbf{W}_t\mathbf{x}_t^{(i)}\right)^T\mathbf{W}_t \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{W}_t\mathbf{x}_t^{(i)}(\mathbf{x}_t^{(i)})^T\mathbf{W}_t^T\mathbf{W}_t \\ &= \mathbf{W}_t\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_t^{(i)}(\mathbf{x}_t^{(i)})^T\right)\mathbf{W}_t^T\mathbf{W}_t \\ &= \mathbf{W}_t\widehat{\Sigma}_{\mathbf{X}_t}\mathbf{W}_t^T\mathbf{W}_t.\end{aligned} \tag{C.16}$$

Substituting Eqs. C.15–C.16 into Eq. C.14 gives

$$\dot{\mathbf{W}}_t = \eta_W\left(\mathbf{W}_t\widehat{\Sigma}_{\mathbf{X}_t} - \mathbf{W}_t\widehat{\Sigma}_{\mathbf{X}_t}\mathbf{W}_t^T\mathbf{W}_t\right), \tag{C.17}$$

which is Eq. C.7. Finally, using

$$\widehat{\Sigma}_{\mathbf{X}_t} = \frac{1}{N}\mathbf{X}_t^T\mathbf{X}_t,$$

we obtain the expanded matrix form

$$\begin{aligned}\dot{\mathbf{W}}_t &= \eta_W\left[\mathbf{W}_t\left(\frac{1}{N}\mathbf{X}_t^T\mathbf{X}_t\right) - \mathbf{W}_t\left(\frac{1}{N}\mathbf{X}_t^T\mathbf{X}_t\right)\mathbf{W}_t^T\mathbf{W}_t\right] \\ &= \frac{\eta_W}{N}\left(\mathbf{W}_t\mathbf{X}_t^T\mathbf{X}_t - \mathbf{W}_t\mathbf{X}_t^T\mathbf{X}_t\mathbf{W}_t^T\mathbf{W}_t\right).\end{aligned} \tag{C.18}$$

This is the finite-sample version of the averaged subspace Oja flow in Eq. 11, with Σ_π replaced by the instantaneous empirical covariance $\widehat{\Sigma}_{\mathbf{X}_t}$. The inhibitory matrix

$$\mathbf{W}_t \widehat{\Sigma}_{\mathbf{X}_t} \mathbf{W}_t^T$$

acts on the latent coordinates and prevents row-space collapse, matching the row-orthogonality mechanism analyzed in Appendix B [10, 76, 77, 81, 91]. ■

This derivation supplies the finite-sample coupled Langevin–subspace Oja dynamics used in Section 4, in particular Eq. 14 of the main text.

C.2 The Joint Lyapunov Functional and Lyapunov Dissipation

Having formulated the finite-population coupled Langevin–subspace Oja system in Section C.1, we now introduce the macroscopic Lyapunov surrogate used in Section 4. The stochastic Brownian term belongs to the Langevin formulation of the fast process and would require Itô correction terms. Therefore, in this subsection we analyze the deterministic drift-only coupled skeleton: the spatial component follows the tangent-projected negative $\mathbf{X}$-gradient of the representation-induced energy on $\Phi(\mathcal{M})^N$, while the parameter component follows the negative row-Stiefel projected gradient direction associated with the finite-sample subspace Oja flow [2, 27, 54, 78].

Although $\mathbf{X}$ is represented as a matrix in $\mathbb{R}^{N\times n}$, its rows are constrained to the embedded manifold $\Phi(\mathcal{M})$. Accordingly, the spatial gradient should be interpreted as the embedded representative of the tangent-projected drift on $\Phi(\mathcal{M})^N$. The Lyapunov identity below is therefore a drift-skeleton identity on the constrained product geometry, not a claim of global boundedness for unconstrained Euclidean $\mathbf{X}$ [1, 45, 63].

Definition 10 (Macroscopic Lyapunov Surrogate) *For a spatial configuration $\mathbf{X} \in \mathbb{R}^{N\times n}$ and a representation matrix $\mathbf{W} \in \mathbb{R}^{m\times n}$, define the following Lyapunov-type energy functional [54, 61]:*

$$\mathcal{E}(\mathbf{X},\mathbf{W}) = \underbrace{-\frac{1}{2N}\mathrm{Tr}\left(\mathbf{X}\mathbf{W}^T\mathbf{W}\mathbf{X}^T\right)}_{\mathcal{H}(\mathbf{X},\mathbf{W})} + \lambda \underbrace{\frac{1}{4}\left\|\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right\|_F^2}_{\mathcal{R}(\mathbf{W})}, \qquad \lambda > 0. \tag{C.19}$$

Equivalently, with the finite-sample covariance

$$\widehat{\Sigma}_{\mathbf{X}} = \frac{1}{N}\mathbf{X}^T\mathbf{X},$$

the interaction term can be written as

$$\mathcal{H}(\mathbf{X},\mathbf{W}) = -\frac{1}{2}\mathrm{Tr}\left(\mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\right). \tag{C.20}$$

This is the appendix form of the macroscopic unsupervised loss in Eq. 13. The first term is the finite-sample variance-maximization objective for the latent features

$$\mathbf{Y} = \mathbf{X}\mathbf{W}^T \in \mathbb{R}^{N\times m},$$

since

$$\frac{1}{N}\|\mathbf{Y}\|_F^2 = \frac{1}{N}\mathrm{Tr}\left(\mathbf{X}\mathbf{W}^T\mathbf{W}\mathbf{X}^T\right).$$

The second term is the row-orthogonality penalty used in Eq. 10, replacing the scalar Frobenius normalization by a row-space constraint compatible with subspace Oja/Sanger lateral inhibition [4, 44, 91].

The factor $1/N$ in $\mathcal{E}$ only rescales the spatial drift and can be absorbed into the positive rate η_X; it does not affect the Lyapunov sign.

We next compute the Fréchet derivatives of $\mathcal{E}$ using standard matrix differential calculus [69]. These derivatives define the deterministic skeleton and its Lyapunov identity; the finite-sample parameter component is interpreted through the negative row-Stiefel projected subspace Oja direction.

Lemma 8 (Matrix Fréchet Gradients on the Product Space) *Let*

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T - \mathbf{I}_m.$$

The Euclidean Fréchet gradients of $\mathcal{E}$ are

$$\nabla_{\mathbf{X}}\mathcal{E} = -\frac{1}{N}\mathbf{X}\mathbf{W}^T\mathbf{W}, \tag{C.21}$$

$$\begin{aligned}\nabla_{\mathbf{W}}\mathcal{E} &= -\frac{1}{N}\mathbf{W}\mathbf{X}^T\mathbf{X} + \lambda\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)\mathbf{W} \\ &= -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} + \lambda\mathbf{S}\mathbf{W}.\end{aligned} \tag{C.22}$$

Proof: Let

$$\langle \mathbf{A}, \mathbf{B}\rangle_F = \mathrm{Tr}(\mathbf{A}^T\mathbf{B})$$

denote the Frobenius inner product [44, 69].

Part 1: Spatial derivative. Only the interaction term depends on $\mathbf{X}$. For a perturbation

$$\delta\mathbf{X} \in \mathbb{R}^{N\times n},$$

write

$$\mathbf{A} = \mathbf{W}^T\mathbf{W} \in \mathbb{R}^{n\times n}.$$

Since $\mathbf{A}$ is symmetric,

$$\begin{aligned}\mathcal{H}(\mathbf{X}+\delta\mathbf{X}, \mathbf{W}) &= -\frac{1}{2N}\mathrm{Tr}\left[(\mathbf{X}+\delta\mathbf{X})\mathbf{A}(\mathbf{X}+\delta\mathbf{X})^T\right] \\ &= -\frac{1}{2N}\mathrm{Tr}\left(\mathbf{X}\mathbf{A}\mathbf{X}^T + \delta\mathbf{X}\mathbf{A}\mathbf{X}^T + \mathbf{X}\mathbf{A}\delta\mathbf{X}^T + \delta\mathbf{X}\mathbf{A}\delta\mathbf{X}^T\right) \\ &= \mathcal{H}(\mathbf{X}, \mathbf{W}) - \frac{1}{2N}\mathrm{Tr}\left(\delta\mathbf{X}\mathbf{A}\mathbf{X}^T + \mathbf{X}\mathbf{A}\delta\mathbf{X}^T\right) + \mathcal{O}(\|\delta\mathbf{X}\|_F^2).\end{aligned} \tag{C.23}$$

Using cyclic invariance of the trace,

$$\mathrm{Tr}\left(\delta\mathbf{X}\mathbf{A}\mathbf{X}^T\right) = \mathrm{Tr}\left(\mathbf{X}^T\delta\mathbf{X}\mathbf{A}\right) = \mathrm{Tr}\left(\mathbf{A}\mathbf{X}^T\delta\mathbf{X}\right),$$

and

$$\mathrm{Tr}\left(\mathbf{X}\mathbf{A}\delta\mathbf{X}^T\right) = \mathrm{Tr}\left[(\mathbf{X}\mathbf{A})\,\delta\mathbf{X}^T\right] = \mathrm{Tr}\left[(\mathbf{X}\mathbf{A})^T\,\delta\mathbf{X}\right] = \mathrm{Tr}\left(\mathbf{A}\mathbf{X}^T\delta\mathbf{X}\right),$$

where the last equality uses $\mathbf{A}^T = \mathbf{A}$. Hence the two linear terms are equal, and Eq. C.23 becomes

$$\begin{aligned}\mathcal{H}(\mathbf{X}+\delta\mathbf{X}, \mathbf{W}) &= \mathcal{H}(\mathbf{X}, \mathbf{W}) - \frac{1}{N}\mathrm{Tr}\left(\mathbf{A}\mathbf{X}^T\delta\mathbf{X}\right) + \mathcal{O}(\|\delta\mathbf{X}\|_F^2) \\ &= \mathcal{H}(\mathbf{X}, \mathbf{W}) - \frac{1}{N}\mathrm{Tr}\left[(\mathbf{X}\mathbf{A})^T\,\delta\mathbf{X}\right] + \mathcal{O}(\|\delta\mathbf{X}\|_F^2) \\ &= \mathcal{H}(\mathbf{X}, \mathbf{W}) + \left\langle -\frac{1}{N}\mathbf{X}\mathbf{W}^T\mathbf{W}, \delta\mathbf{X}\right\rangle_F + \mathcal{O}(\|\delta\mathbf{X}\|_F^2).\end{aligned} \tag{C.24}$$

Thus

$$\nabla_{\mathbf{X}}\mathcal{E} = -\frac{1}{N}\mathbf{X}\mathbf{W}^T\mathbf{W}.$$

Part 2: Representation derivative of the variance term. For

$$\delta\mathbf{W} \in \mathbb{R}^{m\times n},$$

we expand

$$\begin{aligned}\mathcal{H}(\mathbf{X}, \mathbf{W}+\delta\mathbf{W}) &= -\frac{1}{2N}\mathrm{Tr}\left[\mathbf{X}(\mathbf{W}+\delta\mathbf{W})^T(\mathbf{W}+\delta\mathbf{W})\mathbf{X}^T\right] \\ &= -\frac{1}{2N}\mathrm{Tr}\left[\mathbf{X}\left(\mathbf{W}^T\mathbf{W} + \delta\mathbf{W}^T\mathbf{W} + \mathbf{W}^T\delta\mathbf{W} + \delta\mathbf{W}^T\delta\mathbf{W}\right)\mathbf{X}^T\right] \\ &= \mathcal{H}(\mathbf{X}, \mathbf{W}) - \frac{1}{2N}\mathrm{Tr}\left(\mathbf{X}\delta\mathbf{W}^T\mathbf{W}\mathbf{X}^T + \mathbf{X}\mathbf{W}^T\delta\mathbf{W}\mathbf{X}^T\right) + \mathcal{O}(\|\delta\mathbf{W}\|_F^2).\end{aligned} \tag{C.25}$$

Let

$$\widehat{\Sigma}_{\mathbf{X}} = \frac{1}{N}\mathbf{X}^T\mathbf{X}.$$

Then the first linear term satisfies

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{X}\delta\mathbf{W}^T\mathbf{W}\mathbf{X}^T\right) &= \mathrm{Tr}\left(\delta\mathbf{W}^T\mathbf{W}\mathbf{X}^T\mathbf{X}\right)\\ &= \mathrm{Tr}\left[\left(\mathbf{W}\mathbf{X}^T\mathbf{X}\right)^T\delta\mathbf{W}\right].\end{aligned} \tag{C.26}$$

For the second linear term,

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{X}\mathbf{W}^T\delta\mathbf{W}\mathbf{X}^T\right) &= \mathrm{Tr}\left(\mathbf{X}^T\mathbf{X}\mathbf{W}^T\delta\mathbf{W}\right)\\ &= \mathrm{Tr}\left[\left(\mathbf{W}\mathbf{X}^T\mathbf{X}\right)^T\delta\mathbf{W}\right],\end{aligned} \tag{C.27}$$

since $\mathbf{X}^T\mathbf{X}$ is symmetric. Thus both linear terms are equal, and

$$\begin{aligned}\mathcal{H}(\mathbf{X},\mathbf{W}+\delta\mathbf{W}) &= \mathcal{H}(\mathbf{X},\mathbf{W}) - \frac{1}{N}\mathrm{Tr}\left[\left(\mathbf{W}\mathbf{X}^T\mathbf{X}\right)^T\delta\mathbf{W}\right] + \mathcal{O}(\|\delta\mathbf{W}\|_F^2)\\ &= \mathcal{H}(\mathbf{X},\mathbf{W}) + \left\langle -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}}, \delta\mathbf{W}\right\rangle_F + \mathcal{O}(\|\delta\mathbf{W}\|_F^2).\end{aligned} \tag{C.28}$$

Therefore,

$$\nabla_{\mathbf{W}}\mathcal{H} = -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} = -\frac{1}{N}\mathbf{W}\mathbf{X}^T\mathbf{X}.$$

Part 3: Representation derivative of the row-orthogonality regularizer. Let

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T - \mathbf{I}_m.$$

The structural regularizer is

$$\mathcal{R}(\mathbf{W}) = \frac{1}{4}\mathrm{Tr}(\mathbf{S}^2).$$

Under $\mathbf{W} \mapsto \mathbf{W} + \delta\mathbf{W}$,

$$\begin{aligned}(\mathbf{W}+\delta\mathbf{W})(\mathbf{W}+\delta\mathbf{W})^T - \mathbf{I}_m &= \mathbf{W}\mathbf{W}^T - \mathbf{I}_m + \delta\mathbf{W}\mathbf{W}^T + \mathbf{W}\delta\mathbf{W}^T + \delta\mathbf{W}\delta\mathbf{W}^T\\ &= \mathbf{S} + \delta\mathbf{S} + \mathcal{O}(\|\delta\mathbf{W}\|_F^2),\end{aligned} \tag{C.29}$$

where

$$\delta\mathbf{S} = \delta\mathbf{W}\mathbf{W}^T + \mathbf{W}\delta\mathbf{W}^T.$$

Then

$$\begin{aligned}\mathcal{R}(\mathbf{W}+\delta\mathbf{W}) &= \frac{1}{4}\mathrm{Tr}\left[(\mathbf{S}+\delta\mathbf{S})^2\right] + \mathcal{O}(\|\delta\mathbf{W}\|_F^2)\\ &= \frac{1}{4}\mathrm{Tr}\left(\mathbf{S}^2 + \mathbf{S}\delta\mathbf{S} + \delta\mathbf{S}\mathbf{S}\right) + \mathcal{O}(\|\delta\mathbf{W}\|_F^2)\\ &= \mathcal{R}(\mathbf{W}) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{S}\delta\mathbf{S}\right) + \mathcal{O}(\|\delta\mathbf{W}\|_F^2).\end{aligned} \tag{C.30}$$

Substituting $\delta\mathbf{S}$,

$$\begin{aligned}\mathrm{Tr}(\mathbf{S}\delta\mathbf{S}) &= \mathrm{Tr}\left[\mathbf{S}\left(\delta\mathbf{W}\mathbf{W}^T + \mathbf{W}\delta\mathbf{W}^T\right)\right]\\ &= \mathrm{Tr}\left(\mathbf{S}\delta\mathbf{W}\mathbf{W}^T\right) + \mathrm{Tr}\left(\mathbf{S}\mathbf{W}\delta\mathbf{W}^T\right).\end{aligned} \tag{C.31}$$

For the first term,

$$\mathrm{Tr}\left(\mathbf{S}\delta\mathbf{W}\mathbf{W}^T\right) = \mathrm{Tr}\left(\mathbf{W}^T\mathbf{S}\delta\mathbf{W}\right) = \mathrm{Tr}\left[(\mathbf{S}\mathbf{W})^T\delta\mathbf{W}\right],$$

where $\mathbf{S}^T = \mathbf{S}$. For the second term,

$$\mathrm{Tr}\left(\mathbf{S}\mathbf{W}\delta\mathbf{W}^T\right) = \mathrm{Tr}\left[(\mathbf{S}\mathbf{W})^T\delta\mathbf{W}\right].$$

Therefore,

$$\mathrm{Tr}(\mathbf{S}\delta\mathbf{S}) = 2\mathrm{Tr}\left[(\mathbf{S}\mathbf{W})^T\delta\mathbf{W}\right],$$

and Eq. C.30 gives

$$\mathcal{R}(\mathbf{W}+\delta\mathbf{W}) = \mathcal{R}(\mathbf{W}) + \langle \mathbf{S}\mathbf{W}, \delta\mathbf{W}\rangle_F + \mathcal{O}(\|\delta\mathbf{W}\|_F^2). \tag{C.32}$$

Thus

$$\nabla_{\mathbf{W}}\mathcal{R} = \left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)\mathbf{W}.$$

Combining the variance and regularization derivatives gives Eq. C.22. ■

The Euclidean gradients in Eqs. C.21–C.22 are useful for computing variations of the Lyapunov surrogate as an ambient smooth extension. The admissible spatial direction, however, must lie in $T_{\mathbf{X}}\Phi(\mathcal{M})^N$, so the spatial component of the skeleton uses the row-wise tangent projection $\Pi_{\mathbf{X}}\nabla_{\mathbf{X}}\mathcal{E}$ defined in Appendix C.1. Likewise, the finite-sample subspace Oja dynamics in Eq. C.7 are not treated as an unconstrained Euclidean gradient flow in $\mathbf{W}$; on the row-orthonormal component, they coincide with the row-Stiefel projected variance-maximization direction. Away from this idealized component, $\mathcal{E}$ should be interpreted as a Lyapunov surrogate rather than as an exact potential generating the full transient trajectory. We therefore use the following projected deterministic skeleton.

Definition 11 (Projected Deterministic Coupled Skeleton) *Let*

$$\mathcal{S}_{m,n} = \left\{\mathbf{W}\in\mathbb{R}^{m\times n} : \mathbf{W}\mathbf{W}^T = \mathbf{I}_m\right\}$$

be the row-Stiefel manifold. For a matrix direction

$$\mathbf{Z}\in\mathbb{R}^{m\times n},$$

the orthogonal projection onto the tangent space $T_{\mathbf{W}}\mathcal{S}_{m,n}$ is

$$\Pi^{\mathcal{S}}_{\mathbf{W}}(\mathbf{Z}) = \mathbf{Z} - \mathrm{Sym}\left(\mathbf{Z}\mathbf{W}^T\right)\mathbf{W}, \qquad \mathrm{Sym}(\mathbf{A}) = \frac{1}{2}\left(\mathbf{A}+\mathbf{A}^T\right)\ [2,\ 27]. \tag{C.33}$$

On $\mathcal{S}_{m,n}$, the regularizer satisfies

$$\mathbf{W}\mathbf{W}^T - \mathbf{I}_m = \mathbf{0},$$

so the projected $\mathbf{W}$-gradient of $\mathcal{E}$ is determined by the variance term. Since

$$\nabla_{\mathbf{W}}\mathcal{E} = -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} \qquad \text{on } \mathcal{S}_{m,n},$$

the row-Stiefel gradient is

$$\begin{aligned}\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\,\mathcal{E} &= \Pi^{\mathcal{S}}_{\mathbf{W}}\left(-\mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\right)\\ &= -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} - \mathrm{Sym}\left(-\mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\right)\mathbf{W}\\ &= -\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} + \mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\mathbf{W},\end{aligned} \tag{C.34}$$

because

$$\mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T$$

is symmetric. Hence the negative row-Stiefel gradient direction is exactly

$$-\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\,\mathcal{E} = \mathbf{W}\widehat{\Sigma}_{\mathbf{X}} - \mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\mathbf{W}, \tag{C.35}$$

which is the finite-sample subspace Oja vector field in Eq. C.7 and Eq. 14.

The deterministic coupled skeleton is therefore

$$\frac{d\mathbf{X}_t}{dt} = -\eta_X\Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}(\mathbf{X}_t,\mathbf{W}_t), \tag{C.36}$$

$$\frac{d\mathbf{W}_t}{dt} = -\eta_W\,\mathrm{grad}^{\mathcal{S}}_{\mathbf{W}}\,\mathcal{E}(\mathbf{X}_t,\mathbf{W}_t), \tag{C.37}$$

with $\eta_X,\eta_W > 0$. We use uppercase subscripts η_X,η_W for continuous-time skeleton rates and lowercase subscripts η_x,η_w for discrete implementation step sizes. To match the drift-only part of Eq. C.6, the spatial time constant is chosen as

$$\eta_X = ND\beta,$$

because Eq. C.21 gives

$$-\eta_X\Pi_{\mathbf{X}}\nabla_{\mathbf{X}}\mathcal{E} = \frac{\eta_X}{N}\Pi_{\mathbf{X}}\left(\mathbf{X}\mathbf{W}^T\mathbf{W}\right) = D\beta\Pi_{\mathbf{X}}\left(\mathbf{X}\mathbf{W}^T\mathbf{W}\right).$$

This matches the tangent-projected spatial drift in Eq. C.6. Thus the skeleton is constrained in $\mathbf{X}$ by the embedded manifold geometry and projected in $\mathbf{W}$ by the row-Stiefel geometry; it is not an unconstrained Euclidean gradient flow on the full ambient product space.

Proposition 13 (Lyapunov Dissipation of the Deterministic Coupled Skeleton) *Consider the drift-only skeleton in Eqs. C.36–C.37. Along this deterministic coupled skeleton,*

$$\frac{d}{dt}\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) = -\eta_X \left\|\Pi_{\mathbf{X}_t} \nabla_{\mathbf{X}} \mathcal{E}\right\|_F^2 - \eta_W \left\|\mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}\right\|_F^2 \leq 0. \tag{C.38}$$

This formalizes Proposition 4 in the main text.

Proof: The proof is a direct chain-rule computation on the product space, but we write out the matrix inner products to make the role of the row-Stiefel projection explicit.

By the chain rule on the constrained product geometry [1, 54, 69],

$$\begin{aligned}\frac{d}{dt}\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) &= d_{\mathbf{X}}\mathcal{E}\left[\frac{d\mathbf{X}_t}{dt}\right] + d_{\mathbf{W}}\mathcal{E}\left[\frac{d\mathbf{W}_t}{dt}\right] \\ &= \left\langle \nabla_{\mathbf{X}}\mathcal{E}, \frac{d\mathbf{X}_t}{dt}\right\rangle_F + \left\langle \mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}, \frac{d\mathbf{W}_t}{dt}\right\rangle_F .\end{aligned} \tag{C.39}$$

The second equality uses the fact that admissible spatial velocities lie in $T_{\mathbf{X}_t}\Phi(\mathcal{M})^N$, while the $\mathbf{W}$-trajectory lies on the row-Stiefel manifold. Thus only the tangent-projected spatial gradient and the row-Stiefel projected $\mathbf{W}$-gradient contribute to variations along admissible directions [2, 27, 63]. Substituting the projected deterministic skeleton,

$$\begin{aligned}\frac{d}{dt}\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) &= \left\langle \nabla_{\mathbf{X}}\mathcal{E}, -\eta_X \Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\rangle_F + \left\langle \mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}, -\eta_W\, \mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}\right\rangle_F \\ &= -\eta_X \left\langle \nabla_{\mathbf{X}}\mathcal{E}, \Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\rangle_F - \eta_W \left\|\mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}\right\|_F^2 \\ &= -\eta_X \left\|\Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\|_F^2 - \eta_W \left\|\mathrm{grad}_{\mathbf{W}}^{\mathcal{S}}\, \mathcal{E}\right\|_F^2 .\end{aligned} \tag{C.40}$$

The last equality uses that $\Pi_{\mathbf{X}_t}$ is a row-wise orthogonal projection, so

$$\left\langle \nabla_{\mathbf{X}}\mathcal{E}, \Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\rangle_F = \left\|\Pi_{\mathbf{X}_t}\nabla_{\mathbf{X}}\mathcal{E}\right\|_F^2 .$$

Because $\eta_X > 0$, $\eta_W > 0$, and Frobenius norms are non-negative, Eq. C.40 implies

$$\frac{d}{dt}\mathcal{E}(\mathbf{X}_t, \mathbf{W}_t) \leq 0.$$

The identity applies to the deterministic skeleton only; the stochastic Langevin system requires the Itô treatment used in Appendices A and B [54, 78]. This completes the appendix derivation of Proposition 4 in the main text. ■

C.3 Stationary State and Principal Component Alignment

This subsection records the stationary finite-sample counterpart of the asymptotic alignment theorem from Appendix B: on the row-orthonormal invariant component, stationary row spaces must be invariant subspaces of the empirical covariance matrix. The principal eigenspace is then selected by the spectral-gap dynamics established in Theorem 3. Thus this subsection gives a conditional stationary characterization of the coupled skeleton, while the dynamical principal-subspace selection is inherited from the averaged-flow analysis in Appendix B.

Theorem 14 (Spectral Characterization of Stationary Row Spaces) *Let*

$$\mathbf{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^* \in \mathbb{R}^{n\times n}$$

be the stationary empirical covariance of the deterministic coupled skeleton. Suppose that the limiting representation lies on the nondegenerate row-orthonormal component

$$\mathbf{W}^*(\mathbf{W}^*)^T = \mathbf{I}_m, \qquad \mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T \succ 0.$$

Then the stationary condition for the finite-sample subspace Oja skeleton implies that

$$\mathrm{Row}(\mathbf{W}^*)$$

is an invariant m-dimensional subspace of $\mathbf{\Sigma}^*$. If, in addition, the spectral-gap condition and nondegenerate initialization assumptions of Theorem 3 hold for $\mathbf{\Sigma}^*$, the attracting stationary component satisfies

$$\mathrm{Row}(\mathbf{W}^*) = \mathrm{Col}(\mathbf{Q}_m^*),$$

where $\mathbf{Q}_m^* \in \mathbb{R}^{n\times m}$ contains the top m eigenvectors of $\mathbf{\Sigma}^*$. Equivalently, up to an orthogonal latent rotation,

$$\mathbf{W}^* = \mathbf{R}(\mathbf{Q}_m^*)^T, \qquad \mathbf{R} \in \mathbb{R}^{m\times m}, \qquad \mathbf{R}\mathbf{R}^T = \mathbf{I}_m.$$

Proof: At a stationary point of the deterministic finite-sample subspace Oja skeleton, the parameter velocity in Eq. C.7 vanishes. We analyze this stationary condition directly.

Step 1: The macroscopic stationary equation.

The finite-sample row-space dynamics are

$$\dot{\mathbf{W}} = \eta_W \left(\mathbf{W}\widehat{\Sigma}_{\mathbf{X}} - \mathbf{W}\widehat{\Sigma}_{\mathbf{X}}\mathbf{W}^T\mathbf{W}\right), \qquad \widehat{\Sigma}_{\mathbf{X}} = \frac{1}{N}\mathbf{X}^T\mathbf{X}.$$

At a stationary state $(\mathbf{X}^*, \mathbf{W}^*)$, define

$$\mathbf{\Sigma}^* = \widehat{\Sigma}_{\mathbf{X}^*} = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^*.$$

The stationary condition $\dot{\mathbf{W}} = \mathbf{0}$ gives

$$\mathbf{0}_{m\times n} = \mathbf{W}^*\mathbf{\Sigma}^* - \mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T\mathbf{W}^*. \tag{C.41}$$

Equivalently,

$$\mathbf{W}^*\mathbf{\Sigma}^* = \mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T\mathbf{W}^*. \tag{C.42}$$

Let

$$\mathbf{B}^* = \mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T \in \mathbb{R}^{m\times m}.$$

Then Eq. C.42 becomes

$$\mathbf{W}^*\mathbf{\Sigma}^* = \mathbf{B}^*\mathbf{W}^*. \tag{C.43}$$

This is not a scalar-eigenvalue equation for individual rows. It is a matrix invariant-subspace equation: the row space of $\mathbf{W}^*$ is mapped back into itself by $\mathbf{\Sigma}^*$ [35, 44].

To see this explicitly, note that the row-orthonormal condition gives

$$\mathbf{W}^*(\mathbf{W}^*)^T = \mathbf{I}_m.$$

Thus the orthogonal projector onto the row space of $\mathbf{W}^*$, viewed as a subspace of $\mathbb{R}^n$, is

$$\mathbf{P}^* = (\mathbf{W}^*)^T\mathbf{W}^* \in \mathbb{R}^{n\times n}.$$

Multiplying Eq. C.42 by $(\mathbf{W}^*)^T$ on the left gives

$$\begin{aligned}(\mathbf{W}^*)^T\mathbf{W}^*\mathbf{\Sigma}^* &= (\mathbf{W}^*)^T\mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T\mathbf{W}^* \\ \mathbf{P}^*\mathbf{\Sigma}^* &= \mathbf{P}^*\mathbf{\Sigma}^*\mathbf{P}^*. \end{aligned} \tag{C.44}$$

Similarly, transposing Eq. C.42 gives

$$\mathbf{\Sigma}^*(\mathbf{W}^*)^T = (\mathbf{W}^*)^T\mathbf{W}^*\mathbf{\Sigma}^*(\mathbf{W}^*)^T,$$

and multiplying on the right by $\mathbf{W}^*$ yields

$$\mathbf{\Sigma}^*\mathbf{P}^* = \mathbf{P}^*\mathbf{\Sigma}^*\mathbf{P}^*. \tag{C.45}$$

Combining Eqs. C.44–C.45, we obtain

$$\mathbf{P}^*\mathbf{\Sigma}^* = \mathbf{\Sigma}^*\mathbf{P}^* = \mathbf{P}^*\mathbf{\Sigma}^*\mathbf{P}^*. \tag{C.46}$$

Hence the row-space projector commutes with $\mathbf{\Sigma}^*$. This is the standard invariant-subspace condition for a symmetric covariance operator [44].

Step 2: Matrix block decomposition.

Let

$$\mathbf{V} = (\mathbf{W}^*)^T = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n\times m}$$

denote the transpose of the row-orthonormal representation matrix. The columns of $\mathbf{V}$ form an orthonormal basis for $\mathrm{Row}(\mathbf{W}^*)$, because

$$\mathbf{V}^T\mathbf{V} = \mathbf{W}^*(\mathbf{W}^*)^T = \mathbf{I}_m.$$

The matrix equation Eq. C.43 can be transposed into

$$\mathbf{\Sigma}^*\mathbf{V} = \mathbf{V}\mathbf{B}^*. \tag{C.47}$$

Expanding Eq. C.47 columnwise gives

$$\begin{aligned}\mathbf{\Sigma}^*\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \\ | & | & & | \end{bmatrix}\begin{bmatrix} B^*_{11} & B^*_{12} & \cdots & B^*_{1m} \\ B^*_{21} & B^*_{22} & \cdots & B^*_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B^*_{m1} & B^*_{m2} & \cdots & B^*_{mm} \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | \\ \sum_{k=1}^m B^*_{k1}\mathbf{v}_k & \sum_{k=1}^m B^*_{k2}\mathbf{v}_k & \cdots & \sum_{k=1}^m B^*_{km}\mathbf{v}_k \\ | & | & & | \end{bmatrix}. \end{aligned} \tag{C.48}$$

Therefore, for each $j = 1, \ldots, m$,

$$\mathbf{\Sigma}^*\mathbf{v}_j = \sum_{k=1}^m B^*_{kj}\mathbf{v}_k \in \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}. \tag{C.49}$$

Thus $\mathrm{Row}(\mathbf{W}^*)$ is invariant under $\mathbf{\Sigma}^*$. Notice that we do not need to assume that each row of $\mathbf{W}^*$ is individually an eigenvector. A different orthonormal basis inside the same invariant subspace may rotate the rows by an arbitrary orthogonal matrix.

Step 3: Spectral decomposition and coordinate projection.

Since

$$\mathbf{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^*$$

is symmetric positive semi-definite, the spectral theorem yields [44, 35]

$$\mathbf{\Sigma}^* = \mathbf{Q}^*\mathbf{\Lambda}^*(\mathbf{Q}^*)^T, \qquad \mathbf{\Lambda}^* = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_n^*), \qquad \lambda_1^* \geq \cdots \geq \lambda_n^* \geq 0. \tag{C.50}$$

Define the spectral-coordinate representation

$$\mathbf{C}^* = \mathbf{W}^*\mathbf{Q}^* \in \mathbb{R}^{m\times n}. \tag{C.51}$$

Because $\mathbf{Q}^*$ is orthogonal and $\mathbf{W}^*$ is row-orthonormal,

$$\mathbf{C}^*(\mathbf{C}^*)^T = \mathbf{W}^*\mathbf{Q}^*(\mathbf{Q}^*)^T(\mathbf{W}^*)^T = \mathbf{W}^*(\mathbf{W}^*)^T = \mathbf{I}_m. \tag{C.52}$$

Projecting Eq. C.43 onto the eigenbasis gives

$$\begin{aligned} \mathbf{W}^*\mathbf{\Sigma}^* &= \mathbf{B}^*\mathbf{W}^* \\ \mathbf{W}^*\mathbf{Q}^*\mathbf{\Lambda}^*(\mathbf{Q}^*)^T &= \mathbf{B}^*\mathbf{W}^* \\ \mathbf{W}^*\mathbf{Q}^*\mathbf{\Lambda}^* &= \mathbf{B}^*\mathbf{W}^*\mathbf{Q}^* \qquad \text{right-multiplying by } \mathbf{Q}^* \\ \mathbf{C}^*\mathbf{\Lambda}^* &= \mathbf{B}^*\mathbf{C}^*. \end{aligned} \tag{C.53}$$

This equation says that the nonzero columns of $\mathbf{C}^*$ can occur only inside eigenspaces whose eigenvalues participate in the invariant subspace. In the absence of eigenvalue degeneracy, this reduces to coordinate support on a set of m eigen-directions. With degeneracy, the support is identifiable only up to rotations within the degenerate eigenspaces.

Step 4: Subspace alignment resolution.

The stationary equation alone identifies invariant subspaces of $\mathbf{\Sigma}^*$, not necessarily the principal one. Principal-subspace selection requires the stability and spectral-gap mechanism proved in Appendix B.

Under the assumptions of Theorem 3, the averaged subspace Oja flow suppresses components outside the leading eigenspace. In the spectral block notation of Appendix B.4, writing

$$\mathbf{C}(\tau) = [\mathbf{C}_1(\tau) \quad \mathbf{C}_2(\tau)],$$

the relative residual coordinate

$$\mathbf{G}(\tau) = \mathbf{C}_1(\tau)^{-1}\mathbf{C}_2(\tau)$$

obeys the gap-driven linear equation

$$\dot{\mathbf{G}} = \gamma\,(\mathbf{G}\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1\mathbf{G}),$$

and therefore

$$\|\mathbf{G}(\tau)\|_F \le \|\mathbf{G}(0)\|_F \exp\left[-\gamma\,(\lambda_m - \lambda_{m+1})\,\tau\right].$$

Thus the residual component outside the top-m eigenspace vanishes asymptotically. Applying this spectral-gap selection to the stationary covariance $\boldsymbol{\Sigma}^*$ yields

$$\mathrm{Row}(\mathbf{W}^*) = \mathrm{Col}(\mathbf{Q}_m^*).$$

Since $\mathbf{W}^*$ is row-orthonormal, any two row-orthonormal bases of the same m-dimensional principal subspace differ by an orthogonal matrix [2, 27]. Hence there exists

$$\mathbf{R} \in \mathbb{R}^{m\times m}, \qquad \mathbf{R}\mathbf{R}^T = \mathbf{I}_m,$$

such that

$$\mathbf{W}^* = \mathbf{R}(\mathbf{Q}_m^*)^T.$$

This proves the claimed stationary spectral characterization and its alignment with the principal subspace selected by the averaged dynamics [2, 27, 44, 77, 91]. This is the stationary row-space characterization used in the feature-decorrelation statement of Section 4. ■

C.4 Latent Feature Decorrelation

The preceding subsection characterized the stationary row space of the deterministic coupled skeleton. We now translate that row-space statement into a covariance statement for the latent representation

$$\mathbf{Y}^* = \mathbf{X}^*(\mathbf{W}^*)^T \in \mathbb{R}^{N\times m}.$$

The result below formalizes covariance decorrelation in the latent coordinates. When task labels are aligned with the leading covariance directions, this decorrelated structure can support downstream linear classification; the supervised effect is therefore treated as a task-dependent empirical consequence of the representation geometry [6, 42, 68].

Here, a latent gauge refers only to the arbitrary choice of orthonormal coordinates inside the learned row space; changing the gauge rotates the latent coordinates without changing the represented subspace.

Theorem 15 (Latent Covariance Decorrelation) *Let*

$$\boldsymbol{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^* \in \mathbb{R}^{n\times n}$$

be the stationary empirical covariance, and suppose that the stationary row space satisfies the conclusion of Theorem 14. That is, for the top-m eigenvector matrix

$$\mathbf{Q}_m^* = \begin{bmatrix} | & | & & | \\ \mathbf{q}_1^* & \mathbf{q}_2^* & \cdots & \mathbf{q}_m^* \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n\times m},$$

there exists an orthogonal latent rotation

$$\mathbf{R} \in \mathbb{R}^{m\times m}, \qquad \mathbf{R}\mathbf{R}^T = \mathbf{I}_m,$$

such that

$$\mathbf{W}^* = \mathbf{R}(\mathbf{Q}_m^*)^T.$$

Then the stationary latent covariance is

$$\boldsymbol{\Sigma}_{\mathbf{Y}^*} = \frac{1}{N}(\mathbf{Y}^*)^T\mathbf{Y}^* = \mathbf{R}\boldsymbol{\Lambda}_m^*\mathbf{R}^T, \tag{C.54}$$

where

$$\mathbf{\Lambda}_m^* = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_m^*)$$

contains the leading m eigenvalues of $\mathbf{\Sigma}^*$. Consequently, in the canonical latent gauge $\mathbf{R} = \mathbf{I}_m$, or after applying the inverse latent rotation $\mathbf{R}^T$, the latent covariance is diagonal:

$$\mathbf{R}^T \mathbf{\Sigma}_{\mathbf{Y}^*} \mathbf{R} = \mathbf{\Lambda}_m^*. \tag{C.55}$$

Thus the stationary representation is decorrelated up to the unavoidable orthogonal non-identifiability of the latent coordinates.

Proof: We compute the empirical covariance of the stationary latent features and then insert the spectral characterization of the stationary row space.

Step 1: Latent covariance as a projected spatial covariance.

By definition,

$$\mathbf{Y}^* = \mathbf{X}^* (\mathbf{W}^*)^T.$$

Therefore,

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{Y}^*} &= \frac{1}{N} (\mathbf{Y}^*)^T \mathbf{Y}^* \\
&= \frac{1}{N} \left(\mathbf{X}^* (\mathbf{W}^*)^T \right)^T \left(\mathbf{X}^* (\mathbf{W}^*)^T \right) \\
&= \frac{1}{N} \left(\mathbf{W}^* (\mathbf{X}^*)^T \right) \left(\mathbf{X}^* (\mathbf{W}^*)^T \right) \\
&= \mathbf{W}^* \left(\frac{1}{N} (\mathbf{X}^*)^T \mathbf{X}^* \right) (\mathbf{W}^*)^T \\
&= \mathbf{W}^* \mathbf{\Sigma}^* (\mathbf{W}^*)^T. \quad [35, 44]
\end{aligned} \tag{C.56}$$

Thus the latent covariance is the compression of the stationary spatial covariance onto the learned row space.

Step 2: Block spectral decomposition of the stationary covariance.

Since $\mathbf{\Sigma}^*$ is symmetric positive semi-definite, it admits an orthogonal eigendecomposition

$$\mathbf{\Sigma}^* = \mathbf{Q}^* \mathbf{\Lambda}^* (\mathbf{Q}^*)^T, \qquad \mathbf{\Lambda}^* = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_n^*), \qquad \lambda_1^* \geq \cdots \geq \lambda_n^* \geq 0. \tag{C.57}$$

Partition the eigenbasis and eigenvalue matrix into the leading and residual blocks:

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q}_m^* & \mathbf{Q}_\perp^* \end{bmatrix}, \qquad \mathbf{\Lambda}^* = \begin{bmatrix} \mathbf{\Lambda}_m^* & \mathbf{0}_{m \times (n-m)} \\ \mathbf{0}_{(n-m) \times m} & \mathbf{\Lambda}_\perp^* \end{bmatrix}, \tag{C.58}$$

where

$$\mathbf{Q}_m^* \in \mathbb{R}^{n \times m}, \qquad \mathbf{Q}_\perp^* \in \mathbb{R}^{n \times (n-m)},$$

and

$$(\mathbf{Q}_m^*)^T \mathbf{Q}_m^* = \mathbf{I}_m, \qquad (\mathbf{Q}_\perp^*)^T \mathbf{Q}_\perp^* = \mathbf{I}_{n-m}, \qquad (\mathbf{Q}_m^*)^T \mathbf{Q}_\perp^* = \mathbf{0}_{m \times (n-m)}.$$

The spectral decomposition can therefore be written explicitly as

$$\mathbf{\Sigma}^* = \mathbf{Q}_m^* \mathbf{\Lambda}_m^* (\mathbf{Q}_m^*)^T + \mathbf{Q}_\perp^* \mathbf{\Lambda}_\perp^* (\mathbf{Q}_\perp^*)^T. \tag{C.59}$$

These are standard consequences of the spectral theorem for symmetric matrices [44].

Step 3: Projection onto the learned row space.

By Theorem 14, the learned stationary row space equals the top-m eigenspace, up to an orthogonal latent rotation:

$$\mathbf{W}^* = \mathbf{R} (\mathbf{Q}_m^*)^T, \qquad \mathbf{R}\mathbf{R}^T = \mathbf{I}_m. \tag{C.60}$$

We first compute how $\mathbf{W}^*$ acts on the full eigenbasis:

$$\begin{aligned}
\mathbf{W}^* \mathbf{Q}^* &= \mathbf{R} (\mathbf{Q}_m^*)^T \begin{bmatrix} \mathbf{Q}_m^* & \mathbf{Q}_\perp^* \end{bmatrix} \\
&= \mathbf{R} \begin{bmatrix} (\mathbf{Q}_m^*)^T \mathbf{Q}_m^* & (\mathbf{Q}_m^*)^T \mathbf{Q}_\perp^* \end{bmatrix} \\
&= \mathbf{R} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m \times (n-m)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R} & \mathbf{0}_{m \times (n-m)} \end{bmatrix}.
\end{aligned} \tag{C.61}$$

Similarly,

$$(\mathbf{W}^*\mathbf{Q}^*)^T = \begin{bmatrix} \mathbf{R}^T \\ \mathbf{0}_{(n-m)\times m} \end{bmatrix}. \tag{C.62}$$

Step 4: Block multiplication of the latent covariance.

Substituting the spectral decomposition into Eq. C.56, we get

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{Y}^*} &= \mathbf{W}^* \left(\mathbf{Q}^* \boldsymbol{\Lambda}^* (\mathbf{Q}^*)^T\right) (\mathbf{W}^*)^T \\
&= \left(\mathbf{W}^*\mathbf{Q}^*\right) \boldsymbol{\Lambda}^* \left(\mathbf{W}^*\mathbf{Q}^*\right)^T \\
&= \begin{bmatrix} \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_m^* & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_\perp^* \end{bmatrix} \begin{bmatrix} \mathbf{R}^T \\ \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}\boldsymbol{\Lambda}_m^* + \mathbf{0}\cdot\mathbf{0} & \mathbf{R}\mathbf{0} + \mathbf{0}\boldsymbol{\Lambda}_\perp^* \end{bmatrix} \begin{bmatrix} \mathbf{R}^T \\ \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}\boldsymbol{\Lambda}_m^* & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}^T \\ \mathbf{0} \end{bmatrix} \\
&= \mathbf{R}\boldsymbol{\Lambda}_m^* \mathbf{R}^T.
\end{aligned} \tag{C.63}$$

This proves Eq. C.54.

Step 5: Canonical decorrelation gauge.

Equation C.63 shows that the latent covariance is diagonal only after fixing the latent rotational gauge. Indeed,

$$\begin{aligned}
\mathbf{R}^T \boldsymbol{\Sigma}_{\mathbf{Y}^*} \mathbf{R} &= \mathbf{R}^T \left(\mathbf{R}\boldsymbol{\Lambda}_m^* \mathbf{R}^T\right) \mathbf{R} \\
&= \left(\mathbf{R}^T\mathbf{R}\right) \boldsymbol{\Lambda}_m^* \left(\mathbf{R}^T\mathbf{R}\right) \\
&= \mathbf{I}_m \boldsymbol{\Lambda}_m^* \mathbf{I}_m \\
&= \boldsymbol{\Lambda}_m^*.
\end{aligned} \tag{C.64}$$

Thus the off-diagonal mass vanishes in the canonical eigenbasis of the latent coordinates:

$$\left[\mathbf{R}^T \boldsymbol{\Sigma}_{\mathbf{Y}^*} \mathbf{R}\right]_{ij} = 0, \qquad i \neq j. \tag{C.65}$$

Without fixing this orthogonal gauge, $\boldsymbol{\Sigma}_{\mathbf{Y}^*} = \mathbf{R}\boldsymbol{\Lambda}_m^*\mathbf{R}^T$ is generally not diagonal when $\mathbf{R} \neq \mathbf{I}_m$. This is not a failure of decorrelation, but the usual rotational non-identifiability of subspace representations on the Grassmannian [2, 27].

Consequently, the coupled dynamics yield a stationary representation whose covariance is diagonalizable by an orthogonal latent rotation and whose eigenvalues are precisely the leading empirical variances. This supports the feature-decorrelation claim used in the main text. It does not by itself imply supervised linear separability; such separability requires additional alignment between labels and the leading covariance directions. ■

C.5 Degenerate Spectra and Subspace Gauge Symmetry

The principal-subspace alignment theorem in Appendix B assumes a strict spectral gap

$$\lambda_m > \lambda_{m+1}.$$

This condition guarantees that the top-m eigenspace is identifiable as a unique m-dimensional subspace. When the covariance spectrum has repeated eigenvalues, two forms of non-uniqueness appear [20, 51]. First, even when the selected subspace is fixed, its row-orthonormal basis is identifiable only up to an orthogonal latent rotation. Second, if the repeated eigenvalue crosses the cutoff between selected and residual directions, the top-m subspace itself is not unique; the limiting m-plane inside the degenerate eigenspace may depend on initialization. This subsection records the corresponding Grassmannian non-identifiability and shows that the latent covariance decorrelation statement remains valid after fixing a canonical latent gauge [2, 20, 27, 51]. Concretely, this means that when an eigenvalue degeneracy crosses the m-dimensional cutoff, the spectrum determines only a degenerate eigenspace, not a unique m-dimensional selected subspace inside it.

Theorem 16 (Degenerate Spectra and Rotational Non-Identifiability) *Let the stationary covariance matrix*

$$\mathbf{\Sigma}^* = \frac{1}{N}(\mathbf{X}^*)^T\mathbf{X}^*$$

have an eigendecomposition

$$\mathbf{\Sigma}^* = \mathbf{Q}^*\mathbf{\Lambda}^*(\mathbf{Q}^*)^T.$$

Assume that an eigenvalue $\lambda^\star$ has multiplicity $r \geq 2$, with associated eigenspace

$$E_\star = \mathrm{span}(\mathbf{Q}^*_\star) \subset \mathbb{R}^n.$$

Let

$$E_> = \bigoplus_{\lambda > \lambda^\star} E_\lambda$$

be the direct sum of eigenspaces with eigenvalues strictly larger than $\lambda^\star$, and let

$$r_> = \dim(E_>).$$

For target dimension m, define

$$s = m - r_>, \qquad 0 \leq s \leq r.$$

Then the following statements hold.

1. *If $s = 0$, the degenerate block lies entirely below the selected subspace and does not affect the top-m row space.*
2. *If $s = r$, the degenerate block is fully included. The selected row space*

   $$E_> \oplus E_\star$$

   is unique as a subspace, but its row-orthonormal basis is not unique.
3. *If $0 < s < r$, the cutoff passes through the degenerate eigenspace. The selected row space has the form*

   $$E_> \oplus L, \qquad L \in \mathrm{Gr}(s, E_\star),$$

   where L is an s-dimensional subspace of the degenerate eigenspace. The particular L is not determined by the spectrum alone and may depend on initialization.

In all cases, if $\mathbf{Q}^*_{\mathrm{sel}} \in \mathbb{R}^{n\times m}$ is any orthonormal basis for the selected stationary row space, then every row-orthonormal stationary representative has the form

$$\mathbf{W}^* = \mathbf{R}(\mathbf{Q}^*_{\mathrm{sel}})^T, \qquad \mathbf{R} \in O(m).$$

The latent covariance satisfies

$$\mathbf{\Sigma}_{\mathbf{Y}^*} = \mathbf{R}\mathbf{\Lambda}^*_{\mathrm{sel}}\mathbf{R}^T,$$

where

$$\mathbf{\Lambda}^*_{\mathrm{sel}} = \mathrm{diag}(\lambda^*_{\mathrm{selected},1}, \ldots, \lambda^*_{\mathrm{selected},m}).$$

Thus the latent covariance is diagonal in the canonical latent gauge and is decorrelated up to orthogonal rotation.

Proof: We prove the result by combining the spectral-coordinate dynamics from Appendix B with a block decomposition of the stationary covariance.

Step 1: Vanishing relative exponent inside a degenerate eigenspace.

In Appendix B.4, the averaged subspace Oja flow is written in the eigenbasis of the covariance as

$$\mathbf{C}(\tau) = \mathbf{W}(\tau)\mathbf{Q},$$

and the relative residual coordinate

$$\mathbf{G}(\tau) = \mathbf{C}_1(\tau)^{-1}\mathbf{C}_2(\tau)$$

satisfies

$$\dot{\mathbf{G}} = \gamma\left(\mathbf{G}\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1\mathbf{G}\right).$$

Entrywise, this gives

$$\dot{G}_{ij} = \gamma\,(\lambda_{m+j} - \lambda_i)\,G_{ij}. \tag{C.66}$$

Hence

$$G_{ij}(\tau) = G_{ij}(0)\exp\left[-\gamma\,(\lambda_i - \lambda_{m+j})\,\tau\right]. \tag{C.67}$$

When $\lambda_i > \lambda_{m+j}$, the corresponding residual component decays exponentially. This is the spectral-gap mechanism used in Theorem 11. However, inside a degenerate eigenspace,

$$\lambda_i = \lambda_{m+j} = \lambda^\star,$$

and Eq. C.67 becomes

$$G_{ij}(\tau) = G_{ij}(0).$$

Thus the flow has no exponential preference between directions with the same eigenvalue. The relative orientation inside $E_\star$ is therefore not selected by the eigenvalue ordering alone; it is determined by the initialization and by the chosen orthonormal gauge inside the degenerate eigenspace [51, 77].

Step 2: Projection operator and Grassmann equivalence.

Let

$$\mathbf{Q}_> \in \mathbb{R}^{n\times r_>}$$

be an orthonormal basis for $E_>$, and let

$$\mathbf{Q}_\star \in \mathbb{R}^{n\times r}$$

be an orthonormal basis for $E_\star$. If $0 < s < r$, choose an s-dimensional subspace

$$L \in \mathrm{Gr}(s, E_\star)$$

with orthonormal basis

$$\mathbf{Q}_L \in \mathbb{R}^{n\times s}.$$

Then an orthonormal basis for the selected m-dimensional row space can be written as

$$\mathbf{Q}_{\mathrm{sel}} = [\mathbf{Q}_> \quad \mathbf{Q}_L] \in \mathbb{R}^{n\times m}, \qquad m = r_> + s. \tag{C.68}$$

The corresponding row-space projector is

$$\mathbf{\Pi}_{\mathrm{sel}} = \mathbf{Q}_{\mathrm{sel}}\mathbf{Q}_{\mathrm{sel}}^T = \mathbf{Q}_>\mathbf{Q}_>^T + \mathbf{Q}_L\mathbf{Q}_L^T. \tag{C.69}$$

If $s = r$, then $L = E_\star$, $\mathbf{Q}_L\mathbf{Q}_L^T = \mathbf{Q}_\star\mathbf{Q}_\star^T$, and the selected projector is uniquely determined:

$$\mathbf{\Pi}_{\mathrm{sel}} = \mathbf{Q}_>\mathbf{Q}_>^T + \mathbf{Q}_\star\mathbf{Q}_\star^T.$$

If $0 < s < r$, then $\mathbf{Q}_L\mathbf{Q}_L^T$ depends on the selected s-plane $L \subset E_\star$. Therefore the spectrum determines the full degenerate eigenspace $E_\star$, but not a unique s-dimensional subspace inside it. The collection of possible choices is the Grassmannian

$$\mathrm{Gr}(s, E_\star),$$

which records the intrinsic non-identifiability under spectral degeneracy [2, 27].

For any row-orthonormal stationary representative $\mathbf{W}^*$, the rows form an orthonormal basis of the selected row space. Hence there exists an orthogonal latent rotation

$$\mathbf{R} \in O(m)$$

such that the following representative relation holds [2, 27, 39]:

$$\mathbf{W}^* = \mathbf{R}\mathbf{Q}_{\mathrm{sel}}^T. \tag{C.70}$$

This is basis non-identifiability: $\mathbf{W}^*$ is not unique as a matrix, even when the selected subspace is fixed.

Step 3: Block matrix gauge transformation and covariance invariance.

We now compute the latent covariance under the stationary representative in Eq. C.70. Let the stationary covariance decompose as

$$\mathbf{\Sigma}^* = \mathbf{Q}_>\mathbf{\Lambda}_>\mathbf{Q}_>^T + \lambda^\star\mathbf{Q}_\star\mathbf{Q}_\star^T + \mathbf{Q}_<\mathbf{\Lambda}_<\mathbf{Q}_<^T, \tag{C.71}$$

where $\mathbf{Q}_<$ spans eigenspaces with eigenvalues strictly smaller than $\lambda^\star$. Since $L \subset E_\star$, we may further decompose the degenerate eigenspace as

$$E_\star = L \oplus L_\star^\perp,$$

with orthonormal bases $\mathbf{Q}_L \in \mathbb{R}^{n\times s}$ and $\mathbf{Q}_{L^\perp} \in \mathbb{R}^{n\times(r-s)}$. Then

$$\mathbf{Q}_\star \mathbf{Q}_\star^T = \mathbf{Q}_L \mathbf{Q}_L^T + \mathbf{Q}_{L^\perp} \mathbf{Q}_{L^\perp}^T.$$

Thus Eq. C.71 can be written as

$$\mathbf{\Sigma}^* = \mathbf{Q}_> \mathbf{\Lambda}_> \mathbf{Q}_>^T + \lambda^\star \mathbf{Q}_L \mathbf{Q}_L^T + \lambda^\star \mathbf{Q}_{L^\perp} \mathbf{Q}_{L^\perp}^T + \mathbf{Q}_< \mathbf{\Lambda}_< \mathbf{Q}_<^T. \tag{C.72}$$

Define the selected eigenvalue block

$$\mathbf{\Lambda}_{\text{sel}} = \begin{bmatrix} \mathbf{\Lambda}_> & \mathbf{0} \\ \mathbf{0} & \lambda^\star \mathbf{I}_s \end{bmatrix} \in \mathbb{R}^{m\times m}. \tag{C.73}$$

Using

$$\mathbf{Q}_{\text{sel}} = [\mathbf{Q}_> \quad \mathbf{Q}_L],$$

we compute the compression of $\mathbf{\Sigma}^*$ onto the selected row space:

$$\begin{aligned} \mathbf{Q}_{\text{sel}}^T \mathbf{\Sigma}^* \mathbf{Q}_{\text{sel}} &= \begin{bmatrix} \mathbf{Q}_>^T \\ \mathbf{Q}_L^T \end{bmatrix} \mathbf{\Sigma}^* \begin{bmatrix} \mathbf{Q}_> & \mathbf{Q}_L \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Q}_>^T \mathbf{\Sigma}^* \mathbf{Q}_> & \mathbf{Q}_>^T \mathbf{\Sigma}^* \mathbf{Q}_L \\ \mathbf{Q}_L^T \mathbf{\Sigma}^* \mathbf{Q}_> & \mathbf{Q}_L^T \mathbf{\Sigma}^* \mathbf{Q}_L \end{bmatrix}. \end{aligned} \tag{C.74}$$

We evaluate the four blocks separately. Since $\mathbf{Q}_>$ consists of eigenvectors with eigenvalue block $\mathbf{\Lambda}_>$,

$$\mathbf{\Sigma}^* \mathbf{Q}_> = \mathbf{Q}_> \mathbf{\Lambda}_>,$$

and therefore

$$\mathbf{Q}_>^T \mathbf{\Sigma}^* \mathbf{Q}_> = \mathbf{Q}_>^T \mathbf{Q}_> \mathbf{\Lambda}_> = \mathbf{\Lambda}_>.$$

Since $L \subset E_\star$,

$$\mathbf{\Sigma}^* \mathbf{Q}_L = \lambda^\star \mathbf{Q}_L,$$

so

$$\mathbf{Q}_L^T \mathbf{\Sigma}^* \mathbf{Q}_L = \lambda^\star \mathbf{Q}_L^T \mathbf{Q}_L = \lambda^\star \mathbf{I}_s.$$

The cross terms vanish by orthogonality of eigenspaces associated with distinct eigenvalues:

$$\mathbf{Q}_>^T \mathbf{\Sigma}^* \mathbf{Q}_L = \lambda^\star \mathbf{Q}_>^T \mathbf{Q}_L = \mathbf{0},$$

and

$$\mathbf{Q}_L^T \mathbf{\Sigma}^* \mathbf{Q}_> = \mathbf{Q}_L^T \mathbf{Q}_> \mathbf{\Lambda}_> = \mathbf{0}.$$

Substituting these identities into Eq. C.74 gives

$$\mathbf{Q}_{\text{sel}}^T \mathbf{\Sigma}^* \mathbf{Q}_{\text{sel}} = \begin{bmatrix} \mathbf{\Lambda}_> & \mathbf{0} \\ \mathbf{0} & \lambda^\star \mathbf{I}_s \end{bmatrix} = \mathbf{\Lambda}_{\text{sel}}. \tag{C.75}$$

Finally, the stationary latent covariance is

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{Y}^*} &= \mathbf{W}^* \mathbf{\Sigma}^* (\mathbf{W}^*)^T \\ &= \left(\mathbf{R}\mathbf{Q}_{\text{sel}}^T\right) \mathbf{\Sigma}^* \left(\mathbf{R}\mathbf{Q}_{\text{sel}}^T\right)^T \\ &= \mathbf{R}\mathbf{Q}_{\text{sel}}^T \mathbf{\Sigma}^* \mathbf{Q}_{\text{sel}} \mathbf{R}^T \\ &= \mathbf{R}\mathbf{\Lambda}_{\text{sel}} \mathbf{R}^T. \end{aligned} \tag{C.76}$$

In the canonical latent gauge, or after applying the inverse rotation $\mathbf{R}^T$, this becomes

$$\mathbf{R}^T \mathbf{\Sigma}_{\mathbf{Y}^*} \mathbf{R} = \mathbf{\Lambda}_{\text{sel}}. \tag{C.77}$$

Thus the covariance is diagonal in a canonical gauge. If the latent rotation $\mathbf{R}$ is block diagonal and rotates only within equal-eigenvalue blocks, then

$$\mathbf{R}\mathbf{\Lambda}_{\text{sel}} \mathbf{R}^T = \mathbf{\Lambda}_{\text{sel}},$$

because

$$\mathbf{O}_\star(\lambda^\star \mathbf{I}_s)\mathbf{O}_\star^T = \lambda^\star \mathbf{O}_\star \mathbf{O}_\star^T = \lambda^\star \mathbf{I}_s,$$

for any orthogonal block $\mathbf{O}_\star \in \mathbb{R}^{s\times s}$. For a general latent rotation mixing unequal selected eigenvalues, the covariance need not be diagonal in the displayed coordinates, but it remains orthogonally diagonalizable by the latent gauge. This is precisely the decorrelation-up-to-rotation statement in Theorem 15.

Therefore spectral degeneracy does not create a supervised linear-separability guarantee. Rather, it introduces rotational and, at the cutoff, Grassmannian non-identifiability. The stable statement is that the selected stationary representation has a latent covariance that is diagonal in an appropriate orthogonal gauge, with equal-eigenvalue blocks invariant under internal rotations [2, 27, 39, 44]. ■

D Algorithmic Implementation and Extended Empirical Protocols

This appendix supports the empirical evaluations in Section 5 by specifying the discrete implementation, resource scaling, ablation protocol, and extended diagnostics used to numerically verify the macroscopic invariants studied in the main text:

$$E_{\mathcal{L}}, \qquad E_{\mathrm{orth}}, \qquad E_{\mathrm{sub}}, \qquad E_{\mathrm{off}}^{\mathrm{can}}.$$

The evaluation remains theory-aligned: it tests generator convergence, row-orthogonality, subspace recovery, and canonical-gauge decorrelation rather than encoder-level benchmark performance. The generator diagnostic $E_{\mathcal{L}}$ is computed by applying the empirical graph generator and the limiting Langevin generator to the same smooth test functions on a fixed evaluation grid, and is reported only for sampled manifold graph processes. Throughout Appendix D, the canonical gauge in $E_{\mathrm{off}}^{\mathrm{can}}$ is computed by the same Procrustes alignment described in Section 5, not by diagonalizing $\widehat{\Sigma}_{\mathbf{Y}}$ separately for each run.

D.1 Atomic Discretization of the Coupled System

Algorithm 1 implements a first-order discretization of the coupled slow-fast system in Eqs. 7–8. Let $t_k = k\Delta t$, and choose

$$\eta_x = \Delta t, \qquad \eta_w = \epsilon\gamma\Delta t, \qquad \eta_w \ll \eta_x,$$

matching the slow-fast scaling used in the stochastic averaging argument of Appendix B [10, 56, 57].

For each agent i, the embedded spatial coordinate is updated by the Euler–Maruyama step

$$\widetilde{\mathbf{x}}_i^{(k+1)} = \mathbf{x}_i^{(k)} + \eta_x D\beta \Pi_{\mathbf{x}_i^{(k)}}\left((\mathbf{W}^{(k)})^T\mathbf{W}^{(k)}\mathbf{x}_i^{(k)}\right) + \sqrt{2D\eta_x}\,\boldsymbol{\xi}_i^{(k)}, \qquad \boldsymbol{\xi}_i^{(k)} = \Pi_{\mathbf{x}_i^{(k)}}\widetilde{\boldsymbol{\xi}}_i^{(k)}, \quad \widetilde{\boldsymbol{\xi}}_i^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n). \tag{D.1}$$

The spatial update is then retracted to the admissible substrate domain:

$$\mathbf{x}_i^{(k+1)} = \Pi_{\mathcal{X}}\left(\widetilde{\mathbf{x}}_i^{(k+1)}\right), \qquad \Pi_{\mathcal{X}}(\mathbf{v}) = \arg\min_{\mathbf{z}\in\mathcal{X}} \|\mathbf{v}-\mathbf{z}\|_2. \tag{D.2}$$

For embedded synthetic manifolds, $\mathcal{X} = \Phi(\mathcal{M})$, and $\Pi_{\mathcal{X}}$ is implemented by the known parametrization or nearest-point projection [2, 63]; tangent noise is obtained by applying the analytic projector $\Pi_{\mathbf{x}}$ to ambient Gaussian noise as in Eq. D.1. For Euclidean controlled distributions with active spatial motion, $\mathcal{X}$ is a compact empirical diagnostic domain rather than all of $\mathbb{R}^n$; in our implementation we use

$$\mathcal{X} = \mathcal{C}_R = \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}-\bar{\mathbf{x}}_0\|_2 \le R_X\},$$

where $\bar{\mathbf{x}}_0$ and R_X are determined from the initial empirical substrate. The identity map is used only for fixed-stream baselines or diagnostics with no active spatial drift. This bounded retraction prevents the representation-induced positive drift from being interpreted as an unconstrained Euclidean flow. For compact notation, we write the update through the empirical covariance $\widehat{\Sigma}^{(k+1)}$; in implementation, the same vector field is accumulated directly from local rank-one Oja increments, without forming a dense $n\times n$ covariance matrix.

After collecting

$$\mathbf{X}^{(k+1)} = \begin{bmatrix} - - - & (\mathbf{x}_1^{(k+1)})^T & - - - \\ & \vdots & \\ - - - & (\mathbf{x}_N^{(k+1)})^T & - - - \end{bmatrix},$$

we write the equivalent finite-sample covariance notation as

$$\widehat{\Sigma}^{(k+1)} = \frac{1}{N}(\mathbf{X}^{(k+1)})^T\mathbf{X}^{(k+1)}. \tag{D.3}$$

The parameter update uses the finite-sample subspace Oja vector field:

$$\mathbf{W}^{(k+1)} = \mathbf{W}^{(k)} + \eta_w \left(\mathbf{W}^{(k)}\widehat{\Sigma}^{(k+1)} - \mathbf{W}^{(k)}\widehat{\Sigma}^{(k+1)}(\mathbf{W}^{(k)})^T\mathbf{W}^{(k)}\right). \tag{D.4}$$

This is the discrete counterpart of Eq. 14 and avoids the scalar normalization term of single-component Oja. The lateral term

$$\mathbf{W}^{(k)}\widehat{\Sigma}^{(k+1)}(\mathbf{W}^{(k)})^T\mathbf{W}^{(k)}$$

is the finite-sample mechanism that prevents row-space collapse, matching the row-orthogonality analysis in Appendix B [76, 77, 81, 91].

D.2 Asynchronous Decentralized Realization

The main experiments use the synchronous implementation in Algorithm 1. For deployment-oriented scalability tests, we also consider an asynchronous local-copy version in which each agent i maintains $\mathbf{W}_i^{(k)} \in \mathbb{R}^{m\times n}$ and communicates over a connected graph

$$\mathcal{G}_c = (\mathcal{V}, \mathcal{E}_c).$$

At each asynchronous round, one oriented active edge $(i, j) \in \mathcal{E}_c$ is selected by a random clock; endpoint i performs the local update before pairwise mixing; and the two endpoint copies are mixed by a pairwise gossip step with weight $\alpha \in (0, 1/2]$ [23, 74]. This yields Algorithm 2, used only for communication/memory scaling; invariant claims remain tied to the synchronous mean-field flow.

Algorithm 2 Asynchronous Decentralized Coupled Representation Learning (A-DCRL)

Require: Communication graph $\mathcal{G}_c$, gossip weight $\alpha \in (0, 1/2]$, steps η_x, η_w, diffusion coefficient D, inverse temperature β.

1: **Initialize:** local states $\mathbf{x}_i^{(0)}$ and local weights $\mathbf{W}_i^{(0)}$ for all agents i.

2: **for** asynchronous update rounds $k = 0, 1, 2, \ldots$ **do**

3: Select an active edge $(i, j) \in \mathcal{E}_c$ by a random clock.

4: **Local spatial step at endpoint i:**

$$\mathbf{y}_i^{(k)} \leftarrow \mathbf{W}_i^{(k)}\mathbf{x}_i^{(k)}, \qquad \mathbf{g}_i^{(k)} \leftarrow \Pi_{\mathbf{x}_i^{(k)}}\left((\mathbf{W}_i^{(k)})^T\mathbf{y}_i^{(k)}\right).$$

$$\widetilde{\boldsymbol{\xi}}_i^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \qquad \boldsymbol{\xi}_i^{(k)} \leftarrow \Pi_{\mathbf{x}_i^{(k)}}\widetilde{\boldsymbol{\xi}}_i^{(k)}.$$

$$\mathbf{x}_i^{(k+1)} \leftarrow \Pi_{\mathcal{X}}\left(\mathbf{x}_i^{(k)} + \eta_x D\beta\mathbf{g}_i^{(k)} + \sqrt{2D\eta_x}\boldsymbol{\xi}_i^{(k)}\right).$$

5: **Local subspace Oja step at endpoint i:**

$$\mathbf{y}_i^{(k+1)} \leftarrow \mathbf{W}_i^{(k)}\mathbf{x}_i^{(k+1)},$$

$$\Delta\mathbf{W}_i^{(k)} \leftarrow \mathbf{y}_i^{(k+1)}(\mathbf{x}_i^{(k+1)})^T - \mathbf{y}_i^{(k+1)}(\mathbf{y}_i^{(k+1)})^T\mathbf{W}_i^{(k)}.$$

$$\widehat{\mathbf{W}}_i^{(k+1)} \leftarrow \mathbf{W}_i^{(k)} + \eta_w\Delta\mathbf{W}_i^{(k)}.$$

6: **Pairwise gossip on edge (i, j):**

$$\mathbf{W}_i^{(k+1)} \leftarrow (1-\alpha)\widehat{\mathbf{W}}_i^{(k+1)} + \alpha\mathbf{W}_j^{(k)}, \qquad \mathbf{W}_j^{(k+1)} \leftarrow \alpha\widehat{\mathbf{W}}_i^{(k+1)} + (1-\alpha)\mathbf{W}_j^{(k)}.$$

7: For all $\ell \notin \{i, j\}$, set

$$\mathbf{W}_\ell^{(k+1)} \leftarrow \mathbf{W}_\ell^{(k)}, \qquad \mathbf{x}_\ell^{(k+1)} \leftarrow \mathbf{x}_\ell^{(k)}.$$

8: **end for**

The pairwise gossip step is local to the activated edge: only endpoints i, j exchange an $m \times n$ representation matrix, while inactive agents keep both their spatial states and parameter copies unchanged. The two-node mixing is doubly stochastic on the input pair $(\widehat{\mathbf{W}}_i, \mathbf{W}_j)$, so the gossip component preserves their average and contracts endpoint disagreement without requiring two-hop communication or inactive-agent variables. The active endpoint uses the same tangent-projected spatial step and rank-one subspace Oja increment as Algorithm 1 and Eq. D.4; A-DCRL therefore changes the synchronization and communication layer, not the local representation update. In this sense, it is an implementation-level decentralization of the same local update primitive, replacing the global synchronous accumulator by local copies and pairwise exchange. This local-copy realization is included to make the communication and memory footprint explicit and to test whether the invariant diagnostics remain numerically stable under a decentralized communication pattern. The formal invariant interpretation, however, remains tied to the synchronous mean-field formulation analyzed in the main text. Accordingly, we do not state a separate stochastic-approximation limit for A-DCRL. Such a result would require additional assumptions controlling the activation process, graph connectivity, and consensus error between local matrix copies. Overall, A-DCRL should be viewed as a communication-level realization of DCRL rather than a new limiting model: it preserves the local update primitive, exposes the cost of pairwise decentralization, and keeps the diagnostics anchored to the macroscopic invariants used throughout the paper.

D.3 Finite-Step Lyapunov Diagnostics

The continuous averaged flow dissipates

$$V_{\mathrm{orth}}(\mathbf{W}) = \frac{1}{4} \left\| \mathbf{W}\mathbf{W}^T - \mathbf{I}_m \right\|_F^2 .$$

The discrete implementation cannot be expected to satisfy exact monotonicity at every finite step, but it should preserve the first-order Lyapunov descent up to a second-order truncation term.

Theorem 17 (Discrete-Time Lyapunov Consistency) *Let*

$$\mathbf{F}_{\widehat{\Sigma}}(\mathbf{W}) = \mathbf{W}\widehat{\Sigma} - \mathbf{W}\widehat{\Sigma}\mathbf{W}^T\mathbf{W}, \qquad \widehat{\Sigma} \succeq 0,$$

and let

$$\mathbf{W}^+ = \mathbf{W} + \eta_w \mathbf{F}_{\widehat{\Sigma}}(\mathbf{W}).$$

On every compact sublevel set of V_{orth}, there exists a constant $C > 0$ such that

$$V_{\mathrm{orth}}(\mathbf{W}^+) - V_{\mathrm{orth}}(\mathbf{W}) = -\eta_w \mathrm{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2 \mathbf{W}\widehat{\Sigma}\mathbf{W}^T \right] + \mathcal{O}(\eta_w^2) \le C\eta_w^2. \tag{D.5}$$

Proof: A second-order Fréchet expansion gives

$$V_{\mathrm{orth}}(\mathbf{W}^+) = V_{\mathrm{orth}}(\mathbf{W}) + \eta_w \left\langle \nabla_{\mathbf{W}} V_{\mathrm{orth}}(\mathbf{W}), \mathbf{F}_{\widehat{\Sigma}}(\mathbf{W}) \right\rangle_F + \mathcal{O}(\eta_w^2 \|\mathbf{F}_{\widehat{\Sigma}}(\mathbf{W})\|_F^2). \tag{D.6}$$

Since

$$\nabla_{\mathbf{W}} V_{\mathrm{orth}}(\mathbf{W}) = \left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)\mathbf{W},$$

the first-order term is exactly the discrete analogue of Eq. 12:

$$\left\langle \nabla_{\mathbf{W}} V_{\mathrm{orth}}, \mathbf{F}_{\widehat{\Sigma}}(\mathbf{W}) \right\rangle_F = -\mathrm{Tr}\left[\left(\mathbf{W}\mathbf{W}^T - \mathbf{I}_m\right)^2 \mathbf{W}\widehat{\Sigma}\mathbf{W}^T \right] \le 0.$$

The remainder is bounded on compact sublevel sets by standard matrix Taylor estimates [54, 69]. This proves Eq. D.5. ■

Table D.1 isolates the finite-step Lyapunov effect of the discrete subspace Oja update. Theorem 17 predicts that one Euler step changes V_{orth} by a non-positive first-order term plus an $\mathcal{O}(\eta_w^2)$ discretization residual. Accordingly, this diagnostic does not re-evaluate endpoint subspace recovery, which is already reported in Table 1; instead, it measures whether the observed positive increments $\Delta V_{\mathrm{orth}}^+$ remain negligible while the average increment stays negative across substrates. This directly checks that the finite-step implementation preserves the descent structure used by the continuous-time Lyapunov analysis.

Table D.1: **Discrete-Step Lyapunov Consistency.** Finite-step simulations measuring the one-step Lyapunov residual in Theorem 17. Each row reports mean $\pm$ standard deviation over independent seeds. The positive residual $\max_k \Delta V^+_{\mathrm{orth}}$ remains negligible, while the average increment is negative across substrates, consistent with the predicted first-order descent up to $\mathcal{O}(\eta_w^2)$ discretization error.

Substrate	$\boldsymbol{\eta_w/\eta_x}$	$\mathbf{max}_k\ \boldsymbol{\Delta V^+_{\mathrm{orth}}}$	$\mathbf{mean}_k[\boldsymbol{\Delta V_{\mathrm{orth}}}]$
Swiss Roll	5.0e−03	7.594e−13 ± 3.037e−13	−2.145e−06 ± 9.532e−07
HD-GMM	5.0e−03	0.000e+00 ± 0.000e+00	−1.005e−03 ± 6.417e−05
Sparse HD	5.0e−03	0.000e+00 ± 0.000e+00	−8.205e−03 ± 1.745e−04
ResNet-512	5.0e−03	0.000e+00 ± 0.000e+00	−2.892e−03 ± 5.220e−05
ViT-768	5.0e−03	0.000e+00 ± 0.000e+00	−4.895e−03 ± 1.709e−04
VGG-4096	5.0e−03	0.000e+00 ± 0.000e+00	−1.026e−01 ± 1.663e−03
BERT-768	5.0e−03	0.000e+00 ± 0.000e+00	−1.961e−03 ± 1.011e−04

D.4 Resource Complexity and Scalability

The implementation need not explicitly form or diagonalize a dense covariance matrix during training. The covariance notation above is used to express the equivalent finite-sample vector field compactly. Each update uses vector-matrix products involving $\mathbf{x}_i$, $\mathbf{W}$, and, in the synchronous implementation, an empirical covariance accumulator.

Theorem 18 (Resource Complexity of DCRL) *Let N be the number of agents, n the ambient dimension, and $m \ll n$ the latent dimension. The synchronous implementation has per-step arithmetic cost*

$$\mathcal{O}(Nnm)$$

and stores $\mathcal{O}(nm)$ representation parameters. The asynchronous local-copy implementation has per-agent memory

$$\mathcal{O}(nm + n) = \mathcal{O}(nm)$$

and per-round communication payload

$$\mathcal{O}(nm)$$

per active edge.

Proof: The bound uses the rank-one accumulated implementation of Eq. D.4; forming $\widehat{\Sigma}$ explicitly would instead incur $\mathcal{O}(Nn^2)$ covariance construction cost. For each agent, computing $\mathbf{y}_i = \mathbf{W}\mathbf{x}_i$ and $\mathbf{W}^T\mathbf{y}_i$ costs $\mathcal{O}(nm)$. The local subspace Oja increment

$$\mathbf{y}_i\mathbf{x}_i^T - \mathbf{y}_i\mathbf{y}_i^T\mathbf{W}$$

also costs $\mathcal{O}(nm)$. Aggregating across N agents gives $\mathcal{O}(Nnm)$. Each local representation copy requires mn floating-point entries, and pairwise gossip communicates one representation matrix along the active edge, giving $\mathcal{O}(nm)$ payload per round. Exact PCA requires forming or accessing a dense covariance and solving an eigensystem, which scales as $\mathcal{O}(Nn^2)$ for covariance construction and up to $\mathcal{O}(n^3)$ for full eigendecomposition [35, 47]. ■

Table D.2 complements Theorem 18 with a fixed-configuration reference footprint. The table separates analytical scaling from measured wall-clock time: the former describes the asymptotic storage, payload, and update structure, while the latter reflects a particular NumPy implementation, hardware setting, BLAS backend, and matrix layout. The measured values should therefore be read as implementation-level reference costs under a shared configuration, not as hardware-independent speedup claims. In particular, batch eigensolvers, streaming updates, synchronous DCRL, and pairwise A-DCRL have different computational granularity: some entries report one-shot costs, whereas others report per-update or per-edge costs. The main comparison is thus the resource profile implied by each update mechanism—whether it requires dense covariance construction, centralized matrix access, or communication of an $m \times n$ representation copy—rather than an absolute runtime ordering.

Table D.2: **Reference Runtime and Payload Footprint.** Time entries report one-shot cost for batch eigensolvers and per-update cost for iterative methods; K DCRL iterations therefore cost $\mathcal{O}(KNnm)$. Reference timings are reported in milliseconds for a NumPy implementation with $N = 2048$, $n = 512$, $m = 64$, and `float32` arithmetic. Memory and payload columns report analytical per-node footprints in MB.

Method	Update Type	Per-step / One-shot Time	Mem. / Node (MB)	Payload / Node (MB)	Ref. Time (ms)
Exact PCA [47]	Batch eigensolver	$\mathcal{O}(Nn^2 + n^3)$	1.00	< 0.01	22.93 ± 1.46
Streaming Oja [76, 77]	Single-stream update	$\mathcal{O}(Nnm)$	0.12	–	61.70 ± 2.51
Subspace Oja/Sanger [91]	Lateral inhibition	$\mathcal{O}(Nnm)$	0.12	–	93.99 ± 4.23
EigenGame-type [34]	Distributed eigen-objective	$\mathcal{O}(Nnm + m^2n)$	0.12	0.12	0.95 ± 0.06
Distributed PCA [53, 66]	Consensus eigensolver	$\mathcal{O}(Nnm)$	0.12	1.00	1.93 ± 1.61
DCRL (sync.)	Coupled synchronous flow	$\mathcal{O}(Nnm)$	0.12	0.12	2.58 ± 0.09
A-DCRL	Pairwise gossip flow	$\mathcal{O}(nm)$ per active edge	0.12	0.12	0.19 ± 0.01

D.5 Relation to Static PCA and Fixed-Stream Baselines

The stationary row space of the averaged dynamics aligns with a principal eigenspace under the assumptions of Theorem 3. The coupled system therefore has a different object from static PCA. PCA computes an eigenspace of a fixed empirical covariance. DCRL instead studies a coupled process in which the spatial stream is generated by Markovian exploration and the representation dynamics evolve on a slower timescale.

This distinction determines the experimental design. Exact PCA is included as an oracle endpoint for subspace recovery; streaming Oja and subspace Oja/Sanger are included because they share the local stochastic-approximation structure; EigenGame-type and decentralized spectral/distributed PCA methods are included because they target related eigenspace objectives [34, 53, 66]. Modern deep self-supervised pipelines are not treated as direct baselines, because they optimize encoder-level contrastive or teacher-student objectives rather than the local Oja-type coupled dynamics analyzed here [13, 16, 37]. Table D.3 summarizes this mapping from reference methods to invariant diagnostics.

Table D.3: **Baseline Alignment with the Theoretical Object.** This table records why each reference method is included and which invariant it is expected to test. It is intended to make clear that the empirical comparison targets the mathematical object analyzed in the paper, rather than broad encoder-level benchmark performance.

Method	Data Stream	Primary Objective	Relevant Diagnostic
Exact PCA	Fixed batch	Principal eigenspace	$E_{\mathrm{sub}}, E_{\mathrm{off}}^{\mathrm{can}}$
Streaming Oja	Externally supplied stream	Online subspace learning	$E_{\mathrm{orth}}, E_{\mathrm{sub}}$
Subspace Oja/Sanger	Externally supplied stream	Row-space learning with inhibition	$E_{\mathrm{orth}}, E_{\mathrm{sub}}$
EigenGame-type	Distributed eigen-objective	Eigenvector game	E_{sub}
Distributed PCA	Distributed fixed data	Consensus eigenspace	Runtime, memory, E_{sub}
DCRL	Markovian generated stream	Coupled slow-fast limit	All invariants

Table D.4 provides a compact fixed-stream calibration on pretrained feature substrates. These results contextualize the invariant metrics against fixed-covariance or externally supplied-stream methods, rather than claiming broad benchmark superiority. Full DCRL trajectories and coupled-flow ablations are reported separately in the main invariant tables and in Tables D.6–D.7.

Interpretation. The oracle PCA rows verify that the reference principal subspaces are numerically well resolved. The fixed-stream baselines maintain small row-orthogonality error after normalization, but their E_{sub} varies substantially across pretrained feature spectra. This is expected: these methods optimize eigenspace objectives on a fixed stream and do not instantiate the coupled Markovian exploration analyzed in the DCRL limit theory. Consequently, Table D.4 should be read as an invariant-level calibration rather than as an encoder-level benchmark or a dominance claim.

Table D.4: **Invariant-Level Diagnostics for Fixed-Stream Baselines.** Endpoint diagnostics use fixed pretrained features extracted from CIFAR-10 [58] and AG News [100], with $m = 8$ and $N_{\max} = 20000$; values are mean $\pm$ standard deviation over three seeds, except deterministic Exact PCA. These fixed-stream baselines calibrate the invariant metrics against theory-aligned eigenspace references, rather than equating DCRL with static PCA, fixed-stream Oja, or supervised encoder-level benchmarks.

Substrate	Method	E_{orth}	E_{sub}	$E_{\mathrm{off}}^{\mathrm{can}}$
ResNet18-CIFAR10	Exact PCA oracle	5.2929e−07	9.3693e−04	9.1151e−08
ResNet18-CIFAR10	Streaming Oja	4.6722e−07 ± 1.5e−07	1.2115e−01 ± 7.0e−03	3.8577e−04 ± 4.2e−05
ResNet18-CIFAR10	Subspace Oja/Sanger	4.8156e−07 ± 1.5e−07	1.2604e−01 ± 7.7e−03	4.4364e−04 ± 1.0e−04
ResNet18-CIFAR10	EigenGame-type	5.3531e−07 ± 4.0e−08	2.4767e−01 ± 1.1e−01	6.6583e−04 ± 3.2e−04
ResNet18-CIFAR10	Distributed PCA	5.3202e−07 ± 1.9e−08	2.7787e−01 ± 1.1e−01	1.4451e−03 ± 6.0e−04
ViT-B-CIFAR10	Exact PCA oracle	3.8613e−07	6.0411e−04	1.1332e−07
ViT-B-CIFAR10	Streaming Oja	3.5206e−07 ± 3.8e−08	8.2307e−01 ± 1.8e−01	2.3375e−03 ± 1.0e−03
ViT-B-CIFAR10	Subspace Oja/Sanger	2.8129e−07 ± 9.5e−08	8.1832e−01 ± 1.4e−01	3.3884e−03 ± 1.7e−03
ViT-B-CIFAR10	EigenGame-type	3.2897e−07 ± 5.7e−08	6.8630e−01 ± 2.6e−01	9.4749e−04 ± 5.7e−04
ViT-B-CIFAR10	Distributed PCA	4.9748e−07 ± 1.4e−07	4.5495e−01 ± 1.7e−01	1.4925e−03 ± 1.5e−03
BERT-AGNews	Exact PCA oracle	3.4553e−07	1.1983e−03	6.3344e−08
BERT-AGNews	Streaming Oja	4.0925e−07 ± 6.7e−08	4.0800e−01 ± 3.1e−03	9.7984e−04 ± 8.7e−04
BERT-AGNews	Subspace Oja/Sanger	3.7090e−07 ± 4.8e−08	3.9044e−01 ± 2.7e−02	1.4378e−03 ± 1.1e−03
BERT-AGNews	EigenGame-type	3.4547e−07 ± 6.0e−08	4.0519e−01 ± 2.5e−03	8.5837e−05 ± 3.2e−07
BERT-AGNews	Distributed PCA	4.3839e−07 ± 4.6e−08	4.3023e−01 ± 7.8e−02	2.2116e−04 ± 4.0e−05
VGG16-CIFAR10	Exact PCA oracle	2.1851e−07	1.9539e−03	4.3324e−07
VGG16-CIFAR10	Streaming Oja	3.5203e−07 ± 7.0e−08	7.8583e−01 ± 1.1e−02	7.0202e−03 ± 1.6e−04
VGG16-CIFAR10	Subspace Oja/Sanger	2.8647e−07 ± 1.4e−07	7.9526e−01 ± 4.7e−03	7.6357e−03 ± 7.4e−04
VGG16-CIFAR10	EigenGame-type	2.7409e−07 ± 3.0e−08	8.0677e−01 ± 2.5e−02	7.3187e−03 ± 3.5e−04
VGG16-CIFAR10	Distributed PCA	2.5216e−07 ± 8.5e−08	8.4039e−01 ± 1.3e−02	1.0308e−02 ± 5.8e−04

D.6 Benchmark Suite and Substrate Diagnostics

The main text reports a compact subset of the evaluation. Here we specify the full substrate suite used for extended diagnostics. The suite spans three complementary regimes: low-dimensional synthetic manifolds for controlled geometric structure, high-dimensional synthetic distributions for stress-testing spectral concentration and anisotropy, and pretrained feature spaces for assessing behavior on realistic representation spectra. Across these settings, we evaluate the same invariant quantities rather than task-specific accuracy, so that the diagnostics remain aligned with the theoretical objects: generator consistency, row-orthogonality, principal-subspace alignment, and canonical-gauge decorrelation. This design tests whether the predicted macroscopic invariants persist across changes in geometry, ambient dimension, covariance spectrum, and feature origin.

Figure D.1 provides a compact visual summary of the benchmark suite, while Table D.5 records the corresponding substrate configurations used in the extended diagnostics. The visual-feature substrates are extracted from CIFAR-10 [58] using pretrained visual encoders and are used as fixed feature-covariance spectra, not as supervised classification benchmarks. They are included to test whether the row-orthogonality and covariance-decorrelation diagnostics remain stable when the ambient representation geometry is inherited from realistic image features rather than from controlled synthetic distributions. The text-feature substrate denoted as BERT-768 in the main text and figures refers to 768-dimensional BERT-base features extracted from AG News [100] using a pretrained BERT-base encoder [22]; the reported $N = 127.6$k corresponds to the combined train and test split used for the feature-spectrum diagnostics. This substrate complements the synthetic and visual-feature regimes with a language-derived covariance spectrum, allowing the same invariant-level quantities to be tested across modality, ambient dimension, spectral decay, and rank structure. In all pretrained-feature cases, the features are treated as fixed substrates for invariant diagnostics; no claim is made about end-to-end encoder training or downstream task optimality. The extended suite is therefore intended to assess robustness of the macroscopic diagnostics across substrate classes, rather than to make supervised benchmark claims.

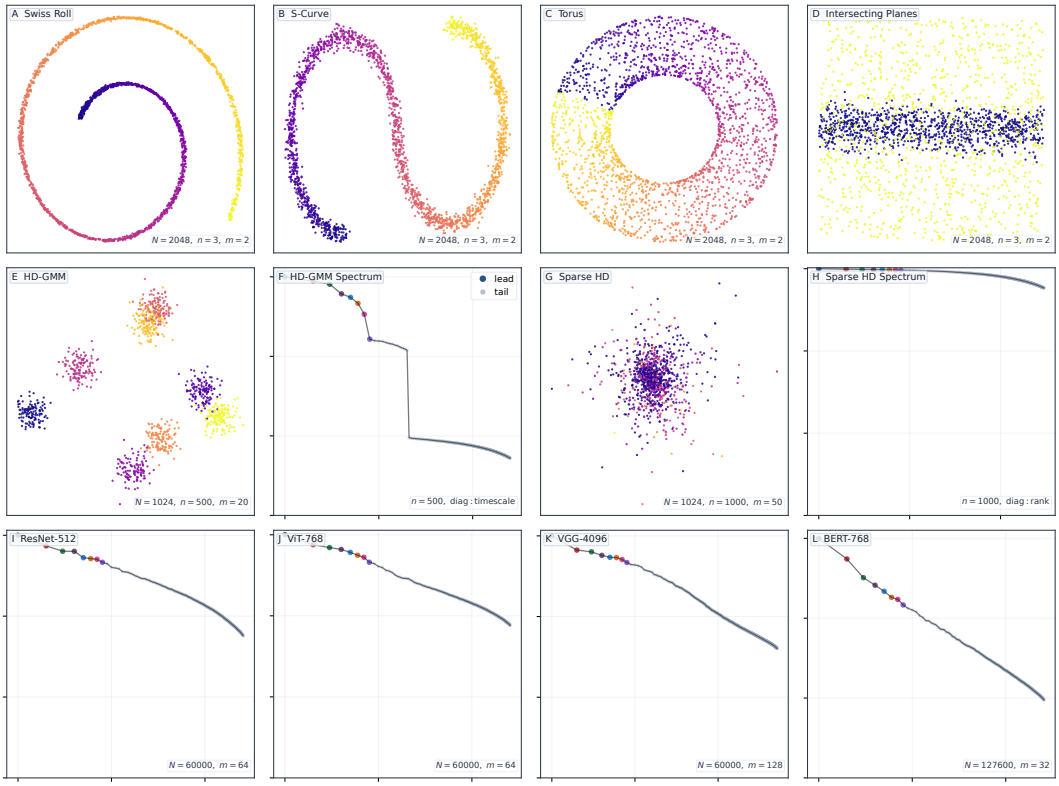

Figure D.1: **Benchmark Substrates for Extended Invariant Diagnostics.** Panels **(A)–(D)** show representative low-dimensional synthetic manifolds, including curved and periodic geometries for testing local graph construction and manifold-sensitive dynamics. Panels **(E)–(H)** show controlled high-dimensional distributions and their spectra, providing stress cases for anisotropy, sparsity, and rank concentration. Panels **(I)–(L)** show spectra from pretrained ResNet, ViT, VGG, and BERT feature spaces, representing realistic visual and text-feature covariance structures. The figure is intended to document the substrate regimes used for invariant diagnostics, rather than to serve as a supervised benchmark comparison. Across these regimes, we measure $E_{\mathcal{L}}$, E_{orth}, E_{sub}, and $E_{\mathrm{off}}^{\mathrm{can}}$. This organization separates geometric, spectral-conditioning, and pretrained-feature effects while keeping the evaluated quantities fixed. It clarifies that the extended suite tests robustness of the same macroscopic diagnostics across substrate classes.

Table D.5: **Full Benchmark Substrate Specification.** Geometry, dimensionality, sample size, target rank, and primary diagnostic role for the extended substrate suite.

Substrate	N	n	m	**Geometry / Spectrum**	**Primary Diagnostic**
Swiss Roll	2048	3	2	Curved 2D manifold	E_{sub}, $E_{\mathrm{off}}^{\mathrm{can}}$
S-Curve	2048	3	2	Curved 2D manifold	E_{sub}, $E_{\mathrm{off}}^{\mathrm{can}}$
Torus	2048	3	2	Periodic manifold	E_{orth}
Intersecting Planes	2048	3	2	Non-smooth controlled structure	Robustness
HD-GMM	1024	500	20	Controlled high-dimensional mixture	Timescale sweep
Sparse HD	1024	1000	50	Sparse high-dimensional signal	Effective rank
ResNet-512	60k	512	64	Pretrained visual features	Spectrum robustness
ViT-768	60k	768	64	Pretrained visual features	Spectrum robustness
VGG-4096	60k	4096	128	High-dimensional feature spectrum	Memory / rank
BERT-768	127.6k	768	32	BERT-base features on AG News	Rank truncation

D.7 Operator Ablations and Timescale Robustness

We isolate the contribution of each component of the coupled flow through controlled operator and timescale ablations. The purpose is to diagnose the role of each theoretical mechanism under controlled perturbations. Static PCA or static Oja can be strong references when the data distribution is fixed and fully observed. Our goal is instead to test whether the mechanisms used in the analysis have the expected numerical roles when the representation is learned from a coupled spatial process.

The ablations are organized around four mechanism-level questions. First, comparing DCRL with Static-$\mathbf{X}$ Oja separates the effect of Markovian spatial exploration from ordinary fixed-stream subspace learning. Second, the Frozen-$\mathbf{W}$ Langevin variant tests whether spatial motion alone is sufficient, or whether adaptive plasticity is needed to recover the induced principal structure. Third, removing lateral inhibition tests the row-orthogonality mechanism predicted by the Lyapunov analysis, especially whether unconstrained Hebbian growth leads to row collapse or parameter blow-up. Finally, the timescale sweep varies η_w/η_x to assess whether the slow-fast hierarchy is numerically necessary for stable finite-step behavior.

Together, these experiments should be read as mechanism diagnostics rather than broad benchmark comparisons: each ablation removes one part of the theoretical construction and records which invariant becomes unstable.

Table D.6 decomposes the coupled flow into operator-level variants. The static-$\mathbf{X}$ arm is included as a fixed-stream Oja reference, while the frozen-$\mathbf{W}$ and no-inhibition arms isolate the failure modes caused by removing plasticity or lateral stabilization.

Table D.6: **Operator Ablation for the Coupled Flow.** DCRL is compared with controlled operator-level variants. The table is intended as a mechanism diagnostic, not a universal dominance claim: fixed-stream Oja serves as a static reference, while frozen-plasticity and no-inhibition variants expose the roles of adaptation and lateral stabilization. The E_{sub} values in this ablation table are computed under perturbed mechanisms and should not be read as numerical replications of the endpoint main-table protocol.

Substrate	Variant	E_{orth}	E_{sub}	$E_{\mathrm{off}}^{\mathrm{can}}$	Effective Rank
Swiss Roll	Full DCRL	$1.540e{-}05 \pm 2.026e{-}05$	$6.244e{-}02 \pm 1.168e{-}01$	$1.051e{-}06 \pm 1.624e{-}06$	$1.57e{+}00 \pm 4.11e{-}01$
Swiss Roll	Static-$\mathbf{X}$ Oja	$6.318e{-}06 \pm 2.270e{-}07$	$0.000e{+}00 \pm 0.000e{+}00$	$4.506e{-}07 \pm 3.626e{-}07$	$1.95e{+}00 \pm 9.97e{-}08$
Swiss Roll	Frozen-$\mathbf{W}$ Langevin	$2.217e{-}08 \pm 2.117e{-}08$	$4.662e{-}01 \pm 1.845e{-}01$	$1.000e{-}01 \pm 8.539e{-}02$	$1.13e{+}00 \pm 1.26e{-}01$
Swiss Roll	No lateral inhibition	Diverged	Diverged	Diverged	Diverged
HD-GMM	Full DCRL	$5.647e{-}05 \pm 5.668e{-}05$	$3.795e{-}01 \pm 4.282e{-}01$	$3.315e{-}01 \pm 7.594e{-}02$	$1.75e{+}01 \pm 1.43e{-}01$
HD-GMM	Static-$\mathbf{X}$ Oja	$8.198e{-}06 \pm 5.027e{-}07$	$3.453e{-}04 \pm 0.000e{+}00$	$1.909e{-}01 \pm 4.713e{-}02$	$1.94e{+}01 \pm 3.59e{-}03$
HD-GMM	Frozen-$\mathbf{W}$ Langevin	$5.822e{-}07 \pm 7.831e{-}08$	$9.995e{-}01 \pm 6.013e{-}04$	$4.679e{-}01 \pm 4.980e{-}02$	$1.43e{+}01 \pm 1.34e{+}00$
HD-GMM	No lateral inhibition	Diverged	Diverged	Diverged	Diverged
VGG-4096	Full DCRL	$1.665e{-}04 \pm 1.776e{-}05$	$9.993e{-}01 \pm 6.605e{-}04$	$6.707e{-}01 \pm 7.260e{-}02$	$3.17e{+}01 \pm 1.14e{-}01$
VGG-4096	Static-$\mathbf{X}$ Oja	$7.929e{-}05 \pm 5.681e{-}06$	$9.585e{-}01 \pm 3.931e{-}02$	$6.433e{-}01 \pm 7.449e{-}02$	$3.81e{+}01 \pm 6.45e{-}02$
VGG-4096	Frozen-$\mathbf{W}$ Langevin	$1.358e{-}06 \pm 1.760e{-}07$	$9.999e{-}01 \pm 8.568e{-}05$	$5.183e{-}01 \pm 3.495e{-}02$	$3.69e{+}01 \pm 2.07e{+}00$
VGG-4096	No lateral inhibition	Diverged	Diverged	Diverged	Diverged

Table D.7 sweeps the ratio η_w/η_x to test the finite-step consequences of weakening the slow-fast hierarchy. The purpose is to identify stable, oscillatory, and unstable discretization regimes, rather than to claim monotonic improvement with a single step-size ratio.

The ablation values are independent mechanism-diagnostic endpoints and are not intended to numerically duplicate the main invariant table. They are evaluated under deliberately perturbed operators or timescale regimes, so their role is to reveal which stability or alignment property fails when a specific component of the coupled construction is removed. Finally, Table D.8 summarizes how the two ablation tables map onto specific theoretical concerns. This mapping separates endpoint invariant validation from mechanism stress tests, and prevents the extended experiments from being read as additional theory claims. The ablations should therefore be interpreted as implementation and failure-mode diagnostics for the coupled slow-fast mechanism, rather than as separate evidence for broader empirical superiority.

Table D.7: **Timescale Separation Sweep.** Sensitivity to η_w/η_x is evaluated to support the stochastic averaging assumption used in Appendix B. Stable regimes show negligible positive Lyapunov increments and controlled endpoint errors; oscillatory or unstable regimes indicate finite-step violations of the slow-fast discretization regime. Here $\max_k \Delta V^+_{\mathrm{orth}}$ is a long-run trajectory diagnostic, not the one-step residual in Table D.1.

Substrate	η_w/η_x	$\max_k \Delta V^+_{\mathrm{orth}}$	E_{orth}	E_{sub}	Observed Regime
HD-GMM	$5.0e{-}03$	$1.188e{-}07 \pm 5.988e{-}09$	$7.257e{-}05 \pm 6.397e{-}05$	$4.838e{-}01 \pm 4.828e{-}01$	oscillatory
HD-GMM	$5.0e{-}02$	$1.260e{-}04 \pm 6.624e{-}06$	$3.843e{-}07 \pm 2.565e{-}08$	$6.905e{-}05 \pm 1.544e{-}04$	oscillatory
HD-GMM	$2.0e{-}01$	$1.691e{-}02 \pm 4.030e{-}03$	$2.077e{-}01 \pm 3.298e{-}02$	$3.453e{-}04 \pm 0.000e{+}00$	oscillatory
VGG-4096	$5.0e{-}03$	$2.359e{-}06 \pm 3.521e{-}07$	$2.119e{-}04 \pm 1.149e{-}05$	$9.951e{-}01 \pm 4.645e{-}03$	unstable
VGG-4096	$5.0e{-}02$	$5.481e{-}03 \pm 3.660e{-}04$	$2.182e{-}06 \pm 9.789e{-}07$	$8.493e{-}03 \pm 1.118e{-}02$	oscillatory
VGG-4096	$2.0e{-}01$	Diverged	Diverged	Diverged	unstable

Table D.8: **Mechanism Coverage of Operator and Timescale Ablations.** This compact summary maps each ablation to the failure mode it is designed to test. It is a navigation aid for Tables D.6 and D.7, not an additional theoretical claim.

Mechanism	Perturbation	Primary Diagnostic	Failure Mode
Spatial exploration	Static-$\mathbf{X}$ Oja	$E_{\mathrm{sub}}, E^{\mathrm{can}}_{\mathrm{off}}$	Coupled stream has no effect beyond fixed-stream PCA
Plasticity	Frozen-$\mathbf{W}$ Langevin	E_{sub}, effective rank	Spatial motion alone fails to recover the principal subspace
Lateral inhibition	No lateral inhibition	E_{orth}, divergence	Hebbian growth causes row collapse or parameter blow-up
Slow-fast hierarchy	Large η_w/η_x	$\max_k \Delta V^+_{\mathrm{orth}}$, endpoint errors	Finite-step dynamics violate the averaging regime

D.8 Extended Invariant Validation and Downstream Sanity Checks

The main text reports a compact set of invariant diagnostics. Here we provide the full endpoint matrix and representative trajectories across the extended substrate suite. The purpose is not to turn every substrate into a downstream benchmark, but to examine how the theory-aligned quantities behave across geometry, dimension, and spectrum shape:

$$E_{\mathcal{L}}, \quad E_{\mathrm{orth}}, \quad E_{\mathrm{sub}}, \quad E^{\mathrm{can}}_{\mathrm{off}}.$$

Here E_{orth} measures whether the lateral-inhibition mechanism keeps the rows on the Stiefel geometry, E_{sub} measures alignment with the principal eigenspace, and $E^{\mathrm{can}}_{\mathrm{off}}$ measures canonical-gauge latent covariance decorrelation. These quantities are invariant-level diagnostics of the coupled dynamics and are not intended to serve as supervised performance claims.

Table D.9 reports additional stress-test diagnostics for the extended substrate suite. Unlike Table 1, which reports the main endpoint invariant protocol, this table focuses on row-orthogonality, canonical-gauge decorrelation, and effective-rank behavior under broader substrate and trajectory settings. For pretrained feature substrates, the principal subspace reference is the top-m eigenspace of the empirical feature covariance; generator diagnostics are not reported because these substrates are fixed feature clouds rather than sampled manifold graph processes.

Figure D.2 visualizes representative trajectories for one low-dimensional manifold, one controlled high-dimensional distribution, and one pretrained feature space, covering geometric curvature, spectral anisotropy, and realistic feature covariance. The panels track row-orthogonality, principal-subspace alignment, canonical-gauge decorrelation, and effective-rank stabilization, while the heatmaps show latent covariance before and after the update. These visualizations are qualitative diagnostics of row-space stabilization and covariance decorrelation, not evidence of supervised separability or benchmark-level superiority.

Remark on downstream linear probes. We do not use supervised linear-probe accuracy as evidence for the theoretical claims. The invariants studied in this work concern generator consistency, row-orthogonality, principal-subspace alignment, and covariance decorrelation, whereas linear sepa-

Table D.9: **Extended Stress Diagnostics Across Substrates.** Additional endpoint diagnostics under the extended substrate suite. Unlike Table 1, this table is used as a stress-test summary and reports row-orthogonality, canonical-gauge covariance decorrelation, and effective-rank behavior rather than repeating the main subspace-alignment protocol. Values are reported as mean ± standard deviation over seeds.

Substrate	$n \to m$	$E_{\mathcal{C}}$	E_{orth}	$E_{\mathrm{off}}^{\mathrm{can}}$	Eff. Rank / Var. Expl.
Swiss Roll	$3 \to 2$	$4.218\mathrm{e}{-02} \pm 0.000\mathrm{e}{+00}$	$4.540\mathrm{e}{-16} \pm 2.230\mathrm{e}{-16}$	$9.794\mathrm{e}{-02} \pm 4.484\mathrm{e}{-02}$	1.98/0.663
S-Curve	$3 \to 2$	$3.522\mathrm{e}{-02} \pm 0.000\mathrm{e}{+00}$	$5.697\mathrm{e}{-16} \pm 2.689\mathrm{e}{-16}$	$5.452\mathrm{e}{-01} \pm 1.775\mathrm{e}{-02}$	1.53/0.823
Torus	$3 \to 2$	$4.482\mathrm{e}{-02} \pm 0.000\mathrm{e}{+00}$	$4.029\mathrm{e}{-16} \pm 2.465\mathrm{e}{-16}$	$6.491\mathrm{e}{-03} \pm 2.680\mathrm{e}{-03}$	2.00/0.945
Intersecting Planes	$3 \to 2$	$3.843\mathrm{e}{-02} \pm 0.000\mathrm{e}{+00}$	$4.212\mathrm{e}{-16} \pm 7.236\mathrm{e}{-17}$	$1.357\mathrm{e}{-01} \pm 1.151\mathrm{e}{-01}$	1.93/0.695
HD-GMM	$500 \to 20$	$6.451\mathrm{e}{-03} \pm 0.000\mathrm{e}{+00}$	$1.153\mathrm{e}{-14} \pm 1.713\mathrm{e}{-15}$	$2.863\mathrm{e}{-01} \pm 2.626\mathrm{e}{-02}$	19.05/0.123
Sparse HD	$1000 \to 50$	$6.113\mathrm{e}{-03} \pm 8.674\mathrm{e}{-19}$	$2.058\mathrm{e}{-14} \pm 1.378\mathrm{e}{-16}$	$3.240\mathrm{e}{-01} \pm 1.300\mathrm{e}{-03}$	47.07/0.996
ResNet-512 [40]	$512 \to 64$	N/A	$2.230\mathrm{e}{-14} \pm 6.696\mathrm{e}{-16}$	$8.616\mathrm{e}{-01} \pm 4.208\mathrm{e}{-03}$	27.72/0.526
ViT-768 [25]	$768 \to 64$	N/A	$2.189\mathrm{e}{-14} \pm 5.933\mathrm{e}{-17}$	$8.595\mathrm{e}{-01} \pm 1.428\mathrm{e}{-03}$	26.89/0.489
VGG-4096 [92]	$4096 \to 128$	N/A	$5.152\mathrm{e}{-14} \pm 1.404\mathrm{e}{-15}$	$9.501\mathrm{e}{-01} \pm 1.324\mathrm{e}{-03}$	20.63/0.355
BERT-AGNews [22]	$768 \to 32$	N/A	$1.411\mathrm{e}{-14} \pm 5.280\mathrm{e}{-16}$	$9.475\mathrm{e}{-01} \pm 5.356\mathrm{e}{-03}$	2.50/0.695

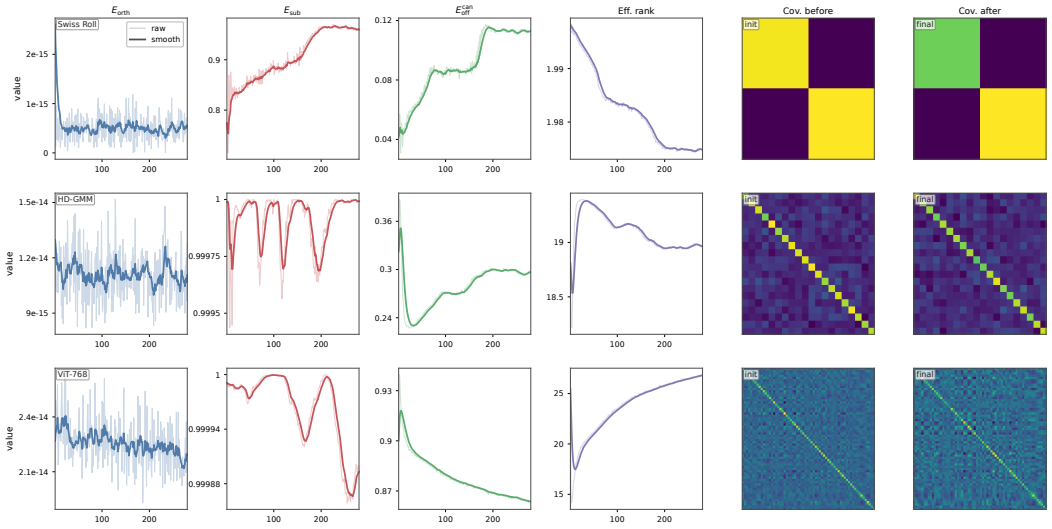

Figure D.2: **Extended Invariant Trajectories and Latent Covariance Diagnostics.** Rows correspond to Swiss Roll, HD-GMM, and ViT-768; columns show trajectory-level E_{orth}, raw $E_{\mathrm{sub}}^{\mathrm{traj}}$, $E_{\mathrm{off}}^{\mathrm{can}}$, effective rank, and latent covariance heatmaps before and after the update. These trajectory diagnostics are not directly comparable to the endpoint E_{sub} protocol in Table 1. The panels compare invariant behavior across representative regimes, without making a supervised benchmark claim.

rability also depends on label geometry, probe optimization, regularization, and the choice of fixed versus dynamically evolved coordinates. In particular, if two representations differ only by an orthogonal rotation on the same fixed samples, the optimum of a linear probe should be invariant; deviations under a coupled spatial process therefore reflect protocol-dependent label alignment rather than the subspace theorem itself. For this reason, downstream probes are omitted from the diagnostic tables and treated as downstream evaluations beyond the invariant-level theory.

D.9 Summary of Diagnostic Coverage

The extended experiments are designed to test the same mathematical quantities used in the theory, rather than to introduce additional empirical claims or benchmark-level conclusions. Appendix D therefore complements the main empirical section by making explicit the implementation and diagnostic layers behind the compact results in the main text. The main tables report endpoint invariant behavior, while the appendix explains how those quantities are produced, how the finite-step algorithm relates to the continuous-time skeleton, and how each additional diagnostic should be interpreted.

Specifically, the appendix first specifies the discrete coupled update used to approximate the slow-fast system, including the spatial retraction, finite-sample covariance, and subspace Oja update. It then checks that the finite-step implementation preserves the Lyapunov descent structure predicted by the continuous averaged flow up to discretization residuals. The resource and communication analysis records the computational footprint of synchronous and asynchronous variants without treating measured runtime as a hardware-independent speedup claim. The fixed-stream baseline calibration places the invariant metrics next to static PCA, streaming Oja, subspace Oja/Sanger, EigenGame-type methods, and distributed PCA, so that the coupled-flow results are not read in isolation from theory-aligned eigenspace references.

The operator and timescale ablations serve a different purpose. They do not constitute additional theory claims; instead, they remove or perturb individual mechanisms used in the analysis and record which diagnostic becomes unstable. Static-$\mathbf{X}$ Oja separates fixed-stream eigenspace learning from coupled Markovian exploration, frozen-$\mathbf{W}$ Langevin isolates spatial motion without adaptive plasticity, no-inhibition variants test the role of lateral stabilization, and timescale sweeps probe the slow-fast hierarchy required by the averaging argument. These experiments therefore address possible alternative explanations, including recovery by ordinary PCA-like dynamics, stability caused only by favorable hyperparameters, or the absence of appropriate eigenspace baselines.

The purpose of this final summary is organizational rather than evidentially additive. It maps each diagnostic to the theorem, assumption, implementation property, or experimental scope it is meant to support, and records the corresponding failure mode it is intended to rule out. Table D.10 provides this mapping as a compact navigation aid for the appendix. In this way, the extended protocols support the main invariant-level interpretation while keeping the claim boundary fixed: the evidence concerns generator consistency, Lyapunov stability, row-space alignment, covariance decorrelation, and implementation robustness of the analyzed coupled dynamics, not broad supervised benchmark superiority.

Table D.10: **Diagnostic Coverage of the Extended Experiments.** Each row links an empirical diagnostic to the theoretical statement, assumption, or experimental scope it supports, together with the failure mode it is designed to rule out. The table summarizes Appendix D and does not introduce an additional claim.

Diagnostic	**Measured Quantity**	**Supports**	**Failure Mode Addressed**
Generator convergence	$E_{\mathcal{L}}$	Theorem 1	Discrete graph flow not approximating Langevin limit
Lyapunov stability	$V_{\mathrm{orth}}, E_{\mathrm{orth}}$	Theorem 2	Parameter divergence or row collapse
Subspace recovery	E_{sub} in Table 1	Theorem 3	Failure to align with the principal eigenspace
Covariance decorrelation	$E_{\mathrm{off}}^{\mathrm{can}}$	Theorem 5	Latent covariance remains entangled
Operator ablation	Ablation deltas	Coupled-flow mechanism	Improvement caused by a single isolated component
Timescale sweep	η_w/η_x sensitivity	Averaging assumption	Stability only due to favorable hyperparameters
Complexity comparison	Memory / payload / time	Algorithmic feasibility	Impractical covariance construction or communication cost
Fixed-stream calibration	$E_{\mathrm{orth}}, E_{\mathrm{sub}}, E_{\mathrm{off}}^{\mathrm{can}}$	Baseline context	Missing comparison to theory-aligned eigenspace references

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction explicitly state our theoretical contributions and their necessary mathematical assumptions.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: We explicitly discuss the compactness, timescale-separation, linear-model, and spectral-gap assumptions that bound the applicability of our results.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

Answer: [Yes]

Justification: All theoretical results include explicitly stated assumptions, with complete formal proofs provided in the Appendix.

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: The integration steps, noise scales, discretization details, and invariant diagnostics are specified in Section 5 and Appendix D.

   Guidelines:

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

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

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

Answer: [No]

Justification: We provide anonymous DCRL/A-DCRL reference code and compact diagnostic scripts; full large-scale experiment code will be released upon publication.

Guidelines:

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

6. **Experimental setting/details**

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

Answer: [Yes]

Justification: All experimental settings, including integration timescales and manifold topologies, are explicitly stated in the text and Appendix.

Guidelines:

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

7. **Experiment statistical significance**

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

Answer: [Yes]

Justification: Standard deviations are reported over independent seeds as specified in the corresponding tables.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: Appendix D reports analytical memory, communication, and reference NumPy timing; compact diagnostics are CPU-runnable, while pretrained-feature diagnostics may require GPU resources.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: This foundational mathematical work does not involve human subjects, sensitive data, or deployable high-risk systems, and conforms to the NeurIPS Code of Ethics.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: This work is foundational theory and invariant-level diagnostics, with no direct deployment, released model, or foreseeable high-risk application.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The paper does not release datasets, pre-trained models, or systems that pose any risk of misuse.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: The experiments use public research assets, including CIFAR-10 and AG News for feature extraction and invariant diagnostics. We use pretrained ResNet-18, ViT-B/16, VGG-16, and BERT-base feature extractors, with Apache-2.0 or original model licenses credited where available; dataset licenses are reported according to their source listings.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [Yes]

    Justification: The supplementary material includes an anonymous reference implementation with a README, requirements file, and run script; no new dataset or deployable model is introduced.

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: No LLM is used as a methodological component; any language-editing or formatting assistance does not affect the core methods, results, or originality.

    Guidelines:

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