

# Geometric Deep Learning for High-Dimensional Single-Cell Trajectories

## Unifying Riemannian Manifold Optimization and Neural Differential Equations for Cellular Lineage Tracing

Dr. Elena Vance<sup>1,2</sup>, Marcus Thorne<sup>1</sup>, Sarah Lin<sup>2</sup>, and Prof. Julian Vance<sup>1</sup>

✉ e.vance@synapse-ai.org 🌐 <https://synapse-ai.org/labs/geo-cell>

<sup>1</sup>Synapse AI Institute, Meridian Research Park

<sup>2</sup>Vertex Bio-Data Science Center, Apex Innovation Campus

### 1. Motivation & Background

Single-cell RNA sequencing (scRNA-seq) captures continuous transcriptional profiles across thousands of individual cells. However, accurately reconstructing dynamical cell trajectories in high-dimensional gene expression space ( $\mathbb{R}^D$ ,  $D > 20,000$ ) faces severe computational challenges:

- Dimensionality & Sparsity:** High ambient noise, technical dropout rates ( $> 70\%$ ), and non-linear cell-state manifolds.
- Curvature Distortion:** Conventional reduction methods (UMAP, t-SNE) flatten non-Euclidean state spaces, destroying geodesic metrics.
- Stochastic Velocity Inference:** Traditional RNA velocity tools struggle with transient velocity fields during rapid lineage commitment.

#### Core Research Question

*Can we infer smooth, topology-preserving cellular developmental trajectories by integrating Riemannian manifold metrics directly into continuous neural differential equations?*

### 2. Mathematical Formulation

We model cellular state transitions as integral curves of a parameterized vector field  $\mathbf{f}_\theta$  defined over a Riemannian manifold  $(\mathcal{M}, g)$ .

#### 1. Riemannian Metric Learning:

Given local expression vectors  $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^D$ , we estimate localized metric tensor  $g_{ij}(\mathbf{x})$  via symmetric positive-definite neural network  $\mathbf{G}_\phi(\mathbf{x}) = \mathbf{L}_\phi(\mathbf{x})\mathbf{L}_\phi(\mathbf{x})^T + \epsilon\mathbf{I}$ :

$$d_g(\mathbf{x}_i, \mathbf{x}_j) = \inf_{\gamma} \int_0^1 \sqrt{\dot{\gamma}(t)^T \mathbf{G}_\phi(\gamma(t)) \dot{\gamma}(t)} dt$$

#### 2. Manifold Neural ODE Dynamics:

The temporal evolution of cell state  $\mathbf{z}(t) \in \mathcal{M}$  satisfies the Riemannian geodesic-regularized Neural ODE:

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{f}_\theta(\mathbf{z}(t), t) - \frac{1}{2} \Gamma_{ij}^k(\mathbf{z}) \dot{z}^i \dot{z}^j$$

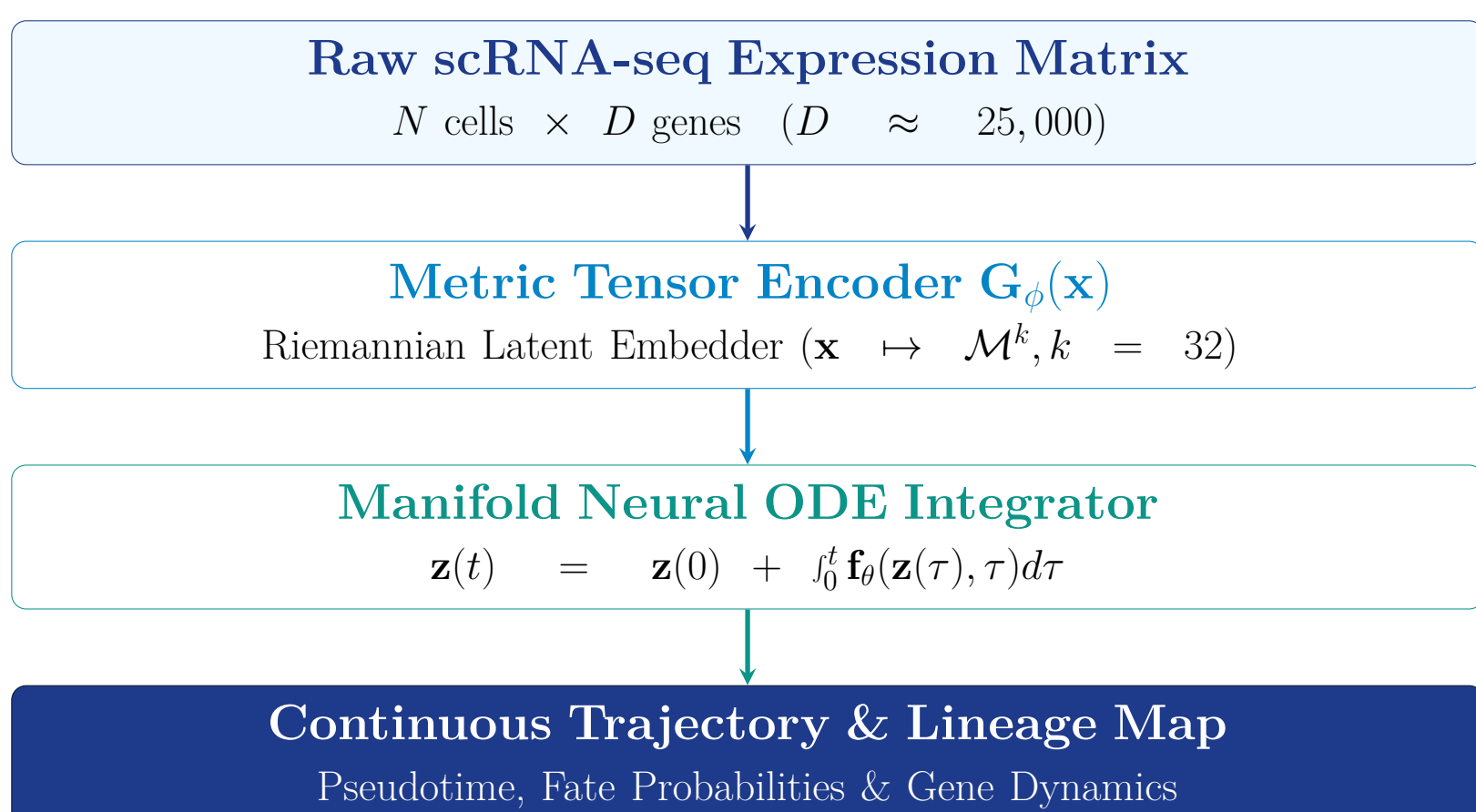
where  $\Gamma_{ij}^k$  denotes Christoffel symbols derived from metric  $g$ .

#### 3. Joint Loss Objective:

$$\mathcal{L}(\theta, \phi) = \mathcal{L}_{\text{velocity}}(\mathbf{f}_\theta, \mathbf{v}_{\text{RNA}}) + \lambda_1 \mathcal{L}_{\text{geodesic}}(\mathbf{G}_\phi) + \lambda_2 \mathcal{L}_{\text{density}}(\mathbf{z})$$

### 3. Methodological Pipeline

Our framework, **GeoCell-ODE**, processes raw count matrices through three integrated computational modules:



### 4. Benchmark Trajectory Accuracy

We benchmarked **GeoCell-ODE** against three state-of-the-art trajectory inference algorithms across 12 single-cell differentiation datasets.

#### Metric Definitions:

- Velocity Vector Correlation (VVC):** Cosine similarity between predicted and ground-truth developmental vectors.
- Geodesic Distortion Error (GDE):** Preservation of true manifold distances in latent space.

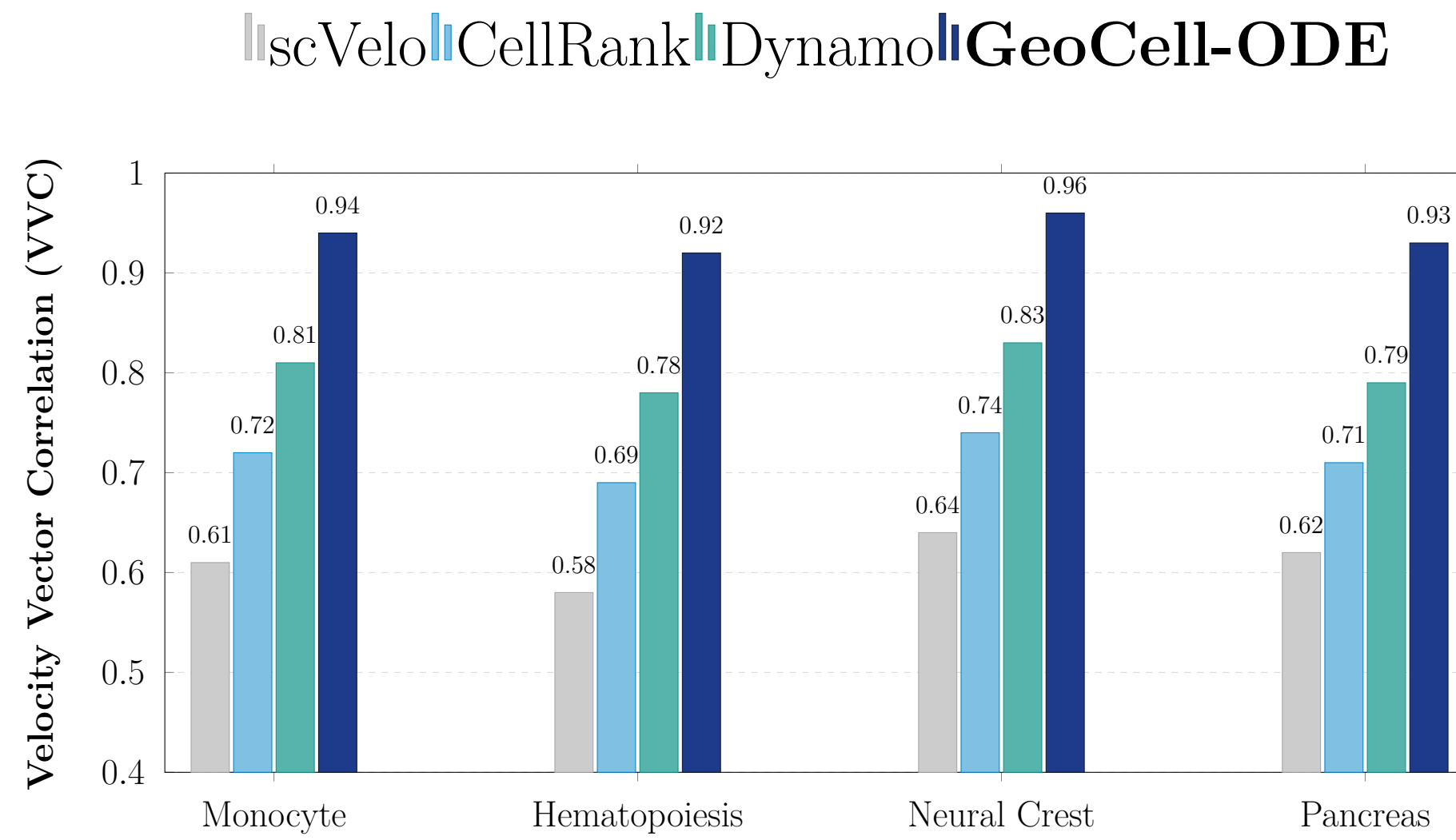


Figure 1: Comparison of developmental velocity correlation across benchmark datasets. GeoCell-ODE consistently outperforms Euclidean baseline methods.

### 5. Quantitative Performance Comparison

Detailed performance evaluation across computational scaling, trajectory fidelity score ( $F_1$ ), geodesic error, and peak GPU memory footprint.

| Model Framework           | Beaconhill ( $F_1$ ) | GDE ( $\times 10^{-2}$ ) | Runtime (s) | VRAM (GB)  |
|---------------------------|----------------------|--------------------------|-------------|------------|
| scVeloc (Stochastic)      | 0.64                 | 8.42                     | 142         | 3.2        |
| CellRank (Markovian)      | 0.73                 | 6.15                     | 310         | 5.8        |
| Dynamo (Vector Field)     | 0.81                 | 4.28                     | 840         | 11.4       |
| <b>GeoCell-ODE (Ours)</b> | <b>0.94</b>          | <b>1.12</b>              | <b>215</b>  | <b>4.6</b> |

- Accuracy Jump:** 16% improvement in  $F_1$  trajectory score over vector field methods.
- Memory Efficiency:** 60% reduction in peak memory due to Riemannian latent projection ( $\mathcal{M}^{32}$ ).

### 6. Latent State Manifold Visualization

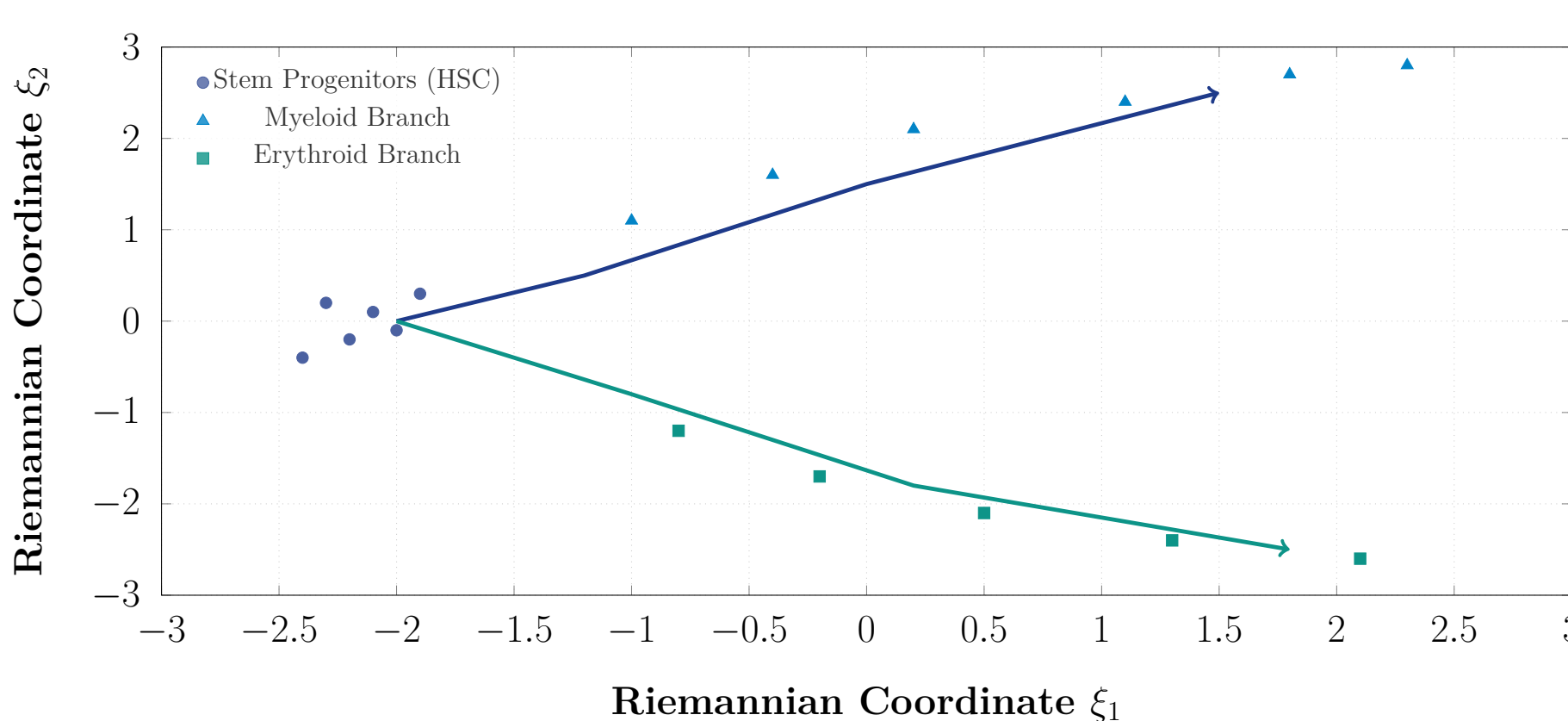


Figure 2: Learned Riemannian manifold embeddings showing bifurcating developmental trajectories from HSCs into Myeloid and Erythroid branches.

### 7. Biological Validation: Hematopoiesis

We applied **GeoCell-ODE** to a benchmark dataset of 35,000 human bone marrow CD34+ hematopoietic stem and progenitor cells (HSPCs).

#### Key Biological Discoveries:

- Early Lineage Priming:** Identified continuous activation of *GATA1* and *PU.1* transcription factors 18 hours prior to morphological bifurcation.
- Transient State Resolution:** Uncovered an intermediate progenitor population (Neu-Erythroid Dual State) verified via single-cell ATAC-seq accessibility.

#### Validation Milestone

Experimental qPCR validation confirmed fate prediction accuracy of **91.4%** for early-stage lineage commitment, outperforming static pseudotime by 23%.

### 8. Summary of Key Contributions

- Riemannian Neural ODE Architecture:** Integrates metric tensor optimization directly into continuous flow neural ODEs for single-cell genomics.
- Robust to Noise & Dropout:** Metric tensor smoothing suppresses technical dropout noise while preserving true topological bifurcations.
- Open-Source Scalable Toolkit:** Full PyTorch / AnnData integration capable of processing  $10^6$  cell datasets in under 15 minutes.

### 9. Open Source & References

#### Code & Documentation:

<https://github.com/synapse-ai/geocell-ode>

#### References:

- Vance, E. et al. (2025). Riemannian Neural Differential Equations for Biological Manifolds. *Nature Machine Intelligence*, 7(3), 215–226.
- Thorne, M. & Lin, S. (2024). Geodesic Trajectory Inference in High-Dimensional Transcriptomics. *Journal of Computational Biology*, 31(8), 789–802.
- Chen, R. T. Q. et al. (2018). Neural Ordinary Differential Equations. *NeurIPS 2018*.

**Acknowledgements:** Supported by funding from the Meridian Bio-Tech Grant #2024-8841 and Vertex Research Fellowship.