Recorded on Oct 22 for NeurIPS 2022

Diffusion Curvature for Estimating Local Curvature in High Dimensional Data

Dhananjay Bhaskar, Kincaid MacDonald, Oluwadamilola Fasina, Dawson Thomas,

Bastian Rieck, Ian Adelstein, Smita Krishnaswamy

NEURAL INFORMATION PROCESSING SYSTEMS

HELMHOLTZ MUNICH

LUX ET VERITAS

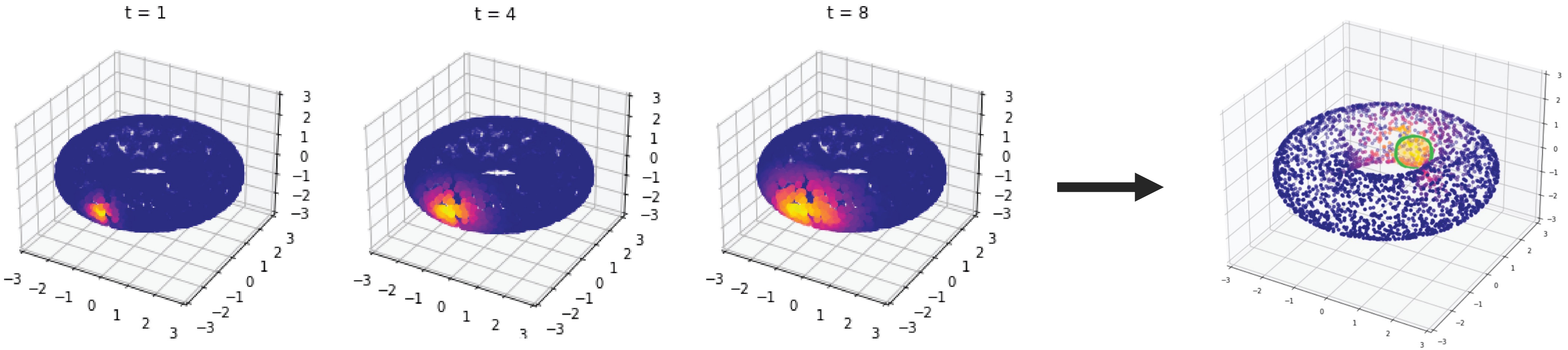
Diffusion Curvature

Definition

The pointwise diffusion curvature $C(x)$ *is the average probability that a random walk starting from a point* x *ends within* $B(x, r)$ *after* t *steps of data diffusion, i.e.,*

$$C(x) = \frac{\sum_{y \in B(x,r)} m_x(y)}{|B(x,r)|}$$

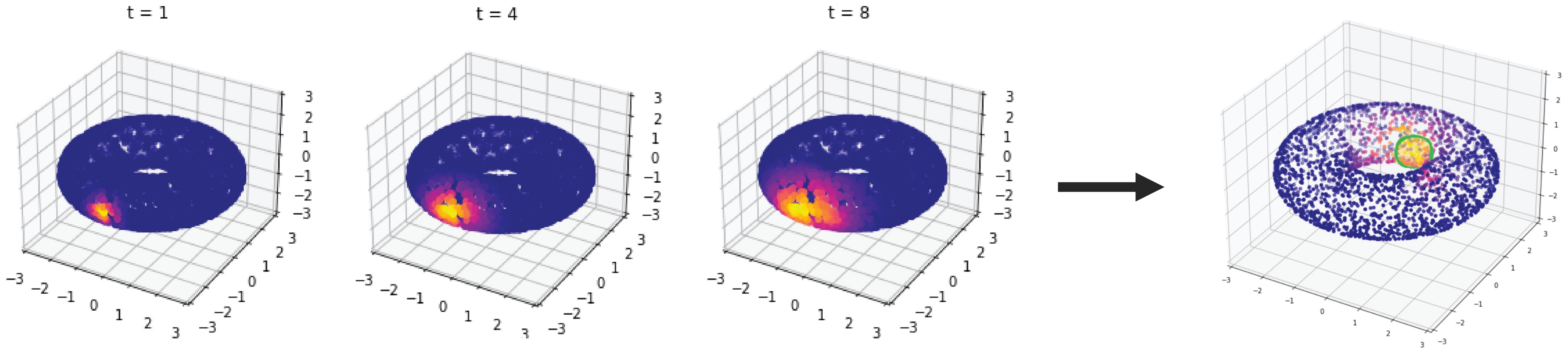
$$B_m(x, r) = \{y \in M : D_m(x, y) \leq r\} \subset M$$



NEURAL INFORMATION PROCESSING SYSTEMS

Diffusion Map (Coifman et al.)

The diffusion map $\Phi_t(x_i) = [\lambda_1^t \phi_1(x_i), \ldots, \lambda_N^t \phi_N(x_i)]^T$ *embeds data into a Euclidean space where the Euclidean distance is equal to the diffusion distance* D_m*, i.e.* $D_m^2(x,y) = \|\Phi_t(x) - \Phi_t(y)\|^2 (1 + O(e^{-\alpha m}))$.



Toy datasets

Diffusion Curvature

$$C(x) = \frac{\sum_{y \in B(x,r)} m_x(y)}{|B(x,r)|}$$

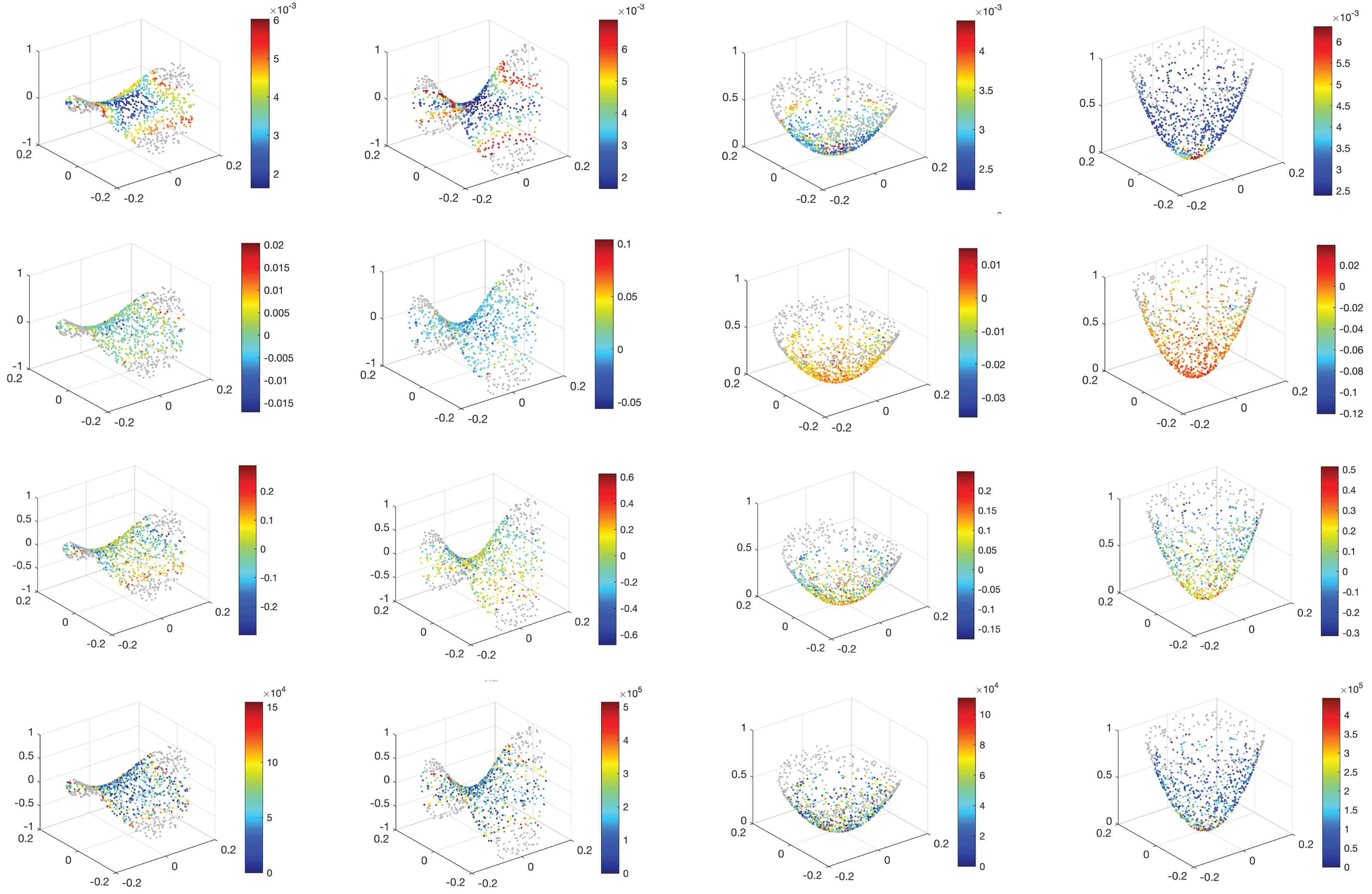
Gaussian Curvature

$$k = \kappa_1 \kappa_2$$

Mean Curvature

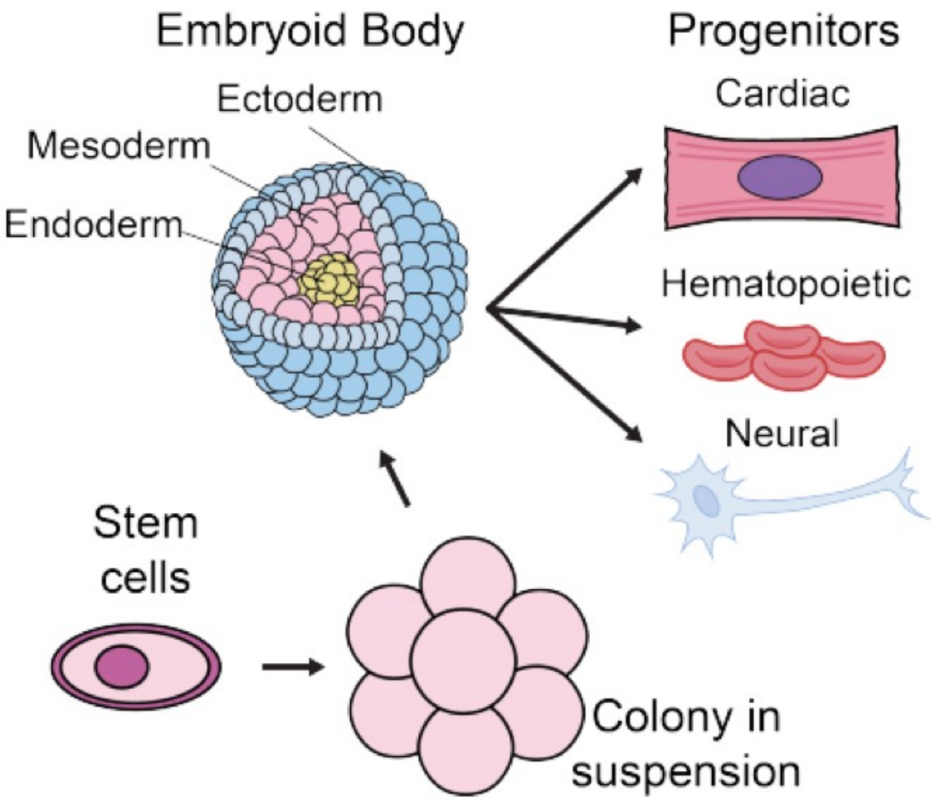
$$H = 1/2(\kappa_1 + \kappa_2)$$

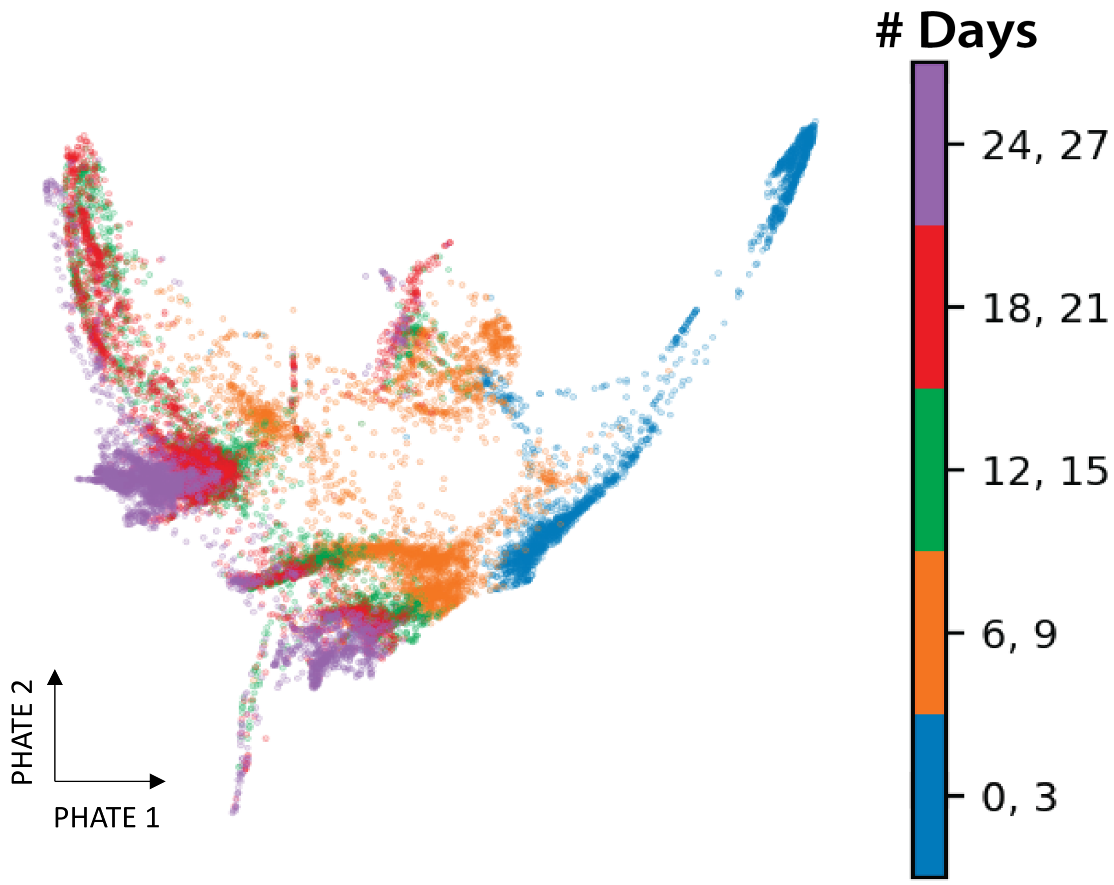
Ollivier Ricci Curvature



NEURAL INFORMATION PROCESSING SYSTEMS

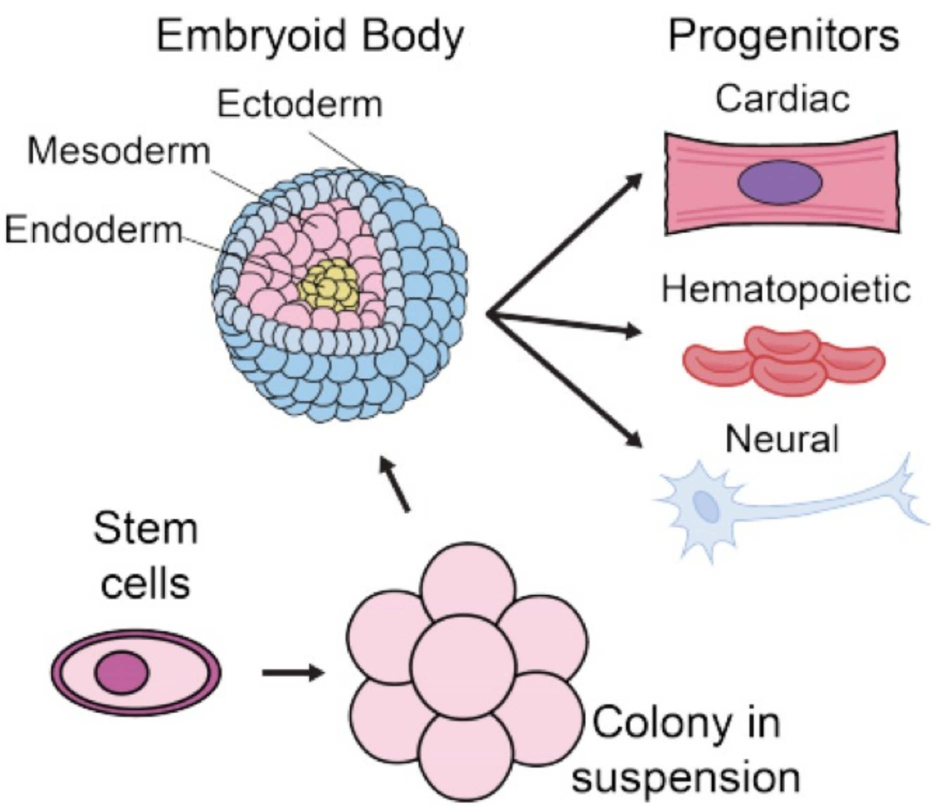
Embryoid stem cell differentiation

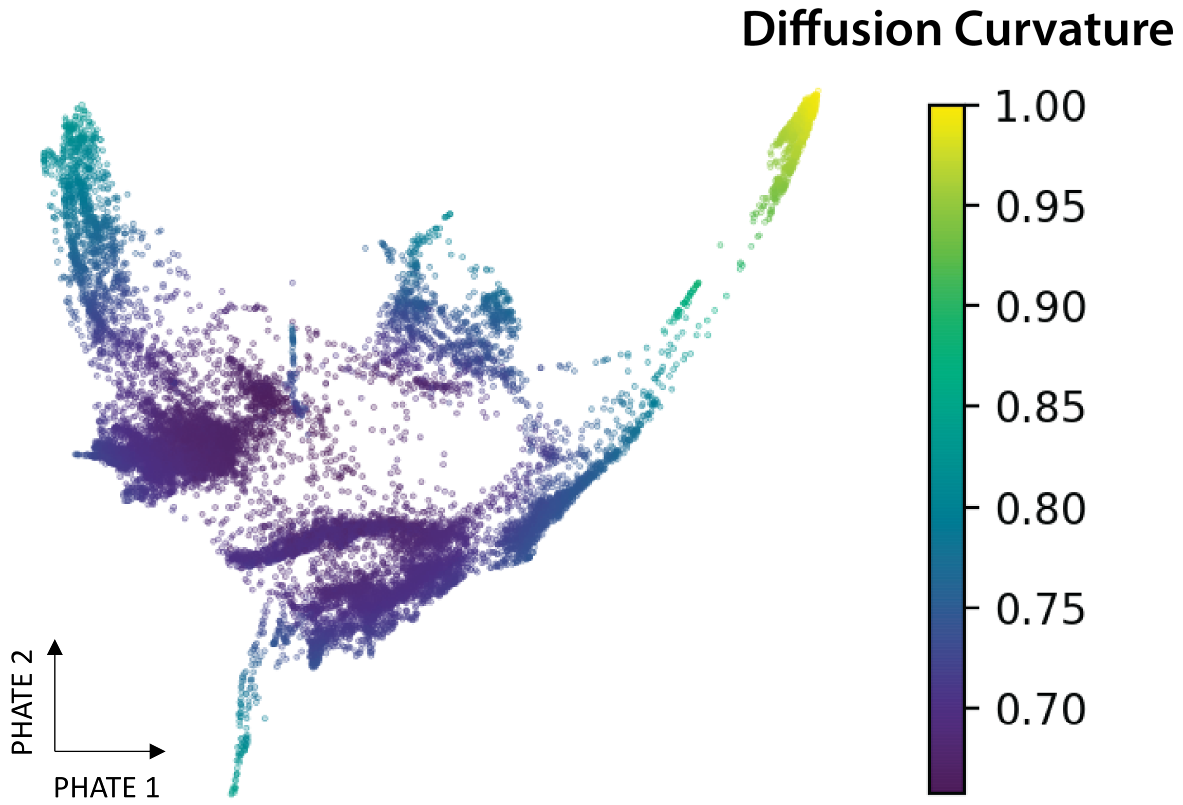




NEURAL INFORMATION
PROCESSING SYSTEMS

Embryoid stem cell differentiation





NEURAL INFORMATION
PROCESSING SYSTEMS

Neural network loss landscape

Loss function:

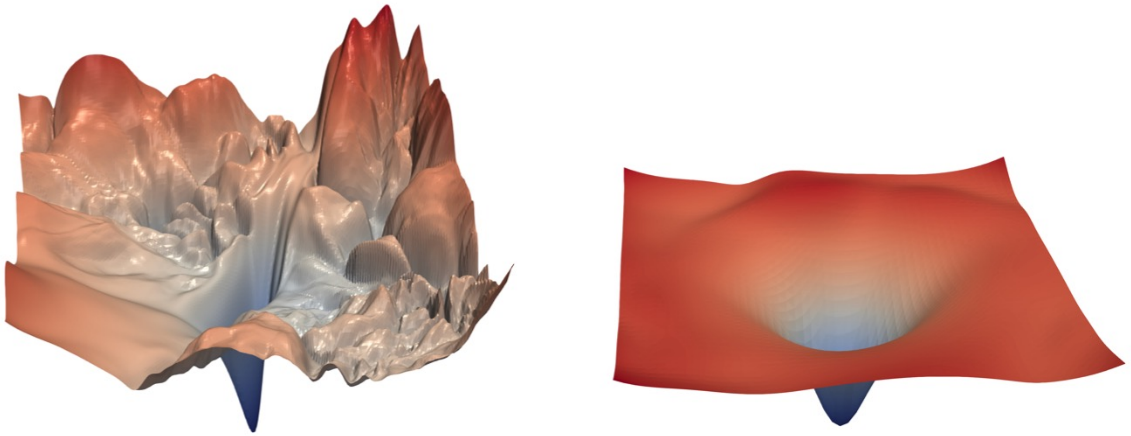
$$\mathcal{L}(X,\theta) = \sum_{x\in X} ||f(x) - y(x)||_2$$

At minima:

$$\nabla\mathcal{L}(\theta_{\mathbf{0}}) = 0,$$

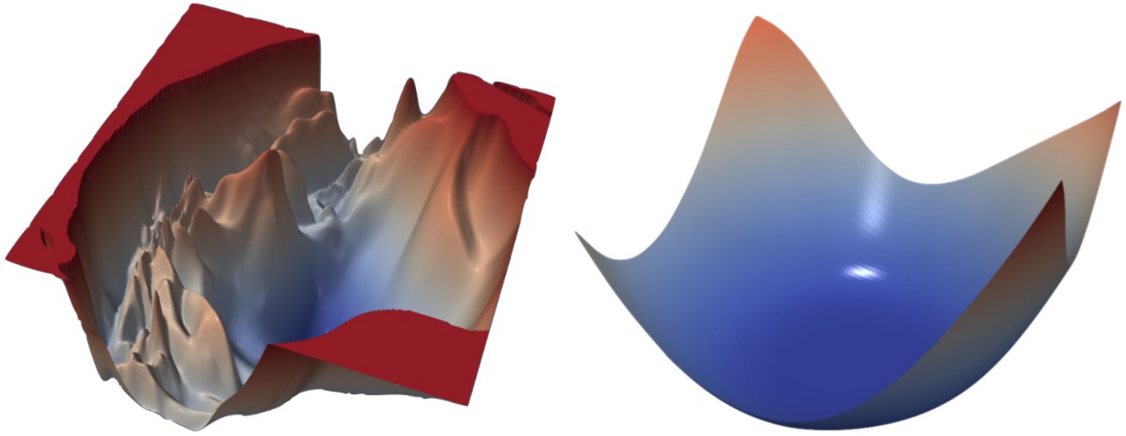
$$\mathcal{L}(\theta) \approx \mathcal{L}(\theta_{\mathbf{0}}) + 1/2\, H\mathcal{L}(\theta - \theta_{\mathbf{0}})$$

$$H\mathcal{L}(\mathbf{v}) = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \begin{pmatrix} \frac{\partial^2\mathcal{L}}{\partial\theta_1\partial\theta_1} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\theta_1\partial\theta_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\mathcal{L}}{\partial\theta_n\partial\theta_1} & \cdots & \frac{\partial^2\mathcal{L}}{\partial\theta_n\partial\theta_n} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$



(a) without skip connections (b) with skip connections

Loss surface for ResNet-56



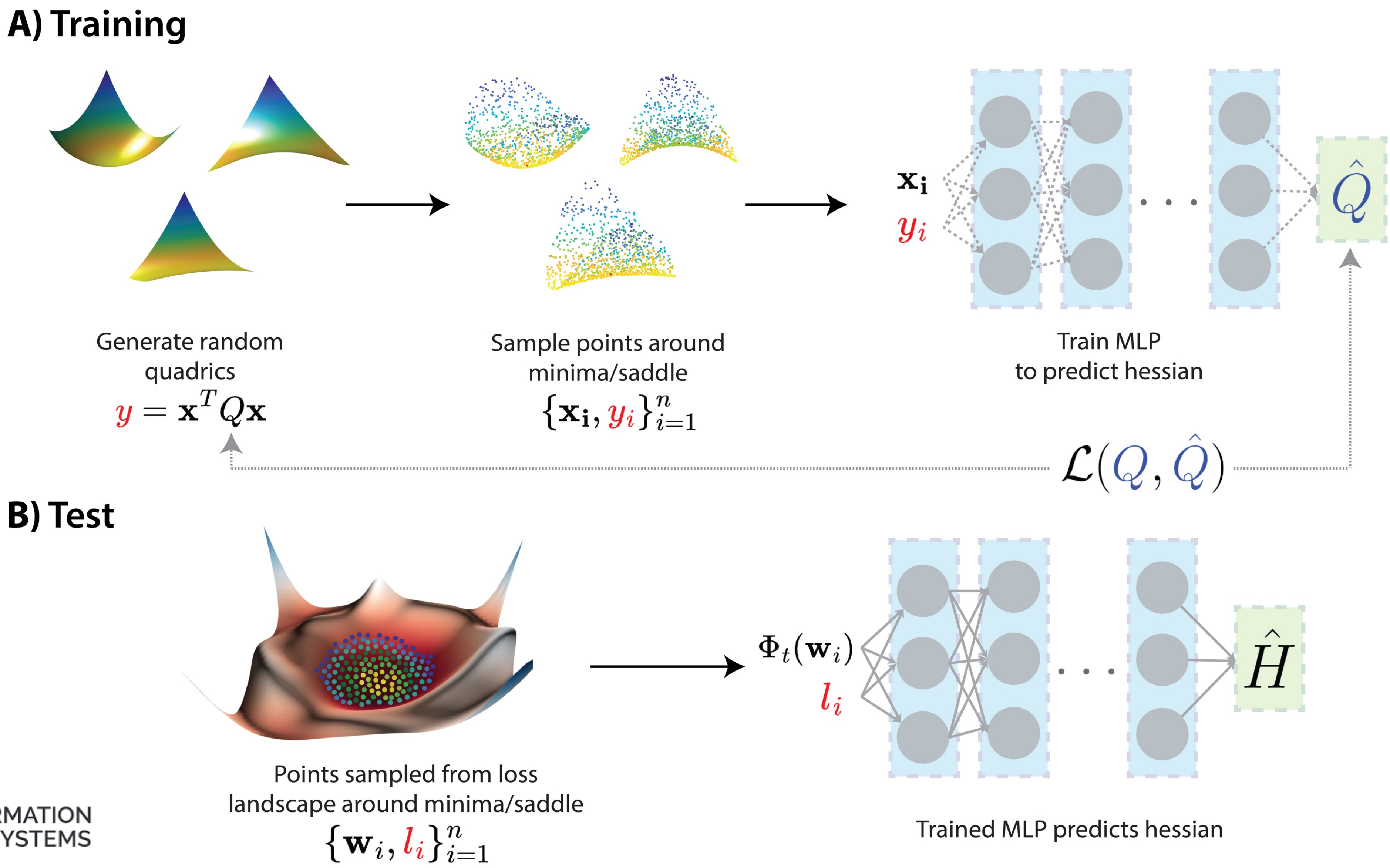
(a) ResNet-110, no skip connections (b) DenseNet, 121 layers

Loss surface for CIFAR-10

NEURAL INFORMATION PROCESSING SYSTEMS

Li et al., NeurIPS 2018

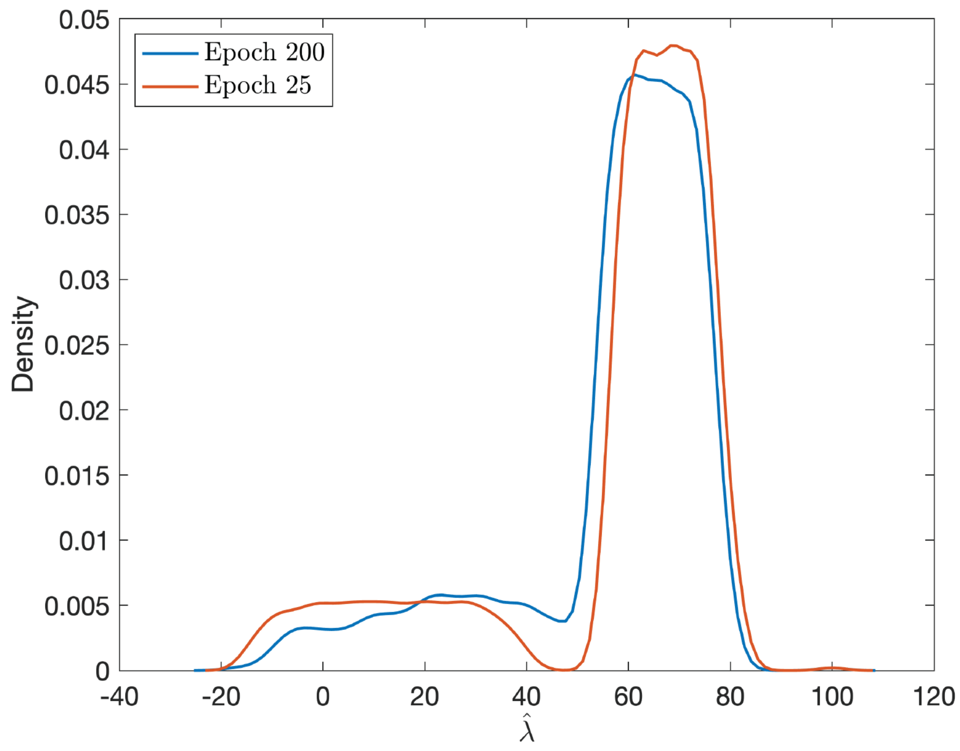
CurveNet



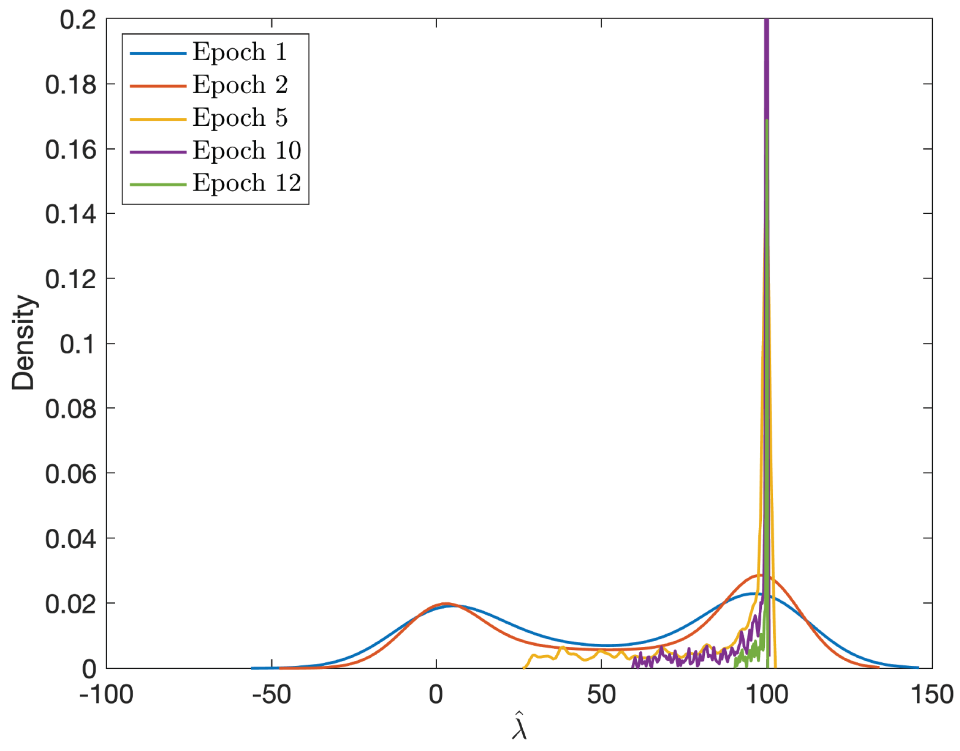
NEURAL INFORMATION PROCESSING SYSTEMS

Eigenspectrum of the loss hessian

MLP



ConvNet



NEURAL INFORMATION PROCESSING SYSTEMS

Acknowledgements

- NIH (R01GM135929, R01GM130847, R01HD100035)
- NSF Career Grant (2047856), NSF DMS-2050398
- Yale-Boehringer Ingelheim Biomedical Data Science Fellowship
- Sloan Fellowship (FG-2021-15883)

NIH

NSF

Boehringer Ingelheim

ALFRED P. SLOAN FOUNDATION

NEURAL INFORMATION PROCESSING SYSTEMS