# Metrics for Deep Generative Models

**Nutan Chen*** **Alexej Klushyn*** **Richard Kurle***
**Xueyan Jiang** **Justin Bayer** **Patrick van der Smagt**
{first name dot last name}@volkswagen.de
AI Research, Data:Lab, Volkswagen Group, Munich, Germany

# CONTENT

TL;DR:
THEY PROPOSE A METHOD TO **VISUALIZE LATENT SPACE** OF NEURAL SAMPLERS BY MEASURING **GEODESIC DISTANCES** ON THE RIEMANNIAN MANIFOLD INDUCED BY THE TRANSFORMATION.

- Motivation
- Method: how to approximate geodesic distance on Riemannian manifold?
- Experiments: Pendulum, binarized MNIST, robot arm, and human motion
- Conclusion

# MOTIVATION

- Manifold hypothesis: observed data is distributed near a **low dimensional manifold** within the observation space.
- The latent subspace of neural samplers are **densely covered.**
- Distant points in the observation space might be close to each other in the latent space! => we need a proper metric to measure the distance
- However, defining a meaningful metric in high dimensional space is hard:
  - Choosing a metric comes with certain assumptions on the data (Minkowski distance with L2 norm is invariant to rotations)
  - These distances become increasingly meaningless for higher dimensions [Aggarwal et al., 2001]

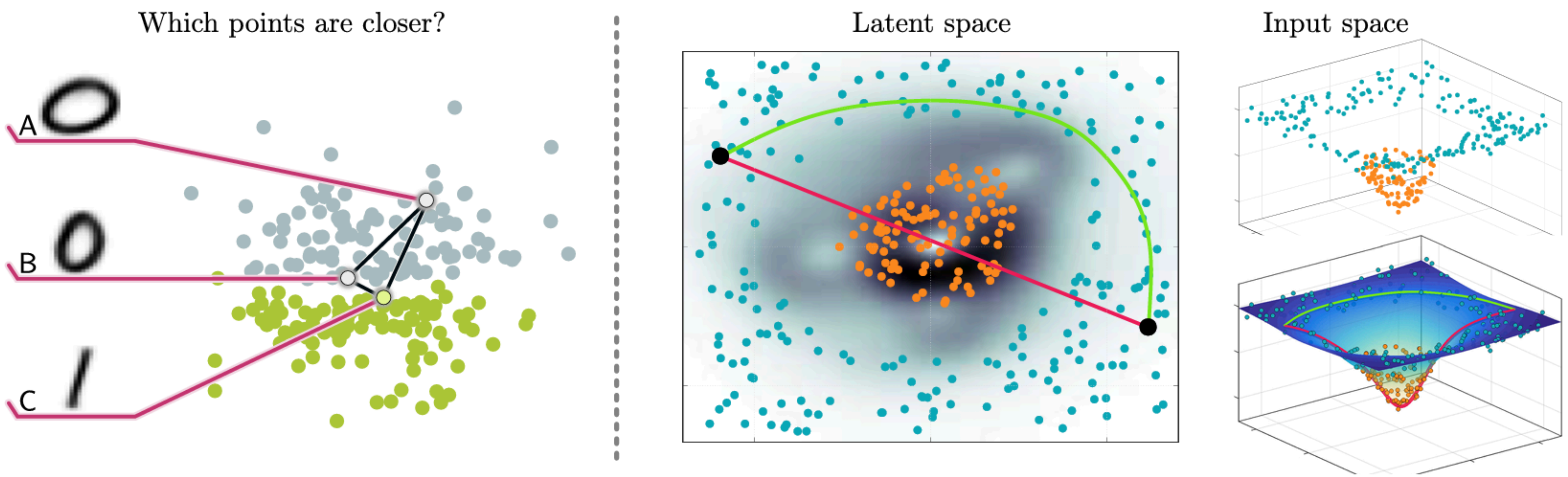
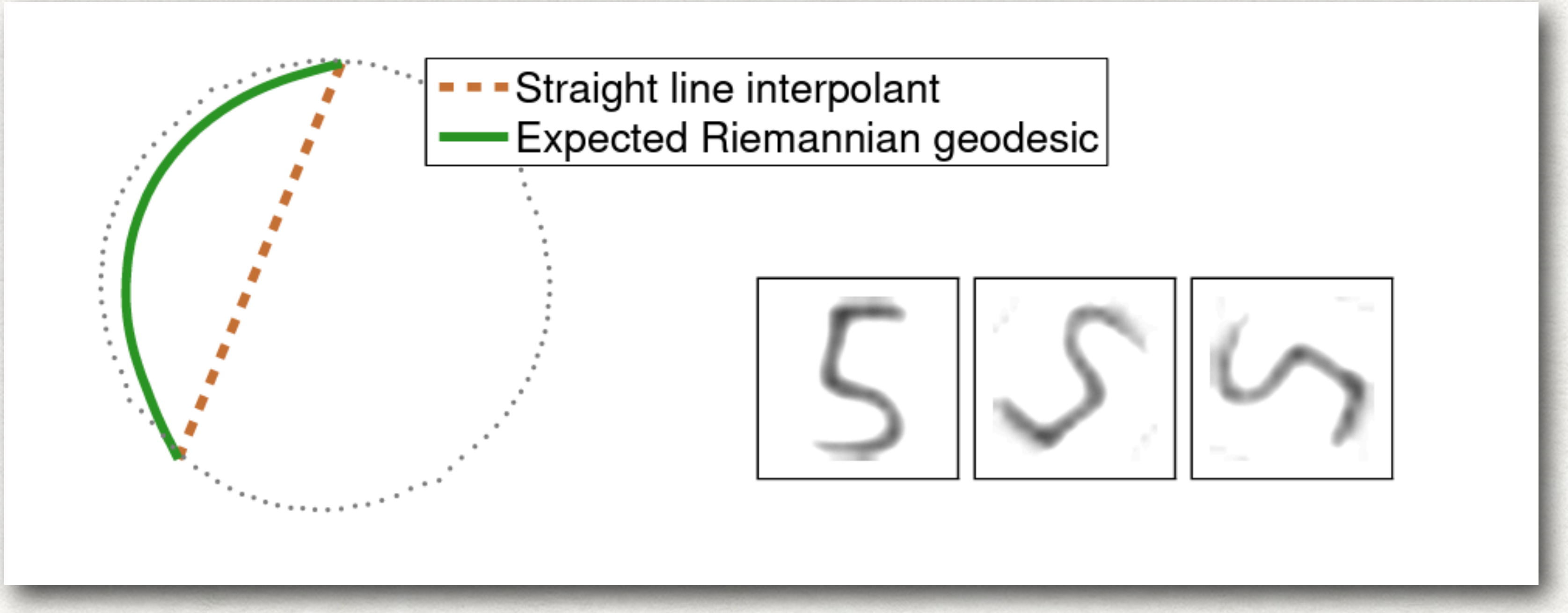
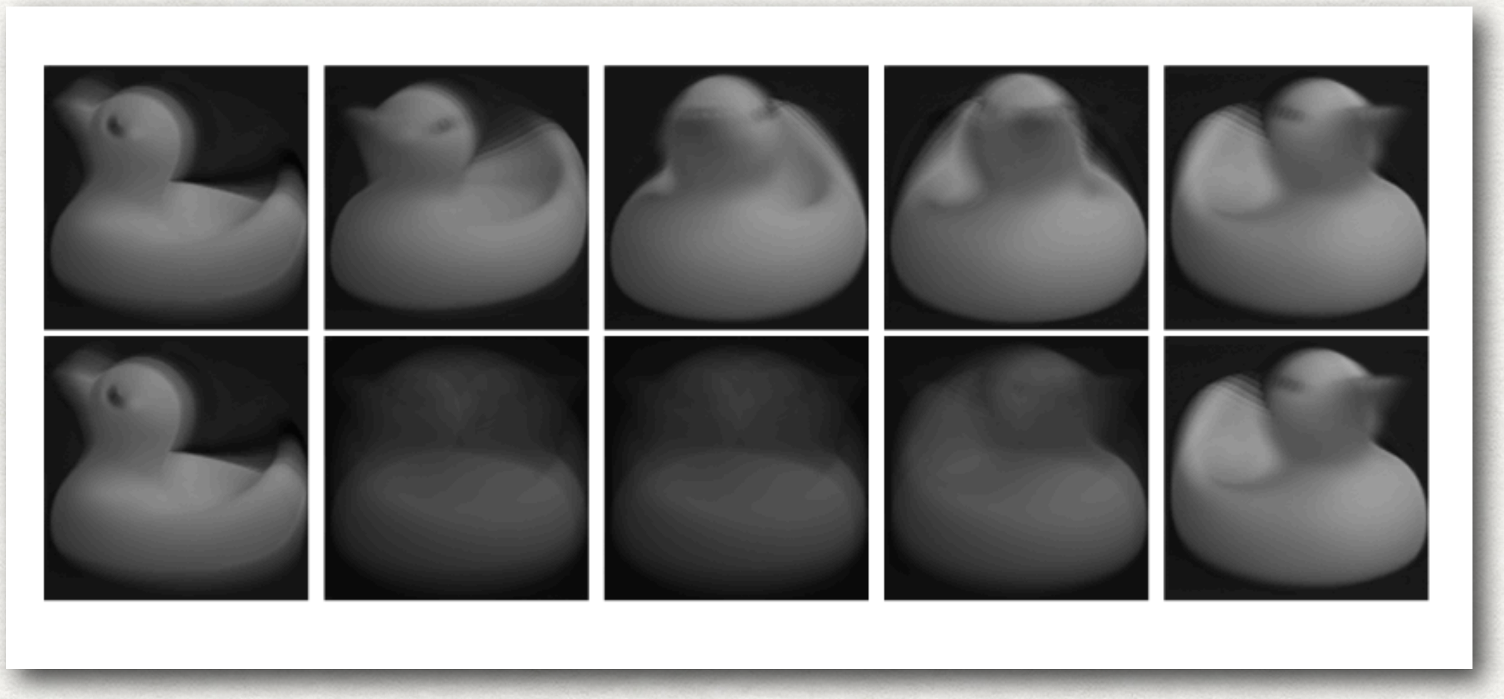



Figure 1: *Left:* An example of how latent space distances do not reflect actual data distances. *Right:* Shortest paths on the surface spanned by the generator do not correspond to straight lines in the latent space, as is assumed by the Euclidean metric.

# MOTIVATION

- Goal: Define a metric on the latent space of deep generative models that respect the underlying data manifold
- Past literature:
  - [Tosi et al., 2014]: perceive the latent space of Gaussian process latent variable models (GP-LVMs) as Riemannian manifold, where the distance between two data points is given as the shortest path along the data manifold.
-



# METHODS

- Riemannian geometry: a differentiable manifold $M$ with a metric tensor $\mathbf{G}$
  - Inner Product: $< \mathbf{z}', \mathbf{z}' > := \mathbf{z}'^{T}\mathbf{G}(\mathbf{z})\mathbf{z}', \quad \mathbf{z}' \in T_{\mathbf{z}}M, z \in M$
  - Curve + continuous transformation: $[0,1] \xrightarrow{\gamma} \mathbb{R}^{N_z} \xrightarrow{f} \mathbb{R}^{N_x}$

    Observation space
  - The length of the curve in the observation space:

$$
\begin{aligned}
\mathscr{L}(\gamma) &:= \int_0^1 ||\frac{\partial f(\gamma(t))}{\partial t}||dt \\
&= \int_0^1 ||\frac{\partial f(\gamma(t))}{\partial \gamma(t)}\frac{\partial \gamma(t)}{\partial t}||dt \\
&= \int_0^1 ||\mathbf{J}\frac{\partial \gamma(t)}{\partial t}||dt
\end{aligned}
$$

Jacobian Matrix

$$\xrightarrow{\mathbf{G}=\mathbf{J}^T\mathbf{J}}$$

$$L(\gamma) = \int_0^1 \sqrt{< \gamma'(t), \gamma'(t) >_{\gamma(t)}}\, dt$$

# METHODS

- What we need: length minimizing curves between samples of generative models
- Approach: Use a neural network $g_w : \mathbb{R} \to \mathbb{R}^{N_z}$ to approximate the curve $\gamma$ in the latent space

$$[0,1] \xrightarrow{\gamma} \mathbb{R}^{N_z} \xrightarrow{f} \mathbb{R}^{N_x}$$

$$L(\gamma) = \int_0^1 \sqrt{< \gamma'(t), \gamma'(t) >_{\gamma(t)}}\, dt$$

$$\mathbb{R} \xrightarrow{g_w} \mathbb{R}^{N_z} \xrightarrow{h^{gen}} \mathbb{R}^{N_x}$$

Velocity $\phi(t)$

$$L(g_w(t)) \approx \frac{1}{n} \sum_{i=1}^{n} \sqrt{< g_w'(t_i), g_w'(t_i) >_{g_w(t_i)}}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sqrt{g_w'(t_i)^T \mathbf{J}^T \mathbf{J} g_w'(t_i)}$$

# METHODS

- Goal: boarder constrained optimization problem

$$L(g_w(t)) \approx \frac{1}{n}\sum_{i=1}^{n}\sqrt{< g'_w(t_i), g'_w(t_i) >_{g_w(t_i)}}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sqrt{g'_w(t_i)^T\mathbf{J}^T\mathbf{J}g'_w(t_i)}$$

$$\underset{w}{min} L(g_w(t)) \quad s.t.\, g_w(0) = \mathbf{z}_0, g_w(1) = \mathbf{z}_1$$

$$\mathbf{z}(t) = \mathbf{A}\hat{\mathbf{z}(t)} - \mathbf{B}$$

$$\mathbf{A} = \frac{\mathbf{z}_0 - \mathbf{z}_1}{\hat{\mathbf{z}}(0) - \hat{\mathbf{z}}(1)}$$

$$\mathbf{B} = \frac{\mathbf{z}_0\hat{\mathbf{z}}(1) - \mathbf{z}_1\hat{\mathbf{z}}(0)}{\hat{\mathbf{z}}(0) - \hat{\mathbf{z}}(1)}$$

Release the boundary condition

# METHODS

- Goal: optimization problem

$$L(g_w(t)) \approx \frac{1}{n}\sum_{i=1}^{n}\sqrt{< g'_w(t_i), g'_w(t_i) >_{g_w(t_i)}}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sqrt{g'_w(t_i)^T \mathbf{J}^T \mathbf{J} g'_w(t_i)}$$

$$\underset{w}{min} L(g_w(t))$$

Smoothing the metric tensor

$$\mathscr{L} = L + \lambda_s ||\mathbf{G}||_2$$

$$\downarrow$$

$$\hat{\mathbf{G}} = \mathbf{U}_r diag\{\frac{s_i^3}{s_i^2 + \lambda_s}\}_{i=1}^{r} \mathbf{V}_r^T$$

# EXPERIMENTS

- Model: Importance Weighted Variational Autoencoder (IWAE, [Burda et al, 2015])
  - Importance sampling
    - $\int p(x)f(x)dx = \mathbb{E}_q[\frac{p(x)}{q(x)}f(x)] \approx \frac{1}{N}\sum_{i=1}^{N}\frac{p(x_i)}{q(x_i)}f(x_i), x_i \sim q(x)$

$$lnp_\theta(X) = \sum_{i=1}^{N} lnp_\theta(\mathbf{x}^{(i)}) \geq \sum_{i=1}^{N} \mathbb{E}_{\mathbf{z}_1^{(i)},\ldots,\mathbf{z}_K^{(i)} \sim q_\phi(\mathbf{z}^{(i)}|\mathbf{x}^{(i)})} ln\frac{1}{K}\sum_{k=1}^{K} w_k^{(i)}$$

$$w_k^{(i)} = \frac{p_\theta(\mathbf{x}^{(i)} \mid \mathbf{z}_k^{(i)})p_\theta(\mathbf{z}_k^{(i)})}{q_\phi(\mathbf{z_k}^{(i)} \mid \mathbf{x}^{(i)})}$$

# EXPERIMENTS

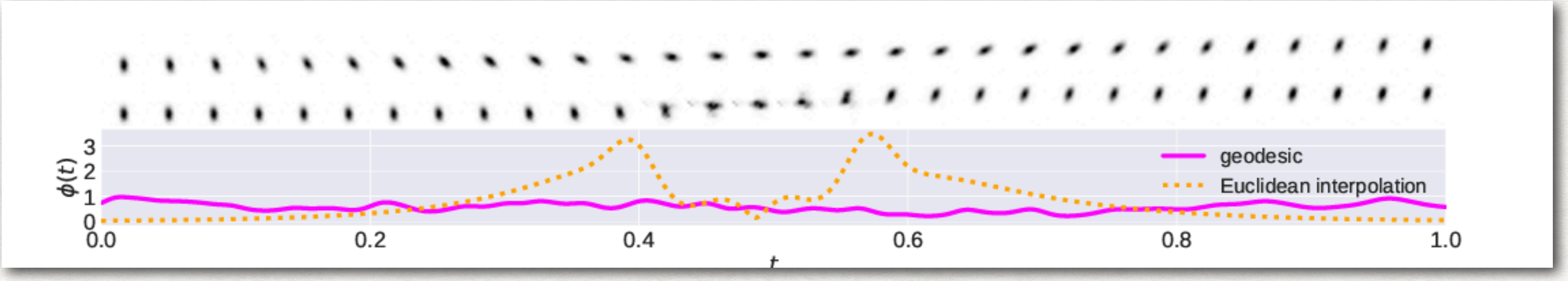
- Model: Importance Weighted Variational Autoencoder (IWAE, [Burda et al, 2015])
- Gaussian prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$
- Inference model with diagonal Gaussian: $q_\phi(\mathbf{z}\,|\,\mathbf{x}) = \mathcal{N}(\mu_\phi(\mathbf{x}, diag(\sigma_\phi^2(\mathbf{x}))))$
- Reconstruction model with likelihood $p_\theta(\mathbf{x}\,|\,\mathbf{z})$ being Bernoulli variable or a Gaussian
- Visualization: the magnification factor [Bishop et al, 1997]

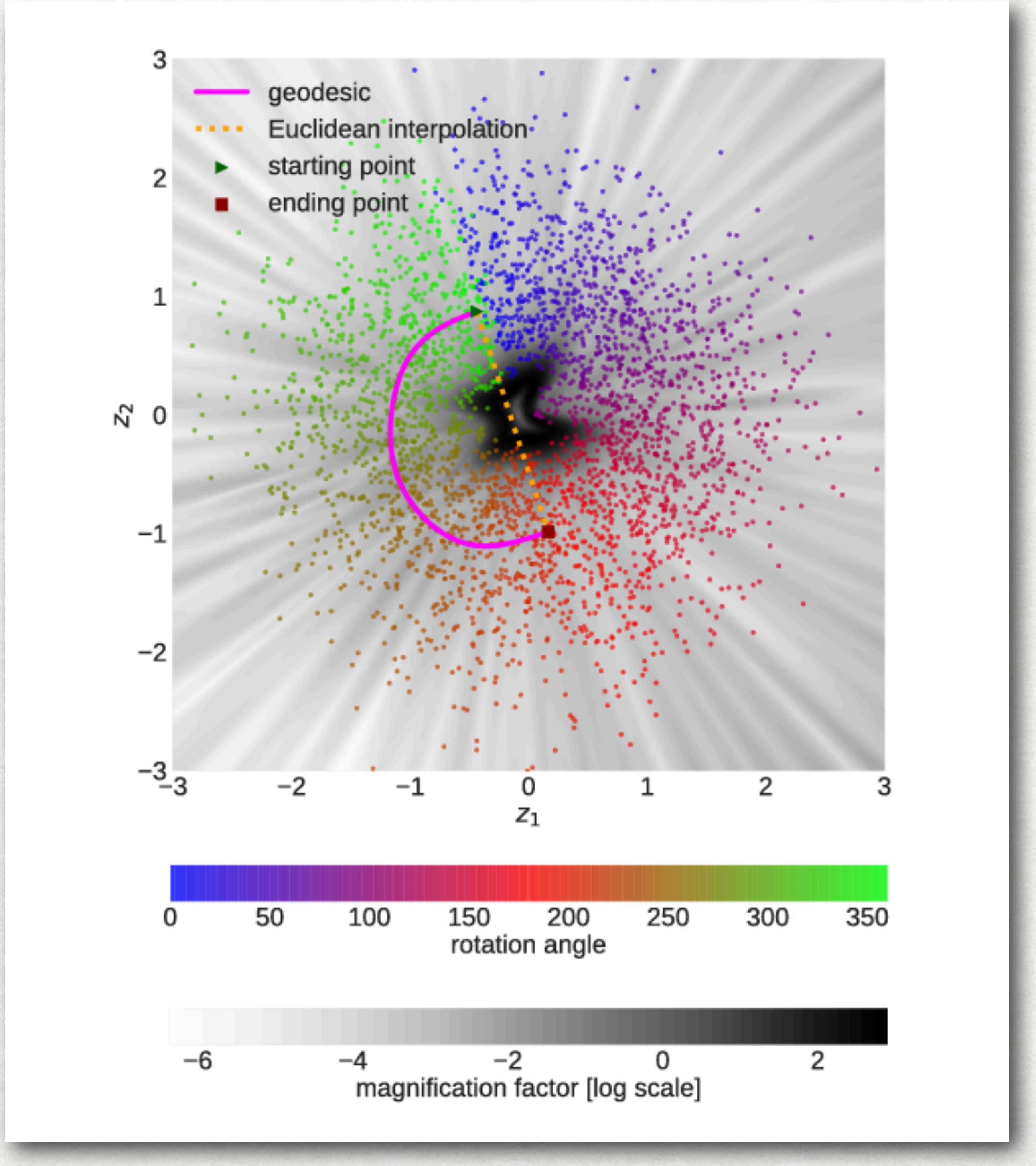
$$MF := \sqrt{det\mathbf{G}}$$

  - Visualize the extent of change of the infinitesimal volume by mapping a point from the Riemannian manifold to the Euclidean space

# EXPERIMENTS

- Pendulum

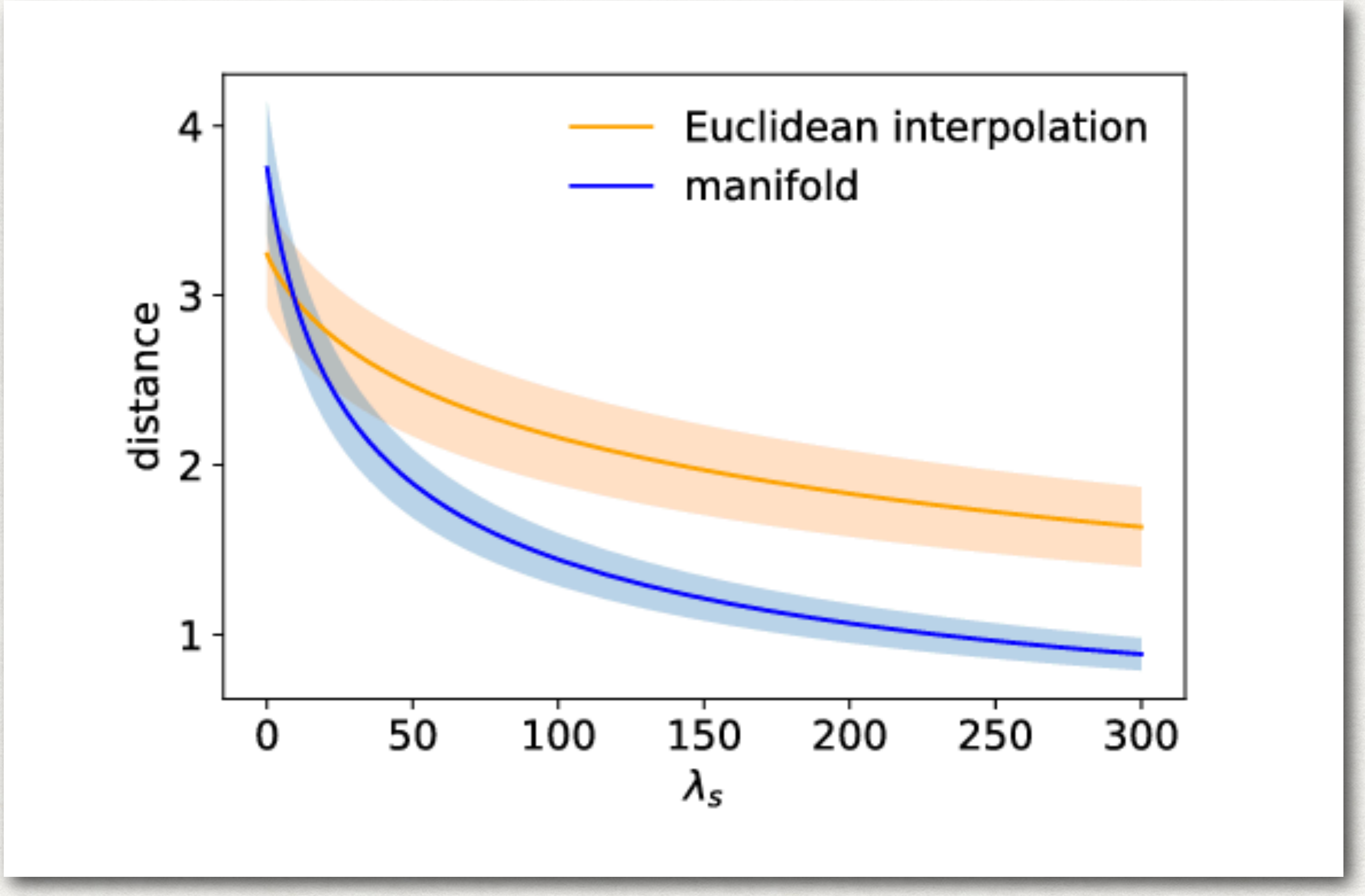


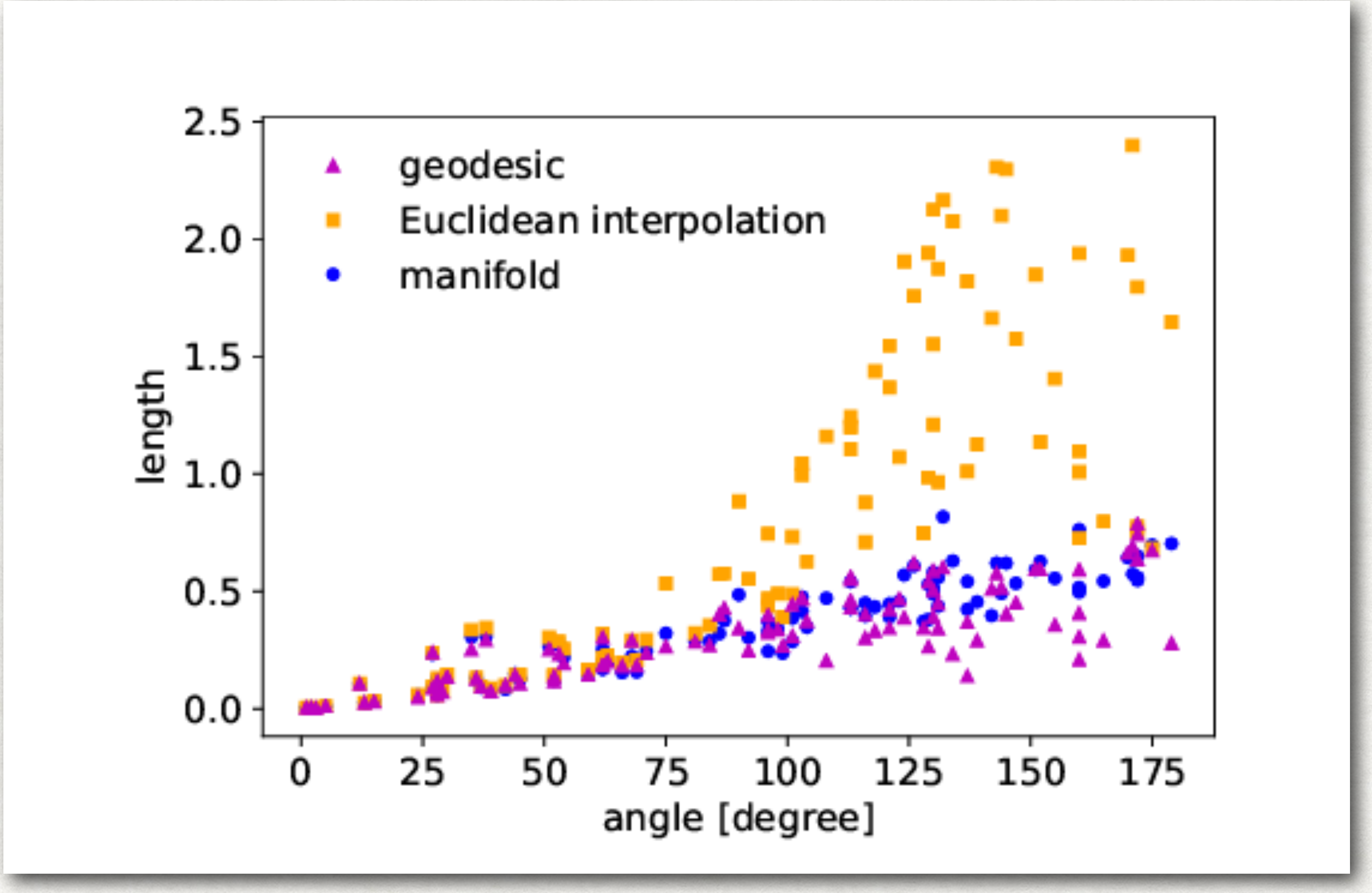



# EXPERIMENTS

- Pendulum

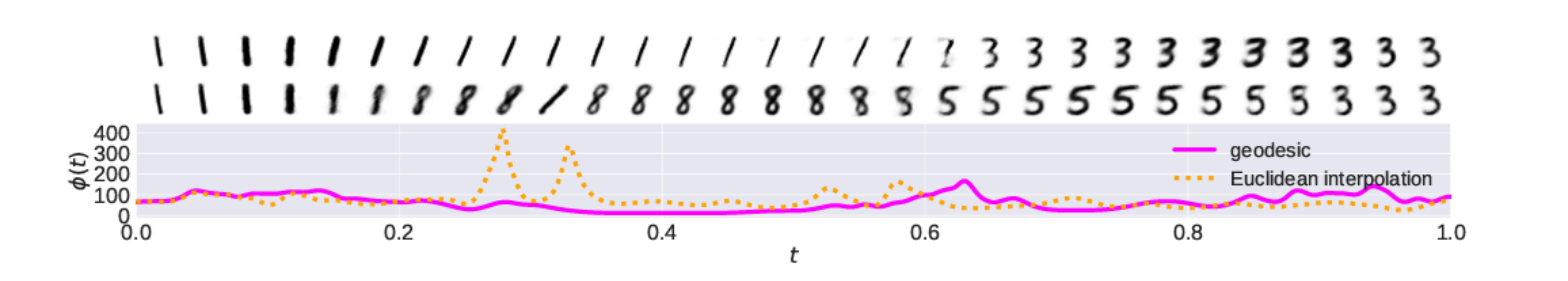
$$\mathscr{L} = L + \lambda_s ||\mathbf{G}||_2$$

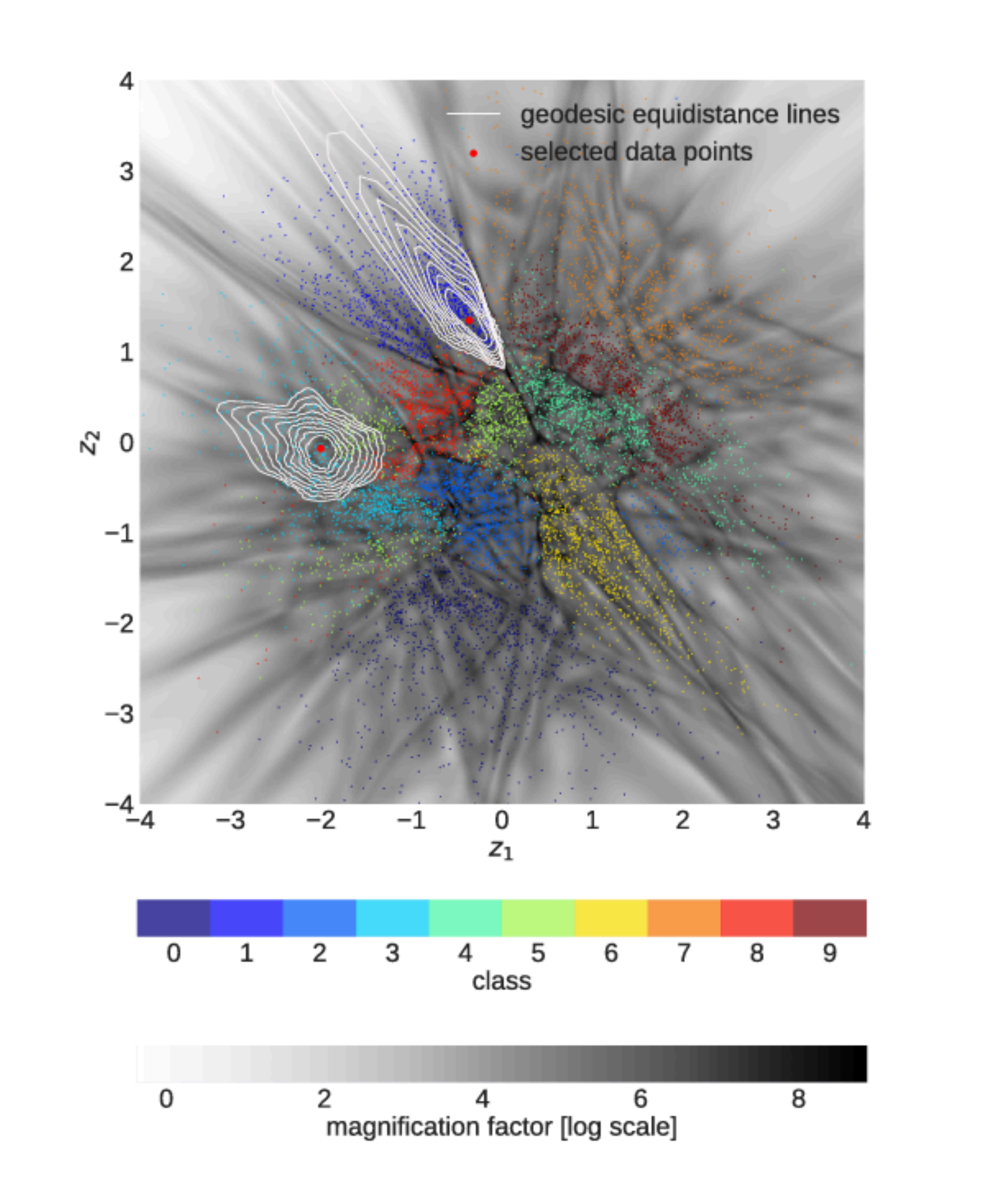


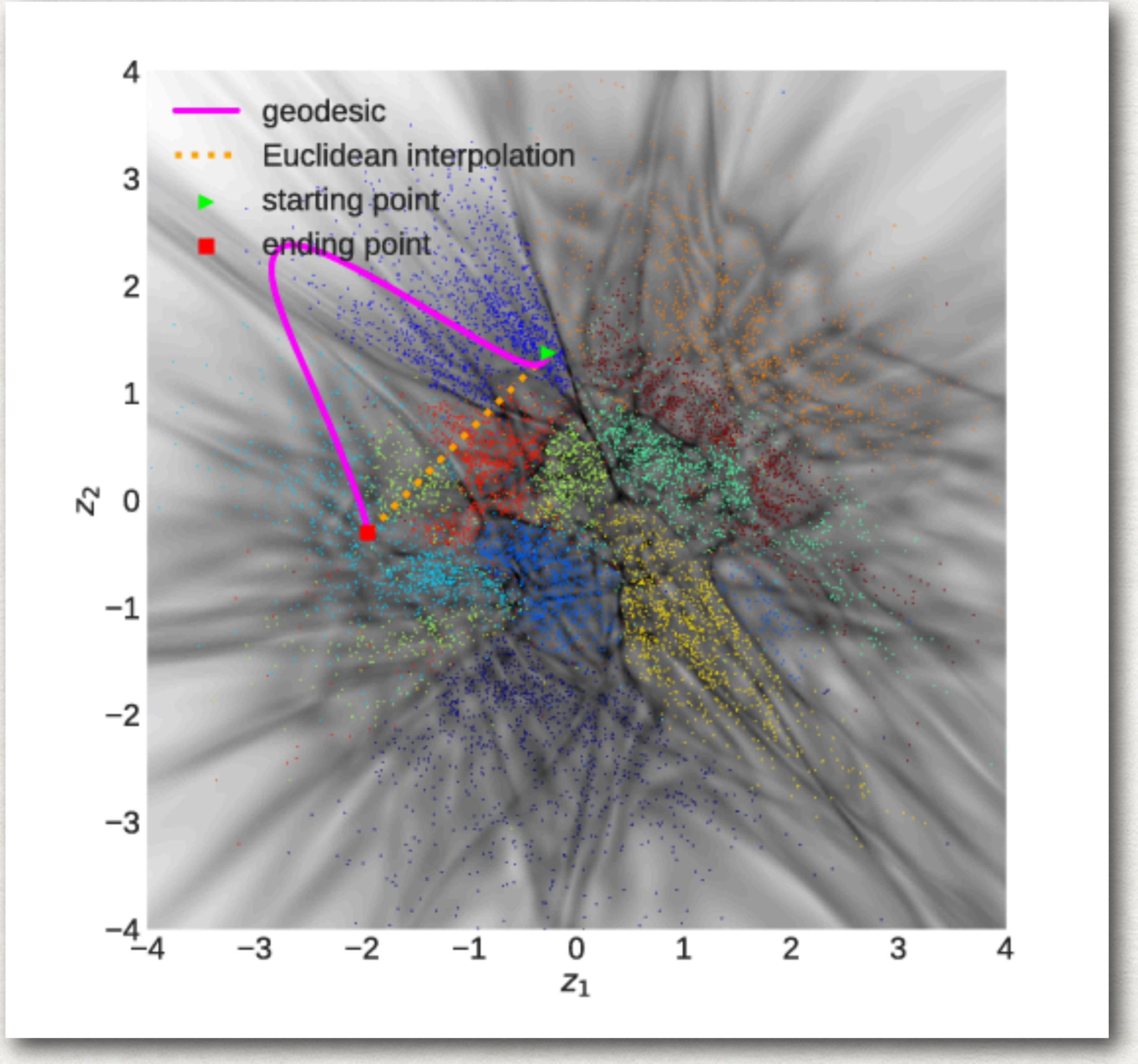



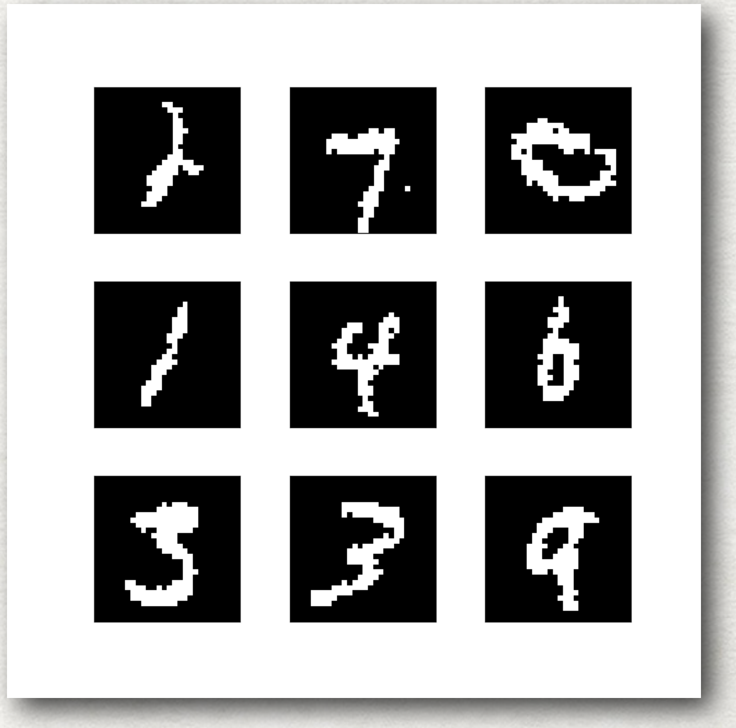
# EXPERIMENTS

- Binarized MNIST



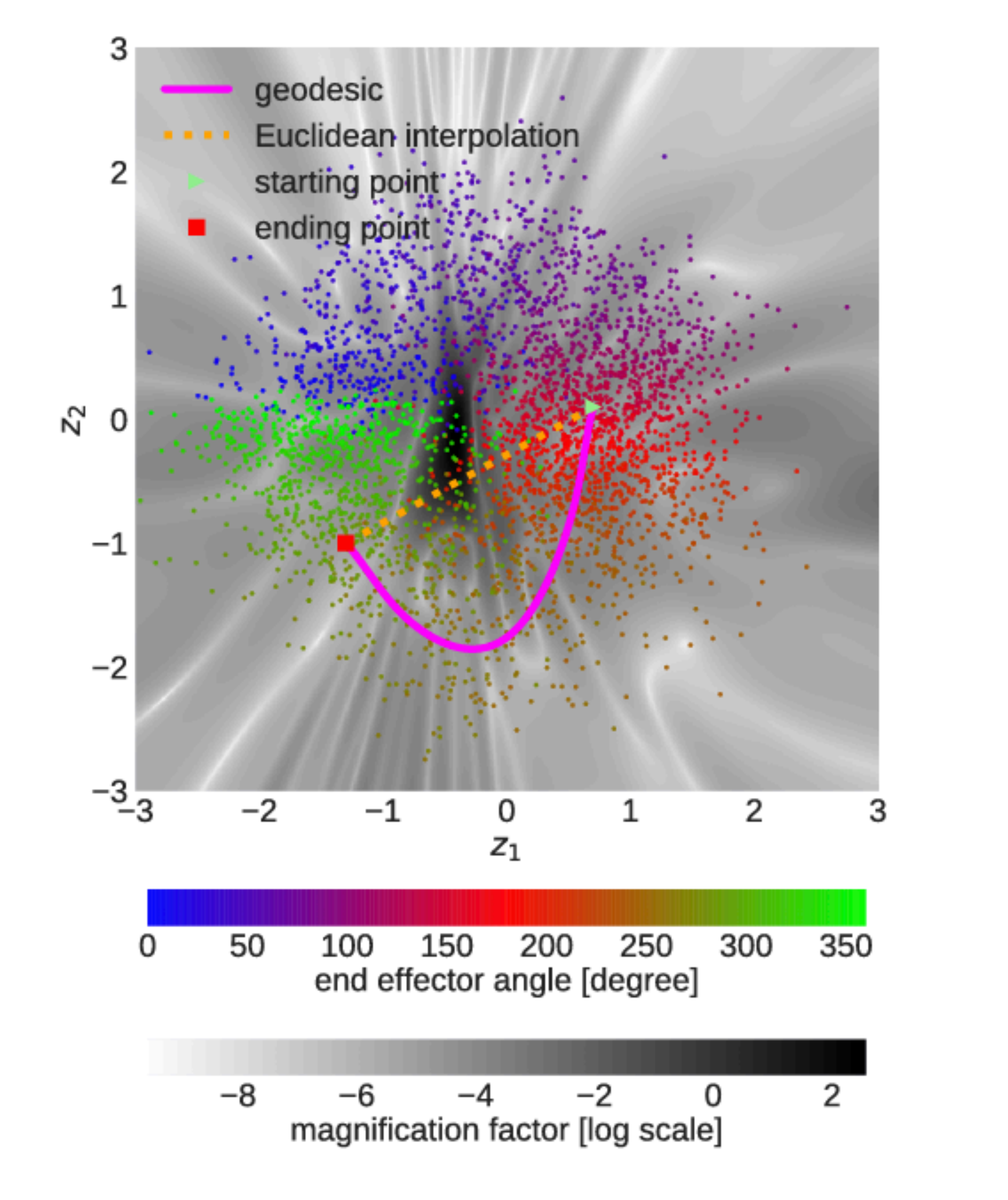



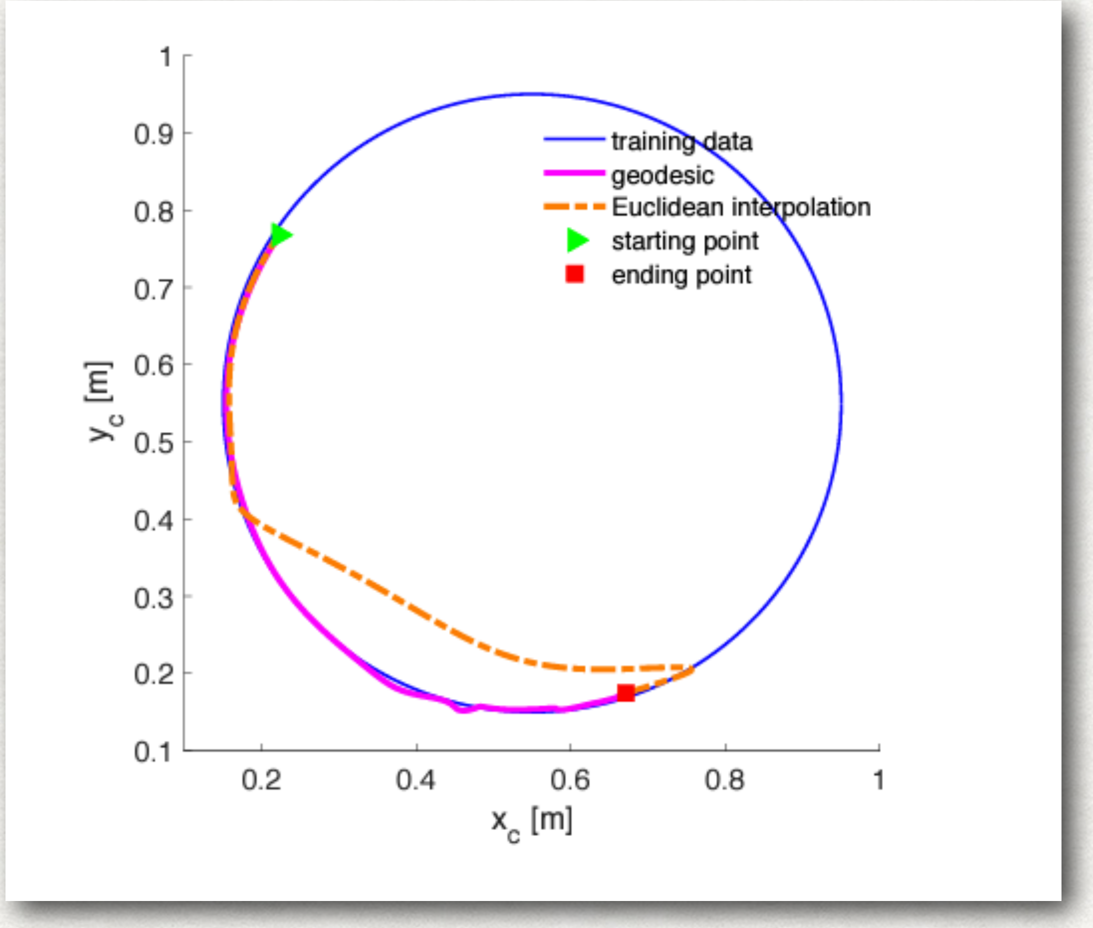



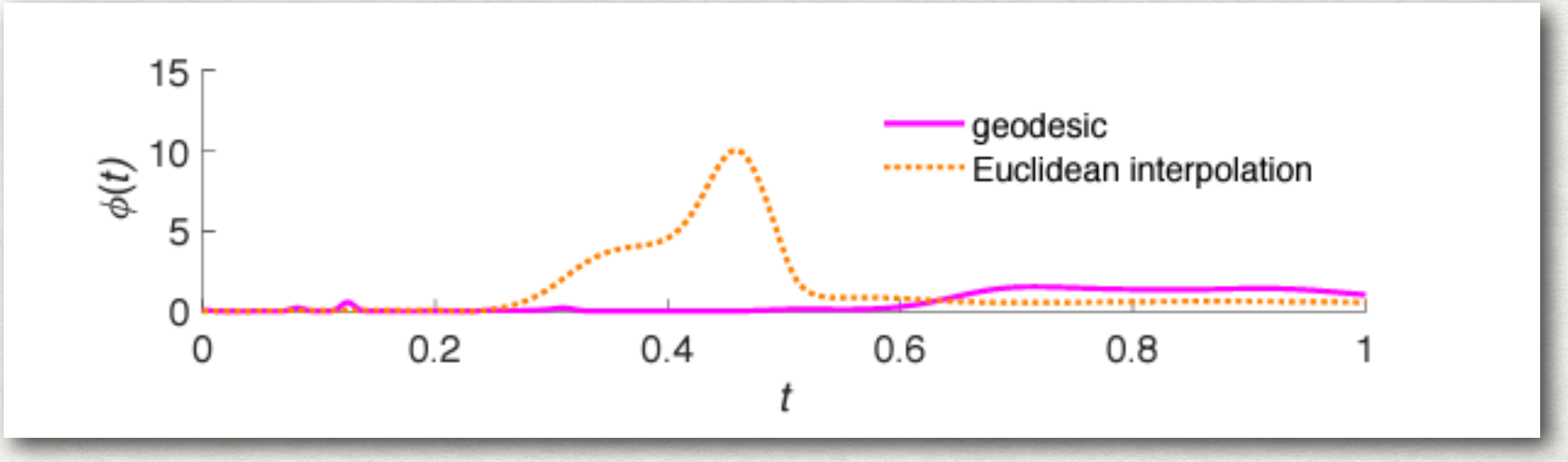



# EXPERIMENTS

- Robot Arm

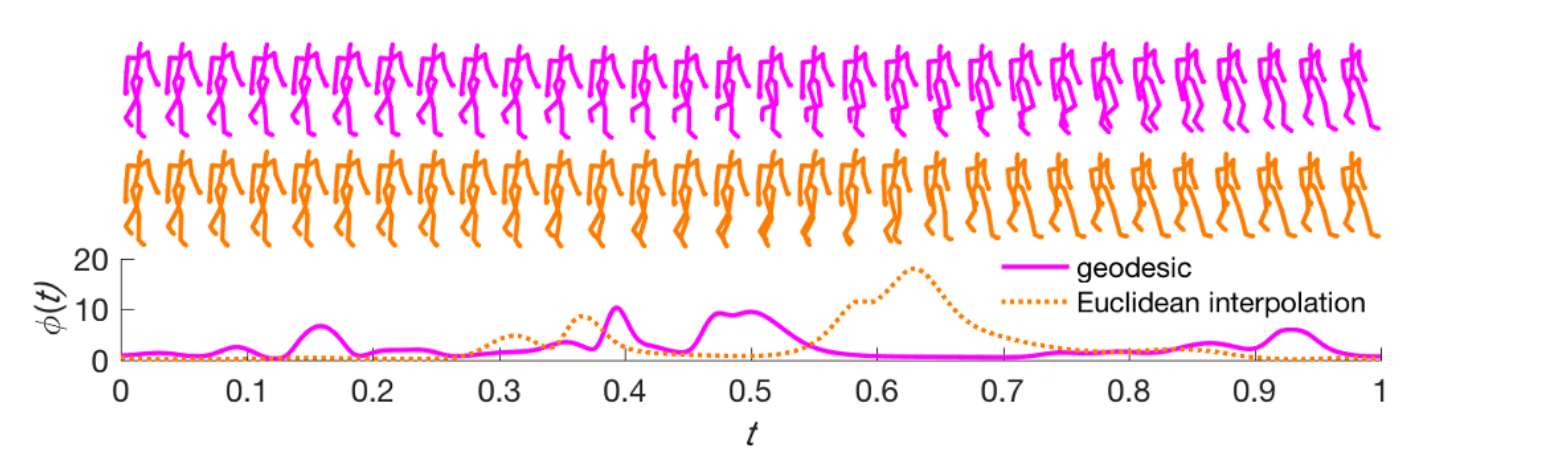


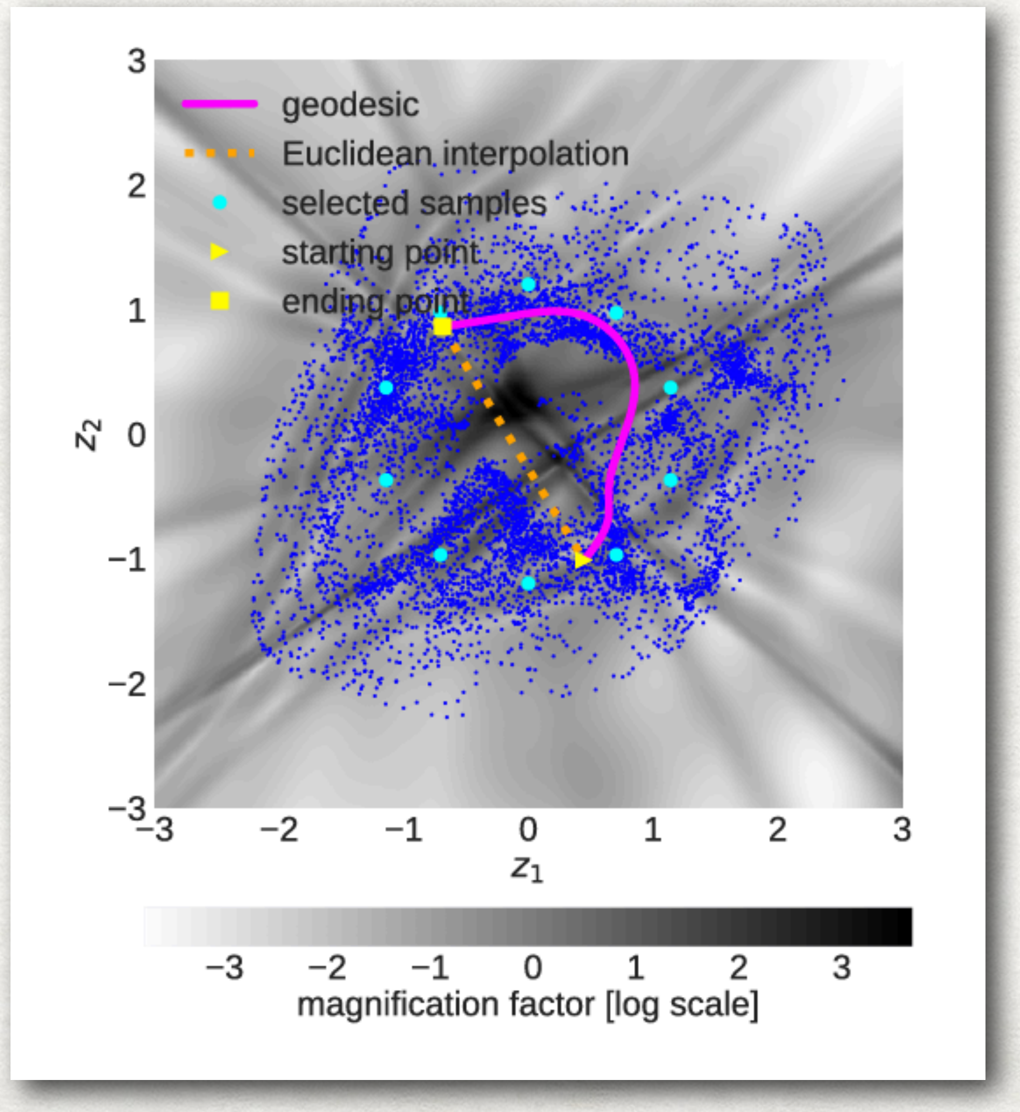



# EXPERIMENTS

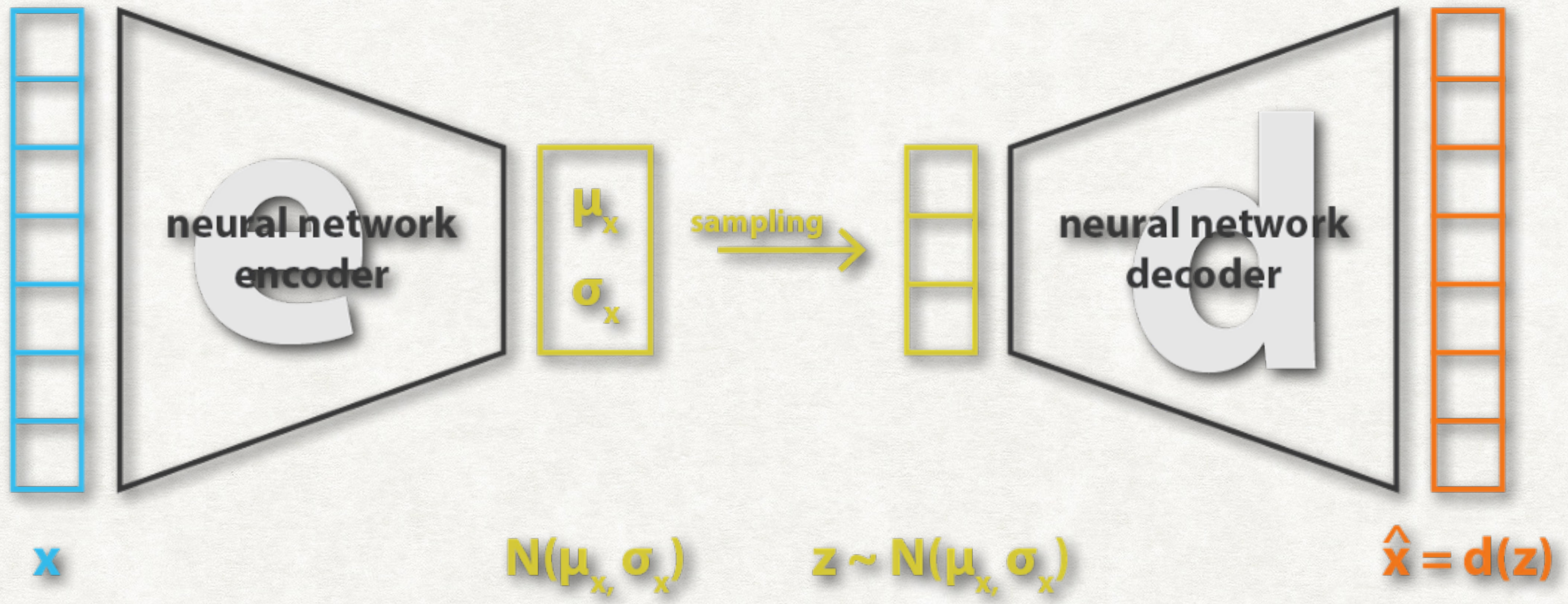
- Human Motion: CMU graphics lab motion capture database




# CONCLUSION

- The distance between points in the latent space in general does not reflect the true similarity of corresponding points in the observation space
- Riemannian distance metric can be used as an alternative metric that takes the underlying manifold into account
- MF can be used as a powerful tool to visualize the magnitude of generative model's distortion of infinitesimal areas in the latent space.




$$\text{loss} = \| x - \hat{x} \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I)\,] = \| x - d(z) \|^2 + \mathrm{KL}[\, N(\mu_x, \sigma_x), N(0, I)\,]$$