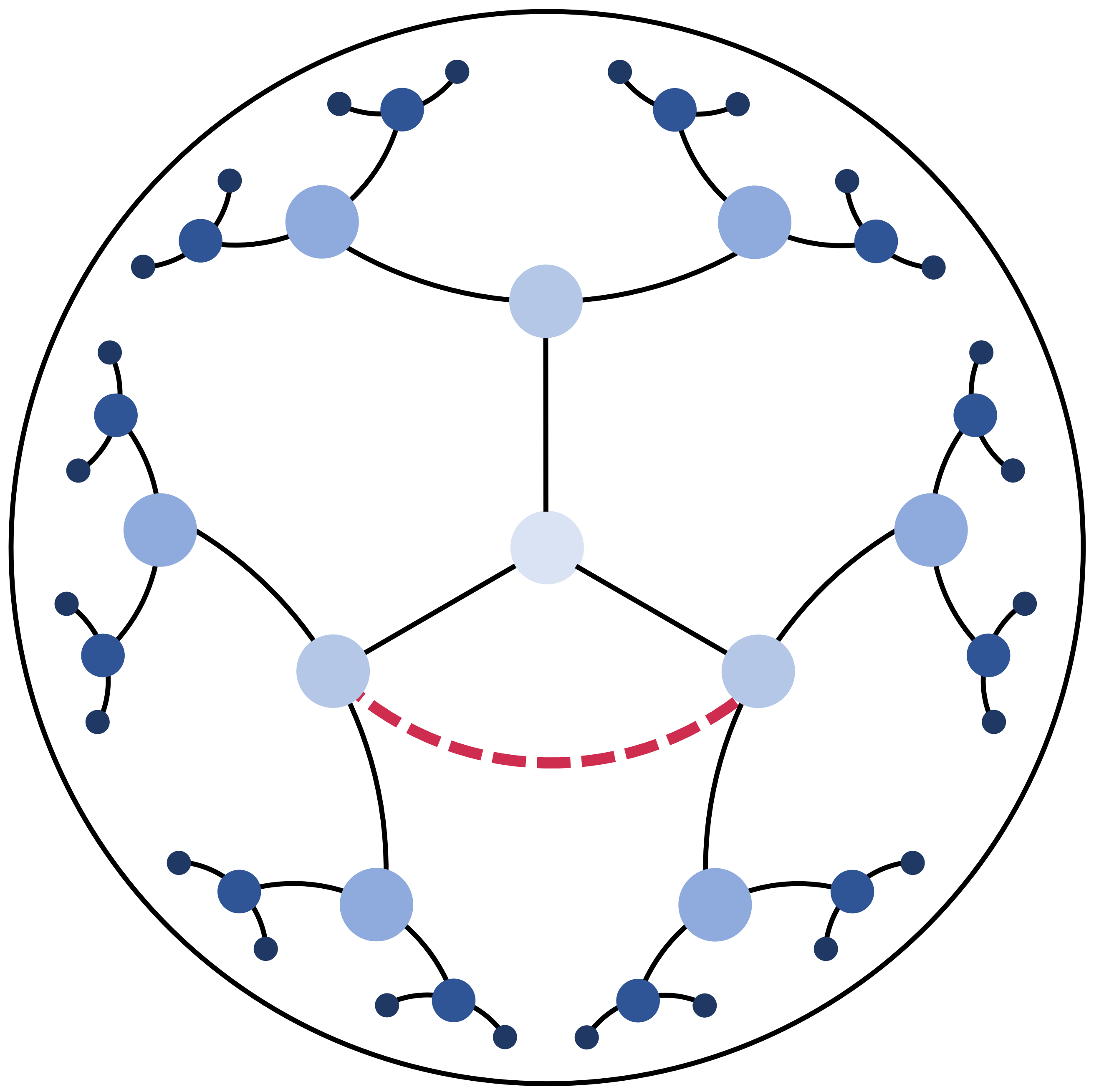
Stable and Fast Optimization on Hyperbolic Space

Seunghyuk Cho, Dongjun Yu

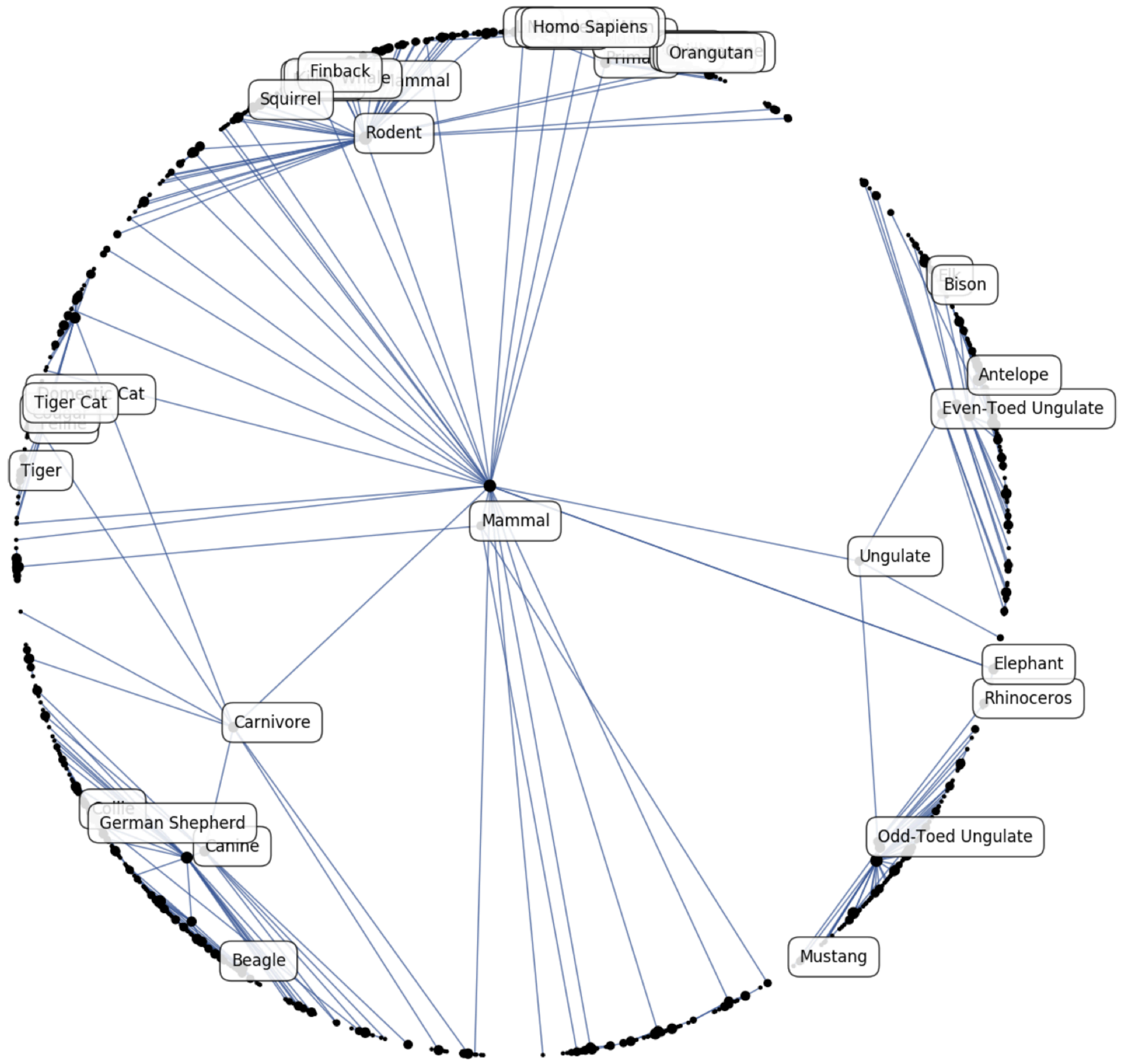
Pohang University of Science and Technology

Hyperbolic Space

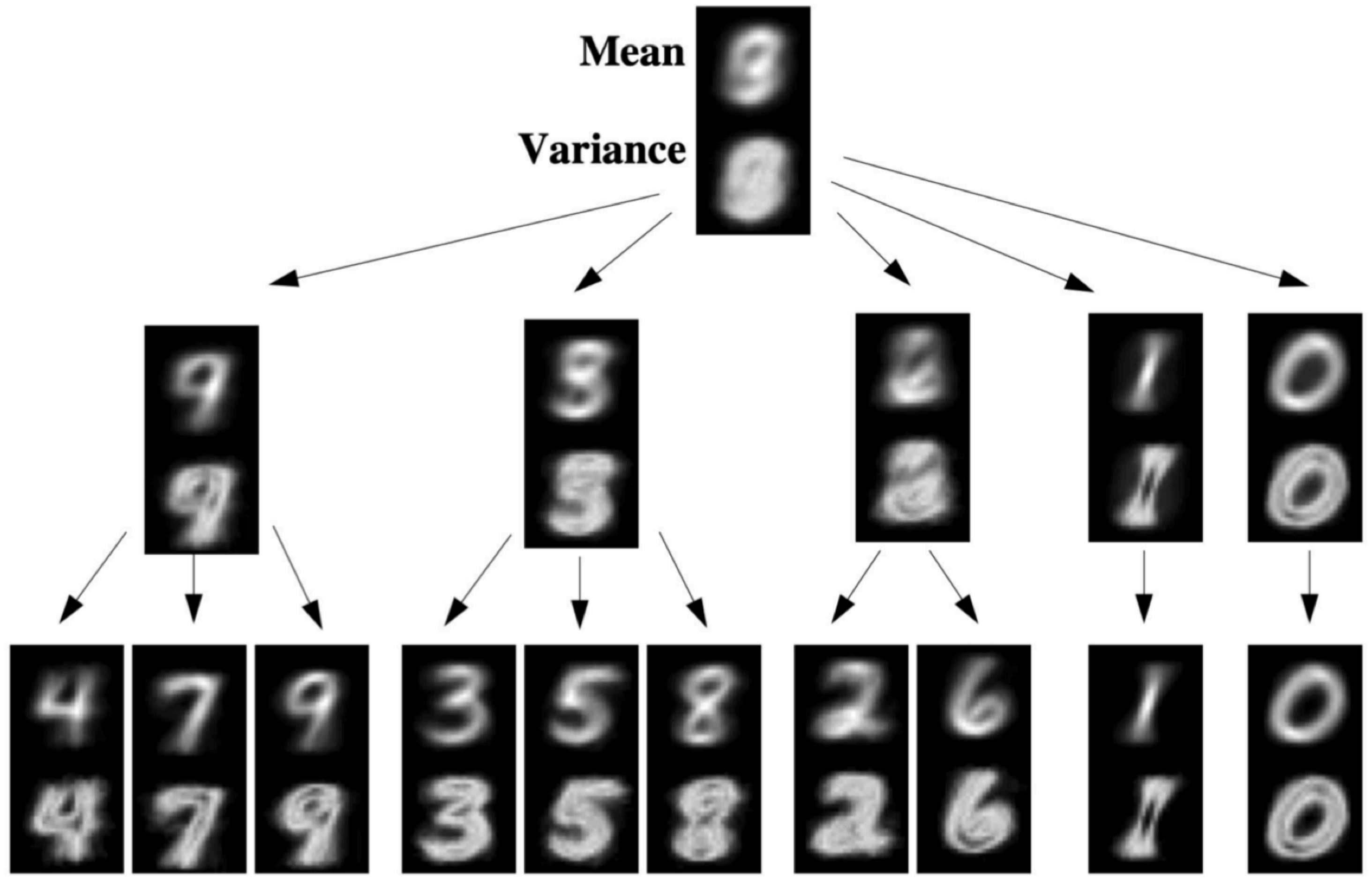
- Hyperbolic space well-captures the hierarchical structure of data.



Hyperbolic space



Word relationships



MNIST

Learning Hyperbolic Embeddings

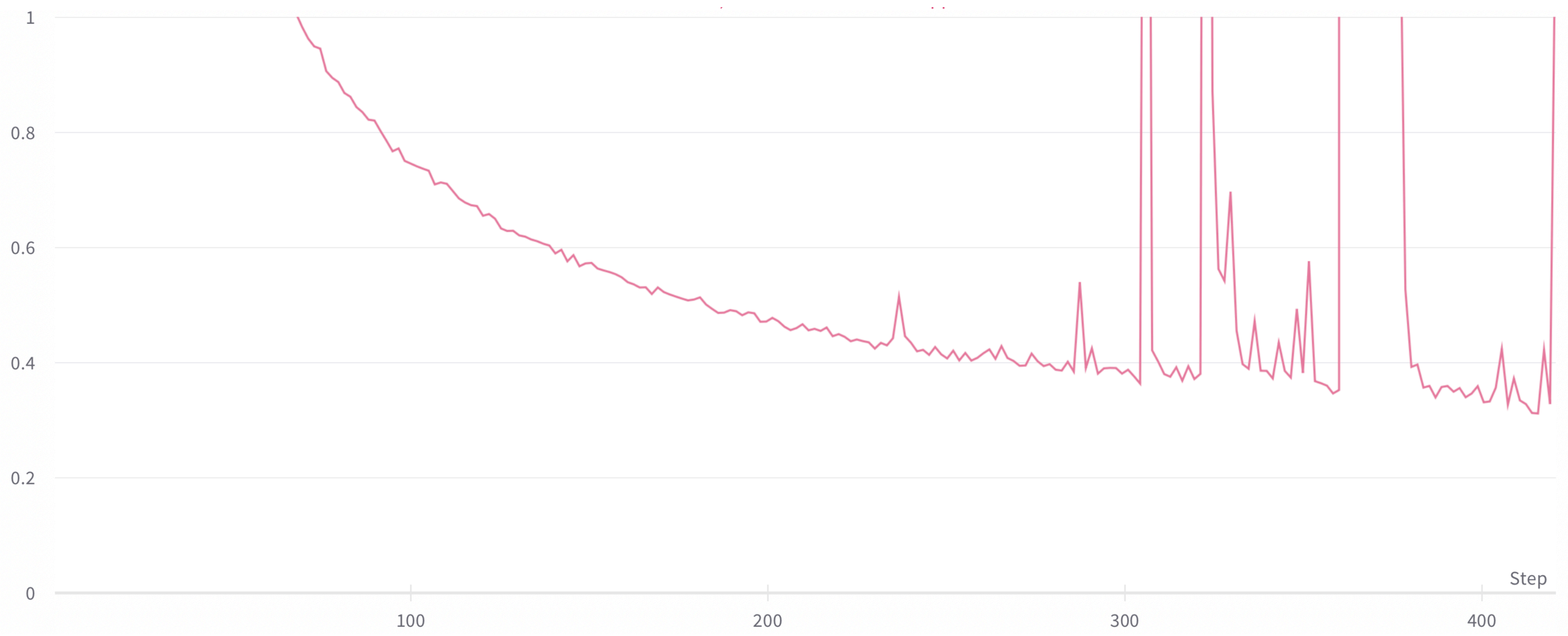
- Learning the embeddings on the hyperbolic space is formulate as:

$$\min_{x \in \mathbb{H}} f(x).$$

- f : the given objective.
- $\mathbb{H}$: the hyperbolic space.

Hyperbolic Embeddings Optimization - Problems

- Learning hyperbolic embeddings is difficult than Euclidean embeddings.
  - Numerical stability issues often appear!



Loss curve of hyperbolic embedding learning

Contributions

(Contribution 1) We analyze the reason why the optimization on the hyperbolic space is difficult.

(Contribution 2) We propose the hyperbolic version of projected gradient descent and show that it is more efficient.

Background

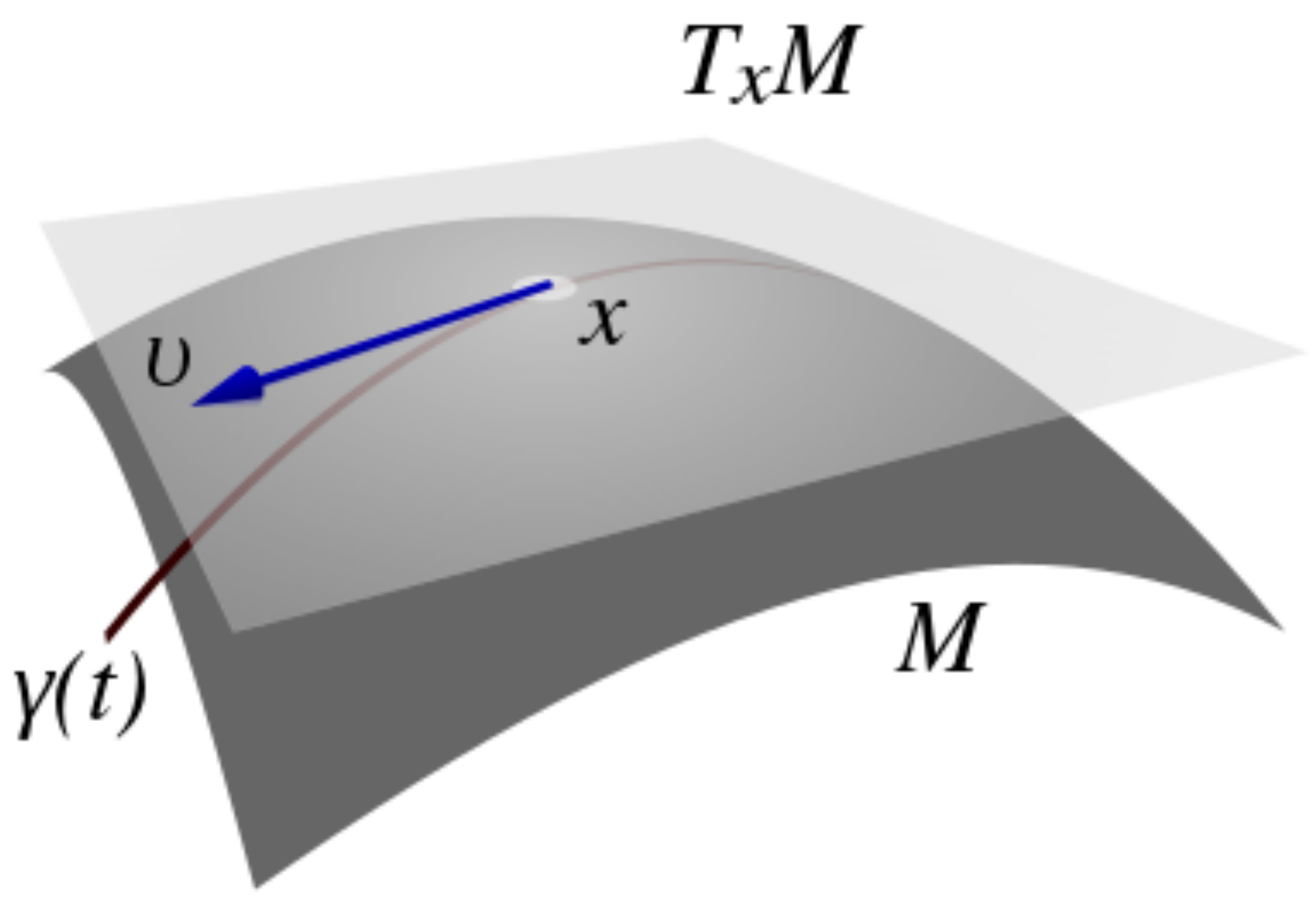
Riemannian Manifold

- Riemannian manifold $(\mathscr{M}, g)$ is a pair of a manifold and a metric tensor.
  - Manifold $\mathscr{M}$: set of points.
  - Metric tensor g: used to define geometric operations, i.e. angle of two vectors and distance.

- Hyperbolic space is a unique, complete, simply connected Riemannian manifold with constant negative sectional curvature.

Tangent Space

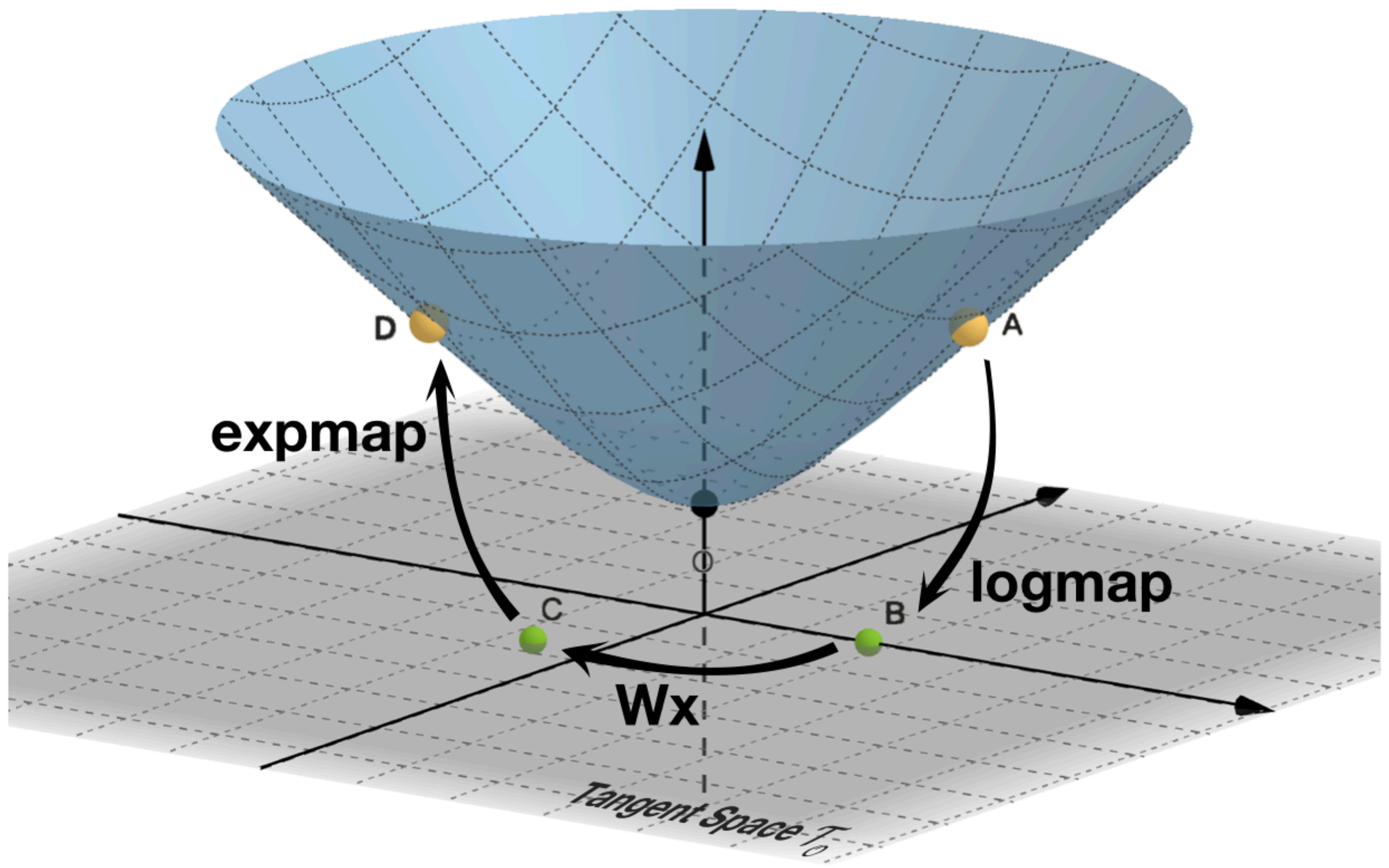
- A tangent space $\mathcal{T}_x\mathcal{M}$ is the set of the vectors tangent to $x \in \mathcal{M}$.
  - Metric tensor: inner product of two tangent vectors.



Tangent space

Exponential Map

- Exponential map $\exp_x(v)$ moves x to a point along the tangent vector v.
  - $(=)$ project a tangent vector to a point of the manifold.



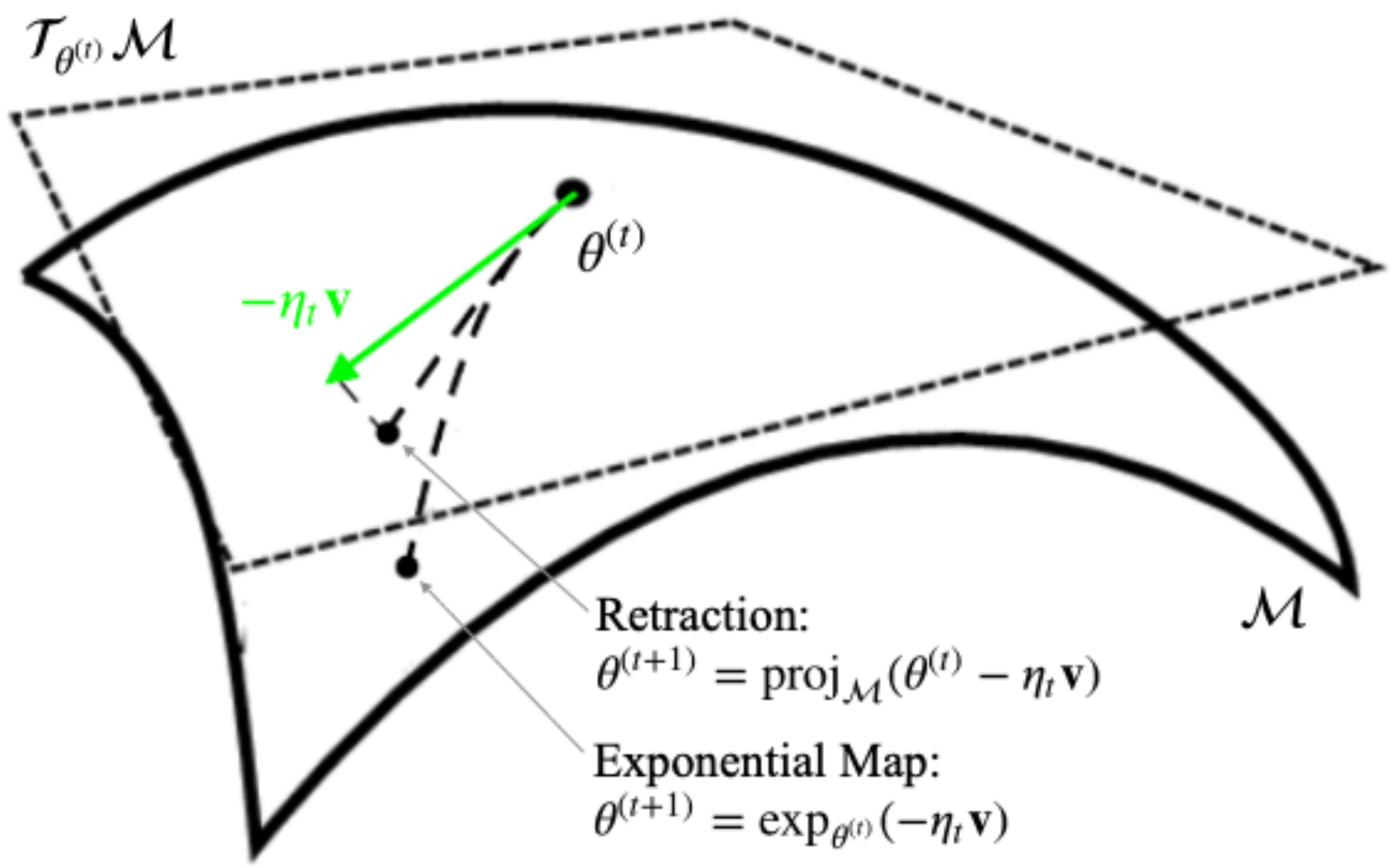
Exponential map and log map of the Riemannian manifold

Riemannian Stochastic Gradient Descent (RSGD)

- We can update the parameters on the Riemannian manifold as:

$$x_{t+1} = \exp_{x_t}\left(-\eta_t H(x_t)\right).$$

- $H(x_t)$: the Riemannian gradient.



Riemannian Stochastic Gradient Descent

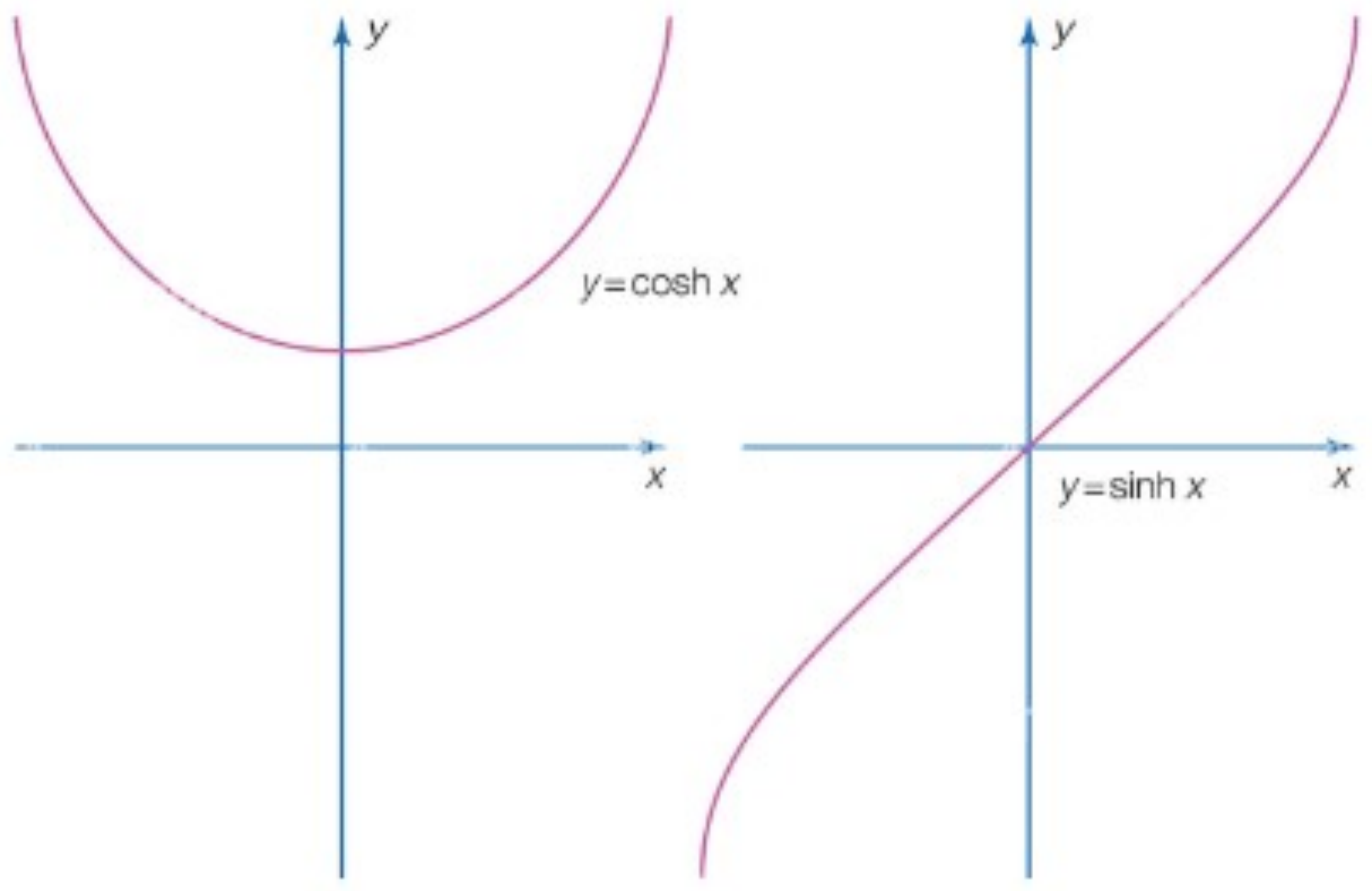
Analysis

RSGD: Issues

- Previous work argue that learning hyperbolic embeddings is difficult.
  - i.e. numerically unstable.
  - Why?

Exponential Map of Hyperbolic Space

- Exponential map of hyperbolic space is consisted of hyperbolic functions.
  - Hyperbolic functions are similar to exponential function which rapidly increases.
  - The identity function $\log_x(\exp_x(v))$ fails to preserve the input with large magnitude.



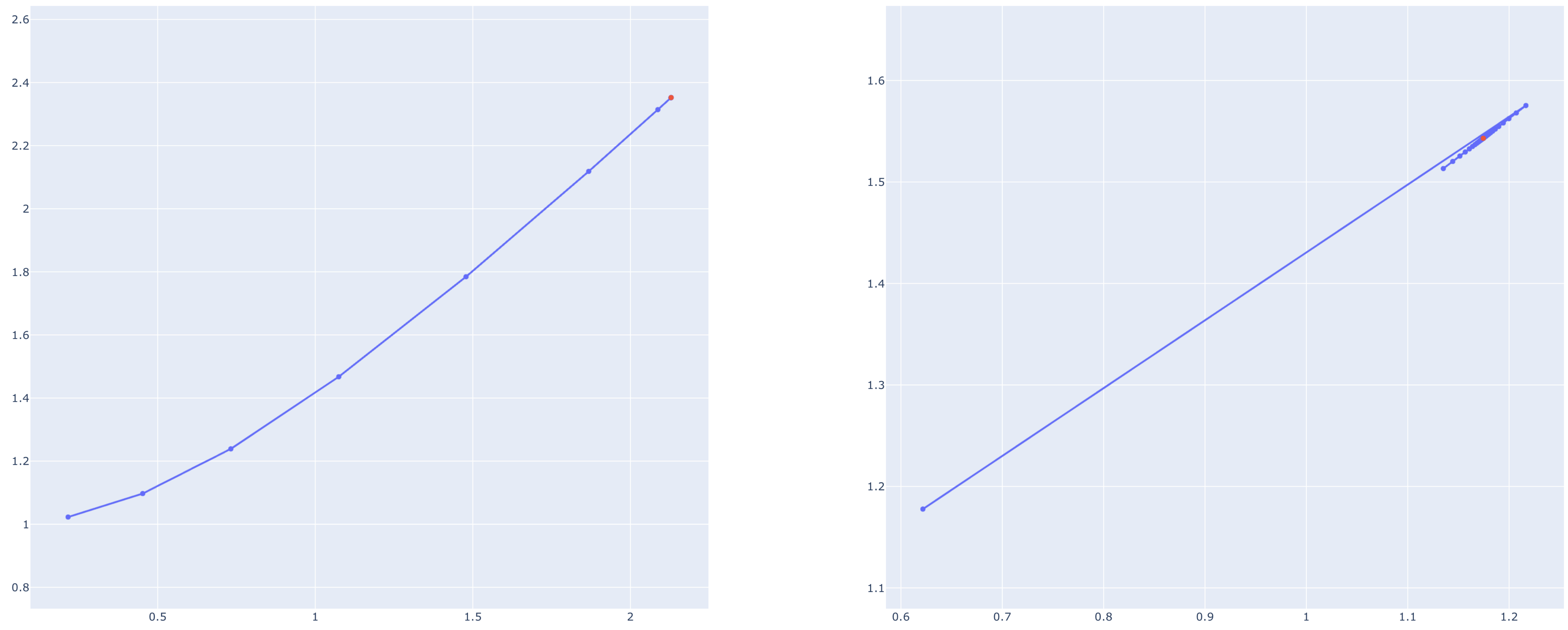
Hyperbolic functions

Norm of input	[-1, 1]	[-10, 10]	[-100, 100]	[-1000, 1000]
Error rate	1.25E-32	4.97E+00	6.15E+03	9.98E+05

Log(Exp(x))

Behavior of Riemannian Gradient

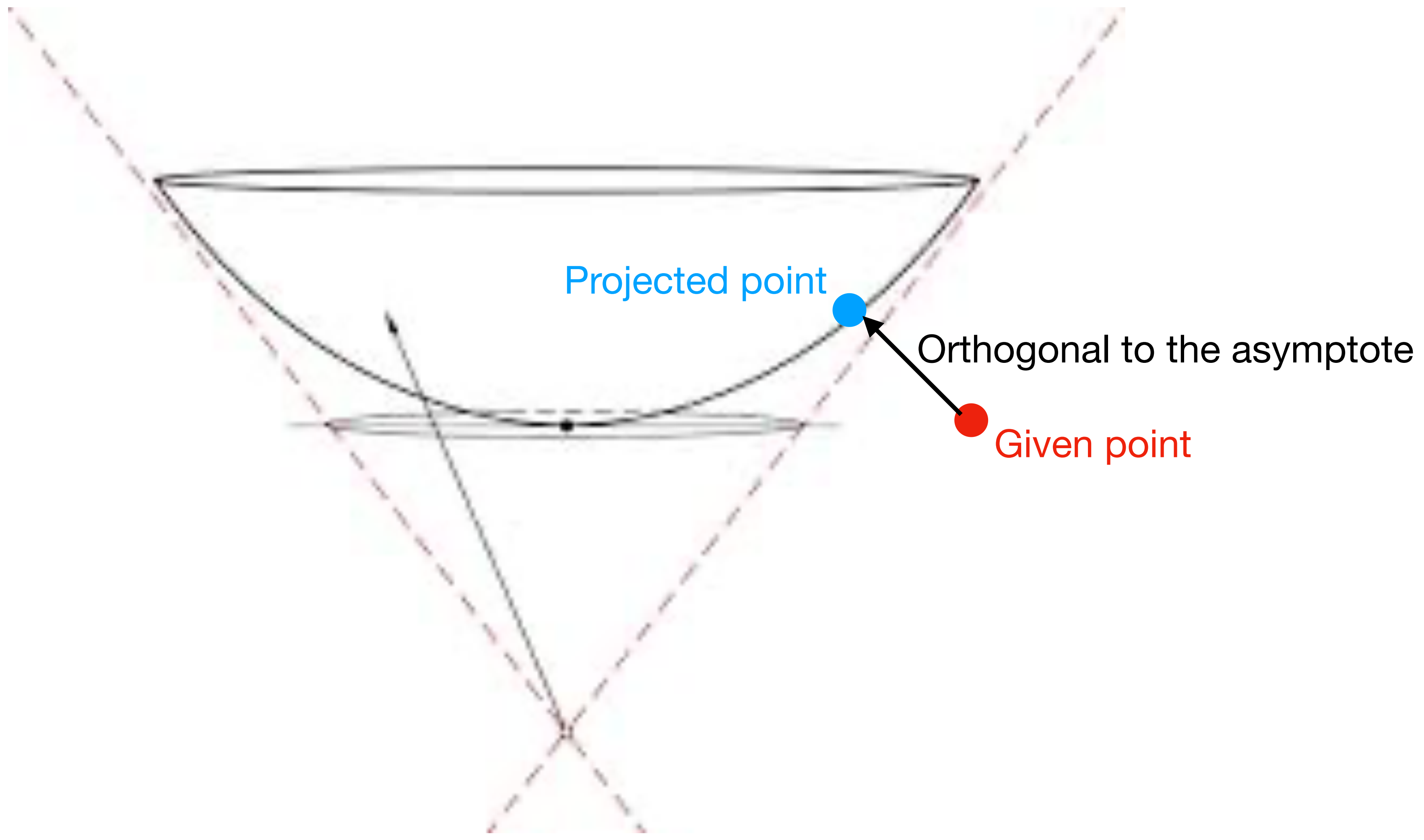
- We empirically show that the Riemannian gradient in the hyperbolic space depends on the norm.
  - The norm of the gradient increases at first and then decreases.
  - Need smaller learning rate to converge.



Trajectory of the learning process

Hyperbolic Projected Gradient Descent (HPGD)

- We propose a hyperbolic projected gradient descent method.
  - Due to the complexity of the projection function, we use the approximated version.



Experiments

Experiments

- We compare the methods in i) synthetic and ii) real-world settings.

Synthetic Setting

- We first sample a ground-truth point $x \in \mathbb{H}$ and define the objective as:

$$L(x, \hat{x}) = \frac{1}{2}\|x - \hat{x}\|_2^2,$$

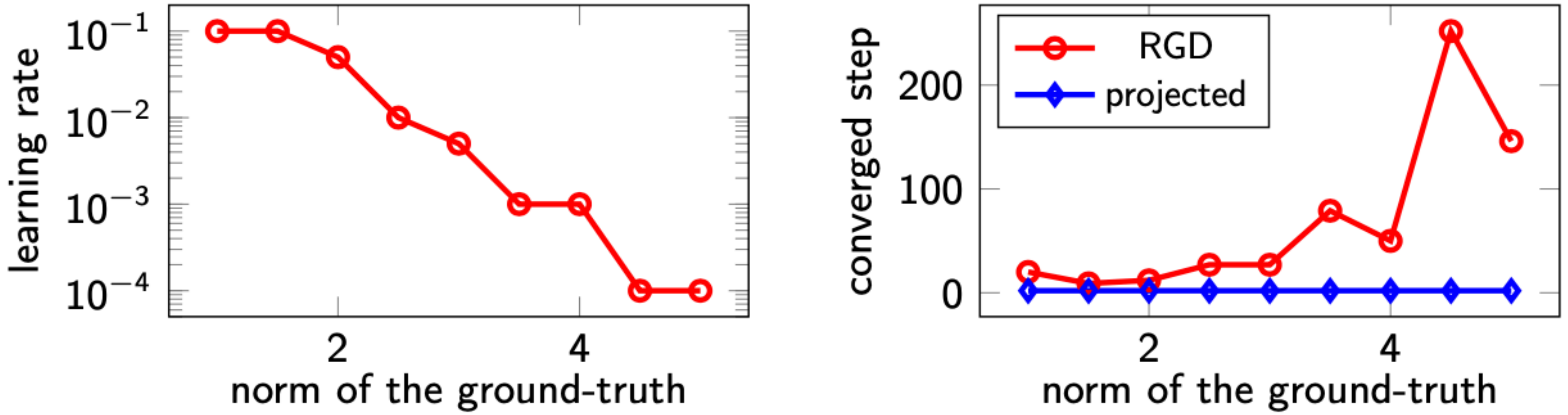
- where $\hat{x}$ is the learned parameter.

Synthetic Setting: Metric

- We compare the steps needed to converge to the ground-truth.
  - We empirically find the learning rate for RGD where it converges.
  - For HPGD, we fix the learning rate to 1.0.

Synthetic Setting: Results

- HPGD is i) stable than RGD w.r.t the choice of learning rate and ii) show faster convergence.



(a) Sufficient learning rate for RGD (b) Number of steps needed to converge

Conclusions

- We provide analysis for the RSGD on the hyperbolic space.
- We show that HPGD converges faster than RSGD due to the stability on the learning rate.

Thank You :)