Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders

Emile Mathieu†, Charline Le Lan†, Chris J. Maddison†*, Ryota Tomioka‡ and Yee Whye Teh†*

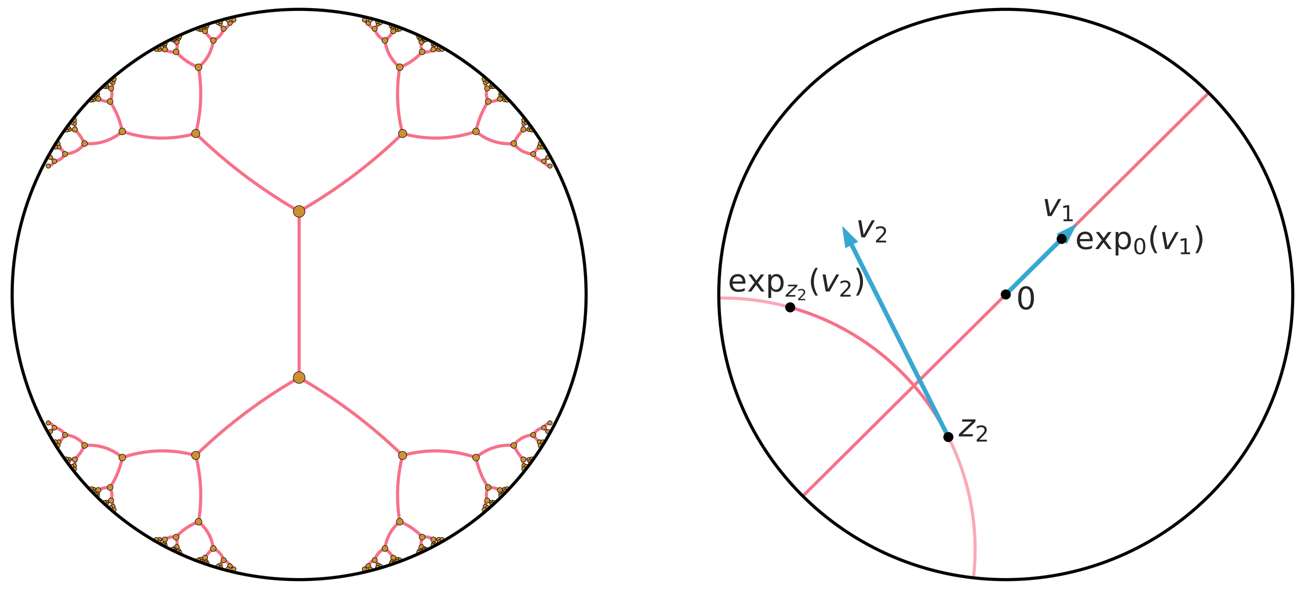
† Department of Statistics, University of Oxford, United Kingdom, * DeepMind, London, United Kingdom, ‡ Microsoft Research, Cambridge, United Kingdom

UNIVERSITY OF OXFORD

Overview

1. We propose efficient and reparametrisable sampling schemes, and calculate the probability density functions, for two canonical Gaussian generalisations defined on the Poincaré ball, namely the maximum-entropy and wrapped normal distributions. These are the required ingredients for learning our variational auto-encoder (VAE)s.
2. We introduce a decoder architecture explicitly taking into account the hyperbolic geometry, which we empirically show to be crucial.
3. We empirically demonstrate that endowing a VAE with a Poincaré ball latent space can be beneficial in terms of model generalisation and can yield more interpretable representations.

The Poincaré Ball model of hyperbolic geometry

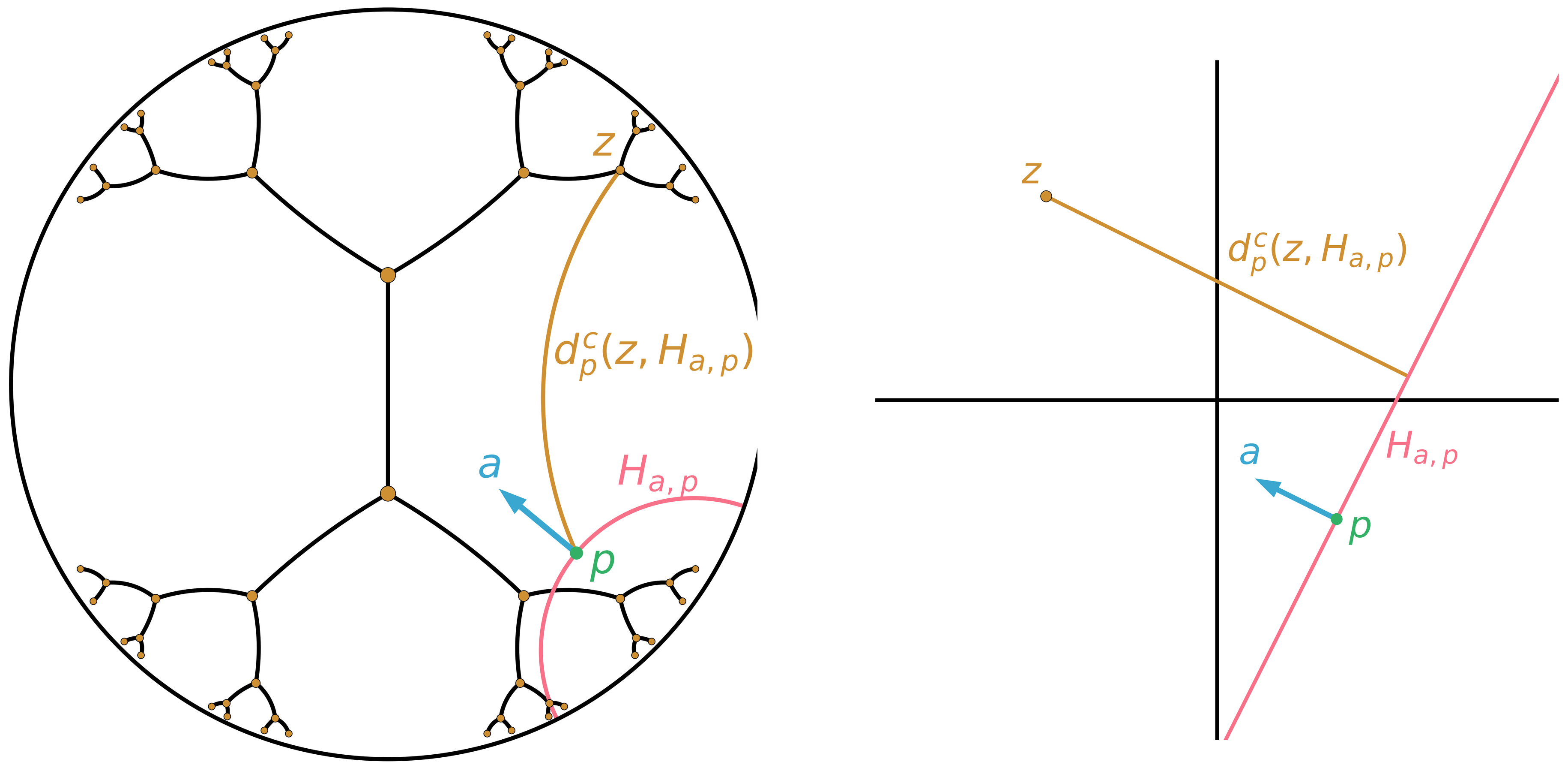


(a) (b)

We are extremely grateful to Adam Foster, Phillipe Gagnon and Emmanuel Chevallier for their help. EM, YWT's research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007- 2013) ERC grant agreement no. 617071 and they acknowledge Microsoft Research and EPSRC for funding EM's studentship, and EPSRC grant agreement no. EP/N509711/1 for funding CL's studentship.

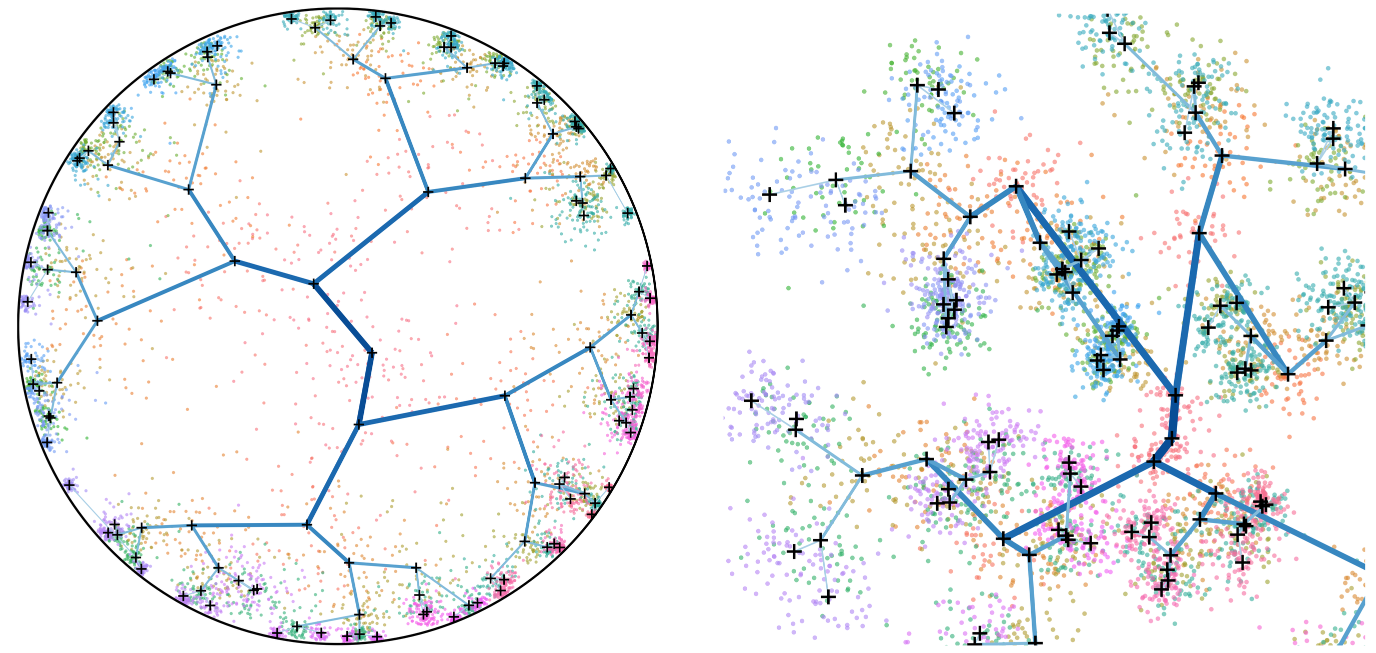
Probability measures on $\mathbb{B}^d$

Hyperbolic decoder



(a) (b)

branching diffusion process



(a) (b)

MNIST digits

Table 1:
Per digit accuracy of a classifier trained on the learned latent 2-d embeddings. Results are averaged over 10 sets of embeddings and 5 classifier trainings.

Digits	0	1	2	3	4	5	6	7	8	9	Avg
$\mathcal{N}$-VAE	89	97	81	75	59	43	89	$\mathbf{78}$	68	$\mathbf{57}$	73.6
$\mathcal{P}^{1.4}$-VAE	$\mathbf{94}$	97	$\mathbf{82}$	$\mathbf{79}$	$\mathbf{69}$	$\mathbf{47}$	$\mathbf{90}$	77	68	53	$\mathbf{75.6}$

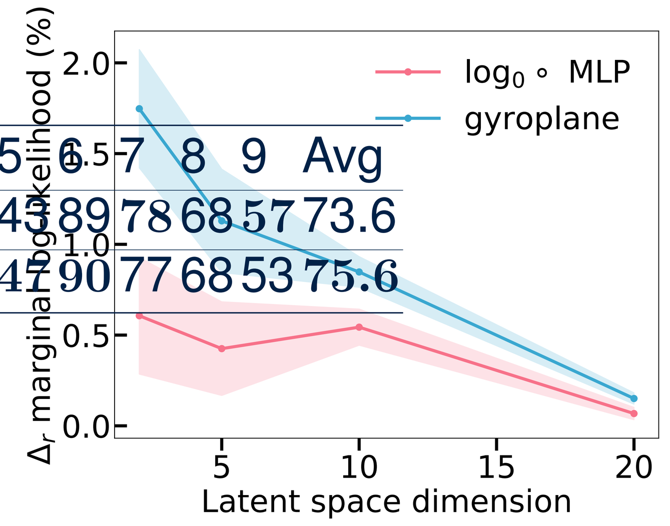


Figure 4: Decoder ablation study on MNIST with *wrapped* normal $\mathcal{P}^1$-VAE. Baseline decoder is a MLP.