Latent Riemannian Diffusion Models with Mixed Curvature

Ria Arora
Universite de Montreal
ria.arora@umontreal.ca

Tommy He
McGill University
tommy.he@mail.mcgill.ca

Abstract

We extend Riemannian diffusion models to learned mixed curvature manifolds and provide a novel comparison of results. Out goal is to infer the manifold for the latent embedding from for the diffusion model to improve sample quality. As conjectured and evidenced by previous research, the quality of representation is better if the geometry of the embedding space and data match. Most of the literature on geometric generative models fix a simple manifold, such as sphere or torus, and run it on synthetic data that matches the structure of the chosen manifold. This project investigates the possibility of using mixed curvature to provide more flexibility for learning the type of manifold for a dataset and posits that the additional expressiveness and lack of bias benefits Riemannian diffusion models.

1 Introduction

Likelihood based generative models have proven to be a stable way to generate state-of-the-art (SOTA) images from an underlying distribution. VAEs, autoregressive models, and more recently diffusion models are all examples of these that achieve the SOTA in various metrics such as Child [2020], which achieves SOTA negative log-liklihood values, and Nichol and Dhariwal [2021], which can achieve SOTA negative log-likelihood values and FID while improving sampling speed. Diffusion models, and in particular DALLE-2 [Ramesh et al., 2021] and its parent model latent diffusion [Rombach et al., 2021] have gained a large amount traction due to their ability to generative a wide diversity of images with high quality.

However, all the previously mentioned models work over Euclidean spaces, whereas much data is best modeled with Riemannian manifolds, such as protein modeling [Shapovalov MV, 2011], robotics [Senanayake and Ramos, 2018], and computer vision [Lui, 2012] among other examples. Therefore, we aim to extend latent diffusion models, which previously operated solely in Euclidean space to manifolds and even to learn a suitable product manifold embedding space.

With this motivation, we extend latent diffusion to Riemmanian manifolds by taking its autoencoder but using Riemmanian diffusion models [Huang et al., 2022] over the latent space of the autoencoder to generate the samples. We use a continuous time formulation with a noising process given by an SDE on a manifold to converge towards a normal distribution wrapped around the manifold. In the generative direction, we uniformly sample on the latent space and take the time-reversal SDE to generate vectors. The vectors are then decoded back to images to form our samples. We compare the quality and investigate the results from simple, fixed manifolds, such as spherical and tori, with learned product manifolds over spherical, Euclidean, hyperbolic, Poincaré ball, and projected hyperspherical manifolds. From this, we demonstrate the importance of finding suitable latent space and the expressivity of the embedding space.

Preprint. Under review.

2 Related Work & Background

2.1 Riemannian Diffusion Model

Diffusion model are probabilistic models that learns a data distribution $p(x, T)$ by gradually adding noise to the data to reach a normal distribution and then use a reverse Markov process of length T to denoise to the actual data again. A continuous-time diffusion model can be defined via SDEs as

$$\mathrm{d}X = \mu dt + \sigma\, \mathrm{d}B_t$$

$$\mathrm{d}Y = (-\mu + \sigma)\, \mathrm{d}s + \sigma\, \mathrm{d}\hat{B}_s$$

where the initial condition X_0 follows some prior p_0, B_t and $\hat{B}_s$ are standard Brownian motion, and μ and σ are the drift and diffusion coefficients. A lower bound on the marginal likelihood can be obtained by

$$\log\ p(x, T) \geq \mathbb{E}\left[\log p_0(Y_T) - \int_0^T \left(\frac{1}{2}\|a(Y_s, s)\|_2^2 + \nabla \cdot \mu(Y_s, T - s)\right) ds \mid Y_0 = x\right]$$

where a is the variational degree of freedom, $\nabla\cdot$ is the divergence operator.

Huang et al. [2022] generalizeed continuous time diffusion model to Riemannian manifolds. Consider Riemannian manifold (M, g), let B_t and $\hat{B}_s$ be w-dimensional Brownian motions, and let X_t , $Y_s \subset M$ be processes solving the SDE

$$\mathrm{d}X = V_0\, \mathrm{d}t + V \circ \mathrm{d}B_t$$

$$\mathrm{d}Y = (-V_0 + (V \cdot \nabla_g)V + Va)\, \mathrm{d}s + V \circ d\hat{B}_s$$

where V_0 are the columns of the diffusion matrix, V are smooth vector fields on M. Let $a : \mathbb{R}^m \times [0, T] \to \mathbb{R}^m$ is the variational degree of freedom. They obtain the Riemannian CT-ELBO

$$\log p(x, T) \geq \mathbb{E}\left[\log p_0(Y_T) - \int_0^T \frac{1}{2}\|a(Y_s, s)\|_2^2 + \nabla_g \cdot \left(V_0 - \frac{1}{2}(V \cdot \nabla_g)V\right) \mathrm{d}s \mid Y_0 = x\right]$$

To compute the Riemannian CT-ELBO, it is required to compute the divergence $\nabla_g\cdot$ of a vector field for which the paper applies the Riemannian divergence identity and uses QR decomposition to find an orthogonal basis for $T_x M$. They experiment on density estimation on sphere, torus, hyperbolic and orthogonal manifolds of constant curvature.

2.2 Mixed Curvature Representations

Gu et al. [2019] discussed learning embeddings in a product manifold combining components of Riemannian manifolds (spherical, hyperbolic, Euclidean), providing a space of heterogeneous and adaptable curvature suitable for a wide variety of structures. Given linearly independent $u, v \in T_p M$, the sectional curvature $K_p(u, v)$ is the Gaussian curvature of the surface $\mathrm{Exp}(U) \subseteq M$. In a product manifold P, the sectional curvature interpolates the sectional curvature of the factors. Let abc be an geodesic triangle in M and m be the geodesic midpoint of bc. Then, they estimate the sectional curvature as

$$\xi_M(a, b, c) := d_M(a, m)^2 + \frac{d_M(b, c)^2}{4} + \frac{d_M(a, b)^2 + d_M(b, c)^2}{4}$$

Each component's curvature is learned jointly with the embedding in the product space via Riemannian optimization. The paper, however, only works with reconstruction tasks on graph embeddings. So, Skopek et al. [2020] extended product manifolds with the addition of stereographic conformal projections Poincaré ball and projective hypersphere to learn the curvature in a sign agnostic way. The paper uses different types of experiments on the curvature of the manifold: fixed curvature, learnable curvature, and universal curvature, which are explained later in section 3.2.2. They experimented on image datasets such as datasets MNIST, CIFAR.

2.3 Latent Diffusion Models

Rombach et al. [2021] proposed to use an autoencoder to encode the input into a latent representation to then apply diffusion. The intuition behind this decision is to lower the computational demands of training diffusion models by processing the input in a lower dimensional space, since diffusion models often have high training costs in a complex pixel space. Especially given our lack of GPUs and strong training power, we want to reduce the complexity of training Riemannian diffusion models with this idea. Additionally, diffusion models bypass perceptually unimportant details, so they propose using an autoencoder first to find the perceptually important details in a smaller latent space, after which to apply a standard diffusion model (U-Net) to generate new data to then get upsampled by a decoder network. They use a perceptual compression network inspired by Esser et al. [2020] with additional regularization to lower variance, whereas we use a relatively simpler autoencoder for lower training time.

3 Methodology

3.1 Latent Riemannian Diffusion Model

We train 2 separate autoencoders for spherical and tori manifolds with MNIST until convergence due to different dimensionality. For both, we take a convolutional autoencoder consisting of 3 convolutional layers and 2 fully connected layers that encodes the MNIST images. For spherical, we encode to a 3 dimensional vector, and for tori, we encode to a 4 dimensional vector to fit on the manifold. For every experiment, to ensure it does not skew results, the encoder is pretrained and frozen. As per Huang et al. [2022], the gap between the Riemannian CT-ELBO and the exact likelihood may be large for one sample, with a similar technique, so we use K samples through importance sampling to get a tighter lower bound.

3.2 Mixed Curvature Riemannian Diffusion Model

3.2.1 Riemannian Diffusion Model with product manifold

In Riemannian Diffusion models [Huang et al., 2022], we have

$$\begin{aligned}\text{Generative SDE:}\quad & dX = V_0 dt + V \circ dB_t \quad X_0 \sim p_0 \\ \text{Inference SDE:}\quad & dY = U_0 dt + V \circ \hat{d}B_s\end{aligned}$$

where V_0 and the columns of the diffusion matrix $V := [V_1, \ldots, V_w] : V_i \in M_i$ are smooth vector fields on $M = M_1 \times M_2 \times \ldots \times M_w$ where a point $p \in M$ with coordinate $p = (p_1, p_2, \ldots, p_w) : p_i \in M_i$, and B_t is a Brownian motion. Here $V \circ dB_t$ can be decomposed to add component wise Riemannian divergence of the vector field of the particular manifold. All other operations defined on the manifolds are element-wise and are then concatenated back to form a representation.

3.2.2 Learning Curvature

For a manifold to be defined, dimensionality and curvature are needed. In this architecture, we fix the dimensionality of the manifold and have 3 possibilities for choosing our manifold:

1. Fixed curvature choses discretely from [0.25,1] as a prior for the training procedure which corresponds to [1,2] radii respectively. The curvature is 0 for Euclidean spaces, negative for hyperboloids, and positive for hypersphere for stability reason as pointed out by Skopek et al. [2020] as there is divergence of points in hyperboloid and hypershere as $K \to 0$.

2. In learning curvature, ELBO is differentiated with respect to to K using gradient based optimisation to learn the curvature across each component in product manifold. In this paper, we mainly compare and contract fixed curvature and learning curvature ideas.

3. Universal Curvature - selects the “partitioning” of our latent space – the number of components and for each of them select the dimension and at least the sign of the curvature of that component to estimate the signature.

4 Experiments

4.1 Latent Riemannian Diffusion Model

Both our autoencoders converge after 30 epochs, at which point the loss stabilizes at around 0.0309 for spherical and 0.0262 for torus shown in figure 1 2, at which point we freeze the model and use it for all other experiments.

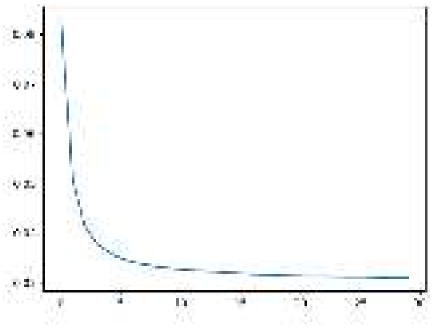


Figure 1: Sphere Autoencoder Loss

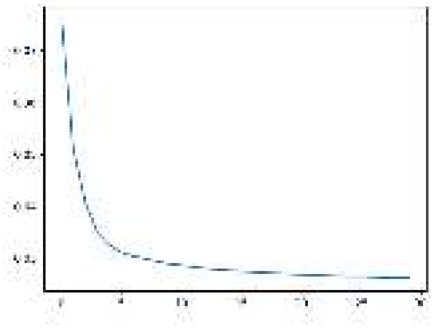


Figure 2: Torus Autoencoder Loss

We pass the result of the encoded images as 3 or 4 dimensional vectors to the Riemannian diffusion model, which was trained for 100,000 iterations until convergence as shown in figure 3 to give a KELBO of -2.148 for spherical and -3.47387 for $K = 10$.

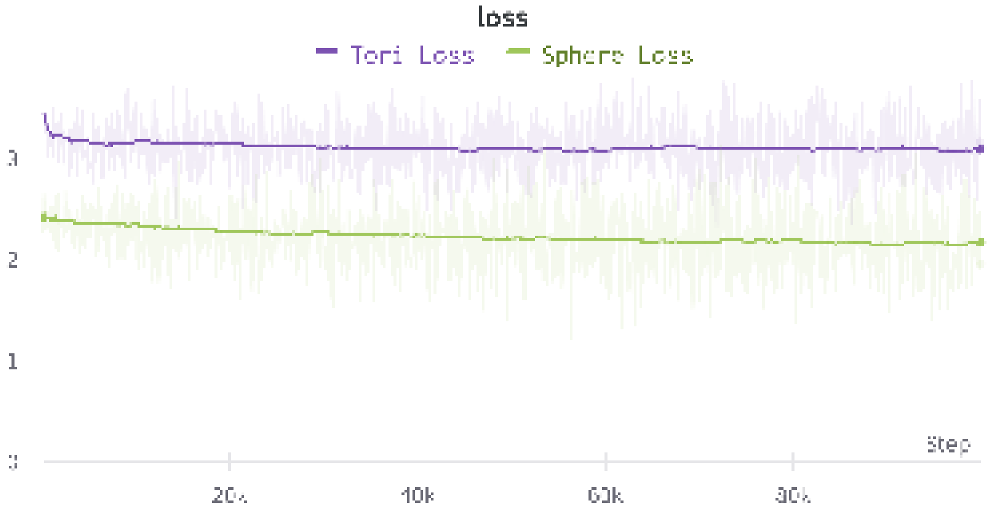


Figure 3: Diffusion Loss

Some generated samples are below in figures 4, 5.

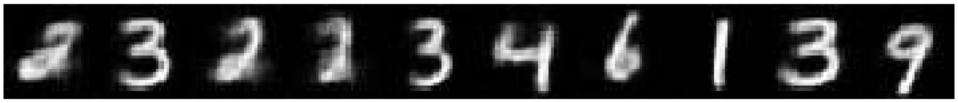


Figure 4: Sphere Samples

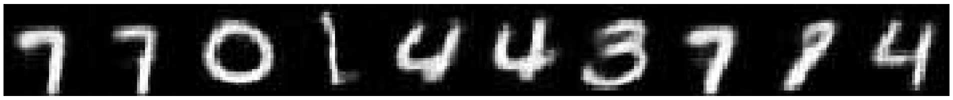


Figure 5: Tori Samples

When plotted on a sphere, the digits don't show much pattern and are quite scattered as in figure 6 and 7. From these plots and the loss curve, we conclude that we needed more expressive geometry to capture digits better, such as more complex mixed curvature.

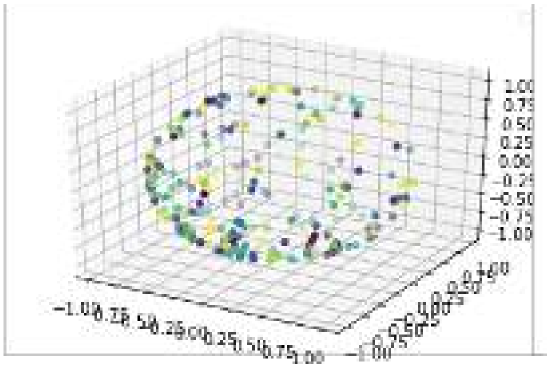


Figure 6: Digits 0-9 on Sphere

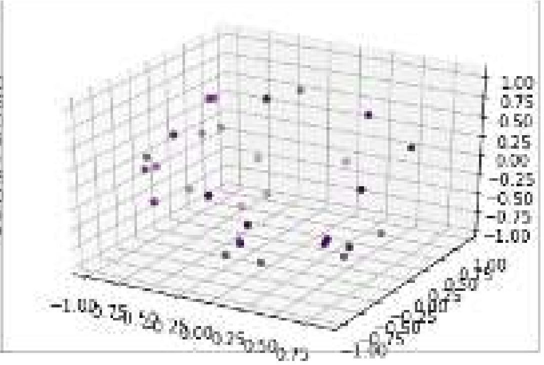


Figure 7: Digit 0 on Sphere

4.2 Mixed Curvature Riemannian Diffusion Model

Riemannian Diffusion Model was trained and evaluated for MNIST dataset on various manifolds for fixed curvature and learnable curvature.

An autoencoder was used to reduce the dimensionality to match the signature of the manifold provided in the case of $\mathbb{E}^2 \times \mathbb{H}^2 \times \mathbb{S}^2$ and then the diffusion model was applied, with the same idea as section 2.2. In the case of $\mathbb{E}^{14} \times \mathbb{H}^{14} \times \mathbb{S}^4$, there was no autoencoder used since the dimensionality of the MNIST image 784 matched the latent space. Observe in the tables 1 and 2 that the learned curvature outperforms the fixed curvature in terms of the KL and the ELBO. Finally, we provide samples in our last figure 7.

Table 1: Fixed Curvature

Manifold Signature	Curvature $(\mathbb{E}, \mathbb{H}, \mathbb{S})$	KL divergence	ELBO
$\mathbb{E}^2 \times \mathbb{H}^2 \times \mathbb{S}^2$	[0,-0.25,0.25]	13.89	-111.45
$\mathbb{E}^{14} \times \mathbb{H}^{14} \times \mathbb{S}^4$	[0,-0.25,0.25]	28.52	-98.23

Table 2: Learning Curvature

Manifold Signature	Curvature $(\mathbb{E}, \mathbb{H}, \mathbb{S})$	KL divergence	ELBO
$\mathbb{E}^2 \times \mathbb{H}^2 \times \mathbb{S}^2$	[0,-0.1385,0.19]	13.77	-112.8
$\mathbb{E}^{14} \times \mathbb{H}^{14} \times \mathbb{S}^4$	[0,-0.045,0.06]	27.95	-99.03

Figure 7: MNIST real

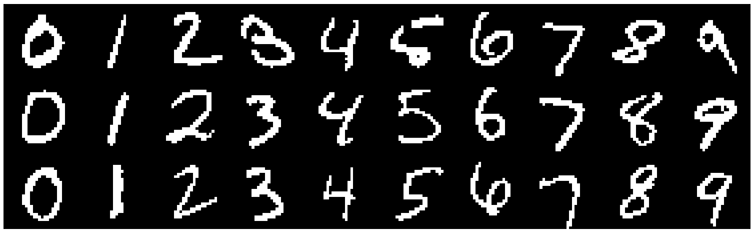


Figure 8: MNIST generated after 60 epochs on $e^2 \times h^2 \times s^2$ with learning the curvature

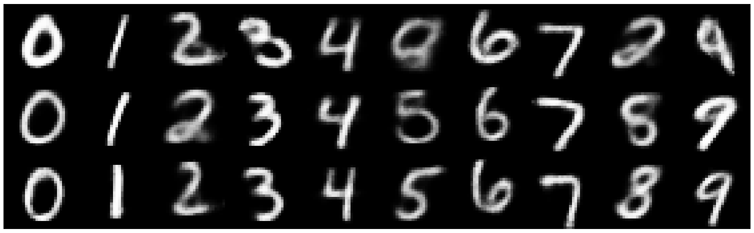


Figure 9: MNIST generated after 60 epochs on $e^2 \times h^2 \times s^2$ with fixed curvature

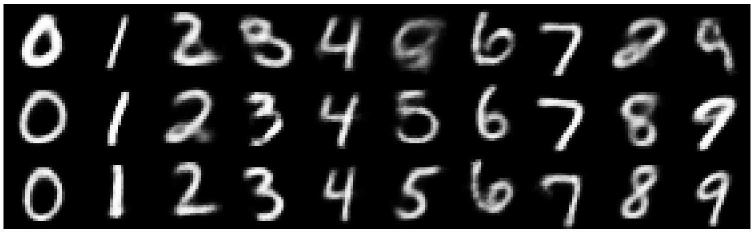


Figure 10: MNIST generated after 30 epochs on $e^{14} \times h^{14} \times s^4$ with learning the curvature

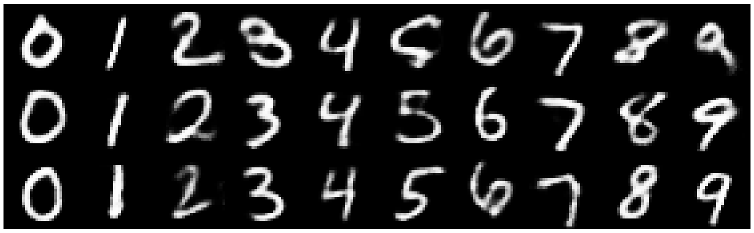
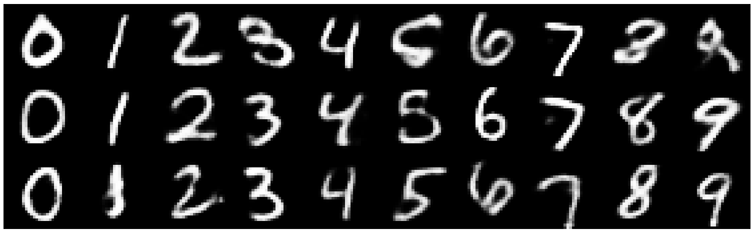


Figure 11: MNIST generated after 30 epochs on $e^{14} \times h^{14} \times s^4$ with fixed curvature



5 Future Work

Experimentally, we would like to extend our work to more geometric data, such as robotic movements that would incorporate Lie groups like $\mathrm{SO}(n)$ and $\mathrm{SE}(n)$ into the product and point clouds of various objects, as we believe it would strengthen the use of manifolds and more complex geometry. Theoretically, our current understanding of curvature is limited to sectional curvature, which can be seen as a function over 2 dimensional subspaces of a tangent place; however, we would like to explore other curvatures like Ricci curvature to get better geometrical understanding of a manifold. Our current method takes a predefined components of a product manifold and learns their curvature value, but it would be interesting to consider how one could learn those values as value, not just as hyperparameters. Investigations into the geometry of the latent space [Chadebec and Allassonnière, 2022] [Manoj, 2018] as well as the dataset [Khrulkov and Oseledets, 2018] are promising ways we can harness the power of more general geometry and, ideally, to learn an optimal low dimensional manifold for the data.

6 Conclusion

We investigate the hypothesis that the quality of representation would be better if the geometry of the embedding space better matches the data. We studied Riemannian diffusion models in a flexible curvature environment and introduced a new architecture incorporated with autocoders to obtain latent Riemannian diffusion models to reduce computational time while maintaining similar quality results. In exploring how to have an adaptable product manifold, we are able to learn the curvature given the dimensionality of the components of the product manifold and produce stronger results on higher dimensional data, like images.

References

C. Chadebec and S. Allassonnière. A geometric perspective on variational autoencoders, 2022. URL https://arxiv.org/abs/2209.07370.

R. Child. Very deep vaes generalize autoregressive models and can outperform them on images, 2020. URL https://arxiv.org/abs/2011.10650.

P. Esser, R. Rombach, and B. Ommer. Taming transformers for high-resolution image synthesis, 2020. URL https://arxiv.org/abs/2012.09841.

A. Gu, F. Sala, B. Gunel, and C. Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HJxeWnCcF7.

C.-W. Huang, M. Aghajohari, J. Bose, P. Panangaden, and A. Courville. Riemannian diffusion models. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=ecevn9kPm4.

V. Khrulkov and I. Oseledets. Geometry score: A method for comparing generative adversarial networks. In *International Conference on Machine Learning*, 2018.

Y. M. Lui. Advances in matrix manifolds for computer vision. *Image Vis. Comput.*, 30:380–388, 2012.

N. Manoj. On the geometry of the latent space of deep generative models. 2018.

A. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models, 2021. URL https://arxiv.org/abs/2102.09672.

A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever. Zero-shot text-to-image generation, 2021. URL https://arxiv.org/abs/2102.12092.

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models, 2021.

R. Senanayake and F. Ramos. Directional grid maps: modeling multimodal angular uncertainty in dynamic environments, 2018. URL https://arxiv.org/abs/1809.00498.

D. R. J. Shapovalov MV. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions, 2011.

O. Skopek, O.-E. Ganea, and G. Bécigneul. Mixed-curvature variational autoencoders. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1g6xeSKDS.