Variational Autoencoders with Riemannian Brownian Motion Priors: A Reproduction

Karol Jurasz [*]
University of Warsaw

Przemysław Pietrzak [†]
University of Warsaw

1 Intro

Variational Autoencoders (VAEs) [1] are a class of generative models based on encoding a sample from the modeled data distribution into a latent representation with an encoder network and decoding this latent representation back into the domain of the modeled data distribution using a decoder network. The model optimization objective is derived from the variational inference framework and requires the notion of an approximate posterior, which is optimized by the encoder network parameters, and prior distributions over the latent space. The prior acts as regularization on the latent structure; however, the training protocol does not guarantee that the modeled latent posterior aligns with the prior. In practice, this mismatch may result in assigning positive densities to regions in the latent space where the encoder does not allocate any samples, causing poor generative capabilities when sampling from the prior.

The latent space is often assumed to have an Euclidean structure, which is convenient because sampling from a Gaussian distribution in such a space can be efficiently performed, yet accuracy is not guaranteed. By assuming a Riemannian structure of the latent space, one can better capture its intrinsic geometry. Using appropriate sampling methods in this structured latent space can generate data more closely resembling the posterior distribution in terms of quality and similarity [2].

This study is a reproduction of [3]. The essential parts of our code, adapted from the original repository, are available at `https://github.com/knum-mimuw/Variational-Autoencoders-with-Riemannian-Brownian-Motion-Priors-A-Reproduction`. We adjusted the original codebase provided by the authors to support the reproduction process by adding run scripts, necessary utility functions, and notebooks to facilitate reproduction, as well as refactoring the plotting utilities for ease of exploratory analysis and improved computational efficiency.

2 Method

The main contribution of the reproduced paper concerns sampling in the latent which is more aligned with the actual latent space geometry with use of a Brownian Motion process reliant on the Riemannian metric.

A Riemannian metric G_z is defined on the latent space Z and induced by the mean μ_θ and variance σ_θ derived from the encoder. The map

$$f : Z \to \mathbb{R}^M \times \mathbb{R}^M_+, \quad z \mapsto (\mu_\theta(z), \sigma_\theta(z))$$

where M is the dimensionality of the manifold, is an immersion that encodes the geometry of both mean and variance. The metric G_z is given by

$$G_z = J_\mu(z)^\top J_\mu(z) + J_\sigma(z)^\top J_\sigma(z),$$

where J_μ and J_σ are the Jacobians of μ_θ and σ_θ. This metric ensures that geodesics in the latent space follow high-density regions, penalizing paths through low-density areas due to the variance term $J_\sigma^\top J_\sigma$ [2].

To avoid costly computations, Brownian motion is used for sampling. In this approach, Brownian motion on a Riemannian manifold is described by a stochastic differential equation, and its density function $p(z)$ corresponds to the heat kernel of the manifold. This density is approximated with use of a zeroth-order expansion, resulting in the following expression:

$$p(z) \approx (2\pi t)^{-d/2} H_0 \exp\left(-\frac{l^2(z,\mu)}{2t}\right),$$

where t is the duration of the Brownian motion, d is the dimensionality of the latent space, and $l(z,\mu)$ is the geodesic distance between z and the center μ of the motion. The term H_0 is a normalization factor that compensates for the differing Riemannian volume measures at the points z and μ, defined as:

$$H_0 = \left(\frac{\det G_z}{\det G_\mu}\right)^{1/2},$$

where G_z and G_μ are the metric tensors at z and μ, respectively. Equipped with this one can choose a starting point on latent space from which the Brownian process will be run.

[*] Worked on: mathematical background, geodesic lines analysis, anomalies analysis, comparisons with original paper, text preparation.

[†] Worked on: codebase adaptation, experiment running, plotting function extension, text preparation.

3 Experiments & Results

We performed a reproduction of the following elements of the original paper: (1) training of multiple VAE and RVAE models on the FMNIST dataset, (2) comparison of the loss performance between VAE and RVAE, (3) comparison of the estimations of the latent informativeness as measured by the classifiers ability to separate classes, (4) examples of geodesic and linear interpolation comparison pointing out successful and failure cases, (5) examples of random process trajectories providing samples from the Brownian motion prior. In experiments we left the default hyperparameters unchanged and manipulated the latent space dimensionality only. We ran less experiments due to computational constraints.

	Neg. ELBO	Rec	KL
VAE			
d = 2	-591.89$_{\pm 0.94}$	-597.57$_{\pm 0.94}$	5.67$_{\pm 0.00}$
d = 5	-671.44$_{\pm 2.34}$	-678.71$_{\pm 2.40}$	7.26$_{\pm 0.05}$
d = 10	-667.06$_{\pm 6.03}$	-674.40$_{\pm 6.19}$	7.34$_{\pm 0.15}$
$\mathcal{R}$-VAE			
d = 2	-934.47$_{\pm 6.81}$	-948.67$_{\pm 6.52}$	14.19$_{\pm 0.29}$
d = 5	-1117.89$_{\pm 1.15}$	-1154.82$_{\pm 21.97}$	24.29$_{\pm 4.02}$
d = 10	-1236.25$_{\pm 16.74}$	-1282.90$_{\pm 10.58}$	51.85$_{\pm 8.90}$

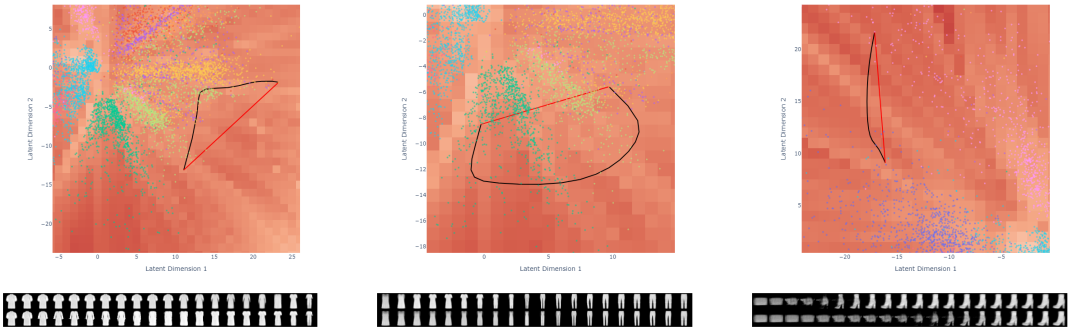
Table 1: Results on FMNIST (mean & std deviation over 10 runs). *Rec* denotes the negative conditional likelihood.

	0	1	2	3	4	5	6	7	8	9	Avg
VAE											
d = 2	0.71	0.93	0.56	0.68	0.38	0.84	0.16	0.82	0.94	0.89	0.69$_{\pm 0.25}$
d = 5	0.77	0.97	0.66	0.82	0.67	0.89	0.44	0.88	0.94	0.90	0.79$_{\pm 0.16}$
d = 10	0.77	0.96	0.70	0.83	0.74	0.92	0.52	0.90	0.95	0.92	0.82$_{\pm 0.13}$
$\mathcal{R}$-VAE											
d = 2	0.71	0.93	0.56	0.68	0.37	0.83	0.21	0.82	0.94	0.89	0.69$_{\pm 0.24}$
d = 5	0.77	0.97	0.66	0.82	0.67	0.89	0.42	0.88	0.94	0.90	0.79$_{\pm 0.16}$
d = 10	0.78	0.96	0.70	0.83	0.74	0.92	0.52	0.90	0.95	0.92	0.82$_{\pm 0.14}$

Table 2: Per class and average F1 score averaged over 5 classifiers on the latent representations derived from different models. The classifier is a MLP with hidden dimension 20 and ReLU.

We show the losses for different models in Table 1. The reconstruction is significantly better for the $\mathcal{R}$-VAE. Table 2 provides the outcomes of the classification based on the latent embeddings derived from VAE and $\mathcal{R}$-VAE. We do not observe superiority of $\mathcal{R}$-VAE in terms of classification metrics for our experimental setup.

Qualitative results regarding sampling from the prior using Brownian motion are presented in the Figure 1 showcasing a comparison of the geometry-informed interpolations and Figure 2 showing sampling from the Brownian motion prior. The background heatmap reflects the value of the Riemannian metric at a specific point.



(a) (b) (c)

Figure 1: Comparison of three estimates of geodesic lines with respect to Riemannian metric (black) with Euclidean line (red) for pairs of points from validation set. Below a decoded images represent changes in decoded images while traversing through Riemannian geodesics (first row) and changes throughout the Euclidean line (second row).

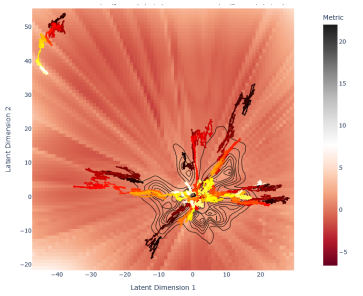


Figure 2: Example Brownian motion process trajectories (temperature corresponds to the step of the process) in the latent space with contours of the kernel density estimation.

In Figure 1 we present examples of geodesic lines estimation compared with Euclidean lines for different pairs of points in latent presenting examples where using geodesic interpolation provides visibly better results. In (a) geodesic line traverses through regions with higher densities, which results in images better resembling FMNIST images than their counterparts generated from points in latent space lying on Euclidean line and corresponds to the theoretically expected properties of geodesic lines. This is not the case in (b) and (c), where geodesics seem to be suboptimal; however, note that in (b) images taken from points lying on geodesic line are better than ones from Euclidean line and close to original ones from FMNIST. In extreme case (c) for points lying on the border of regions with positive densities on the latent space, both sets of images are of low quality, yet still images generated by points taken from geodesic line yield better images, which could undergo further numerical investigation. We point out the discrepancy between the intuitively expected shape of geodesic lines following the regions of locally high density between points [2] lying in the regions of small density and the estimated ones, which may be connected with empirical inefficiency of the method in areas of small densities and should undergo further numerical studies including curve energy computation.

In Figure 2 we present Brownian motion processes starting from chosen points in the latent space. It is worth noting that processes traverse through regions with high densities. This method enables sampling from the prior distribution when the processes are ran long enough.

References

[1] Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, and Pieter Abbeel. Variational lossy autoencoder, 2017.

[2] Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent space oddity: on the curvature of deep generative models, 2021.

[3] Dimitris Kalatzis, David Eklund, Georgios Arvanitidis, and Søren Hauberg. Variational autoencoders with riemannian brownian motion priors, 2020.