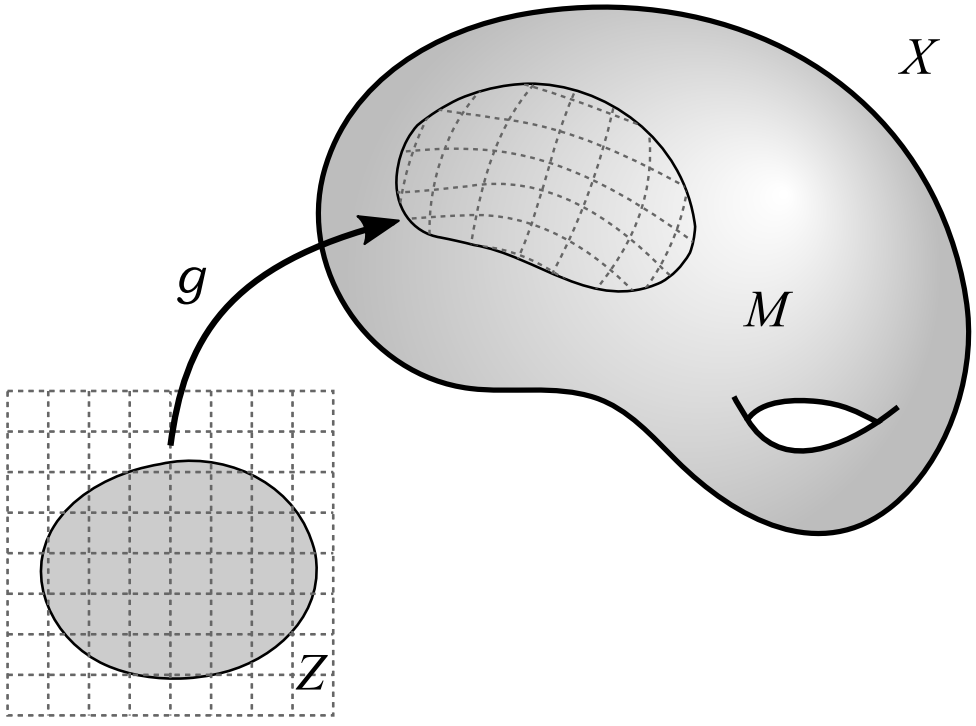
ICML
International Conference
On Machine Learning

Hessian Geometry of Latent Space in Generative Models.

Alexander Lobashev, Dmitry Guskov, Maria Larchenko, Mikhail Tamm.

June 16, 2025

Generative modeling: Manifold Hypothesis

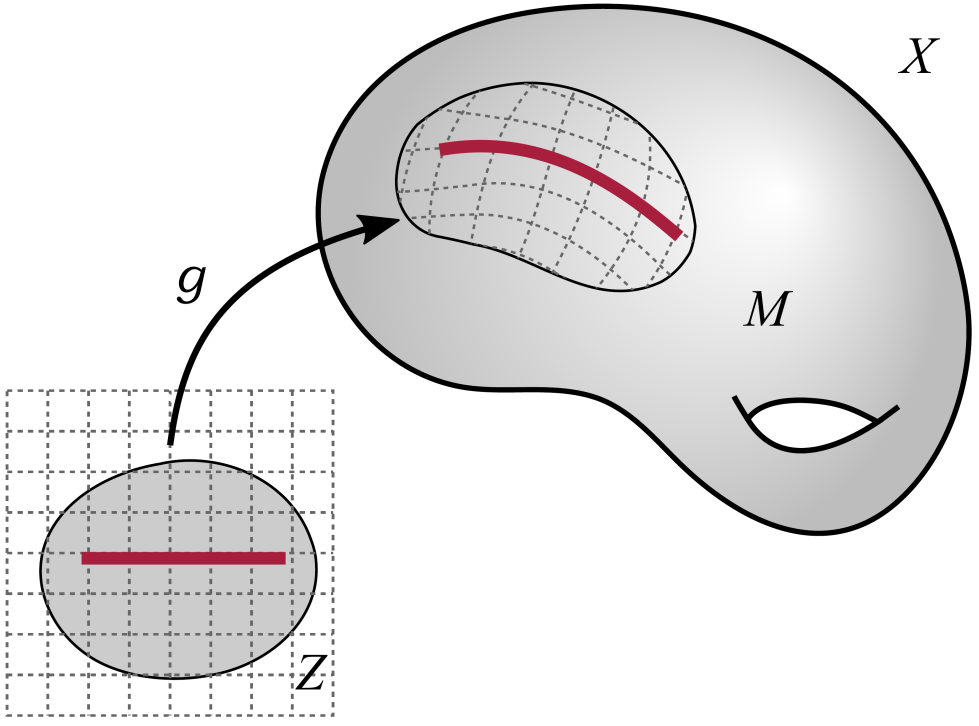


- X - pixel space
- M - lower-dimensional manifold of images
- Z - latent space
- $g : Z \rightarrow M$ - generative mapping

Adapted from: Shao, Hang, Abhishek Kumar, and P. Thomas Fletcher. **The Riemannian Geometry of Deep Generative Models**. CVPR Workshops, 2018

Metric and geodesic curve

Length of a curve:



Generative mapping g induces a metric $\mathfrak{g}$ and defines a curve length in data space. **Geodesic curve** is a curve in the minimal length between two points in data space.

Deterministic generative models

Geodesic curves in generative models:

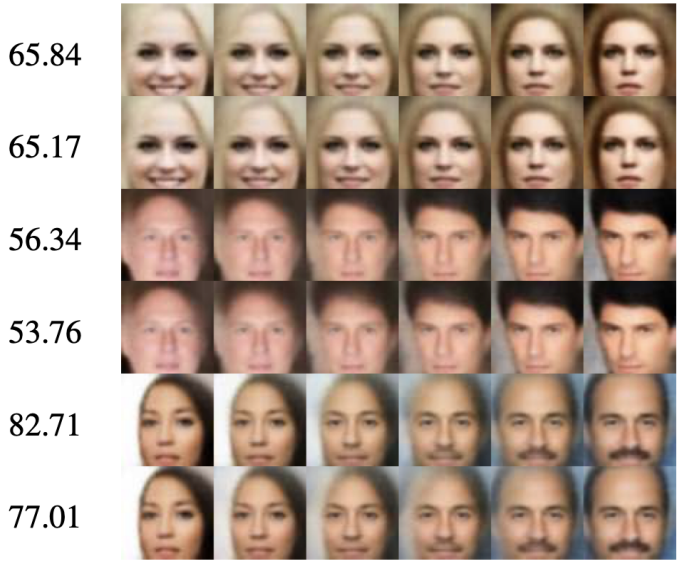


Figure 3: Linear and geodesic interpolation results for CelebA dataset. Rows 1, 3, 5: linear; Rows 2, 4, 6: geodesic; Column 1: arc length.

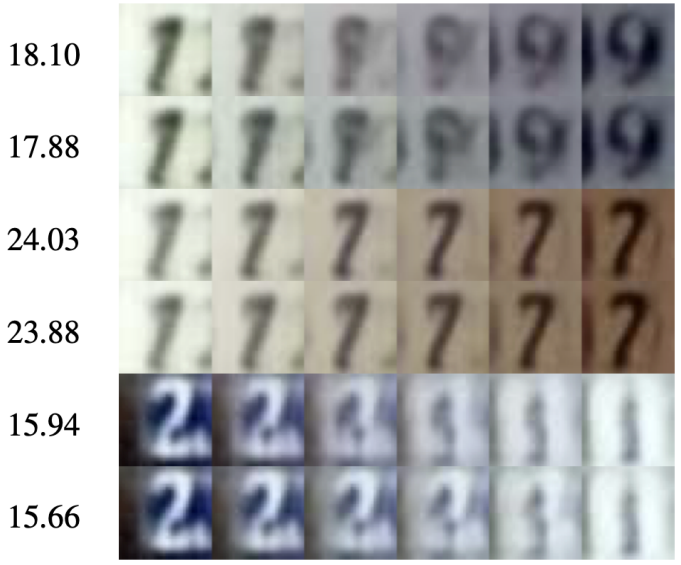
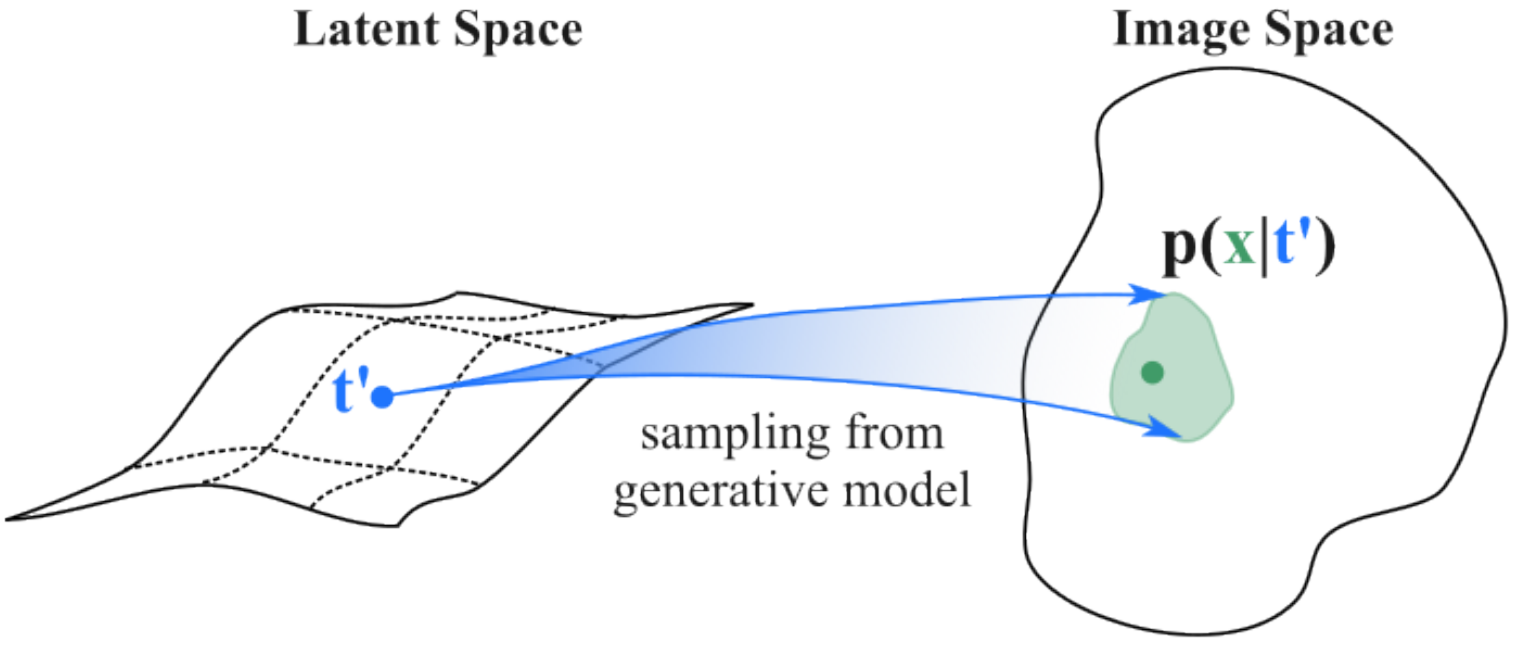


Figure 4: Linear and geodesic interpolation results for SVHN dataset. Rows 1, 3, 5: linear; Rows 2, 4, 6: geodesic; Column 1: arc length.

In the works (Shao et. al., 2018) (Wang & Ponce, 2021) it was observed that **geodesics are close to the linear interpolation** in the latent space of a variational autoencoder (VAE) and Generative Adversarial Networks (GANs). This means implicitly learn almost flat geometry.

Stochastic generative models

The key feature of diffusion models is a stochastic sampling.

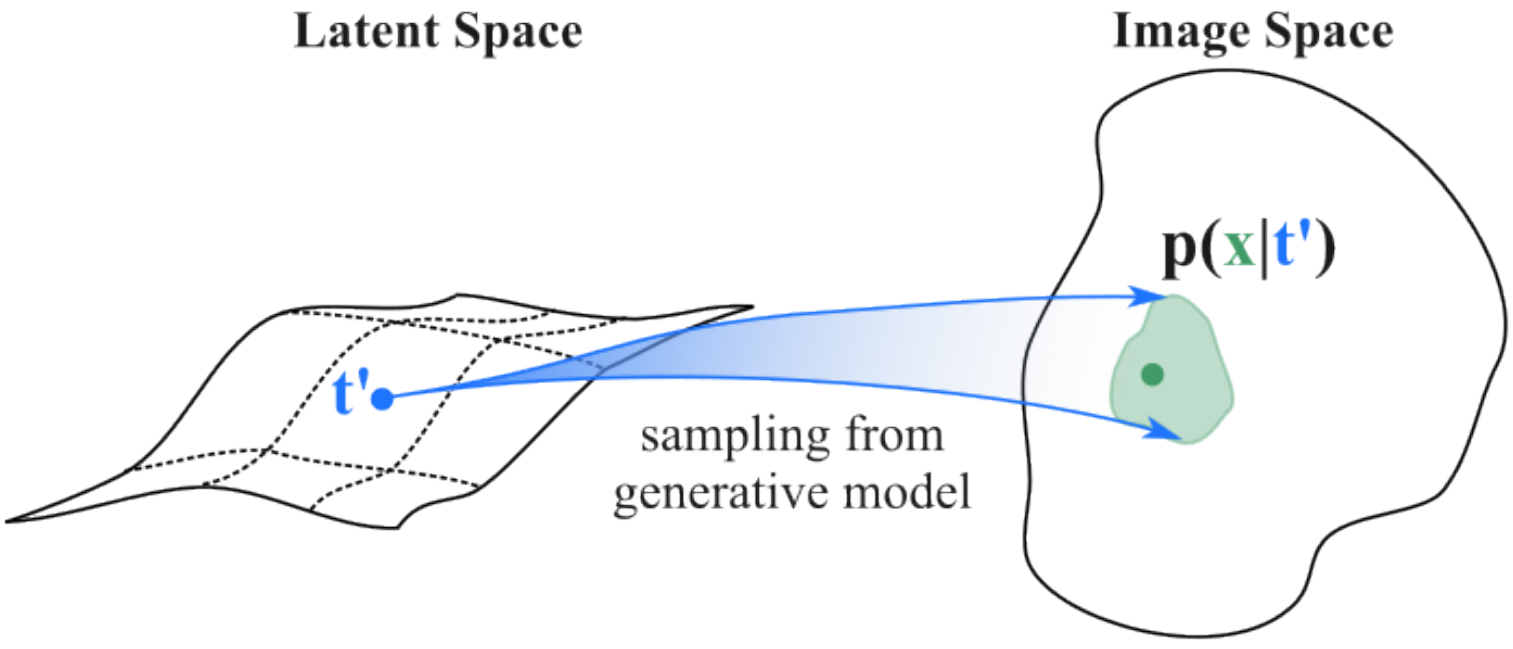


$g : Z \to M \subset X$ - generative mapping is now stochastic and has form of a conditional probability distribution $p(x|t')$, $t' \in Z$.

Previous approaches for metric estimation cannot be applied for stochastic generation.

Stochastic generative models

Stochastic generation:



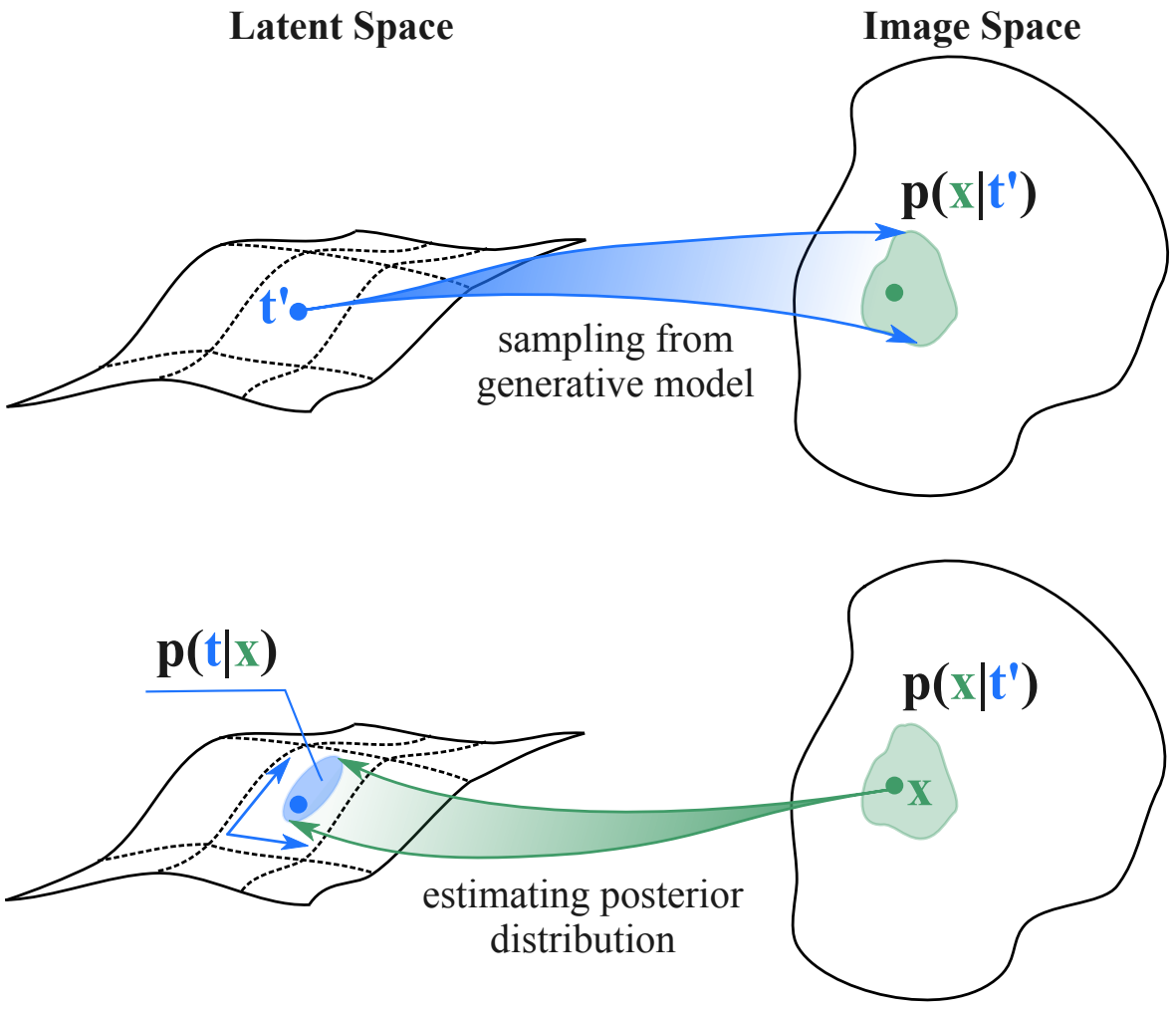
For stochastic generative models natural metric is Fisher information metric.

Fisher metric for a distribution $p(x|t)$ is defined as

$$g_F(t) = \int_{\mathcal{X}} p(x|t) \nabla_t \log p(x|t) (\nabla_t \log p(x|t))^T dx \tag{1}$$

How to obtain a metric for a generative model?

We estimate the metric from the samples by training convex MLP.



Exponential family distributions

We assume that $p(x|t)$ approximates by exponential family

$$p(x|t) = e^{\langle f(x),t\rangle - \log Z(t)}, \tag{2}$$

where the partition function $Z(t)$ is given by

$$Z(t) = \int_{\mathcal{X}} e^{\langle f(x),t\rangle} dx. \tag{3}$$

For exponential families their Fisher metric is a Hessian of $\log Z$:

$$g_F(t) = \sum_{i,j=1}^{N} \frac{\partial^2 \log Z(t)}{\partial t^i \partial t^j} dt^i dt^j = \nabla^2 \log Z(t) \tag{4}$$

How to obtain Hessian metric for a generative model?

Theorem 1 (Lobashev et. al., ICML 2025) Let $\mathcal{X}$ be a space of data samples $x \in \mathcal{X}$, and $S \subset \mathbb{R}^n$ be a compact domain with the continuous prior distribution $p(t)$ supported on S. Suppose the conditional distribution of data samples given parameter t is an exponential family

$$p(x|t) = e^{\langle t, f(x)\rangle - \log Z(t)}, \tag{5}$$

where

$$Z(t) = \int_{\mathcal{X}} e^{\langle t, f(x)\rangle} dx \tag{6}$$

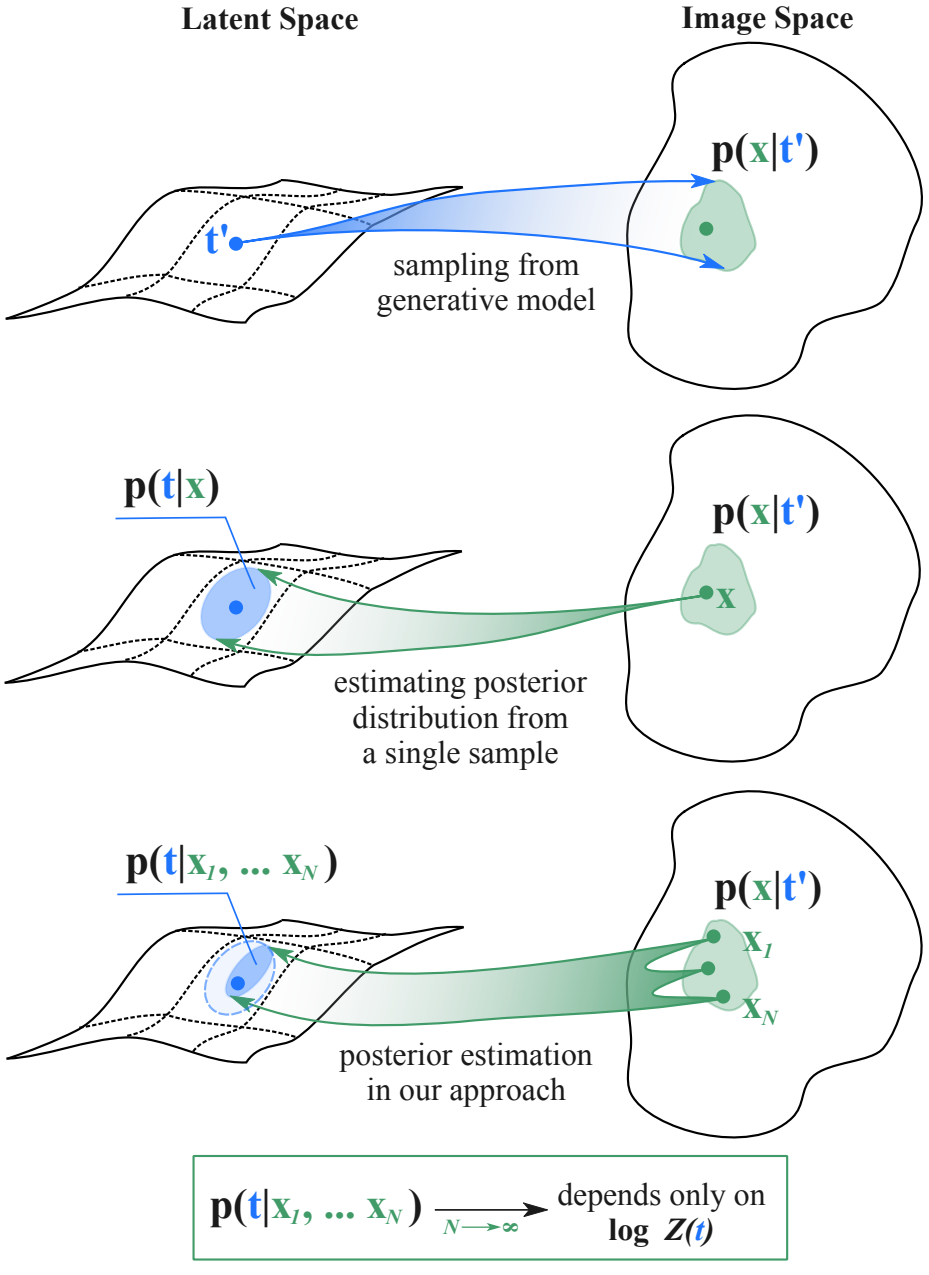
converges for all $t \in S$. Let $x_1, \ldots, x_N \sim p(x|t')$. Then, as $N \to \infty$ the posterior distribution satisfies:

$$\lim_{N\to\infty} \left(p(t|x_1, \ldots, x_N)\right)^{1/N} \overset{\text{a.s.}}{=} e^{-D_{\log Z(t)}(t,t')} \tag{7}$$

where $D_{\log Z(t)}(t, t')$ is the Bregman divergence between exponential family distributions

$$\begin{aligned} &D_{\log Z(t)}(t, t') = \\ &= \log Z(t) - \log Z(t') - \langle \nabla_{t'} \log Z(t'), t - t' \rangle \end{aligned} \tag{8}$$

How to obtain Hessian metric for a generative model?



Hessian geometry of latent diffusion models

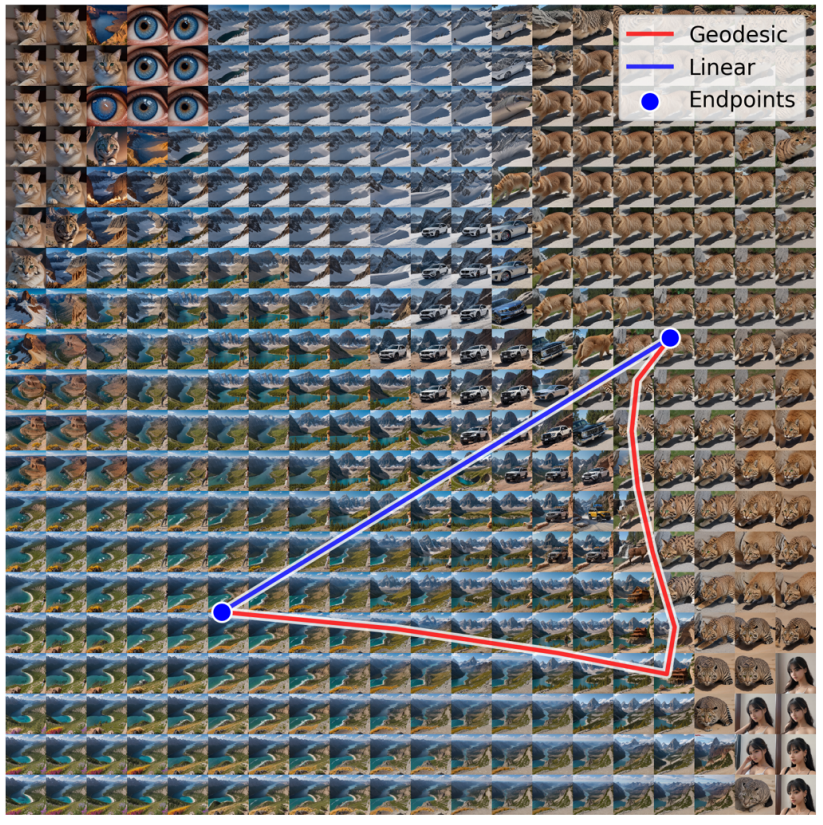
The experiments with diffusion are based on StableDiffusion 1.5 (Dreamshaper8) with DDIM scheduler. For our generation we use 50 inference steps, classifier free guidance scale set to 5. Prompt is chosen as “High quality picture, 4k, detailed” and negative prompt “blurry, ugly, stock photo”.

To build a 2 dimensional latent space section of the diffusion model we generate 3 random initial latents z_0, z_1, z_2. We use interpolation between latent representations:

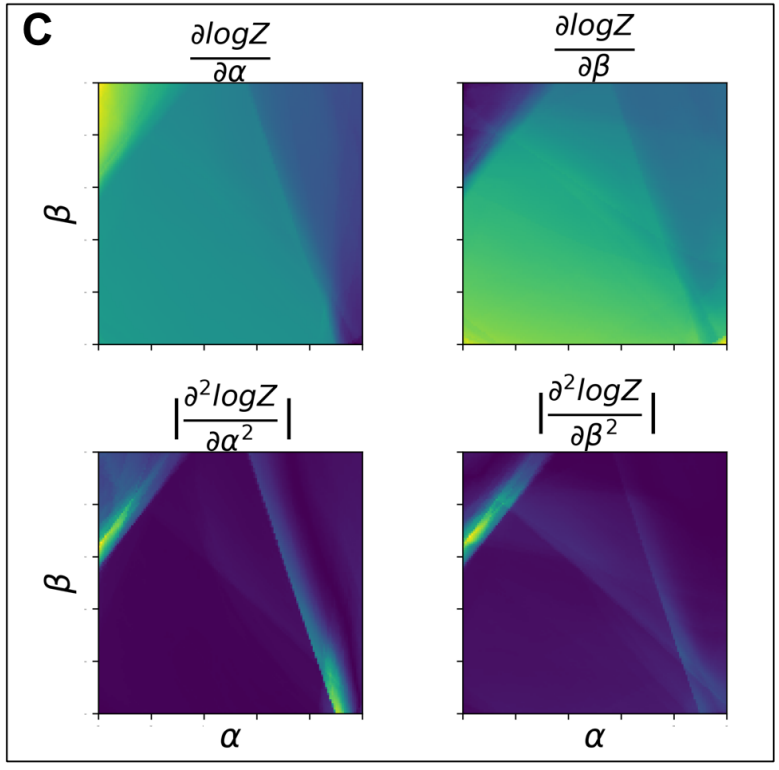
$$z = z_0 + \alpha(z_1 - z_0) + \beta(z_2 - z_0),$$

where α and β are uniformly sampled from $U[0, 1]$.

Geodesic (ours) and linear interpolation between images

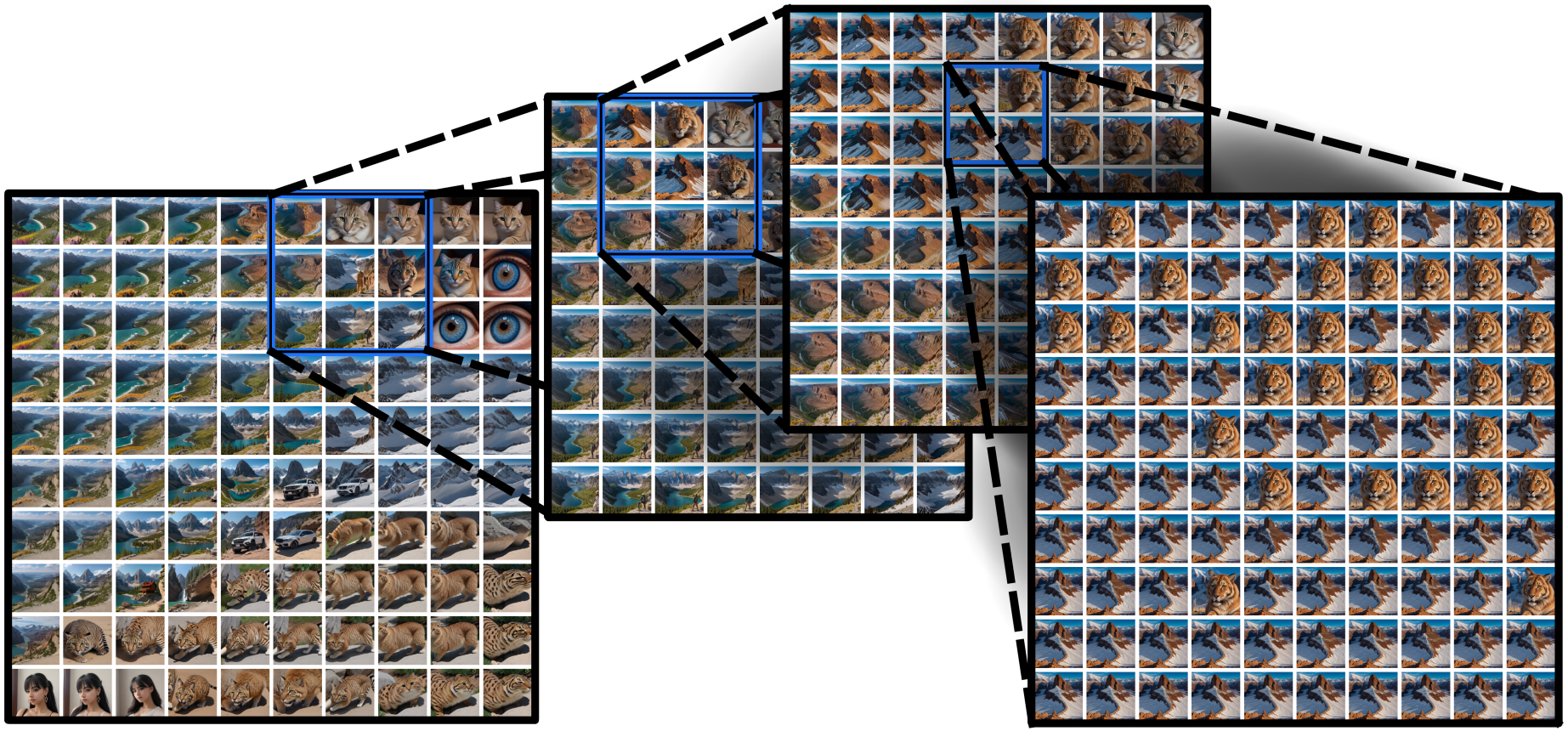


Inside a phase the geodesic
is approximately straight



First and second order derivatives
of log Z indicate phase boundaries

The fractal structure of phase boundary



The fractal structure of phase boundary in the interpolation landscape of diffusion model. The last plot represents the parameter variations 10^{-5} between neighboring images that cause the switch between a mountain and a lion.

The fractal structure of phase boundary

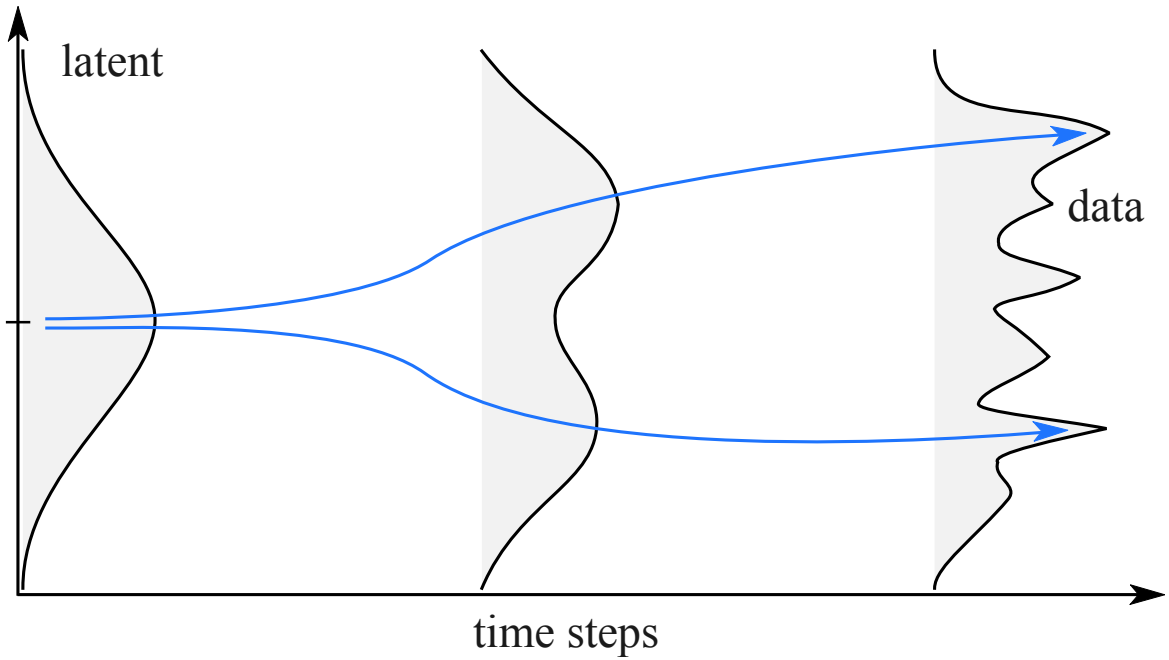


Figure 1: In diffusion models, generation begins from a high-dimensional Gaussian latent distribution. The reverse ODE process maps this distribution onto disjoint, lower-dimensional manifolds corresponding to distinct image modes. Such a transformation—from a unimodal latent space to a multimodal data space with disjoint supports—may result in a diverging Lyapunov exponent or, equivalently, a diverging Lipschitz constant in the generative mapping, indicating phase transitions.