# The Geometry of Diffusion Models: Tubular Neighbourhoods and Singularities

**Kotaro Sakamoto** [*1] **Ryosuke Sakamoto** [*2] **Masato Tanabe** [*2] **Masatomo Akagawa** [*2] **Yusuke Hayashi** [*3]
**Manato Yaguchi** [*1] **Masahiro Suzuki** [1] **Yutaka Matsuo** [1]

**Editors:** *S. Vadgama, E.J. Bekkers, A. Pouplin, S.O. Kaba, H. Lawrence, R. Walters, T. Emerson, H. Kvinge, J.M. Tomczak, S. Jegelka*

## Abstract

Diffusion generative models have been a leading approach for generating high-dimensional data. The current research aims to investigate the relation between the dynamics of diffusion models and the tubular neighbourhoods of a data manifold. We propose an algorithm to estimate the injectivity radius, the supremum of radii of tubular neighbourhoods. Our research relates geometric objects such as curvatures of data manifolds and dimensions of ambient spaces, to singularities of the generative dynamics such as emergent critical phenomena or spontaneous symmetry breaking.

## 1. Introduction

Generative modelling addresses the challenge of approximating and sampling from probability distributions. Some recent studies report that diffusion models, a class of generative models, exhibit critical phenomena during sampling where particular features of data emerge at the final stage of generation process. We delve into this symmetry breaking phenomena through a geometrical perspective. Our research begins with elucidating the connections between the diffusion and generation dynamics of diffusion models and the injectivity radius of tubular neighbourhoods of a given data manifold. The injectivity radius of a given data manifold, a geometrically crucial parameter, dictates the supremum extent to which a neighbourhood of the manifold succeeds to be without singularities (critical loci and self-intersections).

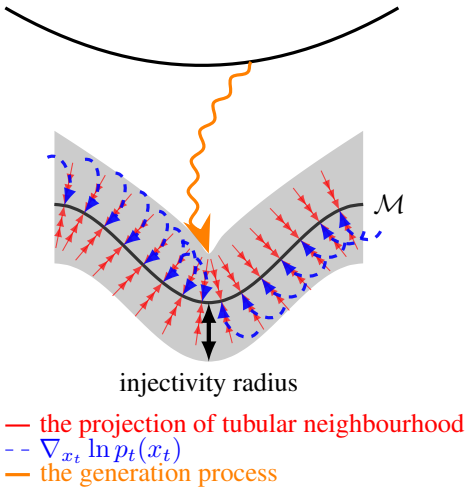



*Figure 1.* Conceptual diagram of our perspective.

We also investigate the behaviour of the score vector at the edge of the injectivity radius. Finally, we speculate that the interplay between the injectivity radius, symmetry breaking, and singularities of the potentials in the Fokker–Planck equation not only influences the performance of diffusion models but also reveals deeper insights into the nature of generative processes in high-dimensional spaces, offering a theoretical bridge between diffusion models, statistical thermodynamics, and Hessian (information) geometry.

**Contributions.** The main results in this paper are as follows.

- We present a geometrical perspective of diffusion models to understand critical phenomena.
- For a given data manifold, we propose an algorithm to estimate the injectivity radius of the tubular neighbourhoods (Section 3).
- We examine behaviours of score vectors around tubular neighbourhoods (Section 4).
- The phenomenon of spontaneous symmetry breaking in the diffusion model is intricately associated with the singularities of noisy manifolds (Section 5).

[*]Equal contribution [1]School of Engineering, The University of Tokyo, 7-chōme-3-1, Hongō, Bunkyo City, Tokyo 113-8654, Japan [2]Department of Mathematics, Graduate School of Science, Hokkaido University, North 10, West 8, Kita-ku, Sapporo 060-0810, Japan [3]AI Alignment Network, 3-chōme-4-12, Higashi-Kanda, Chiyoda City, Tokyo 101-0031, Japan. Correspondence to: Kotaro Sakamoto <kotaro.sakamoto@weblab.t.u-tokyo.ac.jp>.

*Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM) at the 41st International Conference on Machine Learning*, Vienna, Austria. PMLR 251, 2024. 

## 2. Preliminaries

In this section, we briefly introduce some basic mathematical concepts related to the paper.

### 2.1. Diffusion models

In Song et al. (2021), score-matching (Hyvärinen, 2005) and diffusion-based (Sohl-Dickstein et al., 2015; Ho et al., 2020) generative models have been unified into a single continuous-time score-based framework where the diffusion is driven by a stochastic differential equation. This framework relies on Anderson's Theorem (Anderson, 1982), which states that under certain Lipschitz conditions on the drift coefficient $f : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$ and on the diffusion coefficient $g : \mathbf{R}^d \times \mathbf{R} \to \mathbf{R}^d \times \mathbf{R}^d$ and an integrability condition on the target distribution $p_0(x_0)$, a forward diffusion process governed by the SDE

$$dx_t = f_t(x_t)dt + g_t(x_t)dw_t \tag{1}$$

has a reverse diffusion process governed by the SDE

$$dx_t = -\left[f_t(x_t) - \frac{g_t(x_t)^2}{2}\nabla_{x_t} \ln p_t(x_t)\right] dt + g_t(x_t)dw_t, \tag{2}$$

where $w_t$ is a standard Wiener process in reverse time. We could derive that probability distribution $p_t(x)$ of SDE satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial t}p_t(x) = -\nabla_x \cdot (p_t(x)f_t(x_t)) + \frac{1}{2}\Delta_x \left[g_t(x_t)^2 p_t(x_t)\right]. \tag{3}$$

Diffusion models are trained by approximating the score function $\nabla_x \ln p_t(x_t)$ with a neural network $s_\theta(x_t, t)$.

### 2.2. From the Manifold Hypothesis to Tubular Neighbourhoods

Data often concentrates around a lower-dimensional manifold, a concept known as the manifold hypothesis (Fefferman et al., 2013; Loaiza-Ganem et al., 2024). We work in this paper based on this hypothesis. For simplicity, we will assume all data manifolds are compact and embedded in the Euclidean space $\mathbf{R}^d$. In principle, any Riemannian manifolds can be isometrically embedded into some Euclidean space (the Nash embedding theorem).

A tubular neighbourhood of a manifold is roughly speaking a set of points near the manifold and every point of the set has a unique projection onto it (see Appendix C.3 for the formal definition). It is theoretically known that every manifold embedded in $\mathbf{R}^d$ has a tubular neighbourhood. In fact if we take a sufficiently small neighbourhood of a manifold, we may find a tubular neighbourhood. On the other hand, it is easy to imagine that we cannot take a too large neighbourhood as a tubular neighbourhood. See also Appendix A for previous studies which inspired our perspective.

## 3. Injectivity radius of a data manifold

In this section, we present how to estimate the supremum of radii of tubular neighbourhoods — the *injectivity radius* — of a given data manifold. Based on the theoretical argument in below, we establish the algorithm for the estimation (see Algorithm 1 in Appendix F). Throughout this section, let $\mathcal{M}$ denote an $n$-dimensional manifold (data manifold) in the Euclidean space $\mathbf{R}^d$. For the terminologies concerned with Manifold Theory, see Appendices C.2 and C.3.

We refer to (Litherland et al., 1999) for some notions and the case where $(n, d) = (1, 3)$, i.e., the manifold $\mathcal{M}$ is a *knot*. The first crucial claim of this section is that many theoretical facts proven in their paper work for general dimensions as well. The second claim is that the quantities appearing in their paper can be estimated from a given data cloud and its data manifold. For simplicity, we will explain the former briefly and focus on the latter.

### 3.1. Endpoint maps and Tubular neighbourhoods

We explain how to realise a tubular neighbourhood of a manifold embedded in the Euclidean space.

**Definition 3.1.** The *$\epsilon$-neighbourhood* of $\mathcal{M}$ in $\mathbf{R}^d$ is the set

$$\mathcal{M}(\epsilon) = \bigcup_{\boldsymbol{x}\in\mathcal{M}} \{\boldsymbol{y} \in \mathbf{R}^d \mid \|\boldsymbol{y} - \boldsymbol{x}\| < \epsilon\}.$$

**Definition 3.2.** The *normal bundle* to $\mathcal{M}$ in $\mathbf{R}^d$ is the set

$$N\mathcal{M} = \{(\boldsymbol{x}, \boldsymbol{v}) \in \mathbf{R}^d \times \mathbf{R}^d \mid \boldsymbol{x} \in \mathcal{M}, \boldsymbol{v} \perp T_{\boldsymbol{x}}\mathcal{M}\},$$

where $T_{\boldsymbol{x}}\mathcal{M}$ denotes the tangent space to $\mathcal{M}$ at $\boldsymbol{x}$.

Notice that the set $N\mathcal{M}$ forms a $d$-dimensional manifold. (The dimensions in the direction to $\mathcal{M}$ and its normal are $n$ and $d - n$, respectively.)

**Definition 3.3.** Consider the summation map

$$E_0 \colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d, \quad (\boldsymbol{x}, \boldsymbol{v}) \mapsto \boldsymbol{x} + \boldsymbol{v}.$$

We call its restriction

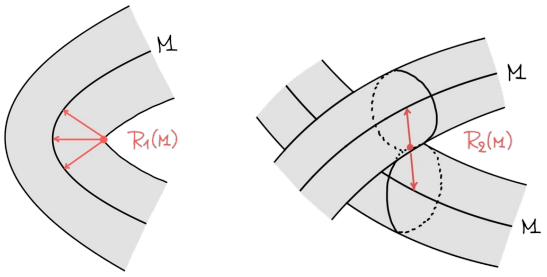
$$E = E_0|_{N\mathcal{M}} \colon N\mathcal{M} \to \mathbf{R}^d, \quad (\boldsymbol{x}, \boldsymbol{v}) \mapsto \boldsymbol{x} + \boldsymbol{v}$$

the *endpoint map* (or the *exponential map*).

**Proposition 3.4.** *Let $\epsilon > 0$ and consider the subset*

$$N\mathcal{M}_\epsilon = \{(\boldsymbol{x}, \boldsymbol{v}) \in N\mathcal{M} \mid \|\boldsymbol{v}\| < \epsilon\} \subset N\mathcal{M}.$$

*Then the image of $N\mathcal{M}_\epsilon$ under the endpoint map $E$ coincides with the $\epsilon$-neighbourhood $\mathcal{M}(\epsilon)$ of $\mathcal{M}$ in $\mathbf{R}^d$. Furthermore, this image forms a tubular neighbourhood of $\mathcal{M}$ if and only if the map $E|_{N\mathcal{M}_\epsilon}$ is an embedding.*




(a) $R_1(\mathcal{M})$ (b) $R_2(\mathcal{M})$

*Figure 2.* First and second injectivity radii

*Proof.* See the proof of Theorem C.11. □

### 3.2. Injectivity radius and its estimation

We consider the following three quantities.

**Definition 3.5.** (0) The *injectivity radius* $R(\mathcal{M})$ of $\mathcal{M}$ is the supremum of numbers $\epsilon > 0$ such that the $\epsilon$-neighbourhood of $\mathcal{M}$ in $\mathbf{R}^d$ is also a tubular neighbourhood. If such $\epsilon$ does not exist, define $R(\mathcal{M}) = 0$.

(1) The *first injectivity radius* $R_1(\mathcal{M})$ of $\mathcal{M}$ is the infimum of the set

$$\left\{ \|\boldsymbol{v}\| \;\middle|\; \begin{array}{l} (\boldsymbol{x},\boldsymbol{v}) \in N\mathcal{M} \text{ is} \\ \text{a critical point of the map } E \\ \text{for some point } \boldsymbol{x} \in \mathcal{M} \end{array} \right\}.$$

(2) The *second injectivity radius* $R_2(\mathcal{M})$ of $\mathcal{M}$ is the infimum of the set

$$\left\{ \frac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \;\middle|\; \begin{array}{l} \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2, \\ \boldsymbol{x}_1 - \boldsymbol{x}_2 \perp T_{\boldsymbol{x}_1}\mathcal{M}, \\ \text{and } \boldsymbol{x}_1 - \boldsymbol{x}_2 \perp T_{\boldsymbol{x}_2}\mathcal{M} \end{array} \right\}.$$

Roughly saying, $R_1(\mathcal{M})$ is the radius that the endpoint map fails to be regular at some point; $R_2(\mathcal{M})$ is the radius that two separated tubes touch each other (see Figure 2).

Thanks to the following assertion, it suffices to estimate $R_1(\mathcal{M})$ and $R_2(\mathcal{M})$.

**Theorem 3.6.** $R(\mathcal{M}) = \min\{R_1(\mathcal{M}), R_2(\mathcal{M})\}$.

*Proof.* See §2 of (Litherland et al., 1999). □

In this paper, the estimation of $R_2(\mathcal{M})$ is performed by definition. See Appendix D.3 for some ideas which may make the estimation easier. Therefore we here argue how to estimate $R_1(\mathcal{M})$. It is simple if we consider the case that $n = 1$ — the manifold $\mathcal{M}$ is a curve in $\mathbf{R}^d$ (see Appendix D.2); in general case, it seems to be difficult. However we show the following (see also Theorem C.7).

**Theorem 3.7.** *Assume that the manifold $\mathcal{M} \subset \mathbf{R}^d$ is expressed by $\mathcal{M} = \boldsymbol{F}^{-1}(\boldsymbol{0}) = \{\boldsymbol{x} \in \mathbf{R}^d \mid \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}\}$, where $\boldsymbol{F}\colon \mathbf{R}^d \to \mathbf{R}^{d-n}$ is a differentiable map of which $\boldsymbol{0} \in \mathbf{R}^k$ is a regular value. In addition, assume that we have vector fields $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_N$ $(n \leq N)$ defined near $\mathcal{M}$ such that for every $\boldsymbol{x} \in \mathcal{M}$ the vectors $\boldsymbol{t}_1(\boldsymbol{x}), \boldsymbol{t}_2(\boldsymbol{x}), \ldots, \boldsymbol{t}_N(\boldsymbol{x})$ are tangent to $\mathcal{M}$ and span the tangent space $T_{\boldsymbol{x}}\mathcal{M}$. Then the first injectivity radius $R_1(\mathcal{M})$ coincides with the infimum of the Euclidean norm $\|\boldsymbol{v}\|$ of the vector $\boldsymbol{v} \perp T_{\boldsymbol{x}}\mathcal{M}$ such that the $d \times (d + N - n)$-matrix*

$$L_{\mathcal{M}}(\boldsymbol{x},\boldsymbol{v}) = \left[ \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}(\boldsymbol{x})^T \left( \frac{\partial \varphi_1}{\partial \boldsymbol{x}}(\boldsymbol{x},\boldsymbol{v}) - \frac{\partial \varphi_1}{\partial \boldsymbol{v}}(\boldsymbol{x},\boldsymbol{v}) \right)^T \cdots \left( \frac{\partial \varphi_N}{\partial \boldsymbol{x}}(\boldsymbol{x},\boldsymbol{v}) - \frac{\partial \varphi_N}{\partial \boldsymbol{v}}(\boldsymbol{x},\boldsymbol{v}) \right)^T \right] \tag{4}$$

*is degenerate for some point $\boldsymbol{x} \in \mathcal{M}$, where*

$$\varphi_i\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}, \quad \varphi_i(\boldsymbol{x},\boldsymbol{v}) = \langle \boldsymbol{t}_i(\boldsymbol{x}), \boldsymbol{v} \rangle$$

*for $i = 1, 2, \cdots, N$.*

This assertion is proven by an application of the Method of Lagrange Multiplier. See Appendix D.1 for its precise proof. We here note some remarks.

**Remark 3.8.** The condition that the matrix $L(\boldsymbol{x},\boldsymbol{v})$ degenerates at $(\boldsymbol{x},\boldsymbol{v}) \in N\mathcal{M}$ is equivalent to that the determinant of the $d \times d$-minor

$$\left[ \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}(\boldsymbol{x})^T \left( \frac{\partial \varphi_{i_1}}{\partial \boldsymbol{x}}(\boldsymbol{x},\boldsymbol{v}) - \frac{\partial \varphi_{i_1}}{\partial \boldsymbol{v}}(\boldsymbol{x},\boldsymbol{v}) \right)^T \cdots \left( \frac{\partial \varphi_{i_n}}{\partial \boldsymbol{x}}(\boldsymbol{x},\boldsymbol{v}) - \frac{\partial \varphi_{i_n}}{\partial \boldsymbol{v}}(\boldsymbol{x},\boldsymbol{v}) \right)^T \right] \tag{5}$$

of $L(\boldsymbol{x},\boldsymbol{v})$ vanishes for every $n$-tuple $(i_1, \ldots, i_n)$ satisfying that $1 \leq i_1 < \cdots < i_n \leq N$. Indeed, the matrix $\dfrac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}(\boldsymbol{x})$ is of full-rank for every point $\boldsymbol{x} \in \mathcal{M} = \boldsymbol{F}^{-1}(\boldsymbol{0})$.

**Remark 3.9.** It is crucial to find vector fields $\boldsymbol{t}_i$ satisfying the above condition. For example, (small extensions of) the gradient vector fields $\boldsymbol{t}_i = \operatorname{grad} x_i$ $(i = 1, \ldots, d)$ satisfies the condition, where $x_i\colon \mathcal{M} \to \mathbf{R}$ denotes the projection to the $i$-th axis in $\mathbf{R}^d$. In general, we have to take the number $N$ greater than $n$.

### 3.3. Example (unit circle $S^1$)

Let us verify Theorem 3.7 through the most typical manifold — the unit circle $S^1$. Define a function $F\colon \mathbf{R}^2 \to \mathbf{R}$ by

$$F(x, y) = x^2 + y^2 - 1.$$

Then we have $S^1 = F^{-1}(0)$. One of the normal vector field on $S^1$ is given as $\operatorname{grad}(F) = (\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}) = (2x, 2y)$,

so $(-y, x)$ is a tangent vector field which spans the tangent space to $S^1$ at each point $(x, y) \in S^1$.

Applying Theorem 3.7, the first injectivity radius $R_1(S^1)$ is calculated as follows. For a point $(x, y) \in S^1$, the matrix

$$L_{S^1}((x, y), (v_1, v_2)) = \begin{bmatrix} 2x & v_2 + y \\ 2y & -v_1 - x \end{bmatrix}$$

is degenerate (i.e., its determinant is zero) if and only if $(v_1, v_2) = (-x, -y)$. Thus, we obtain

$$R_1(S^1) = \sqrt{(-x)^2 + (-y)^2} = 1.$$

By definition, $R_2(S^1)$ is also equal to 1, so the injectivity radius $R(S^1)$ is equal to 1.

### 3.4. A pilot numerical experiment to validate the algorithm

We perform a pilot experiment to verify the algorithm. The detailed setting and the results are present in the Appendix F.1. The estimated $R$ for $S^1$ is $0.999 \pm 0.006$.

## 4. Some observable relations between the behaviour of the score function and a tubular neighbourhood

As in the previous section, let $\mathcal{M}$ denote an $n$-dimensional manifold (data manifold) embedded in the Euclidean space $\mathbf{R}^d$. In this section, we discuss relation between the behaviour of the score function and the tubular neighbourhood.

### 4.1. Curvatures of the data manifolds and the score functions within the tubular neighbourhoods

Let us explain a property of the score vectors in a tubular neighbourhood of a manifold by an Example 4.1. It is observed that the score vectors $\nabla_x \ln p_t(x)$ within the tubular neighbourhood changes its direction in proportional to the curvature at the uniquely projected point $\pi(x)$. Similar phenomena are observed in prior research ((Sidorova et al., 2004)[Lemma 11]). In (Batzolis et al., 2022)[Theorem 5.1], the authors proved that when $t \to 0$ the score vectors $\nabla_x \ln p_t(x)$ converges to the normal space at $\pi(x)$ if $x$ is in the tubular neighbourhood of the data manifold. This mathematical fact explains that the score vectors change its direction proportional to the curvature if the Brownian motion has a component tangent to the manifold within the tubular neighbourhood when $\sigma_t \to 0$.

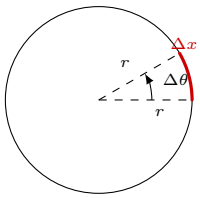



*Figure 3.* Arc segment

**Example 4.1.** Let us consider a circle of radius $r$ (see Figure 3). In this case we have

$$\lim_{t \to 0} \nabla_x \ln p_t(x) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{6}$$

for some $0 \leq \theta < 2\pi$. Here $x$ is a point in a tubular neighbourhood and $\theta$ is the angle of $\pi(x)$. Let $\Delta x$ be an infinitesimal arc length. Then we have $\Delta\theta = \frac{\Delta x}{r}$ by definition. Moreover, taking Taylor's series, we may compute the change rate tangent to the circle of the score vectors (6) as:

$$\lim_{\Delta x \to 0} \frac{1}{\Delta x} \begin{bmatrix} \cos(\theta + \Delta\theta) - \cos(\theta) \\ \sin(\theta + \Delta\theta) - \sin(\theta) \end{bmatrix} = \frac{1}{r} \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}.$$

### 4.2. Analysis of the behaviour of the score vector field at the boundary of the tubular neighbourhood

In (Batzolis et al., 2022)[Appendix. D], the authors re-expressed the score vectors as:

$$\nabla_x \ln p_t(x) = \frac{1}{\sigma_t^2 p_t(x)} \int_{\mathcal{M}} (y - x) N(y|x, \sigma_t^2 I) p_0(y) dy,$$

where $N(y|x, \sigma_t^2 I)$ is a normal distribution and $p_t(x) = \int_{\mathcal{M}} N(y|x, \sigma_t^2 I) p_0(y) dy$. $p_0$ is a smooth function on $\mathcal{M}$ such that $\int_{\mathcal{M}} p_0(y) dy = 1$. $dy$ is a volume form on $\mathcal{M}$ ($dy$ is a nowhere vanishing $n$-form on $\mathcal{M}$, $\int_{\mathcal{M}} dy = \text{Vol}(\mathcal{M})$). As pointed out by the authors, it means that the score is the weighted average of vectors pointing from $x$ to $y$ over all choices of points $y$ on the manifold, with weights given by $w := N(y|x, \sigma_t^2 I) p_0(y)$. One may notice from the expression (7) that if $x$ is far enough from the data manifold $\mathcal{M}$ and $\sigma_t$ is large, the score vectors points toward the centre of gravity. If $p_0(y)$ is symmetric with respect to $(0, 0)$, it is clear that if for example $\mathcal{M} = S^1$ in $\mathbf{R}^2$ with its centre at the origin and let $x = (0, 0)$. We find:

$$\nabla_x \ln p_t(x) = 0.$$

On the other hand, $(0, 0)$ is a point of the boundary of the tubular neighbourhood. We understand that the points that the score $\nabla_x \ln p_t(x)$ vanishes is important because the second term of (2) becomes dominant at this point. Therefore we presume the behaviour of the score vector field at the boundary of the tubular neighbourhood has a significant influence on a trajectory of the diffusion model.

**Conjecture 4.2.** *Suppose $\mathcal{M}$ is a compact oriented manifold embedded in $\mathbf{R}^d$. We predict the following observation: Let $\epsilon > 0$ be the injectivity radius. Let $\boldsymbol{n}$ be a unit outward pointing normal vector to $\partial\mathcal{M}(\epsilon)$. Assume $\epsilon > \sqrt{d}\sigma_t$, $x \in \partial\mathcal{M}(\epsilon)$ and $p_0(y)$ is constant $C$ greater than 0 on $\mathcal{M}$. Assume moreover the following conditions:*

- *(i) For any $y \in \mathcal{M}$ with $(y - x) \cdot \boldsymbol{n} > 0$, there exists $y' \in \mathcal{M}$ and some $c > 0$ such that $-c(y - x) = (y' - x)$.*
- *(ii) Assume that for each $y \in \mathcal{M}$ such that $(y - x) \cdot \boldsymbol{n} < 0$, there exists $\tilde{y} \in \mathcal{M}$ and $c > 0$ such that $-c(\tilde{y} - x) = (y - x)$. Then $c \leq 1$.*
- *(iii) For any $y \in \mathcal{M}$, $\{c(y - x) | c > 0\} \cap \mathcal{M}$ is a finite set.*

*Then:*

$$\nabla_x p_t(x) \cdot \boldsymbol{n} \leq 0.$$

The verification of this conjecture for the case $d = 2$ and $M$ being a curve can be found in Appendix G.2.

**Remark 4.3.** The condition $\epsilon > \sqrt{d}\sigma_t$ in Conjecture 4.2 tells us how to control $\nabla_x p_t(x)$. These three values determine the behaviour of directions of the score vectors at the boundary of the tubular neighbourhood. For example if the dimension $d$ is much larger than the radius $\epsilon$, you have to take very small $\sigma_t$ to have $\nabla_x p_t(x) \cdot \boldsymbol{n} < 0$. A related work can be found in (Chen et al., 2023b)[Proposition 2.].

### 4.3. Escaping time from the tubular neighbourhood

Let $\epsilon > 0$. Let $\mathcal{M}(\epsilon)$ be the $\epsilon$-neighbourhood of a compact oriented manifold $\mathcal{M}$ in the Euclidean space $\mathbf{R}^d$ as defined in Definition 3.1. Assume $\mathcal{M}(\epsilon)$ is a tubular neighbourhood. Suppose $p_t(x)$ is a smooth solution to the Fokker-Planck equation (3) with an initial condition $p_0(x) = \delta_{\mathcal{M}}(x)$ here $\delta_{\mathcal{M}}(x)$ is Dirac's density function with its support $\mathcal{M}$. We define a function $\Gamma_{\mathcal{M}(\epsilon)}(t)$ as follows:

$$\Gamma_{\mathcal{M}(\epsilon)}(t) := \int_{\mathcal{M}(\epsilon)} p_t(x) dx. \tag{7}$$

**Remark 4.4.** If the data manifold $\mathcal{M}$ is the $n$-sphere $S^n$ the blue lines in Subsection 6.2 represents the graphs of $\Gamma_{\mathcal{M}(\epsilon)}(t)$.

**Proposition 4.5.** *Assume $\beta(t) : \mathbf{R}_{\geq 0} \to \mathbf{R}$ is a smooth function and $f(t,x) = \beta(t)\frac{f(x)}{2}$, $g(t,x) = \sqrt{\beta(t)}$ in (3) ($f(x)$ is some smooth vector field). We have:*

$$\lim_{t\to 0} \frac{\partial}{\partial t}\Gamma_{\mathcal{M}(\epsilon)}(t) = 0 \text{ and } \lim_{t\to\infty} \frac{\partial}{\partial t}\Gamma_{\mathcal{M}(\epsilon)}(t) = 0.$$

*Thus there exists at least one $t_c$ in $(0, +\infty)$ such that $\frac{\partial^2}{\partial t^2}\Gamma_{\mathcal{M}(\epsilon)}(t_c) = 0$. Moreover if $\beta(t) > 0$ and*

$$(\nabla_x \ln p_t(x) - f(x)) \cdot \boldsymbol{n} < 0 \tag{8}$$

*for any $x \in \partial\mathcal{M}(\epsilon)$ and any $t \in \mathbf{R}_{>0}$ then $\Gamma_{\mathcal{M}(\epsilon)}(t)$ is strictly monotonically decreasing. Here $\boldsymbol{n}$ is a unit outward pointing normal vector field along $\partial\mathcal{M}(\epsilon)$.*

The verification of this can be found in Appendix G.3.

**Remark 4.6** (Returning time into the tubular neighbourhood)**.** Suppose $p_t(x)$ is a solution to the Fokker-Planck equation (12) associated to the reverse diffusion process with an initial value condition $p_0(x)$ equals to some simple distribution. We predict we may prove there exists $t_c$ such that the second derivative of

$$\tilde{\Gamma}_{\mathcal{M}(\epsilon)}(t) = \int_{\mathcal{M}(\epsilon)} p_t(x) dx$$

at $t_c$ vanishes in a similar way as above. We think that the main problem is an estimation of score vectors and we could make use of the fact like (Bortoli, 2022a)[Lemma C.1.].

## 5. Evolution of latent structures during diffusion processes

The geometry of diffusion models can be conceptualised as a set of noise manifolds spreading in layers around a data manifold. This involves the space of latent variables $x_t$ created by centring the data manifold and adding stochastic noise (See Figure 4). We first elucidate the relationship between the stochastic Riemannian metric of the noisy manifold and the number of particles within the tubular neighbourhood in Subsection 5.1. Subsequently, in Subsection 5.2, we recapitulate the perspective that the continuous-time diffusion model can be interpreted as an infinitely deep hierarchical $\beta$-VAE (Huang et al., 2021; Luo, 2022; Kingma & Gao, 2023), a type of variational autoencoder with an inverse temperature hyperparameter $\beta$ varying at each layer. We confirm that the equivalent of spontaneous symmetry breaking in the latent space of a particular layer of this hierarchical $\beta$-VAE occurs in a trained $\beta$-VAE by observing the training error of the trained $\beta$-VAE. In Subsection 5.3, we experimentally confirm that a phase transition occurs in the geometric structure of the latent space, using $\beta$-VAE with temperature parameters, corresponding to diffusion steps in the diffusion models.

### 5.1. Geometrical structure

Initially, we can deduce that the probability distribution $p_t(x_t)$ of the stochastic differential equation (SDE) of the diffusion process adheres to the Fokker-Planck equation (see Appendix H for details):

$$dx_t = f_t(x_t)dt + g_t dw_t, \tag{9}$$

$$\frac{\partial}{\partial t} p_t(x_t) = -\nabla_{x_t} \cdot \left[\nabla_{x_t} u_t(x_t) p_t(x_t)\right], \tag{10}$$

where

$$u_t(x_t) := \int_{x_0}^{x_t} f_t(z)dz - \frac{g_t^2}{2} \ln q_t(x_t). \tag{11}$$

Here, we introduce the potential function $u_t(x_t)$ (Raya & Ambrogioni, 2023). On the other hand, the probability distribution $q_t(x_t)$ of the SDE for the reverse diffusion process also adheres to the another Fokker-Planck equation:

$$dx_t = -\nabla_{x_t} u_t(x_t)dt + g_t dw_t, \tag{12}$$

$$\frac{\partial}{\partial t} q_t(x_t) = \nabla_{x_t} \cdot [\nabla_{x_t} f_t(x_t) q_t(x_t)] \tag{13}$$

with $q_t(x_t) = p_{T-t}(x_{T-t})$, where $T$ is the number of diffusion steps at which the diffusion process terminates (Franzese et al., 2023). Hence, the fixed point $\bar{x}_t$

of the potential function leads to the following relationship:

$$\nabla_{x_t} u_t(\bar{x}_t) = 0 \implies \frac{\partial}{\partial t} p_{\text{eq}}(\bar{x}_t) = 0, \tag{14}$$

$$\implies f_t(\bar{x}_t) = \frac{g_t^2}{2} \nabla_{x_t} \ln p_{\text{eq}}(\bar{x}_t). \tag{15}$$

where $p_{\text{eq}}(x_t)$ is the solution of $\frac{\partial}{\partial t} p_t(x_t) = 0$. This property is observed in the equilibrium thermodynamic formulation of diffusion models, previously discussed in the work of Ambrogioni (2023), where stochastic fluctuations of physical quantities can be ignored. Non-equilibrium thermodynamics (stochastic thermodynamics), on the other hand, where such stochastic fluctuations cannot be ignored, is given by $\frac{\partial}{\partial t} p_t(x_t) \neq 0$. In other words, the trajectory on the fixed points $\bar{x}_t$ consists only of states in which equilibrium thermodynamics holds.

Utilising the potential function $u_t(x_t)$, the second derivative of the function $\Gamma_{\mathcal{M}(\epsilon)}(t)$, as introduced in the section 4, can be expressed as follows (see Appendix H):

$$\frac{\partial^2}{\partial t^2} \Gamma_{\mathcal{M}(\epsilon)}(t) = \int_{\partial \mathcal{M}(\epsilon)} (2\Delta u_t(z) + \nabla_z u_t(z) \cdot \nabla_z \ln p_t(z)) \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z) dz. \tag{16}$$

As the diffusion step $t$ progresses, the latent space structured by the latent variable $x_t$ undergoes notable transformations. In the following, we explore these transformations, drawing on the approach of Arvanitidis et al. (2021), who analysed the geometric structure of the latent space in VAE through the Jacobian $J_{\beta_t}(x_t) := \nabla_{x_t} \ln p_t(x_t)$ and the stochastic Riemannian metric $G_{\beta_t}(x_t) := J_{\beta_t}^T(x_t) J_{\beta_t}(x_t)$.

The stochastic Riemannian metric, initially introduced for analysing the geometric structure of the latent space in VAEs, offers significant insights into the geometric configuration at various diffusion steps within diffusion models. Two typical drift terms are variance-preserving $f_t(x_t) = -\frac{g_t^2 x_t}{2}$ and variance-exploding $f_t(x_t) = 0$. Then, (Eq. 16) simplifies into the following tabular form (see Appendix H):

$$\nu_t(x_t) := \Delta u_t(z) - \frac{g_t^2}{4} \left( x_t \cdot \nabla_{x_t} \ln p_t(x_t) + G_{\beta_t}(z) \right), \tag{17}$$

$$\frac{\partial^2}{\partial t^2} \Gamma_{\mathcal{M}(\epsilon)}(t) = \int_{\partial \mathcal{M}(\epsilon)} 2\nu_t(x_t) \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z) dz. \tag{18}$$

From (Eq. 8), we have $\nabla_z u_t(x_t) \cdot \boldsymbol{n} > 0$. Therefore, the following relationship holds between the diffusion steps $t_c$ at which the second-order derivative of the function $\Gamma_{\mathcal{M}(\epsilon)}(t)$ vanishes:

$$\frac{\partial^2}{\partial t^2} \Gamma_{\mathcal{M}(\epsilon)}(t_c) = 0 \implies \int_{\partial \mathcal{M}(\epsilon)} \nu_t(z) \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z) dz = 0 \tag{19}$$

As stated in the previous study (Raya & Ambrogioni, 2023), there exists a specific diffusion time $t_c$ at which the second-order derivative of the potential function $\Delta u_t(x_t)$ becomes zero, indicating a moment when spontaneous symmetry breaking occurs within the latent space. This equation clearly elucidates that spontaneous symmetry breaking in the diffusion model is intricately linked to the singularities of the stochastic Riemannian metric of noisy manifolds.

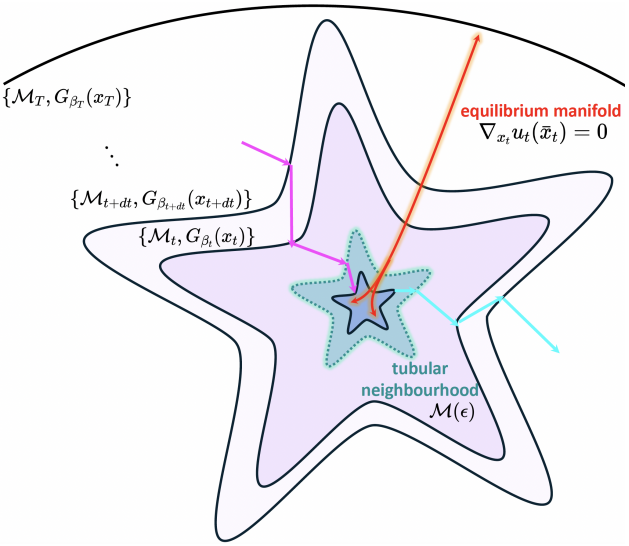



*Figure 4.* Geometric schematic of the diffusion process: The star-shaped figure at the center represents the data manifold. The diffusion process, where noise is incrementally added to the data, is depicted by light blue Brownian particles (arrows). Conversely, the reverse diffusion process, where noise is gradually removed from the data, is illustrated by magenta Brownian particles (arrows). At each diffusion step $t$, a noise manifold $\mathcal{M}_t$ is formed by the latent variable $x_t$ and its stochastic Riemannian metric $G_{\beta_t}(x_t)$.

### 5.2. Correspondence with $\beta$-VAE

As noted at the beginning of this section, the continuous-time diffusion model can be viewed as an infinitely deep hierarchical VAE. Furthermore, the dependence on the number of layers in each hierarchical VAE layer can be expressed by varying $\beta_t$ in the $\beta$-VAE. The relationship between the standard diffusion model and $\beta$-VAE can be described by

the following SDE, $g_t = \sqrt{\frac{2}{\beta_t}}$:

$$dx_t = f_t(x_t)dt + \sqrt{\frac{2}{\beta_t}}dw_t, \tag{20}$$

$$dx_t = \left[-f_t(x_t) + \frac{1}{\beta_t}\nabla_{x_t} \ln q_t(x_t)\right] dt + \sqrt{\frac{2}{\beta_t}}dw_t, \tag{21}$$

where the inverse temperature $\beta_t = \frac{1}{\sigma^2 t}$ decreases monotonically as the diffusion step $t$ proceeds. The objective function at each layer of the U-net in the diffusion model coincides with the objective function of the $\beta$-VAE. The diffusion and reverse diffusion processes are represented by the following probability model (Watanabe, 2010).

$$q_{\beta_t}(x_{t+dt}, x_t) = \frac{q(x_t)q^{\beta_t}(x_t \mid x_{t+dt})}{Z_q(\beta_t)}, \tag{22}$$

$$p_{\beta_t}(x_{t+dt}, x_t) = \frac{p^{\beta_t}(x_t)p(x_{t+dt} \mid x_t)}{Z_p(\beta_t)}. \tag{23}$$

The objective function of $\beta$-VAE is defined as follows.

$$\begin{aligned}\mathcal{L}_{\beta-\mathrm{VAE}} &:= \mathbb{E}_{q(x_{t+dt},x_t)}\left[\ln p(x_{t+dt} \mid x_t)\right] \\ &\quad -\beta_t \mathbb{E}_{q(x_t)}\left[D_{\mathrm{KL}}\left(q(x_t \mid x_{t+dt}) \| p(x_t)\right)\right].\end{aligned} \tag{24}$$

Now, the Jacobian of $\beta$-VAE is $J_{p,\beta_t} := \nabla_{x_t} \ln p^{\beta_t}(x_t)$. Then, the stochastic Riemanian metric of $\beta$-VAE is $G_{p,\beta_t} := J_{p,\beta_t}^T J_{p,\beta_t}$. The magnification factor $m_{p,\beta_t}$ of $\beta$-VAE (Arvanitidis et al., 2021) is

$$m_{p,\beta_t} := \sqrt{\det G_{p,\beta_t}} = \beta_t^d m_{p,1}. \tag{25}$$

Hence, the magnification factor $m_{p,\beta_t}$, which represents the local curvature of the latent space, exhibits a strong dependence on the inverse temperature. By our definition, the inverse temperature $\beta_t = \frac{1}{\sigma^2 t}$ decreases progressively as the diffusion step $t$ proceed. As the diffusion steps increases, the magnification factor approaches flatness. This indicates that the spatial threshold for embedding the input data into different regions in the latent space based on its characteristics tends to disappear.

### 5.3. Numerical Experiment

Figure 5 is a graph showing the $\beta_t$ dependence of training error, validation error measured after training the $\beta$-VAE for various inverse temperatures $\beta_t$. In this experiment, the $\beta$-VAE was trained for 3000 epochs on MNIST.

## 6. Numerical analysis

In this section, we present experiments investigating the behaviour of tubular neighbourhoods in diffusion models.

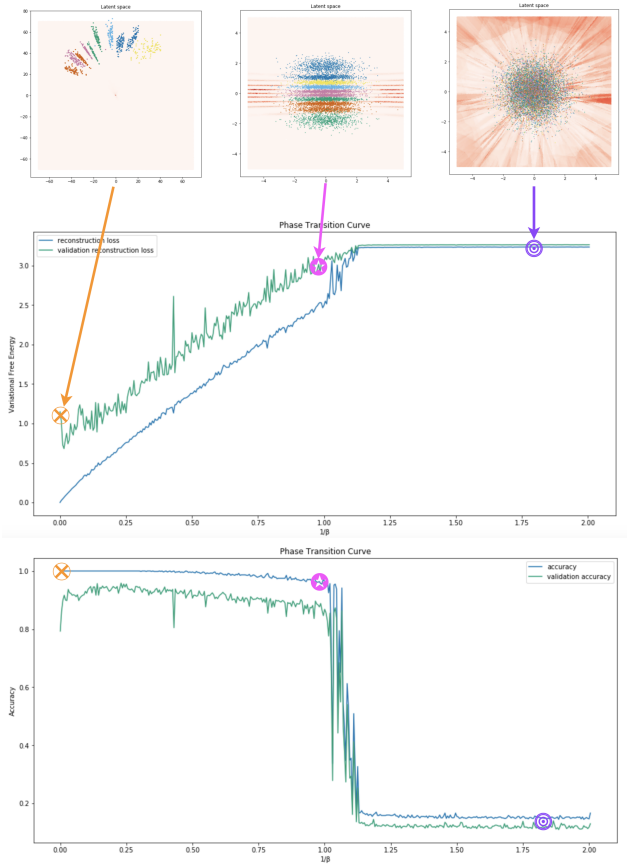



*Figure 5.* As the inverse temperature $\beta$ increases, there exists discontinuity in the performance. A phase transition occurs in the curvature (magnification factor) of the latent space, depicted by the red heatmap in the background, which shows a scatterplot of the embedded hand-written digits from 0 to 9.

### 6.1. Experiment setting

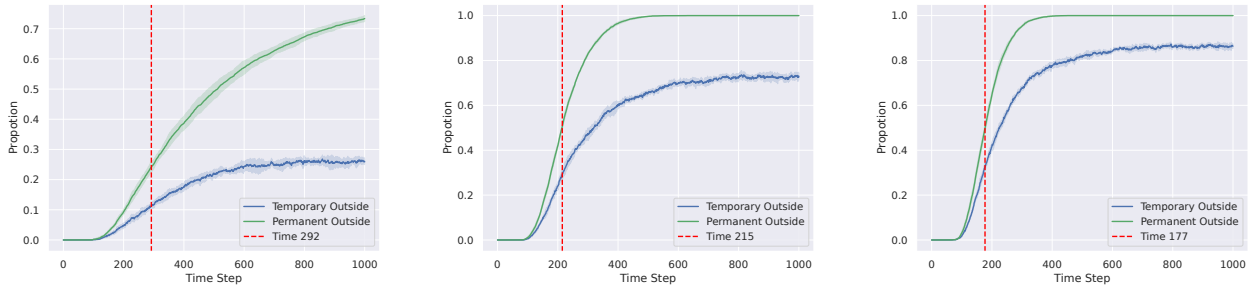
The particles on each of unit spheres $S^0$, $S^1$, and $S^2$ are diffused and again reversed. We count the proportions inside and outside the injectivity radius. The detailed description of the experiment setting is described in Appendix I.

### 6.2. Results

The results are shown in Figure 6 for the forward process and Figure 7 for the backward process. In both graphs, it can be observed that there are inflection points in the proportion outside the tubular neighbourhoods.

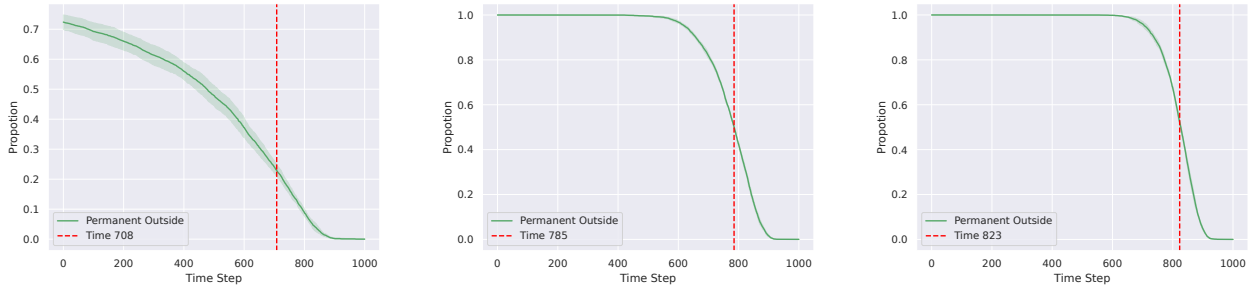
## 7. Future works

We can extend the current research in which we present some basic arguments on tubular neighbourhoods and around to have a chance to deepen them employing the knowledge of Geometry — both Differential Geometry and




(a) Forward 1D Diffusion Result $S^0$ (b) Forward 2D Diffusion Result $S^1$ (c) Forward 3D Diffusion Result $S^2$

*Figure 6.* The Proportion Outside the Tubular Neighbourhoods Over Time with Symmetry Breaking Time Point. Forward Diffusion Results for Initial States $S^0$, $S^1$, and $S^2$.




(a) Backward 1D Diffusion Result $S^0$ (b) Backward 2D Diffusion Result $S^1$ (c) Backward 3D Diffusion Result $S^2$

*Figure 7.* The Proportion Outside the Tubular Neighbourhoods Over Time with Symmetry Breaking Time Point. Backward Diffusion Results for Initial States $S^0$, $S^1$, and $S^2$.

Topology — as follows.

### 7.1. Reconsidering how to estimate the injectivity radius

The injectivity radius of a data manifold is a constant defined as the infimum of norms of normal vectors, which make the endpoint map critical, to the manifold. However, the radius of a tubular neighbourhood should be able to be taken smaller and larger at each point. For instance, the tubular neighbourhood of $S^1 \subset \mathbf{R}^2$ can only extend inwards close to the centre, but outwards as far as possible.

Besides, although we used the endpoint map to define the tubular neighbourhood, it can be defined more generally by the flow of a vector field, which is called a *spray*. Because the score function deviates from the normal direction as it moves away from data manifold, we could be able to estimate the injectivity radius much more accurately if we obtain a spray that well approximates the score function.

According to this geometrical picture, we might be able to control generation results of diffusion models by varying the scaling of radii of the tubular neighbourhood and the normal direction at each point.

### 7.2. From the viewpoint of Singularity Theory

Singularity Theory is the research area which studies properties of singularities (or critical points) appearing in spaces, functions, and maps to investigate the geometry of them (cf., e.g., (Arnold et al., 1985)). Since mid 20th century, it has been discovered that singularities have important data of geometric objects from many aspects of mathematics. Being based on Singularity Theory, we have a chance to estimate how much the $\epsilon$-neighbourhood $\mathcal{M}(\epsilon)$ of a given data manifold $\mathcal{M}$ fails to be tubular for a given radius $\epsilon > 0$, by analysing the singularities appearing in the boundary $\partial\mathcal{M}(\epsilon)$ and the critical points of the endpoint map $E\colon N\mathcal{M} \to \mathbf{R}^d$.

We also note that Singularity Theory has numerous applications — they are nothing but the revival and new generation of Thom's Catastrophe Theory (Arnold, 1992). Specifically, a powerful generalisation of the Amari–Nagaoka Theory (Amari & Nagaoka, 2000) to singular models is suggested (Nakajima & Ohmoto, 2021).

### 7.3. Connection with stochastic thermodynamics and information geometry

Based on the findings in the literature (Ambrogioni, 2024), which highlight the connection between non-equilibrium thermodynamics and diffusion models, we can analyse spontaneous symmetry breaking in the latent space as the diffusion step increases using the tools of physics. Furthermore, we propose that the equilibrium equations presented in this paper may have more general counterparts applicable to non-equilibrium states. This assertion is supported by the insights from (Ito, 2023), which elucidate the close relationship between non-equilibrium thermodynamics and the rapidly advancing field of information geometry.

## Acknowledgements

The authors express special thanks to the anonymous reviewers whose comments led to valuable improvements of this paper. Part of this work is supported by projects commissioned by JSPS KAKENHI Grant Number 23H04974 as well as JST SPRING, Grant Number JPMJSP2119.

## Author Contributions

R.S. developed the original idea and together with K.S. led the conceptualisation and initiated the project. M.T. and M.A. proved the main theorem. Y.H. provided an additional yet profound perspective. M.Y. and K.S. conducted the numerical experiments. K.S., R.S., M.T., M.A., Y.H., and M.Y contributed critically to the drafts. All authors reviewed and gave final approval for publication.

## References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., ė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., dek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, May 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://doi.org/10.1038/s41586-024-07487-w.

Amari, S. and Nagaoka, H. *Methods of Information Geometry*. American Mathematical Society, 2000. URL https://doi.org/10.1090/mmono/191.

Ambrogioni, L. The statistical thermodynamics of generative diffusion models. *CoRR*, abs/2310.17467, 2023. doi: 10.48550/ARXIV.2310.17467. URL https://doi.org/10.48550/arXiv.2310.17467.

Ambrogioni, L. The statistical thermodynamics of generative diffusion models: Phase transitions, symmetry breaking and critical instability, 2024.

Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(82)90051-5. URL https://www.sciencedirect.com/science/article/pii/0304414982900515.

Arnold, V. I. *Catastrophe Theory*. Springer Berlin, Heidelberg, 1992. ISBN 978-3-642-58124-3. doi: 10.1007/978-3-642-58124-3. URL https://doi.org/10.1007/978-3-642-58124-3.

Arnold, V. I., Guseĭn-Zade, S. M., and Varchenko, A. N. *Singularities of differentiable maps. Vol. I*, volume 82 of *Monographs in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1985. ISBN 0-8176-3187-9. doi: 10.1007/978-1-4612-5154-5. URL https://doi.org/10.1007/978-1-4612-5154-5. The classification of critical points, caustics and wave fronts, Translated from the Russian by Ian Porteous and Mark Reynolds.

Arvanitidis, G., Hansen, L. K., and Hauberg, S. Latent space oddity: on the curvature of deep generative models, 2021.

Aurell, E., Mejía-Monasterio, C., and Muratore-Ginanneschi, P. Optimal protocols and optimal transport in stochastic thermodynamics. *Physical Review Letters*, 106(25), June 2011. ISSN 1079-7114. doi: 10.1103/physrevlett.106.250601. URL http://dx.doi.org/10.1103/PhysRevLett.106.250601.

Batzolis, G., Stanczuk, J., and Schönlieb, C. Your diffusion model secretly knows the dimension of the data manifold. *CoRR*, abs/2212.12611, 2022. doi: 10.48550/ARXIV.2212.12611. URL https://doi.org/10.48550/arXiv.2212.12611.

Benton, J., Bortoli, V. D., Doucet, A., and Deligiannidis, G. Linear convergence bounds for diffusion models via stochastic localization. *CoRR*, abs/2308.03686, 2023a. doi: 10.48550/ARXIV.2308.03686. URL https://doi.org/10.48550/arXiv.2308.03686.

Benton, J., Deligiannidis, G., and Doucet, A. Error bounds for flow matching methods. *CoRR*, abs/2305.16860, 2023b. doi: 10.48550/ARXIV.2305.16860. URL https://doi.org/10.48550/arXiv.2305.16860.

Biroli, G., Bonnaire, T., Bortoli, V. D., and Mézard, M. Dynamical regimes of diffusion models. *CoRR*, abs/2402.18491, 2024. doi: 10.48550/ARXIV.2402.18491. URL https://doi.org/10.48550/arXiv.2402.18491.

Block, A., Mroueh, Y., and Rakhlin, A. Generative modeling with denoising auto-encoders and langevin sampling. *CoRR*, abs/2002.00107, 2020. URL https://arxiv.org/abs/2002.00107.

Bond-Taylor, S., Leach, A., Long, Y., and Willcocks, C. G. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(11):7327–7347, 2022. doi: 10.1109/TPAMI.2021.3116668. URL https://doi.org/10.1109/TPAMI.2021.3116668.

Bortoli, V. D. Convergence of denoising diffusion models under the manifold hypothesis. *ArXiv*, abs/2208.05314, 2022a. URL https://api.semanticscholar.org/CorpusID:251468296.

Bortoli, V. D. Convergence of denoising diffusion models under the manifold hypothesis. *Trans. Mach. Learn. Res.*, 2022, 2022b. URL https://openreview.net/forum?id=MhK5aXo3gB.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. Diffusion schrödinger bridge with applications to score-based generative modeling. In Ranzato, M., Beygelzimer, A., Dauphin, Y. N., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 17695–17709, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/940392f5f32a7ade1cc201767cf83e31-Abstract.html.

Chen, D., Zhou, Z., Mei, J., Shen, C., Chen, C., and Wang, C. A geometric perspective on diffusion models. *CoRR*, abs/2305.19947, 2023a. doi: 10.48550/ARXIV.2305.19947. URL https://doi.org/10.48550/arXiv.2305.19947.

Chen, D., Zhou, Z., Mei, J.-P., Shen, C., Chen, C., and Wang, C. A geometric perspective on diffusion models, 2023b.

Chen, H., Lee, H., and Lu, J. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 4735–4763. PMLR, 2023c. URL https://proceedings.mlr.press/v202/chen23q.html.

Chen, N., Zhang, Y., Zen, H., Weiss, R. J., Norouzi, M., and Chan, W. Wavegrad: Estimating gradients for waveform generation. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=NsMLjcFaO8O.

Chen, S., Chewi, S., Lee, H., Li, Y., Lu, J., and Salim, A. The probability flow ODE is provably fast. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023d. URL http://papers.nips.cc/paper_files/paper/2023/hash/d84a27ff694345aacc21c72097a69ea2-Abstract-Conference.html.

Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023e. URL https://openreview.net/pdf?id=zyLVMgsZ0U_.

Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., and Yoon, S. Perception prioritized training of diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 11462–11471. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01118. URL https://doi.org/10.1109/CVPR52688.2022.01118.

Chung, H., Sim, B., Ryu, D., and Ye, J. C. Improving diffusion models for inverse problems using manifold constraints. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/a48e5877c7bf86a513950ab23b360498-Abstract-Conference.html.

Croitoru, F., Hondru, V., Ionescu, R. T., and Shah, M. Diffusion models in vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(9):10850–10869, 2023. doi: 10.1109/TPAMI.2023.3261988. URL https://doi.org/10.1109/TPAMI.2023.3261988.

Dhariwal, P. and Nichol, A. Q. Diffusion models beat gans on image synthesis. In Ranzato, M., Beygelzimer, A., Dauphin, Y. N., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 8780–8794, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.html.

Duan, J., Kong, F., Wang, S., Shi, X., and Xu, K. Are diffusion models vulnerable to membership inference attacks? In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 8717–8730. PMLR, 2023. URL https://proceedings.mlr.press/v202/duan23b.html.

Dubinski, J., Kowalczuk, A., Pawlak, S., Rokita, P., Trzcinski, T., and Morawiecki, P. Towards more realistic membership inference attacks on large diffusion models. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2024, Waikoloa, HI, USA, January 3-8, 2024*, pp. 4848–4857. IEEE, 2024. doi: 10.1109/WACV57701.2024.00479. URL https://doi.org/10.1109/WACV57701.2024.00479.

Fefferman, C., Mitter, S., and Narayanan, H. Testing the manifold hypothesis, 2013.

Franzese, G., Rossi, S., Yang, L., Finamore, A., Rossi, D., Filippone, M., and Michiardi, P. How much is enough? A study on diffusion times in score-based generative models. *Entropy*, 25(4):633, 2023. doi: 10.3390/E25040633. URL https://doi.org/10.3390/e25040633.

Fu, W., Wang, H., Gao, C., Liu, G., Li, Y., and Jiang, T. A probabilistic fluctuation based membership inference attack for diffusion models. *CoRR*, abs/2308.12143, 2023. doi: 10.48550/ARXIV.2308.12143. URL https://doi.org/10.48550/arXiv.2308.12143.

Georgiev, K., Vendrow, J., Salman, H., Park, S. M., and Madry, A. The journey, not the destination: How data guides diffusion models. *CoRR*, abs/2312.06205, 2023. doi: 10.48550/ARXIV.2312.06205. URL https://doi.org/10.48550/arXiv.2312.06205.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Ho, J., Salimans, T., Gritsenko, A. A., Chan, W., Norouzi, M., and Fleet, D. J. Video diffusion models. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/39235c56aef13fb05a6adc95eb9d8d66-Abstract-Conference.html.

Huang, C.-W., Lim, J. H., and Courville, A. A variational perspective on diffusion-based generative models and score matching, 2021.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005. ISSN 1532-4435,1533-7928.

Ito, S. Geometric thermodynamics for the fokker–planck equation: stochastic thermodynamic links between information geometry and optimal transport. *Information Geometry*, 7(S1):441–483, March 2023. ISSN 2511-249X. doi: 10.1007/s41884-023-00102-3. URL http://dx.doi.org/10.1007/s41884-023-00102-3.

Kingma, D. P. and Gao, R. Understanding diffusion objectives as the ELBO with simple data augmentation. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/ce79fbf9baef726645bc2337abb0ade2-Abstract-Conference.html.

Kong, F., Duan, J., Ma, R., Shen, H., Zhu, X., Shi, X., and Xu, K. An efficient membership inference attack for the diffusion model by proximal initialization. *CoRR*, abs/2305.18355, 2023. doi: 10.48550/ARXIV.2305.18355. URL https://doi.org/10.48550/arXiv.2305.18355.

Kong, Z., Ping, W., Huang, J., Zhao, K., and Catanzaro, B. Diffwave: A versatile diffusion model for audio synthesis. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=a-xFK8Ymz5J.

Lee, H., Lu, J., and Tan, Y. Convergence for score-based generative modeling with polynomial complexity. In

Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/8ff87c96935244b63503f542472462b3-Abstract-Conference.html. http://papers.nips.cc/paper_files/paper/2022/hash/8ff87c96935244b63503f542472462b3-Abstract-Conference.html.

Lee, H., Lu, J., and Tan, Y. Convergence of score-based generative modeling for general data distributions. In Agrawal, S. and Orabona, F. (eds.), *International Conference on Algorithmic Learning Theory, February 20-23, 2023, Singapore*, volume 201 of *Proceedings of Machine Learning Research*, pp. 946–985. PMLR, 2023. URL https://proceedings.mlr.press/v201/lee23a.html.

Lee, J. M. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013. ISBN 978-1-4419-9982-5(eBook). doi: 10.1007/978-1-4419-9982-5. URL https://doi.org/10.1007/978-1-4419-9982-5.

Li, G., Wei, Y., Chen, Y., and Chi, Y. Towards faster non-asymptotic convergence for diffusion-based generative models. *CoRR*, abs/2306.09251, 2023. doi: 10.48550/ARXIV.2306.09251. URL https://doi.org/10.48550/arXiv.2306.09251.

Li, G., Huang, Z., and Wei, Y. Towards a mathematical theory for consistency training in diffusion models. *CoRR*, abs/2402.07802, 2024. doi: 10.48550/ARXIV.2402.07802. URL https://doi.org/10.48550/arXiv.2402.07802.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/pdf?id=PqvMRDCJT9t.

Litherland, R., Simon, J., Durumeric, O., and Rawdon, E. Thickness of knots. *Topology and its Applications*, 91 (3):233–244, 1999. ISSN 0166-8641. doi: https://doi.org/10.1016/S0166-8641(97)00210-1. URL https://www.sciencedirect.com/science/article/pii/S0166864197002101.

Liu, H., Chen, Z., Yuan, Y., Mei, X., Liu, X., Mandic, D. P., Wang, W., and Plumbley, M. D. Audioldm: Text-to-audio generation with latent diffusion models. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 21450–21474. PMLR, 2023. URL https://proceedings.mlr.press/v202/liu23f.html.

Liu, X., Wu, L., Ye, M., and Liu, Q. Let us build bridges: Understanding and extending diffusion generative models. *CoRR*, abs/2208.14699, 2022. doi: 10.48550/ARXIV.2208.14699. URL https://doi.org/10.48550/arXiv.2208.14699.

Loaiza-Ganem, G., Ross, B. L., Hosseinzadeh, R., Caterini, A. L., and Cresswell, J. C. Deep generative models through the lens of the manifold hypothesis: A survey and new connections. *CoRR*, abs/2404.02954, 2024. doi: 10.48550/ARXIV.2404.02954. URL https://doi.org/10.48550/arXiv.2404.02954.

Lou, A., Meng, C., and Ermon, S. Discrete diffusion language modeling by estimating the ratios of the data distribution. *CoRR*, abs/2310.16834, 2023. doi: 10.48550/ARXIV.2310.16834. URL https://doi.org/10.48550/arXiv.2310.16834.

Luo, C. Understanding diffusion models: A unified perspective. *CoRR*, abs/2208.11970, 2022. doi: 10.48550/ARXIV.2208.11970. URL https://doi.org/10.48550/arXiv.2208.11970.

Matsumoto, T., Miura, T., and Yanai, N. Membership inference attacks against diffusion models. In *2023 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, May 25, 2023*, pp. 77–83. IEEE, 2023. doi: 10.1109/SPW59333.2023.00013. URL https://doi.org/10.1109/SPW59333.2023.00013.

Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J., and Ermon, S. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=aBsCjcPu_tE.

Nakajima, N. and Ohmoto, T. The dually flat structure for singular models. *Information Geometry*, 4(1), 2021. doi: 10.1007/s41884-021-00044-8. URL https://link.springer.com/article/10.1007/s41884-021-00044-8.

Pang, Y. and Wang, T. Black-box membership inference attacks against fine-tuned diffusion models. *CoRR*, abs/2312.08207, 2023. doi: 10.48550/ARXIV.2312.08207. URL https://doi.org/10.48550/arXiv.2312.08207.

Pang, Y., Wang, T., Kang, X., Huai, M., and Zhang, Y. White-box membership inference attacks against diffusion models. *CoRR*, abs/2308.06405, 2023. doi: 10.48550/ARXIV.2308.06405. URL https://doi.org/10.48550/arXiv.2308.06405.

Park, Y., Kwon, M., Choi, J., Jo, J., and Uh, Y. Understanding the latent space of diffusion models through the lens of riemannian geometry. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/4bfcebedf7a2967c410b64670f27f904-Abstract-Conference.html.

Pidstrigach, J. Score-based generative models detect manifolds. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/e8fb575e3ede31f9b8c05d53514eb7c6-Abstract-Conference.html.

Raya, G. and Ambrogioni, L. Spontaneous symmetry breaking in generative diffusion models. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/d0da30e312b75a3fffd9e9191f8bc1b0-Abstract-Conference.html.

Sclocchi, A., Favero, A., and Wyart, M. A phase transition in diffusion models reveals the hierarchical nature of data. *CoRR*, abs/2402.16991, 2024. doi: 10.48550/ARXIV.2402.16991. URL https://doi.org/10.48550/arXiv.2402.16991.

Sidorova, N. A., Smolyanov, O. G., v. Weizsäcker, H., and Wittich, O. The surface limit of brownian motion in tubular neighborhoods of an embedded riemannian manifold. *Journal of Functional Analysis*, 206 (2):391–413, 2004. ISSN 0022-1236. doi: https://doi.org/10.1016/S0022-1236(03)00067-3. URL https://www.sciencedirect.com/science/article/pii/S0022123603000673.

Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., and Taigman, Y. Make-a-video: Text-to-video generation without text-video data. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/pdf?id=nJfylDvgzlq.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In Bach, F. R. and Blei, D. M. (eds.), *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pp. 2256–2265. JMLR.org, 2015. URL http://proceedings.mlr.press/v37/sohl-dickstein15.html.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Tang, S., Wu, Z. S., Aydöre, S., Kearns, M., and Roth, A. Membership inference attacks on diffusion models via quantile regression. *CoRR*, abs/2312.05140, 2023. doi: 10.48550/ARXIV.2312.05140. URL https://doi.org/10.48550/arXiv.2312.05140.

Tong, A., FATRAS, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Verdecchia, R., Sallou, J., and Cruz, L. A systematic review of green AI. *WIREs Data. Mining. Knowl. Discov.*, 13 (4), 2023. doi: 10.1002/WIDM.1507. URL https://doi.org/10.1002/widm.1507.

Watanabe, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory, 2010.

Wenliang, L. K. and Moran, B. Score-based generative models learn manifold-like structures with constrained mixing. *CoRR*, abs/2311.09952, 2023. doi: 10.48550/ARXIV.2311.09952. URL https://doi.org/10.48550/arXiv.2311.09952.

Wibisono, A. and Yang, K. Y. Convergence in KL divergence of the inexact langevin algorithm with application to score-based generative models. *CoRR*, abs/2211.01512,

2022. doi: 10.48550/ARXIV.2211.01512. URL https://doi.org/10.48550/arXiv.2211.01512.

Xing, Z., Feng, Q., Chen, H., Dai, Q., Hu, H., Xu, H., Wu, Z., and Jiang, Y. A survey on video diffusion models. *CoRR*, abs/2310.10647, 2023. doi: 10.48550/ARXIV.2310.10647. URL https://doi.org/10.48550/arXiv.2310.10647.

Zheng, H., He, P., Chen, W., and Zhou, M. Truncated diffusion probabilistic models and diffusion-based adversarial auto-encoders. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/pdf?id=HDxgaKk956l.

## A. Related works

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021) have emerged as a powerful class of generative models (Bond-Taylor et al., 2022), demonstrating remarkable performance in various domains, image synthesis (Dhariwal & Nichol, 2021; Croitoru et al., 2023), audio generation (Chen et al., 2021; Kong et al., 2021; Liu et al., 2023), video generation (Ho et al., 2022; Singer et al., 2023; Xing et al., 2023), natural language processing (Lou et al., 2023), robot manipulations, and protein interactions (Abramson et al., 2024). These models define a forward stochastic process that progressively adds noise to the data until it reaches a Gaussian distribution, followed by a generative process that denoises the data through the approximation of the gradient of the forward logarithmic density, known as the Stein score.

**Motivations and related works.** Our work is motivated by several recent theoretical advancements and practical challenges:

- **Optimisation of Diffusion Time**: Some empirical studies report existence of an optimal diffusion time that enhances model efficiency (Franzese et al., 2023).
- **Critical Phenomena and Statistical Thermodynamics of Diffusion Models**: There are some empirical studies report heterogeneity/non-uniformity, critical phenomena during generation (Ho et al., 2020; Meng et al., 2022; Choi et al., 2022; Zheng et al., 2023; Raya & Ambrogioni, 2023; Georgiev et al., 2023; Sclocchi et al., 2024; Biroli et al., 2024).
- **Geometrical approaches**: There are some geometrical perspectives on diffusion models inspired our work (Chung et al., 2022; Wenliang & Moran, 2023; Chen et al., 2023a; Park et al., 2023).
- **Other theories to understand diffusion and generation processes**: A deeper understanding of these processes is essential for advancing theoretical research and practical applications, such as generation control through prompting and interpolation. Recent studies have delved into the underlying mechanisms of diffusion and generation trajectories to identify optimal intervention points during the generation process, which can help achieve desired data outputs. While flow-matching algorithms have shown promise, in the practical user cases, diffusion models surprisingly sometimes outperform the flow-matching, underscoring the need to understand the factors contributing to this superior performance. There are several works on convergence guarantees for diffusion models (Bortoli et al., 2021; Bortoli, 2022b; Block et al., 2020; Chen et al., 2023c; Lee et al., 2022; Liu et al., 2022; Pidstrigach, 2022; Wibisono & Yang, 2022; Chen et al., 2023e; Lee et al., 2023; Li et al., 2023; Benton et al., 2023a;b; Chen et al., 2023d; Li et al., 2024)
- **Flow matching techniques**: Flow matching algorithms (Lipman et al., 2023; Tong et al., 2024) are yet another prominent techniques in generative modelling. They are closely related to diffusion models as flow matching often leverages diffusion paths for training, in which optimal transport via ODEs yields straighter trajectories. It is very interesting to consider the influence on the quality and diversity of generated samples or critical dynamics such as spontaneous symmetry breaking. Our method may have the potential to analyse these aspects. Such generative models considering a transport from one distribution to another are expected to continue to develop, and geometric interpretations will further contribute to improving interpretability, efficiency, and control to ensure safety.

## B. Social Impacts

- **Green AI (Environmental Impact)**: Reducing the high energy consumption of diffusion models during both training and generation is crucial. The exponential increase in computational demands due to the growing use of diffusion models in industry poses significant environmental concerns. Optimising these models can lead to more sustainable AI practices, addressing the urgent need for eco-friendly AI technologies. Recent studies emphasise the need for environmental sustainability in AI, focusing on reducing the energy consumption and carbon footprint of AI models (Verdecchia et al., 2023).
- **AI Safety and Alignment**: Ensuring AI safety and alignment is critical. This includes improving the mechanistic interpretability of diffusion models, optimising control to prevent undesirable behaviours, and mitigating risks such as hallucinations and adversarial attacks. Effective control mechanisms and interpretability can enhance trust and safety in AI applications. Matsumoto et al. (2023) report that the diffusion time is the crucial for mitigating the membership inference attacks on diffusion models (Pang et al., 2023; Pang & Wang, 2023; Duan et al., 2023; Tang et al., 2023; Fu et al., 2023; Dubinski et al., 2024; Kong et al., 2023)

## C. Mathematical Supplementaries

In this appendix, we quickly recall basic mathematical concepts and facts concerned with Linear Algebra and Manifold Theory. See, e.g., (Lee, 2013) for a detail of Manifold Theory.

### C.1. Formal operations in Linear Algebra

For the Euclidean space $\mathbf{R}^d$ and its linear subspace $V \subset \mathbf{R}^d$, let $V^\perp$ denote the orthogonal complement of $V$ in $\mathbf{R}^d$.

**Proposition C.1.** *Let $V$ and $W$ be subspaces of $\mathbf{R}^n$. Then the following hold:*

*(1) $V \subset W$ if and only if $V^\perp \supset W^\perp$;*

*(2) $V^\perp \cap W^\perp = (V + W)^\perp$.*

### C.2. Differentiable manifolds

In this paper, as *manifolds*, we treat only 'submanifolds of the Euclidean space $\mathbf{R}^d$'. So we adapt the following definition.

**Definition C.2.** A subset $\mathcal{M}$ of $\mathbf{R}^d$ is called an *$n$-dimensional manifold*, if for each point $\boldsymbol{x} \in \mathcal{M}$, there is an open neighbourhood $U$ of $\boldsymbol{x}$ in $\mathbf{R}^d$, an open subset of $V$ in $\mathbf{R}^d = \mathbf{R}^n \times \mathbf{R}^{d-n}$, and a diffeomorphism $\phi\colon U \to V$ such that $\phi(\mathcal{M} \cap U) = V \cap (\mathbf{R}^n \times \{\mathbf{0}\})$. We call the map $\phi$ a *chart* on $\mathcal{M}$ around $\boldsymbol{x}$.

**Definition C.3.** Let $\mathcal{M} \subset \mathbf{R}^d$ be a manifold and $\boldsymbol{x} \in \mathcal{M}$ be a point. Then the *tangent space* $T_{\boldsymbol{x}}\mathcal{M}$ to $\mathcal{M}$ at $\boldsymbol{x}$ is defined as the set consisting of all velocity vectors of curves on $\mathcal{M}$ through $\boldsymbol{x}$, that is,

$$T_{\boldsymbol{x}}\mathcal{M} = \left\{ \frac{d\gamma}{dt}(0) \;\middle|\; \gamma\colon (-\epsilon, \epsilon) \to \mathcal{M},\ \gamma(0) = \boldsymbol{x} \right\}.$$

Notice that the tangent space forms a linear subspace of $\mathbf{R}^d$.

**Definition C.4.** Let $\mathcal{M} \subset \mathbf{R}^d$ and $\mathcal{M}' \subset \mathbf{R}^{d'}$ be manifolds, and let $F\colon \mathcal{M} \to \mathcal{M}'$ be a differentiable map (i.e., there is an extension $\tilde{F}\colon U \to \mathbf{R}^{d'}$ of $F$ which is a differentiable map on an open set $U$ of $\mathbf{R}^d$). Then the *differential* $dF_{\boldsymbol{x}}$ of $F$ at $\boldsymbol{x}$ is defined as the linear map

$$dF_{\boldsymbol{x}}\colon T_{\boldsymbol{x}}\mathcal{M} \to T_{F(\boldsymbol{x})}\mathcal{M}', \quad dF_{\boldsymbol{x}}\left(\frac{d\gamma}{dt}(0)\right) = \frac{d(F \circ \gamma)}{dt}(0).$$

**Remark C.5.** Take charts $\phi\colon U \to V$ and $\psi\colon U' \to V'$ on $\mathcal{M}$ and $\mathcal{M}'$, respectively. Also let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_{n'})$ denote the coordinate on $V \subset \mathbf{R}^n$ and $V' \subset \mathbf{R}^{n'}$, respectively. Then the diffenrential $dF_{\boldsymbol{x}}$ is represented by the Jacobi matrix

$$\frac{\partial(\psi \circ F \circ \phi^{-1})}{\partial \boldsymbol{x}}(\boldsymbol{x}) = \left[\frac{\partial(\psi \circ F \circ \phi^{-1})_i}{\partial x_j}(\boldsymbol{x})\right]$$

of the map $\psi \circ F \circ \phi^{-1}\colon V \to V'$ at the point $\boldsymbol{x} \in \mathcal{M}$.

**Definition C.6.** Let $F\colon \mathcal{M} \to \mathcal{M}'$ be a differentiable map between manifolds. A point $\boldsymbol{x} \in \mathcal{M}$ is called a *regular point* (resp. a *critical point*) if the differential $dF_{\boldsymbol{x}}\colon T_{\boldsymbol{x}}\mathcal{M} \to T_{F(\boldsymbol{x})}\mathcal{M}'$ is surjective (resp. not surjective). A point $\boldsymbol{y} \in \mathcal{M}'$ is called a *regular value* (resp. a *critical value*) if every point $\boldsymbol{x} \in \mathcal{M}$ satisfying that $F(\boldsymbol{x}) = \boldsymbol{y}$ is a regular point of $F$ (resp. or not).

The following is essentially a consequence of Implicit Function Theorem.

**Theorem C.7** (cf. (Lee, 2013)[Corollary 5.14])**.** *Let $\boldsymbol{F}\colon \mathbf{R}^d \to \mathbf{R}^{d'}$ be a differentiable map and $\boldsymbol{y} \in \mathbf{R}^{d'}$ a regular value of $\boldsymbol{F}$. Then the level set*

$$\boldsymbol{F}^{-1}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathbf{R}^d \mid \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{y}\} \subset \mathbf{R}^d$$

*forms a $(d - d')$-dimensional manifold.*

**Remark C.8** (explicit description of the tangent spaces to a manifold)**.** Consider the same setup of Theorem C.7 and denote $\boldsymbol{F} = (F_1, \ldots, F_{d'})$. Then the normal to the tangent space $T_{\boldsymbol{x}}\mathcal{M}$ coincides with

$$\left\langle \frac{\partial F_1}{\partial \boldsymbol{x}}(\boldsymbol{x})^T, \cdots, \frac{\partial F_{d'}}{\partial \boldsymbol{x}}(\boldsymbol{x})^T \right\rangle_{\mathbf{R}},$$

which is spanned by the gradient vectors of components of $\boldsymbol{F}$. Therefore the tangent space itself is noting but its orthogonal complement, i.e.,

$$T_{\boldsymbol{x}}\mathcal{M} = \left\langle \frac{\partial F_1}{\partial \boldsymbol{x}}(\boldsymbol{x})^T, \cdots, \frac{\partial F_{d'}}{\partial \boldsymbol{x}}(\boldsymbol{x})^T \right\rangle_{\mathbf{R}}^{\perp}.$$

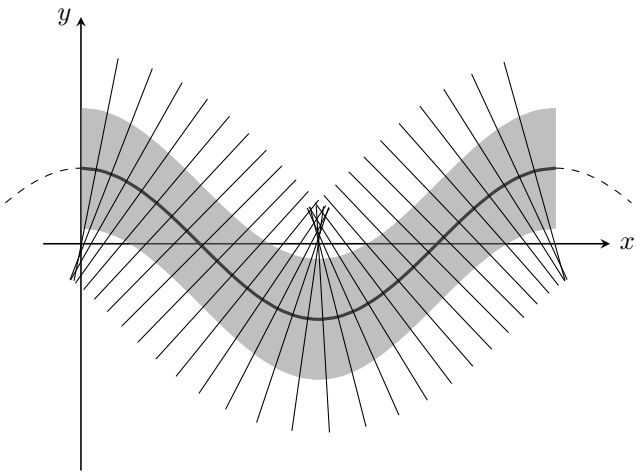



*Figure 8.* Image under $E$ and tubular neighbourhood of the cosine curve in $\mathbf{R}^2$

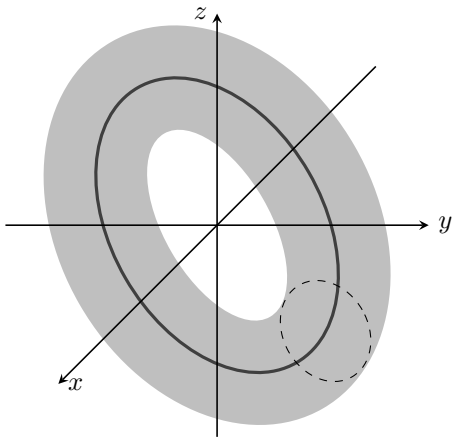



*Figure 9.* Tubular neighbourhood of $S^1$ embedded in $\mathbf{R}^3$

**Definition C.9.** A differentiable map $F\colon \mathcal{M} \to \mathcal{M}'$ is called an *embedding* if its differential $dF\colon T_{\boldsymbol{x}}\mathcal{M} \to T_{F(\boldsymbol{x})}\mathcal{M}'$ is injective for every point $\boldsymbol{x} \in \mathcal{M}$ and the restriction $F\colon \mathcal{M} \to f(\mathcal{M})$ is a topological homeomorphism (i.e. there is the inverse map $F^{-1}$, and both $F$ and $F^{-1}$ are continuous).

### C.3. Tubular neighbourhoods

Let $\mathcal{M} \subset \mathbf{R}^d$ be a manifold. Recall the normal bundle

$$N\mathcal{M} = \{(\boldsymbol{x}, \boldsymbol{v}) \in \mathbf{R}^d \times \mathbf{R}^d \mid \boldsymbol{x} \in \mathcal{M}, \boldsymbol{v} \perp T_{\boldsymbol{x}}\mathcal{M}\}$$

to $\mathcal{M}$ and the endpoint map

$$E\colon N\mathcal{M} \to \mathbf{R}^d, \quad E(\boldsymbol{x}, \boldsymbol{v}) = \boldsymbol{x} + \boldsymbol{v},$$

which are defined in §3 (Definitions 3.2 and 3.3).

**Definition C.10** (Tubular neighbourhood)**.** A tubular neighbourhood of $\mathcal{M}$ is a neighbourhood of $\mathcal{M}$ in $\mathbf{R}^d$ that is the diffeomorphic image under $E$ of an open subset $V \subset N\mathcal{M}$ of the form

$$V = \{(\boldsymbol{x}, \boldsymbol{v}) \in N\mathcal{M} \mid \|\boldsymbol{v}\| < \delta(\boldsymbol{x})\}$$

for some positive continuous function $\delta : \mathcal{M} \to \mathbf{R}$.

**Theorem C.11** (Theorem 6.24 in (Lee, 2013))**.** *Every manifold embedded in $\mathbf{R}^d$ has a tubular neighbourhood.*

*Proof.* Let $\mathcal{M}_0$ denote the subset $\{(\boldsymbol{x}, 0) \mid \boldsymbol{x} \in \mathcal{M}\} \subset N\mathcal{M}$. Fix a point $\boldsymbol{x} \in \mathcal{M}$. Since both differentials $dE|_{T_{(\boldsymbol{x},0)}\mathcal{M}_0} : T_{(\boldsymbol{x},0)}\mathcal{M}_0 \to T_{\boldsymbol{x}}\mathcal{M}$ and $dE|_{N_{\boldsymbol{x}}\mathcal{M}} : N_{\boldsymbol{x}}\mathcal{M} \to N_{\boldsymbol{x}}\mathcal{M}$ are isomorphisms, we have that $dE\colon T_{(\boldsymbol{x},0)}N\mathcal{M} \to \mathbf{R}^d$ is also an isomorphism. By Inverse Function Theorem, the map $E$ is a diffeomorphism on a neighbourhood of $(\boldsymbol{x}, 0) \in N\mathcal{M}$. We can take the neighbourhood to be of the form $V_\delta(\boldsymbol{x}) = \{(\boldsymbol{x}', \boldsymbol{v}') \in N\mathcal{M} \mid \|\boldsymbol{x} - \boldsymbol{x}'\| < \delta, \|\boldsymbol{v}'\| < \delta\}$ for some $\delta > 0$. Let $\rho(\boldsymbol{x})$ denote the supremum of all such $\delta < 1$. We can prove that the function $\rho\colon \mathcal{M} \to \mathbf{R}$ is positive and continuous.

Now consider the open subset $V = \{(\boldsymbol{x}, \boldsymbol{v}) \in N\mathcal{M} \mid \|\boldsymbol{v}\| < \frac{1}{2}\rho(\boldsymbol{x})\}$ of $N\mathcal{M}$. Then the map $E$ is injective on $V$, and hence $E|_V : V \to \mathbf{R}^d$ is a smooth embedding. Thus $E(V)$ is a tubular neighbourhood of $\mathcal{M}$. □

## D. Theoretical supplementaries of Section 3

### D.1. Proof of Theorem 3.7

Under the setup of Theorem 3.7, put $k = d - n$ and we define a map $\boldsymbol{\varphi}\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^{N+k}$ by

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{v}) = (\boldsymbol{F}(\boldsymbol{x}), \varphi_1(\boldsymbol{x}, \boldsymbol{v}), \cdots, \varphi_N(\boldsymbol{x}, \boldsymbol{v})).$$

Notice that the normal bundle $N\mathcal{M}$ to $\mathcal{M}$ is expressed by

$$N\mathcal{M} = \boldsymbol{\varphi}^{-1}(\mathbf{0}) = \{(\boldsymbol{x}, \boldsymbol{v}) \in \mathbf{R}^d \times \mathbf{R}^d \mid \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{v}) = \mathbf{0}\}.$$

Hence the tangent space $T_{(\boldsymbol{x},\boldsymbol{v})}N\mathcal{M} \subset \mathbf{R}^d \times \mathbf{R}^d$ to $N\mathcal{M}$ at a point $(\boldsymbol{x}, \boldsymbol{v})$ coincides with the orthogonal complement of

$$\left\langle \begin{bmatrix} \frac{\partial F_1}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial F_d}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \frac{\partial \varphi_1}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_1}{\partial \boldsymbol{v}}^T \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial \varphi_N}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_N}{\partial \boldsymbol{v}}^T \end{bmatrix} \right\rangle_{\mathbf{R}}$$

in $\mathbf{R}^d \times \mathbf{R}^d$ (cf. Remark C.8).

We now employ the Method of Lagrange multiplier. That is, we paraphrase the condition that a point $(\boldsymbol{x}, \boldsymbol{v}) \in N\mathcal{M}$ is a critical point of the endpoint map

$$E = E_0|_{N\mathcal{M}} \colon N\mathcal{M} \to \mathbf{R}^d$$

(i.e., the differential $dE_{(\boldsymbol{x},\boldsymbol{v})} : T_{(\boldsymbol{x},\boldsymbol{v})}N\mathcal{M} \to \mathbf{R}^d$, which is a linear map, is degenerate) as follows. First, the condition is equivalent to that there exists a non-zero vector of $T_{(\boldsymbol{x},\boldsymbol{v})}N\mathcal{M}$ which vanishes by the differential $(dE_0)_{(\boldsymbol{x},\boldsymbol{v})} : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$, i.e.,

$$T_{(\boldsymbol{x},\boldsymbol{v})}N\mathcal{M} \cap \mathrm{Ker}(dE_0)_{(\boldsymbol{x},\boldsymbol{v})} \supsetneq \{\mathbf{0}\}.$$

Moreover, we have the following:

$$\begin{aligned}
&T_{(\boldsymbol{x},\boldsymbol{v})}N\mathcal{M} \cap \mathrm{Ker}(dE_0)_{(\boldsymbol{x},\boldsymbol{v})} \supsetneq \{\mathbf{0}\} \\
\iff &\left\langle \begin{bmatrix} \frac{\partial F_1}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial F_k}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \frac{\partial \varphi_1}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_1}{\partial \boldsymbol{v}}^T \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial \varphi_N}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_N}{\partial \boldsymbol{v}}^T \end{bmatrix} \right\rangle_{\mathbf{R}}^{\perp} \cap \left\langle \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_1 \end{bmatrix}, \cdots, \begin{bmatrix} \boldsymbol{e}_d \\ \boldsymbol{e}_d \end{bmatrix} \right\rangle_{\mathbf{R}}^{\perp} \supsetneq \{\mathbf{0}\} \\
\iff &\left\langle \begin{bmatrix} \frac{\partial F_1}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial F_k}{\partial \boldsymbol{x}}^T \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \frac{\partial \varphi_1}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_1}{\partial \boldsymbol{v}}^T \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial \varphi_N}{\partial \boldsymbol{x}}^T \\ \frac{\partial \varphi_N}{\partial \boldsymbol{v}}^T \end{bmatrix} \right\rangle_{\mathbf{R}} + \left\langle \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_1 \end{bmatrix}, \cdots, \begin{bmatrix} \boldsymbol{e}_d \\ \boldsymbol{e}_d \end{bmatrix} \right\rangle_{\mathbf{R}} \subsetneq \mathbf{R}^d \times \mathbf{R}^d,
\end{aligned}$$

where $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$ denotes the standard basis of $\mathbf{R}^d$. Here we used a property on orthogonal complements (see Appendix C.1).

Finally, it is equivalent to that the matrix

$$\begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}^T & \frac{\partial \varphi_1}{\partial \boldsymbol{x}}^T & \cdots & \frac{\partial \varphi_N}{\partial \boldsymbol{x}}^T & E_d \\ O_{n,d} & \frac{\partial \varphi_1}{\partial \boldsymbol{v}}^T & \cdots & \frac{\partial \varphi_N}{\partial \boldsymbol{v}}^T & E_d \end{bmatrix}$$

is degenerate. Performing elementary row and column operations, and by the definition of $R_1(\mathcal{M})$, the conclusion of Theorem 3.7 follows. □

### D.2. Curvature and the first injectivity radius of a curve

Let $\mathcal{M}$ be a curve in $\mathbf{R}^d$, i.e., a one-dimensional manifold embedded in $\mathbf{R}^d$. We see that, in this case, the first injectivity radius $R_1(\mathcal{M})$ is derived from the curvature of $\mathcal{M}$ as follows.

**Definition D.1.** Let $\gamma \colon \mathbf{R} \to \mathbf{R}^d$ be an arc-length parametrization of the curve $\mathcal{M}$, i.e., $\left\| \dfrac{d\gamma}{ds} \right\| \equiv 1$. Then the curvature $\kappa$ of $\mathcal{M}$ at a point $p = \gamma(s) \in \mathcal{M}$ is defined by the Euclidean norm of the second order derivative $\dfrac{d^2\gamma}{ds^2}(s)$.

**Proposition D.2.** *Assume that $n = 1$. Let $\gamma \colon \mathbf{R} \to \mathbf{R}^d$ be an arbitrary regular parametrization of the curve $\mathcal{M}$. Then the curvature $\kappa$ of $\mathcal{M}$ is computed by*

$$\kappa(\gamma(u)) = \frac{\sqrt{\|\gamma'(u)\|^2 \|\gamma''(u)\|^2 - \langle \gamma'(u), \gamma''(u) \rangle^2}}{\|\gamma'(u)\|^3}, \tag{26}$$

*where $'$ denotes the differential by $u$.*

Although this is a well-known fact, we show it briefly as follows.

*Proof.* Let $s$ and $u$ denote an arc-length parameter and an arbitrary regular parameter of the curve $\mathcal{M}$. Since it holds that

$$\gamma' = s' \frac{d\gamma}{ds}, \tag{27}$$

we also have that

$$\gamma'' = s'' \cdot \frac{d\gamma}{ds} + (s')^2 \kappa \cdot \nu, \tag{28}$$

where $\nu$ denotes the normalization of the vector $\dfrac{d^2\gamma}{ds^2}$. Since two vectors $\dfrac{d\gamma}{ds}$ and $\nu$ form an orthonormal frame of the curve $\mathcal{M}$, it holds that

$$\|\gamma''\|^2 = (s'')^2 + (s')^4 \cdot \kappa^2. \tag{29}$$

Now notice the following: it holds that

$$\|\gamma'\|^2 = (s')^2 \tag{30}$$

by Equation (27), and hence

$$\langle \gamma', \gamma'' \rangle = s' \cdot s''. \tag{31}$$

Applying Equations (30) and (31) to Equation (29), we have the claim. □

**Theorem D.3.** *Assume that $n = 1$. Let $\kappa$ denote the curvature of $\mathcal{M}$. Then $R_1(\mathcal{M})$ coincides with the infimum of radii of curvature $1/\kappa$.*

*Proof.* See Lemma 1 of (Litherland et al., 1999). □

### D.3. Comments on the computation of the second injectivity radius

In this paper we used the definition of $R_2(\mathcal{M})$ as-is for the numerical estimation.

We note that one can weaken the condition appearing to the definition of $R_2(\mathcal{M})$ as follows.

**Proposition D.4.** The second injectivity radius $R_2(\mathcal{M})$ coincides with the infimum of the set

$$\left\{ \frac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \;\middle|\; \begin{array}{l} \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{M}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2, \\ \text{and } \boldsymbol{x}_1 - \boldsymbol{x}_2 \perp T_{\boldsymbol{x}_1}\mathcal{M} \end{array} \right\}.$$

*Proof.* See §4 of (Litherland et al., 1999). □

We also have a comment on $R_2(\mathcal{M})$. Numerically, it seems to be possible to compute $R_2(\mathcal{M})$ by using the *persistent homology* of the given data cloud. Indeed, the topology of the $\epsilon$-neighbourhood of the data cloud might change when two tubes touch each other.

## E. Other examples of injectivity radii

We have already seen that Theorem 3.7 works in the case that a data manifold is the unit circle $S^1 \subset \mathbf{R}^2$. In this appendix, we verify the theorem by observing other typical manifolds.

### E.1. Torus $T^2$

Let $r' > r > 0$, and define a function $F\colon \mathbf{R}^3 \to \mathbf{R}$ by

$$F(x,y,z) = (\sqrt{x^2+y^2} - r')^2 + z^2 - r^2.$$

Then we have a torus $T^2 = F^{-1}(0)$ embedded in $\mathbf{R}^3$. We can see that vector fields

$$\boldsymbol{t}_1 = (-y, x, 0), \quad \boldsymbol{t}_2 = (xz, yz, r'\sqrt{x^2+y^2} - x^2 - y^2)$$

satisfy the assumption in Theorem 3.7. Then the matrix $L_{T^2}((x,y,z),(v_1,v_2,v_3))$ is calculated as follows:

$$\begin{aligned} &L_{T^2}((x,y,z),(v_1,v_2,v_3)) \\ &= \begin{bmatrix} 2(\sqrt{x^2+y^2}-r')\frac{x}{\sqrt{x^2+y^2}} & v_2+y & zv_1 - 2xv_3 - xz + \frac{r'xv_3}{\sqrt{x^2+y^2}} \\ 2(\sqrt{x^2+y^2}-r')\frac{y}{\sqrt{x^2+y^2}} & -v_1-x & zv_2 - 2yv_3 - yz + \frac{r'yv_3}{\sqrt{x^2+y^2}} \\ z & 0 & xv_1 + yv_2 + x^2 + y^2 - r'\sqrt{x^2+y^2} \end{bmatrix}. \end{aligned}$$

We now parametrise the torus $T^2$ by $(x,y,z) = ((r'+r\cos t)\cos u, (r'+r\cos t)\sin u, \cos t)$ of $T^2 \subset \mathbf{R}^3$. Then the vector $(v_1,v_2,v_3)$ makes $L_{T^2}((x,y,z),(v_1,v_2,v_3))$ degenerate if and only if

$$\begin{aligned} &(v_1,v_2,v_3) = -(r\cos t\cos u, r\cos t\sin u, r\sin t) \quad \text{or} \\ &(v_1,v_2,v_3) = -\frac{r'+r\cos t}{r\cos t}(r\cos t\cos u, r\cos t\sin u, r\sin t) \quad \left(t \neq \pm\frac{\pi}{2}\right). \end{aligned}$$

Hence we obtain $R_1(T^2) = \min\{r, r'-r\}$. Moreover we can see that $R_2(T^2) = \min\{r, r'-r\}$. Thus the injectivity radius is $R(T^2) = \min\{r, r'-r\}$.

### E.2. Unit Sphere $S^2$

Define a function $F\colon \mathbf{R}^3 \to \mathbf{R}$ by

$$F(x,y,z) = x^2 + y^2 + z^2 - 1.$$

Then we have $S^2 = F^{-1}(0)$. Considering the rotation in $\mathbf{R}^3$ around coordinate axes, we see that vector fields

$$\boldsymbol{t}_1 = (-y, x, 0), \quad \boldsymbol{t}_2 = (-z, 0, x), \quad \boldsymbol{t}_3 = (0, -z, y).$$

satisfy the assumption of Theorem 3.7. (Here notice that the number of vector fields which we desire is needed to be greater than 2, by topological reason.) Then the matrix $L_{S^2}((x,y,z),(v_1,v_2,v_3))$ is calculated as follows:

$$\begin{aligned} &L_{S^2}((x,y,z),(v_1,v_2,v_3)) \\ &= \begin{bmatrix} 2x & v_2+y & v_3+z & 0 \\ 2y & -v_1-x & 0 & v_3+z \\ 2z & 0 & -v_1-x & -v_2-y \end{bmatrix}. \end{aligned}$$

This matrix is degenerate on $((x,y,z),(v_1,v_2,v_3)) \in NS^2$ if and only if $(v_1,v_2,v_3) = (-x,-y,-z)$. Hence we obtain $R_1(S^2) = \sqrt{(-x)^2+(-y)^2+(-z^2)} = 1$. Moreover it is clear that $R_2(S^2) = 1$. Thus the injectivity radius is $R(S^2) = 1$.

### E.3. Unit $n$-Sphere $S^n$

As the final example, we observe the unit $n$-sphere. Define a function $F\colon \mathbf{R}^{n+1} \to \mathbf{R}$ by

$$F(x_1, x_2, \ldots, x_{n+1}) = x_1^2 + x_2^2 + \cdots + x_{n+1}^2 - 1.$$

Then we have $S^n = F^{-1}(0)$. Considering gradient vector fields of the height functions $(x_1, x_2, \ldots, x_{n+1}) \mapsto x_j$ $(j = 1, 2, \ldots, n+1)$, we see that vector fields

$$\boldsymbol{t}_j = (-x_1x_j, \ldots, -x_{j-1}x_j, 1-x_j^2, -x_{j+1}x_j, \ldots, -x_{n+1}x_j) \quad (j = 1, 2, \ldots, n+1)$$

*Table 1.* Estimated injectivity radii of various manifolds.

| DATA SET | $R_1$ | $R_2$ |
|---|---|---|
| $S^1$ | 1.005± 0.003 | 0.999±0.006 |
| $S^2$ | 1.063± 0.032 | 0.997±0.038 |
| $S^{128}$ | 1.068± 0.023 | 0.922±0.056 |

satisfy the assumption of Theorem 3.7. Then the matrix $L_{S^n}(\boldsymbol{x}, \boldsymbol{v})$ is calculated as follows:

$$L_{S^n}(\boldsymbol{x}, \boldsymbol{v}) = \begin{bmatrix} 2x_1 & -2x_1 - \sum_{i\neq 1} x_i v_i + x_1^2 - 1 & & -x_{n+1}v_1 + x_{n+1}x_1 \\ \vdots & -x_1 v_2 + x_1 x_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -x_{n+1}v_n + x_{n+1}x_n \\ 2x_{n+1} & -x_1 v_{n+1} + x_1 x_{n+1} & & -2x_{n+1}v_{n+1} - \sum_{i\neq n+1} x_i v_i + x_{n+1}^2 - 1 \end{bmatrix},$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_{n+1}), \boldsymbol{v} = (v_1, v_2, \cdots, v_{n+1})$. Now notice that for a point $\boldsymbol{x} \in S^n$ and a normal vector $\boldsymbol{v}$ to $\boldsymbol{x}$, there exists a scalar $c \in \mathbf{R}$ such that $\boldsymbol{v} = c\boldsymbol{x}$. Using it and performing the elementary row and column operations, the matrix $L_{S^n}(\boldsymbol{x}, \boldsymbol{v})$ is transformed as follows:

$$\begin{bmatrix} x_1 & -c\sum_i x_i^2 - 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ x_{n+1} & 0 & \cdots & 0 & -c\sum_i x_i^2 - 1 \end{bmatrix}.$$

Hence the vector $\boldsymbol{v}$ makes the matrix $L_{S^n}(\boldsymbol{x}, \boldsymbol{v})$ degenerate if and only if $c = -1$. Hence we obtain $R_1(S^n) = \| - \boldsymbol{x}\| = 1$. Moreover it is clear that $R_2(S^n) = 1$. Thus the injectivity radius is $R(S^n) = 1$.

## F. Algorithm for Estimating the injectivity radius

In this appendix, we show the pseudo-algorithm for estimating the injectivity radius (see Algorithm 1) and some preliminary numerical experiments to verify the proposed algorithm.

### F.1. Numerical experiments to validate AIER

For the $S^1$, $S^2$, $S^{128}$ cases, the estimated $R_1$ and $R_2$ using the proposed algorithm are shown in Table 1. We first generate dataset using the exact generative equations and add some Gaussian noise. The $\boldsymbol{F}$ is then approximated using a neural network. The following Step 1 to Step 4 are executed using the neural network approximation $\boldsymbol{F}$. We note that we use the cosine similarity instead of inner products for the discrimination condition defined in the Step 4.

## G. Calculation and verification in Section 4

### G.1. Example in Section 4

**Example G.1.** Let us consider a circle of radius $r$ (see Figure 3). In this case

$$\lim_{t\to 0} \nabla_x \ln p_t(x) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{32}$$

for some $0 \leq \theta < 2\pi$. Here $x$ is a point in a tubular neighbourhood and $\theta$ is the angle of $\pi(x)$. Let $\Delta x$ be an infinitesimal arc length. By definition, $\Delta\theta = \frac{\Delta x}{r}$. And by taking Taylor's series we may compute the change rate tangent to the circle of

**Algorithm 1** Algorithm for estimating the injectivity radius (AEIR)

**Input:** data $\mathcal{D} \subset \mathbf{R}^d$

**Step 0:** Estimate a map $\boldsymbol{F} = (F_1, \dots, F_{d-n})\colon \mathbf{R}^d \to \mathbf{R}^{d-n}$ such that the point $\mathbf{0}$ is a regular value of $\boldsymbol{F}$ and the manifold $\boldsymbol{F}^{-1}(\mathbf{0}) \subset \mathbf{R}^d$ approximates data $\mathcal{D}$. Put $\mathcal{M} \coloneqq \boldsymbol{F}^{-1}(\mathbf{0})$.

**Step 1:** Estimate vector fields $\boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_N$ ($n \leq N$) defined near $\mathcal{M}$ such that for every $\boldsymbol{x} \in \mathcal{M}$ the vectors $\boldsymbol{t}_1(\boldsymbol{x}), \boldsymbol{t}_2(\boldsymbol{x}), \dots, \boldsymbol{t}_N(\boldsymbol{x})$ span the tangent space $T_{\boldsymbol{x}}\mathcal{M}$.

**Step 2:** Put $g_i\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$, $g_i(\boldsymbol{x}, \boldsymbol{v}) \coloneqq \langle \boldsymbol{v}, \boldsymbol{t}_i(\boldsymbol{x}) \rangle$ ($i = 1, 2, \dots, N$). Calculate the matrix

$$[A_1, \cdots, A_{d-n}, B_1, \cdots, B_N] \coloneqq \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_{d-n}}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_1} - \frac{\partial \varphi_1}{\partial v_1} & \cdots & \frac{\partial \varphi_N}{\partial x_1} - \frac{\partial \varphi_N}{\partial v_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_1}{\partial x_d} & \cdots & \frac{\partial F_{d-n}}{\partial x_d} & \frac{\partial \varphi_1}{\partial x_d} - \frac{\partial \varphi_1}{\partial v_d} & \cdots & \frac{\partial \varphi_N}{\partial x_d} - \frac{\partial \varphi_N}{\partial v_d} \end{bmatrix},$$

where $\boldsymbol{x} = (x_1, \cdots, x_d), \boldsymbol{v} = (v_1, \cdots, v_d)$.

**Step 3:** Collect sufficient amount of samples from the set

$$\left\{ (\boldsymbol{x}, \boldsymbol{v}) \in \mathbf{R}^d \times \mathbf{R}^d \;\middle|\; \begin{array}{l} \boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}, g_i(\boldsymbol{x}, \boldsymbol{v}) = 0 \ (i = 1, 2, \cdots, N), \\ \det[A_1, \dots, A_{d-n}, B_{i_1}, \dots, B_{i_n}] = 0 \\ (1 \leq i_1 < \cdots < i_n \leq N) \end{array} \right\},$$

and estimate $\min \|\boldsymbol{v}\|$ on the set. Put this value $R_1$.

**Step 4:** Collect sufficient amount of samples from the set

$$\left\{ (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbf{R}^d \times \mathbf{R}^d \;\middle|\; \begin{array}{l} \boldsymbol{F}(\boldsymbol{x}_1) = \boldsymbol{F}(\boldsymbol{x}_2) = \mathbf{0}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2, \\ \langle \boldsymbol{x}_1 - \boldsymbol{x}_2, \boldsymbol{t}_i(\boldsymbol{x}_1) \rangle = \langle \boldsymbol{x}_1 - \boldsymbol{x}_2, \boldsymbol{t}_i(\boldsymbol{x}_2) \rangle = 0 \\ (i = 1, 2, \cdots, N) \end{array} \right\},$$

and estimate $\min \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ on the set. Put this value $R_2$.

**Step 5:** Calculate $R \coloneqq \min\{R_1, R_2\}$.

**Output:** $R$, which estimates $R(\mathcal{M})$.

the score vectors (32) as:

$$\begin{aligned} & \lim_{\Delta x \to 0} \frac{1}{\Delta x} \begin{bmatrix} \cos(\theta + \Delta\theta) - \cos(\theta) \\ \sin(\theta + \Delta\theta) - \sin(\theta) \end{bmatrix} \\ = & \lim_{\Delta x \to 0} \frac{1}{\Delta x} \begin{bmatrix} \cos(\theta + \frac{\Delta x}{r}) - \cos(\theta) \\ \sin(\theta + \frac{\Delta x}{r}) - \sin(\theta) \end{bmatrix} \\ = & \lim_{\Delta x \to 0} \frac{1}{\Delta x} \begin{bmatrix} \frac{\Delta x}{r}\theta - \frac{1}{3}\frac{\Delta x}{r}\theta^3 + \cdots \\ \frac{\Delta x}{r} - \frac{1}{2!}\frac{\Delta x}{r}\theta^2 + \cdots \end{bmatrix} \\ = & \frac{1}{r} \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}. \end{aligned}$$

**Remark G.2.** We suspect the above discussion could be generalised in the language of differential geometry (covariant derivative, connection, scalar curvature and etc...). We would also like to highlight that there are significant mathematical studies showing that curvature has rich connections with the topology of a manifold, with the Gauss-Bonnet theorem being one of them.

### G.2. Verification of Conjecture 4.2

We have the following re-expression:

$$\begin{aligned} & \nabla_x \ln p_t(x) \\ = \quad & \frac{\nabla_x p_t(x)}{p_t(x)} \\ = \quad & \frac{1}{\sigma_t^2 p_t(x)} \int_{\mathcal{M}} (y-x) N(y|x, \sigma_t^2 I) p_0(y) dy. \end{aligned} \tag{33}$$

**Conjecture G.3.** *Suppose $\mathcal{M}$ is a compact oriented manifold embedded in $\mathbf{R}^d$. We predict the following observation: Let $\epsilon > 0$ be the injectivity radius. Let $\boldsymbol{n}$ be an unit outward pointing normal vector to $\partial\mathcal{M}(\epsilon)$. Assume $\epsilon > \sqrt{d}\sigma_t$, $x \in \partial\mathcal{M}(\epsilon)$ and $p_0(y)$ is constant $C$ and greater than $0$ on $\mathcal{M}$. Assume moreover the following condition.*

*(i) For any $y \in \mathcal{M}$ with $(y-x)\cdot \boldsymbol{n} > 0$ there exists $y' \in \mathcal{M}$ and some $c > 0$ such that $-c(y-x) = (y'-x)$.*

*(ii) Assume that for each $y \in \mathcal{M}$ such that $(y-x)\cdot \boldsymbol{n} < 0$, there exists $\tilde{y} \in \mathcal{M}$ and $c > 0$ such that $-c(\tilde{y}-x) = (y-x)$. Then $c \leq 1$.*

*(iii) For any $y \in \mathcal{M}$, $\{c(y-x) | c > 0\} \cap \mathcal{M}$ is a finite set.*

*Then:*

$$\nabla_x p_t(x) \cdot \boldsymbol{n} \leq 0.$$

*Proof.* (this proof is yet informal. Although we only perform this proof for the case $d = 2$ and $\mathcal{M}$ is a curve, we hope it can be done in general dimensions). Performing a change of variables $w = \frac{y-x}{\sigma_t}$ we have:

$$\begin{aligned} \nabla_x p_t(x) \cdot \boldsymbol{n} &= \frac{1}{\sigma_t^2} \int_{\mathcal{M}} (y-x) N(y; x, \sigma_t^2 I) p_0(y) dy \cdot \boldsymbol{n} \\ &= \int_{\frac{\mathcal{M}-x}{\sigma_t}} w N(w; 0, I) p_0(x + \sigma_t w) dw \cdot \boldsymbol{n} \\ &= \int_{\frac{\mathcal{M}-x}{\sigma_t}} \frac{w}{|w|} \cdot \boldsymbol{n} |w| N(w; 0, I) p_0(x + \sigma_t w) dw \\ &= \int_{N_-} \frac{w}{|w|} \cdot \boldsymbol{n} |w| N(w; 0, I) p_0(x + \sigma_t w) dw + \int_{N_+} \frac{w}{|w|} \cdot \boldsymbol{n} |w| N(w; 0, I) p_0(x + \sigma_t w) dw \\ &= \int_{\mathbf{R}^2} \frac{z}{|z|} \cdot \boldsymbol{n} |z| N(z; 0, I) p_0(x + \sigma_t z) \delta_{N_-}(z) dz + \int_{\mathbf{R}^2} \frac{z}{|z|} \cdot \boldsymbol{n} |z| N(z; 0, I) p_0(x + \sigma_t z) \delta_{N_+}(z) dz, \quad (!) \end{aligned}$$

where $\frac{\mathcal{M}-x}{\sigma_t}$ is the image of the manifold $\mathcal{M}$ by a diffeomorphism $y \mapsto \frac{y-x}{\sigma_t}$ and $N_-$ (resp. $N_+$) is $\{w \in \frac{\mathcal{M}-x}{\sigma_t}; w \cdot \boldsymbol{n} < 0 (\text{resp.} > 0)\}$. $dz$ is a volume form of $\mathbf{R}^d$. Let $\theta$ be the angle between $z/|z|$ and $\boldsymbol{n}$. If we use the polar coordinates $(|z_\theta|, \theta) \in (0, \infty] \times [0, 2\pi)$, since $\cos(\theta + \pi) = -\cos(\theta)$, (put $N_z(\theta) := \{(|z|, \theta) \in (0, \infty] \times [0, 2\pi); z \in N$ for some $\theta$ s.t. $\cos\theta = \frac{z}{|z|} \cdot \boldsymbol{n})\}$) we may estimate (!) as follows:

$$\begin{aligned} (!) &= \int_{\pi/2}^{-\pi/2} \cos\theta \left( \int_0^\infty |z_\theta|^2 N(z_\theta : 0, I) \delta_{N_z(\theta)}(|z_\theta|) d|z| \right) d\theta + \int_{-\pi/2}^{\pi/2} \cos\theta \left( \int_0^\infty |z_\theta|^2 N(z_\theta : 0, I) \delta_{N_z(\theta)}(|z_\theta|) d|z| \right) d\theta \\ &= \int_{-\pi/2}^{\pi/2} \cos\theta \left( \sum |z_{\theta_+}|^2 N(z_{\theta_+} : 0, I) - \sum |z_{\theta_-}|^2 N(z_{\theta_-} : 0, I) \right) d\theta \\ &\leq C' \int_{-\pi/2}^{\pi/2} \cos\theta \left( |z_{\theta_+}|^2 N(z_{\theta_+} : 0, I) - |z_{\theta_-}|^2 N(z_{\theta_-} : 0, I) \right) d\theta, \quad (\int) \end{aligned}$$

where $z_{\theta_+} \in N^+$, $z_{\theta_-} \in N^-$ and $z_{\theta_+} = -c_\theta z_{\theta_-}$ for some $c_\theta > 0$. If there is no such $z_{\theta_+}$, we set $z_{\theta_+} = 0$. Also we set $|z_{\theta_+}| N(z_{\theta_+} : 0, I) := \max\{|z_{\theta_+}|^2 N(z_{\theta_+} : 0, I)\}$ and $|z_{\theta_-}| N(z_{\theta_-} : 0, I) := \min\{|z_{\theta_-}|^2 N(z_{\theta_-} : 0, I)\}$. Thus by the

assumption (ii) we may obtain ($\int$). This integral ($\int$) is negative or zero if

$$\left(|z_{\theta_+}|^2 N(z_{\theta_+} : 0, I) - |z_{\theta_-}|^2 N(z_{\theta_-} : 0, I)\right) \leq 0 \tag{34}$$

for any $\theta$. Since $x \in \partial\mathcal{M}(\epsilon)$ and by the assumption (ii), $|z_{\theta_+}| \geq |z_{\theta_-}| \geq \frac{\epsilon}{\sigma_t}$ holds. Since $|z|^2 N(z : 0, I)$ is strictly monotonically decreasing if $|z| \geq \sqrt{2}$, the inequality holds for $|z_{\theta_+}| \geq |z_{\theta_-}| \geq \sqrt{2}$. Thus when $\frac{\epsilon}{\sigma_t} \geq \sqrt{2}$ the result follows. □

### G.3. Escaping time from the tubular neighbourhood

Let $\epsilon > 0$. Let $\mathcal{M}(\epsilon)$ be the $\epsilon$-neighbourhood of a compact oriented Riemannian manifold $\mathcal{M}$ embedded in the Euclidean space $\mathbf{R}^d$ as defined in Definition 3.1. Assume $\mathcal{M}(\epsilon)$ is a tubular neighbourhood. Suppose $p_t(x)$ is a solution to the Fokker-Planck equation (3) with an initial condition $p_0(x) = \delta_M(x)$ here $\delta_M(x)$ is Dirac's density function with its support $\mathcal{M}$. We define a function $\Gamma_{\mathcal{M}(\epsilon)}(t)$ as follows:

$$\Gamma_{\mathcal{M}(\epsilon)}(t) := \int_{\mathcal{M}(\epsilon)} p_t(x) dx. \tag{35}$$

**Proposition G.4.** *Assume $\beta(t) : \mathbf{R}_{\geq 0} \to \mathbf{R}$ is a smooth function and $f(t, x) = \frac{1}{2}\beta(t)f(x)$, $g(t, x) = \sqrt{\beta(t)}$ in (3) ($f(x)$ is some smooth vector field). We have:*

$$\lim_{t\to 0} \frac{\partial}{\partial t} \Gamma_{\mathcal{M}(\epsilon)}(t) = 0$$

*and*

$$\lim_{t\to \infty} \frac{\partial}{\partial t} \Gamma_{\mathcal{M}(\epsilon)}(t) = 0.$$

*Thus there exists at least one $t_c$ in $(0, +\infty)$ such that $\frac{\partial^2}{\partial t^2} \Gamma_{\mathcal{M}(\epsilon)}(t_c) = 0$. Moreover if $\beta(t) > 0$ and*

$$\nabla_x p_t(x) \cdot \boldsymbol{n} - p_t(x) f(x) \cdot \boldsymbol{n} < 0 \tag{36}$$

*for any $x \in \partial\mathcal{M}(\epsilon)$ and any $t \in \mathbf{R}_{>0}$ then $\Gamma_{\mathcal{M}(\epsilon)}(t)$ is strictly monotonically decreasing. Here $\boldsymbol{n}$ a unit outward pointing normal vector field along $\partial M(\epsilon)$.*

*Proof.* (informal) We may compute for $t > 0$:

$$\begin{aligned}
\frac{\partial}{\partial t} \Gamma_{\mathcal{M}(\epsilon)}(t) &= \int_{\mathcal{M}(\epsilon)} \frac{\partial}{\partial t} p_t(x) dx \\
&= \beta(t) \int_{\mathcal{M}(\epsilon)} \left(\nabla_x \cdot p_t(x) f(x) + \Delta_x p_t(x)\right) dx \\
&= \beta(t) \left( - \int_{\partial\mathcal{M}(\epsilon)} p_t(x) f(x) \cdot \boldsymbol{n} ds + \int_{\partial\mathcal{M}(\epsilon)} \nabla_x p_t(x) \cdot \boldsymbol{n} ds \right) \\
&\xrightarrow{t\to 0} 0.
\end{aligned} \tag{37}$$

The second equality follows since $p_t(x)$ satisfies the Fokker-Planck equation (3). The third equality follows from the divergence theorem where $\boldsymbol{n}$ is the unit outward pointing normal vector field along $\partial\mathcal{M}(\epsilon)$. The last limit follows since $\lim_{t\to 0} p_t(x) = \delta_{\mathcal{M}}(x)$ and in particular $\lim_{t\to 0} p_t(x) = 0$ and $\lim_{t\to 0} \nabla_x p_t(x) = 0$ in $\partial\mathcal{M}(\epsilon)$. To be more precise the convergence

of the limit we could make use of the following chain of inequalities:

$$
\begin{aligned}
&\left|\beta(t)\left(-\int_{\partial\mathcal{M}(\epsilon)} p_t(x)f(x)\cdot n ds+\int_{\partial\mathcal{M}(\epsilon)}\nabla_x p_t(x)\cdot\boldsymbol{n}ds\right)\right| \\
&\leq|\beta(t)|\left(\max_{x\in\partial\mathcal{M}(\epsilon)}\{|f(x)|\}\int_{\partial\mathcal{M}(\epsilon)}|p_t(x)|ds+\int_{\partial\mathcal{M}(\epsilon)}|\nabla_x p_t(x)\cdot\boldsymbol{n}|ds\right) \\
&\leq|\beta(t)|\left(\max_{x\in\partial\mathcal{M}(\epsilon)}\{|f(x)|\}\sup_{x\in\partial\mathcal{M}(\epsilon)}|p_t(x)|\int_{\partial\mathcal{M}(\epsilon)}1ds+\sup_{x\in\partial\mathcal{M}(\epsilon)}|\nabla_x p_t(x)|\int_{\partial\mathcal{M}(\epsilon)}1ds\right) \\
&=|\beta(t)|\int_{\partial\mathcal{M}(\epsilon)}1ds\left(\max_{x\in\partial\mathcal{M}(\epsilon)}\{|f(x)|\}\sup_{x\in\partial\mathcal{M}(\epsilon)}|p_t(x)|+\sup_{x\in\partial\mathcal{M}(\epsilon)}|\nabla_x p_t(x)|\right).
\end{aligned}
$$

When $t\to\infty$, $p_t(x_t)$ tends to be stationary i.e., $\lim_{t\to\infty}\frac{\partial}{\partial t}p_t(x_t)=0$. Therefore

$$
\lim_{t\to\infty}\frac{\partial}{\partial t}\Gamma_{\mathcal{M}(\epsilon)}(t)=0.
$$

The existence of an inflection point follows from the mean value theorem. Finally let us show it is strictly monotonically decreasing. The negativity of $\frac{\partial}{\partial t}\Gamma_{\mathcal{M}(\epsilon)}(t)$ follows from (37) and (36). □

## H. Second-order derivative of the function $\Gamma_{\mathcal{M}(\epsilon)}(t)$

The Fokker-Planck equation for a forward diffusion process can be reformulated utilizing the potential function $u_t(x_t)$ as follows.

$$
\begin{aligned}
\frac{\partial}{\partial t}p_t(x_t)&=-\nabla_{x_t}\cdot[f_t(x_t)p_t(x_t)]+\frac{g_t^2}{2}\Delta_{x_t}p_t(x_t), &(38)\\
&=-\nabla_{x_t}\cdot\left[f_t(x_t)p_t(x_t)-\frac{g_t^2}{2}\nabla_{x_t}p_t(x_t)\right], &(39)\\
&=-\nabla_{x_t}\cdot\left[\left(f_t(x_t)-\frac{g_t^2}{2}\nabla_{x_t}\ln p_t(x_t)\right)p_t(x_t)\right], &(40)\\
&=-\nabla_{x_t}\cdot[\nabla_{x_t}u_t(x_t)p_t(x_t)]. &(41)
\end{aligned}
$$

Furthermore, the second-order derivative of the function $\Gamma_{\mathcal{M}(\epsilon)}(t)$ can be expressed in terms of the potential function as

follows.

$$\frac{\partial^2}{\partial t^2}\Gamma_{\mathcal{M}(\epsilon)}(t) = \frac{\partial}{\partial t}\int_{\mathcal{M}(\epsilon)} \frac{\partial}{\partial t} p_t(z)dz, \tag{42}$$

$$= -\frac{\partial}{\partial t}\int_{\mathcal{M}(\epsilon)} \nabla_z \cdot \left[\nabla_z u_t(z) p_t(z)\right] dz, \tag{43}$$

$$= -\frac{\partial}{\partial t}\int_{\partial\mathcal{M}(\epsilon)} \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z)dz, \tag{44}$$

$$= \int_{\partial\mathcal{M}(\epsilon)} -\left[\frac{\partial \nabla_z u_t(z)}{\partial t} \cdot \boldsymbol{n}\right] p_t(z) - \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] \frac{\partial p_t(z)}{\partial t} dz, \tag{45}$$

$$= \int_{\partial\mathcal{M}(\epsilon)} -\left[\nabla_z \frac{\partial u_t(z)}{\partial t} \cdot \boldsymbol{n}\right] p_t(z) + \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] \nabla_z \cdot \left[\nabla_z u_t(z) p_t(z)\right] dz, \tag{46}$$

$$= \int_{\partial\mathcal{M}(\epsilon)} \left[\nabla_z \frac{\|\nabla_z u_t(z)\|^2}{2} \cdot \boldsymbol{n}\right] p_t(z) + \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] \left(\Delta u_t(z) + \nabla_z u_t(z) \cdot \nabla_z \ln p_t(z)\right) p_t(z)dz, \tag{47}$$

$$= \int_{\partial\mathcal{M}(\epsilon)} \left(2\Delta u_t(z) + \nabla_z u_t(z) \cdot \nabla_z \ln p_t(z)\right) \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z)dz, \tag{48}$$

$$= \int_{\partial\mathcal{M}(\epsilon)} \left(2\Delta u_t(z) + f_t(z) \cdot \nabla_z \ln p_t(z) - \frac{g_t^2 G_{\beta_t}(z)}{2}\right) \left[\nabla_z u_t(z) \cdot \boldsymbol{n}\right] p_t(z)dz \tag{49}$$

We here applied $\frac{\partial u_t(z)}{\partial t} + \frac{\|\nabla_z u_t(z)\|^2}{2} = 0$ in (Aurell et al., 2011) Eq.(14).

# I. The detailed description of the experiments

## I.1. Experimental Setup

In our experiments, diffusion was performed for T=1000 steps according to Equation 1, which represents the forward dynamics. Taking the initial states as $S^0, S^1, S^2$, we counted the propotion of particles outside the injectivity radius at each time step. Here, a particle is considered outside if it is located at a point that is at or beyond the injectivity radius from its initial position. There are two methods for counting the proportion of particles. The green line represents the method where once a particle exits the injectivity radius, it is considered outside for all subsequent time steps. The blue line represents the method where a particle is counted as inside if it re-enters the injectivity radius after having exited. Next, using the pre-trained DDPM, backward diffusion was performed according to Equation 2, which represents the backward dynamics. The expected final states are $S^0, S^1, S^2$, and the proportion of particles outside the injectivity radius was counted at each time step. The red dashed line in the figure indicates the theoretical time point at which spontaneous symmetry breaking occurs. To examine the relationship with spontaneous symmetry breaking, we predict that the green line method for counting particles outside the injectivity radius is more appropriate. Therefore, only the green line method is shown for the backward case.

## I.2. Rationale for the Setup

In this experiment, we used the injectivity radius from each point on the manifold to count the particles outside the tubular neighbourhoods. There is room for discussion regarding whether this counting method accurately reflects the concept of tubular neighbourhoods, which we will examine here.

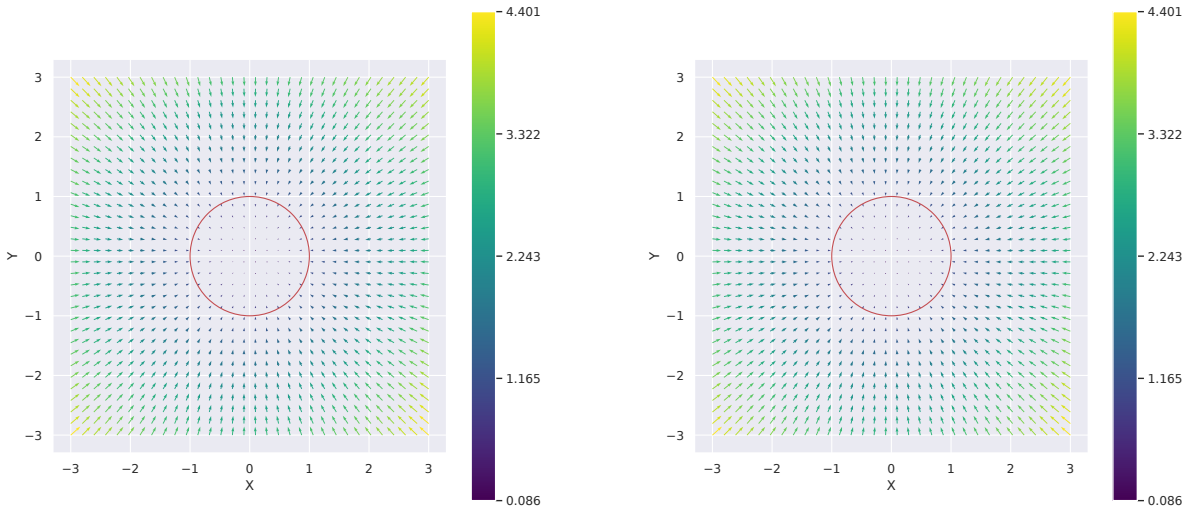
First, as explained in Section 3, a tubular neighbourhood is defined for a manifold $\mathcal{M}$. However, in this experiment, we consider the injectivity radius at each point on the manifold and categorize particles as inside or outside based on whether they are within the injectivity radius from each point. These two concepts do not necessarily coincide. This discrepancy arises because the condition for our experimental setup to reflect the concept of tubular neighbourhoods is that the vector representing the displacement of the particles must be orthogonal to any vector constituting the basis of the tangent plane.

Therefore, the counting method using the injectivity radius in this experiment can be considered a technique that reflects the concept of tubular neighbourhoods in practical applications.

# J. Additional experiments

## J.1. Score Vector field

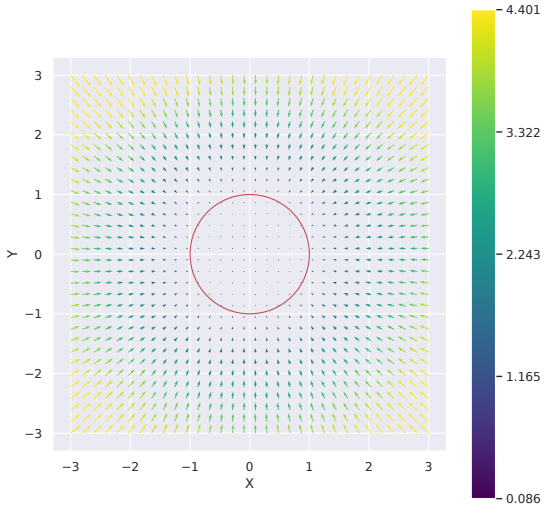
We present additional experiments detailing the score vectors of DDPM. This section includes two experimental setups concerning the score vector field. Firstly, for the 2D $S^1$ case, the experimental setup includes a grid size of $32 \times 32$ and a trained DDPM with $T = 1000$. The training data is $S^1$, with the red circle at the centre representing $S^1$.
Secondly, for the $3D$ $S^2$ case, the experimental setup includes a grid size of $16 \times 16 \times 16$ and a trained DDPM with $T = 1000$. The training data is $S^2$. Except for the grid size and training data, all other settings remain the same.



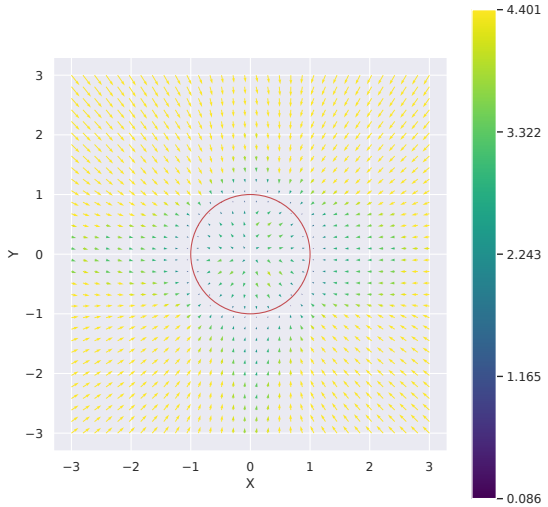

(a) T=0, $S^1$




(b) T=500, $S^1$

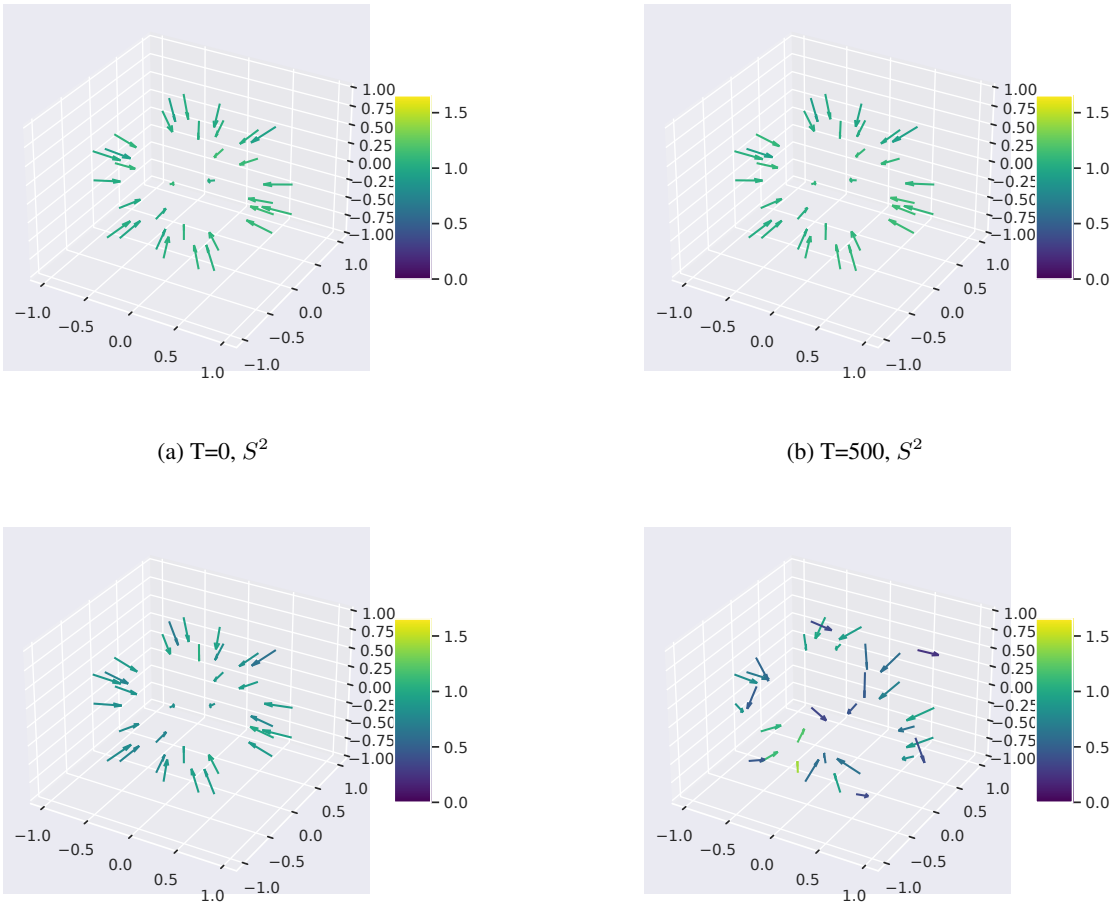



(c) T=785, $S^1$




(d) T=1000, $S^1$

*Figure 10.* Time evolution of score vectors in the backward process of DDPM, $S^1$



(a) T=0, $S^2$

(b) T=500, $S^2$

(c) T=823, $S^2$

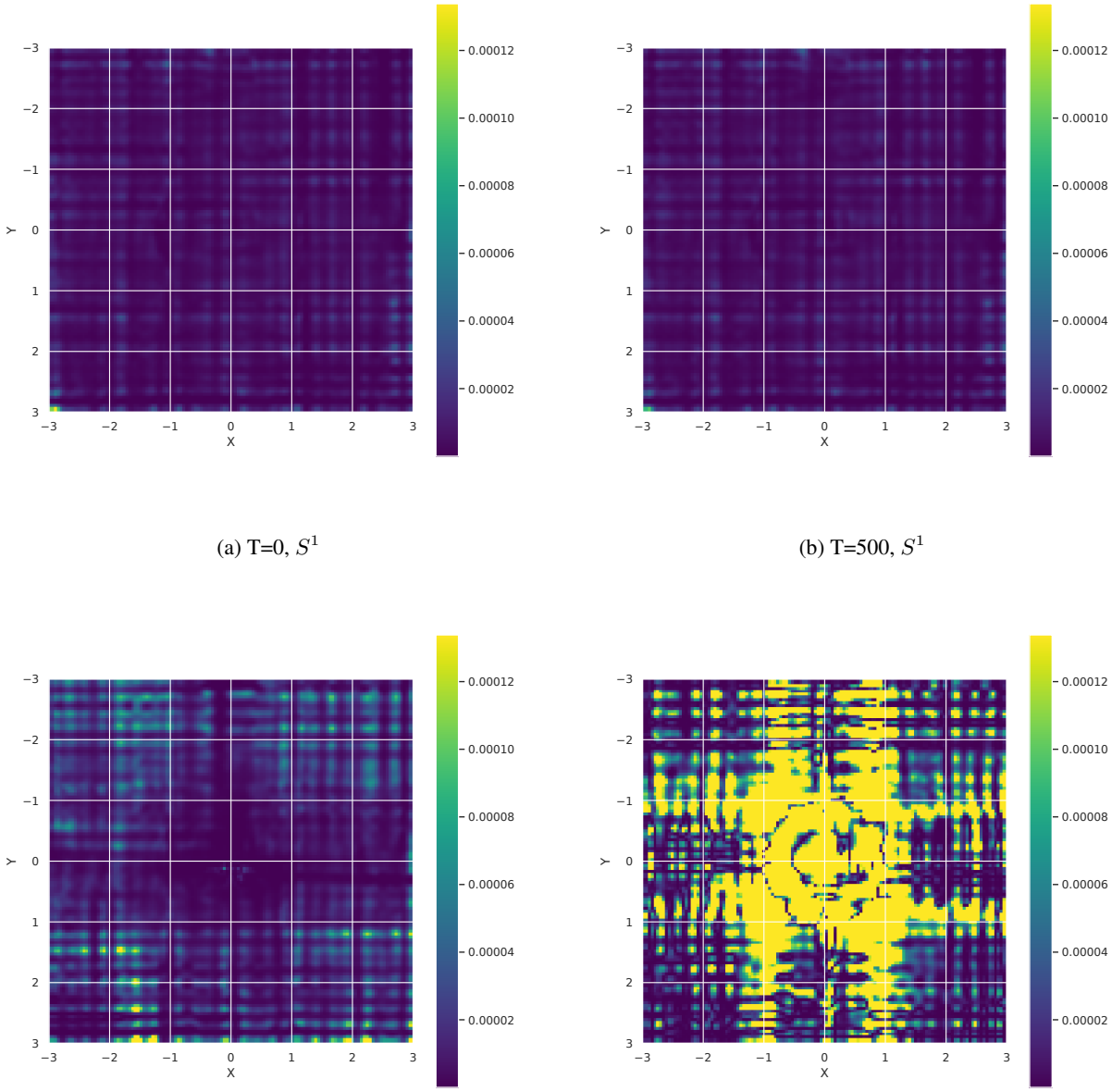
(d) T=1000, $S^2$

*Figure 11.* Time evolution of score vectors in the backward process of DDPM, $S^2$

### J.2. Square of the Jacobian J of the Score Vector Field

In this section, we extend our analysis to the square of the Jacobian $J$ of the score vector field. We utilize updated experimental setups for both the 2D $S^1$ and the 3D $S^2$ cases. For the 2D $S^1$ case, the grid size is $128 \times 128$ with a trained DDPM using $T = 1000$. The training data remains $S^1$, and we compute and analyze the square of the Jacobian of the score vector field for this setup.

Similarly, for the $3D$ $S^2$ case, the grid size is $128 \times 128 \times 128$ with a trained DDPM using $T = 1000$. The training data remains $S^2$, and we compute and analyze the square of the Jacobian of the score vector field for this setup.



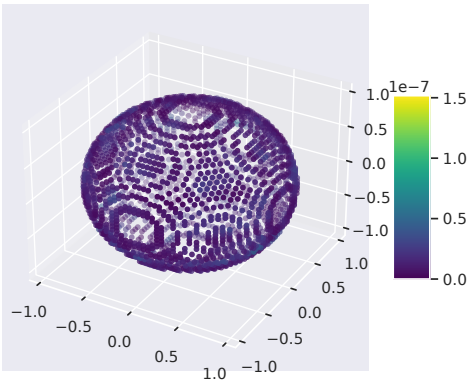

(a) T=0, $S^1$

(b) T=500, $S^1$

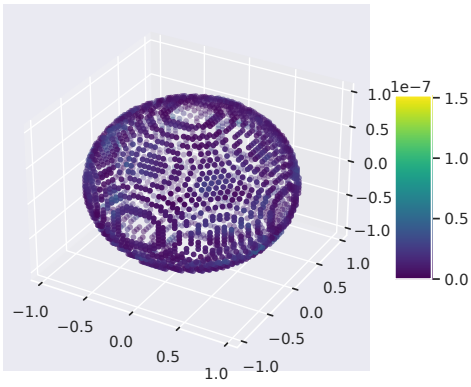
(c) T=785, $S^1$

(d) T=1000, $S^1$

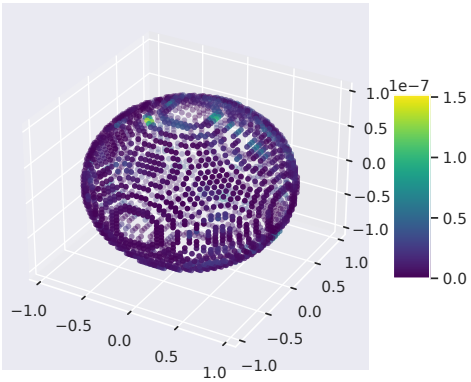
*Figure 12.* Time evolution of the squared Jacobian of score vectors in the backward process of DDPM, $S^1$



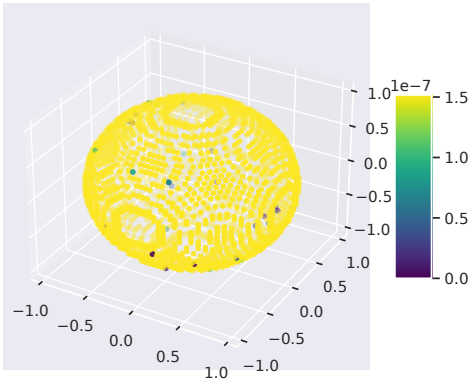

(a) T=0, $S^2$




(b) T=500, $S^2$




(c) T=823, $S^2$

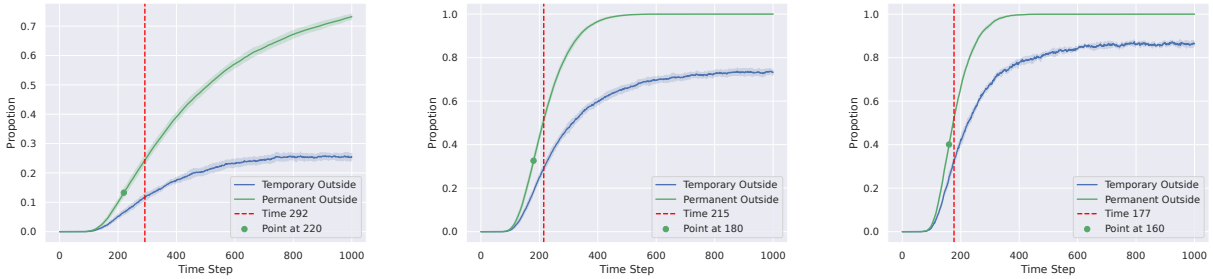



(d) T=1000, $S^2$

*Figure 13.* Time evolution of the squared Jacobian of score vectors in the backward process of DDPM, $S^2$

### J.3. Inflection Points in the Proportion within Tubular neighbourhoods

We analyzed the inflection points in the graph of the green line representing the proportion within tubular neighbourhoods. Each forward process was analyzed for $S^0$, $S^1$, and $S^2$. For the analysis, we first applied smoothing to the data using the Savitzky-Golay filter. Specifically, we processed the data with a window width of 51 and a cubic polynomial. Afterwards, we computed the moving average of the second derivative with a window size of 5, and then estimated the values visually.

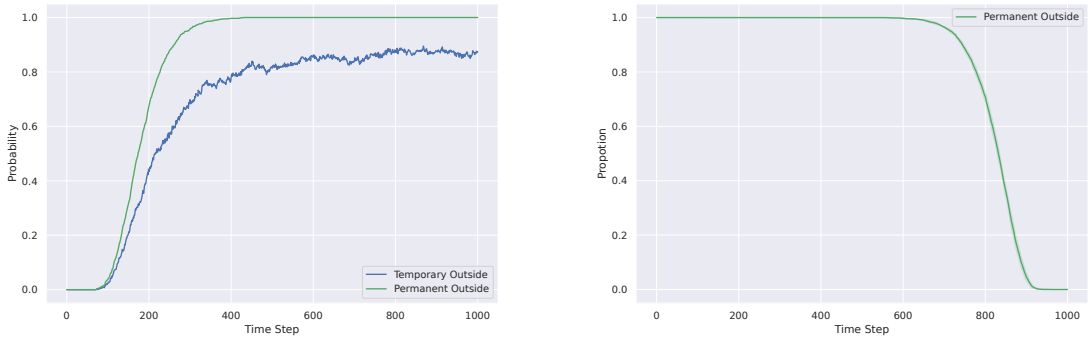



(a) forward diffusion process $S^0$ (b) forward diffusion process $S^1$ (c) forward diffusion process $S^2$

*Figure 14.* The Proportion Outside the Tubular Neighbourhoods Over Time with Symmetry Breaking Time Point and Inflection Points

### J.4. Proportion Analysis Outside the Tubular Neighbourhoods on a Torus

As an extension of the experiment conducted in Section 6, we investigated the proportion of particles outside the tubular neighbourhood using a torus. Here, the torus is defined with a major radius $r' = 2$ and a minor radius $r = 1$. Thus, the supremum of the tubular neighbourhood is $\min\{r, r' - r\} = 1$. For both the forward and backward processes, we plotted the proportion of particles outside the tubular neighbourhood at each time step on the vertical axis. In the transition of the backward process, we used a model trained with DDPM where $T = 1000$.

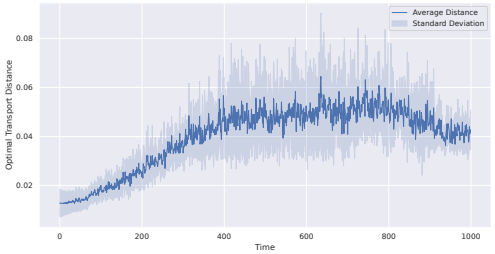



(a) forward diffusion process Torus (b) backward diffusion process Torus

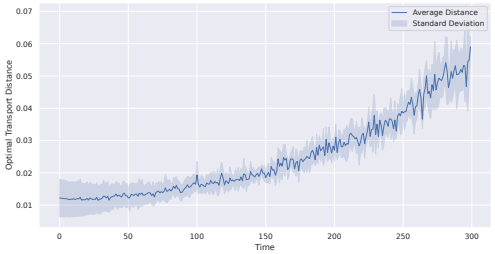
*Figure 15.* Proportion of Particles Outside the Tubular Neighbourhoods Over Time for Both Forward and Backward Processes Using Data on a Torus

### J.5. Evaluation of Reconstruction Error Using the Wasserstein Distance

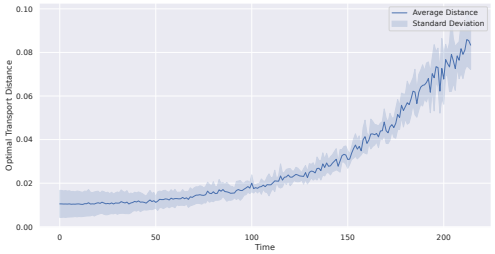
We evaluated the reconstruction error of the data with late initialization using the Wasserstein distance, utilizing a DDPM trained on $S^1$ data. The experiments involved delaying the initialization with a Gaussian distribution by 0, 700, 785, and 900 steps during inference, i.e., the backward process. Subsequently, we prepared 1000 points for each time step and calculated the Wasserstein distance for both the forward process and the backward process.



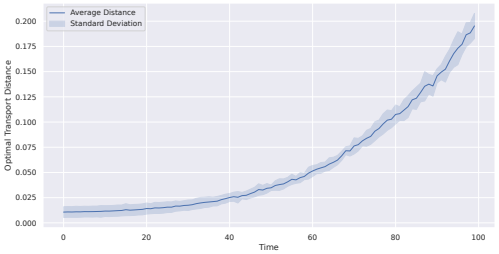

(a) Wasserstein distance for the case where late initialization time is 0.




(b) Wasserstein distance for the case where late initialization time is 700.




(c) Wasserstein distance for the case where late initialization time is 785.




(d) Wasserstein distance for the case where late initialization time is 900.

*Figure 16.* Evaluation of reconstruction error using the Wasserstein distance for $S^1$ data.