Diffusion Models under Group Transformations

Haoye Lu* **Spencer Szabados*** **Yaoliang Yu**
University of Waterloo, Vector Institute

Abstract

In recent years, diffusion models have become the leading approach for distribution learning. This paper focuses on structure-preserving diffusion models (SPDM), a specific subset of diffusion processes tailored for distributions with inherent structures, such as group symmetries. We complement existing sufficient conditions for constructing SPDMs by proving complementary necessary ones. Additionally, we propose a new framework that considers the geometric structures affecting the diffusion process. Leveraging this framework, we design a structure-preserving bridge model that maintains alignment between the model's endpoint couplings. Empirical evaluations on equivariant diffusion models demonstrate their effectiveness in learning symmetric distributions and modeling transitions between them. Experiments on real-world medical images confirm that our models preserve equivariance while maintaining high sample quality. We also showcase the practical utility of our framework by implementing an equivariant denoising diffusion bridge model, which achieves reliable equivariant image noise reduction and style transfer, irrespective of prior knowledge of image orientation.

1 INTRODUCTION

Diffusion models (Song and Ermon, 2019; Ho et al., 2020; Song et al., 2021a,b; Rombach et al., 2022; Karras et al., 2022; Song et al., 2023) have become the leading method in a plethora of generative modelling tasks including image generation (Song and Ermon, 2019; Ho et al., 2020; Song et al., 2021a), audio synthesis (Kong et al., 2021), image segmentation (Baranchuk et al., 2022; Wolleb et al., 2022), image editing and style transfer (Meng et al., 2022; Zhou et al., 2024).

In many generation tasks, the data involved often exhibit inherent "structures" that are invariant – or the mappings between them being equivariant – under a set of transformations. For example, it is commonly assumed that the distribution of photographic images is invariant under horizontal flipping. In tasks such as image denoising or inpainting, where the orientation of an image is not provided, it is natural to require the denoised or inpainted image to retain the same orientation as the input. Namely, the denoising or inpainting processes should exhibit equivariance under rotations and flipping (see Fig 1).

In critical applications, such as medical imaging analysis, these properties must not only be desired but also theoretically guaranteed to ensure model output consistency and prevent the introduction of biases or errors. Diagnostic images, such as X-rays or biopsy assays, are often captured from different orientations (Lafarge et al., 2021; Shao et al., 2023), which has led many diagnostic methods to be designed with invariance to image transformations. These methods are used in tasks such as distinguishing between benign and malignant breast lesions in digitized mammograms (Pohlman et al., 1996; Rangayyan et al., 1997), analyzing blood cells (Lin et al., 1998), and performing digital pathology segmentation (Veeling et al., 2018). Since these methods often require high-resolution images for precise segmentation (Pohlman et al., 1996), denoising techniques are applied to improve image quality. For these techniques to function effectively, they must exhibit perfect equivariance—without it, the overall method's invariance may be compromised.

This paper investigates structure-preserving diffusion models (SPDM), a family of diffusion processes that preserve the group-invariant properties of the distributions. Our framework extends previous research (Yim

*Equal contribution. Haoye proposed the theoretical results, while both Spencer and Haoye dedicated substantial time to model implementation and evaluation. Yaoliang supervised the project and provided valuable guidance.

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

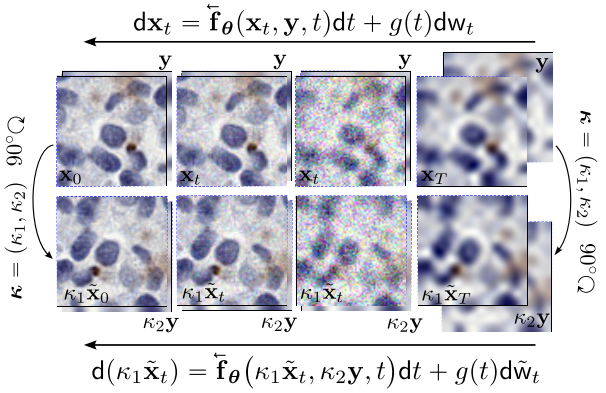


Figure 1: Equivariant inference trajectory: if $(\mathbf{x}_T, \mathbf{y})$ undergos a 90° rotation (via operator $\boldsymbol{\kappa}$), the denoised output precisely mirrors this rotation.

et al., 2023; Xu et al., 2022; Hoogeboom et al., 2022; Qiang et al., 2023; Martinkus et al., 2023) on drift equivariance by incorporating additional factors influencing diffusion trajectories, not just noisy samples; allowing us to characterize structure preserving properties beyond those developed for classical diffusion processes, see Sec 4. Fig 1 illustrates an example of this, for denoising diffusion bridge models (DDBM) (Zhou et al., 2024; Bortoli et al., 2023), where variables $\mathbf{y}$ correspond to starting point $\mathbf{x}_T$ of the backward sampling process. The extension allows us to build a bridge model that captures the equivariant coupling precisely between $\mathbf{x}_T$ and $\mathbf{x}_0$ where a rotation on $\mathbf{x}_T$ results in the same rotation on $\mathbf{x}_0$.

Building upon this generalized framework, we establish an equivalence relationship between the group equivariance of the drifts and the structural preservation of the distributions of induced flows. This development complements existing discussions around structural diffusion models, which have predominantly focused on identifying sufficient conditions. We exemplify the utility of our framework by presenting two equivariant score-based models that achieve theoretically guaranteed capabilities for invariant data generation and equivariant data editing: (1) SPDM+WT, using a weight-tied implementation to reduce training and sampling costs at the expense of image quality, and (2) SPDM+FA, employing Framing-Averaging (Puny et al., 2022) to combine outputs from conventionally trained diffusion models, attaining the same theoretical guarantees while achieving a sample quality comparable, or superior, to standard diffusion models.

Unlike other equivariant implementations of diffusion models that incorporate FA during training (Martinkus et al., 2023; Duval et al., 2023), our method applies FA only during inference, greatly reducing training costs. Empirical studies on both artificial and medical image datasets support our claims. Additionally, we demonstrate the effectiveness of our method for equivariant denoising and style transfer, as shown in Fig 1.

Our code is available at `https://github.com/watml/SPDM`.

2 RELATED WORK

The problem of conditioning neural networks to respect group-invariant (or equivariant) distributions (Shawe-Taylor, 1993) a longstanding issue within the domains of physical modelling, computer vision, and, generative modelling. This is underscored by the widespread utilization of diverse forms of data augmentation, that seek to increase model robustness to perturbation. However, achieving true group invariance (or equivariance) through data augmentation alone is impractical, as it would require an infinite number of samples to guarantee invariance. Consequently, models conditioned solely by data augmentation often fail to fully capture the desired properties (Elesedy and Zaidi, 2021; Gao et al., 2022). As our primary topic is diffusion models, we will devote this section to these works.

The study of group invariance within diffusion models (Song and Ermon, 2019; Song et al., 2021a; Ho et al., 2020; Karras et al., 2022; Kim et al., 2024; Yim et al., 2023), and diffusion bridge models (Bortoli et al., 2021; Liu et al., 2023; Zhou et al., 2024; Lee et al., 2024), has focused primarily on applications in molecule generation (e.g., molecular conformation, and protein backbone generation) (Shi et al., 2021; Xu et al., 2022; Hoogeboom et al., 2022; Yim et al., 2023; Jing et al., 2022; Corso et al., 2023; Martinkus et al., 2023). Most of these approaches can be broadly described as conditioning the diffusion process on a graph prior that represents the unconformed molecule and employing a transformation (applied to the inner molecular atomic distances - such as the relative torsion angle coordinates (Jing et al., 2022)) that produces a group-invariant form (or one that is more robust to the selected group transformations). This thereby, results in a representation that is sufficient to ensure the diffusion process is equivariant. More generally, De Bortoli et al. (2022); Mathieu et al. (2023), and Yim et al. (2023), investigate distribution invariance over more general geometries (e.g., Riemannian manifolds generated by Lie groups). The study of distribution invariance comes about naturally as a result of finding a limiting probability distribution over the geometry in these settings, a requirement for the diffusion process to be well-defined. These methods, which are most similar to ours, operate by designing Gaussian kernels, those that define the diffusion process, that are invariant to select linear isometry groups.

In this work we extend existing theoretical results, developed within the forgoing works, by providing a complete characterization of the necessary and sufficient conditions of the drift and diffusion terms in order to ensure a diffusion process is invariant under linear isometry groups. The diffusion processes we consider encapsulates both regular diffusion models and diffusion bridge models with and without conditioning, treating the conditioning variable as a tensor admitting its own group operations. Our work is expressly more general than existing results which focus primarily on providing sufficient conditions with a single conditioning variable.

3 PRELIMINARY

3.1 Diffusion Processes and Diffusion Bridges

Let $\{\mathbf{x}_t\}_{t=0}^T$ denote a set of time-indexed random variables in $\mathbb{R}^d$ such that $\mathbf{x}_t \sim p_t$, where p_t are the marginal distributions of an underlying diffusion process defined by a stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x}_t = \mathbf{u}(\mathbf{x}_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t, \quad \mathbf{x}_0 \sim p_0(\mathbf{x}_0). \tag{1}$$

Here, $\mathbf{u} : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ is the *drift*, $g : [0,T] \to \mathbb{R}$ is a scalar *diffusion coefficient*, and $\mathsf{w}_t \in \mathbb{R}^d$ denotes a Wiener process. In generative diffusion models, we take $p_0 = p_{\text{data}}$ and $p_T = p_{\text{prior}}$; thereby, the diffusion process constructs a path p_t from p_{data} to p_{prior}.

In practice, $\mathbf{u}$ and g are chosen to accelerate the sampling of $\mathbf{x}_t$ in (1). Table 4 in Appx A lists two popular choices of $\mathbf{u}$ and g, corresponding to the variance preserving (VP, Ho et al. 2020; Song et al. 2021a) and variance exploding (VE, Song et al. 2021b) SDEs.

For these selections, $p(\mathbf{x}_t|\mathbf{x}_0)$ is an easy-to-sample spherical Gaussian, and the sampling of $\mathbf{x}_t$ is carried out by first picking $\mathbf{x}_0 \sim p_0$ and then sampling from $p(\mathbf{x}_t|\mathbf{x}_0)$.

For any SDE, there is a corresponding reverse SDE with the same marginal distribution p_t for all $t \in [0,T]$ (Anderson, 1982). In fact, the forward SDE in (1) has a family of reverse-time SDEs (Zhang and Chen, 2023):

$$\begin{aligned}\mathrm{d}\mathbf{x}_t =& \Big[\mathbf{u}(\mathbf{x}_t, t) - \tfrac{1+\lambda^2}{2} g^2(t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\Big]\,\mathrm{d}t \\ &+ \lambda g(t)\,\mathrm{d}\mathsf{w}_t, \quad \text{for all } \lambda \geq 0.\end{aligned} \tag{2}$$

Setting $\lambda = 1$ in (2) simplifies the equation to the original reverse SDE as derived in (Anderson, 1982). For $\lambda = 0$, the process transforms into a deterministic ODE process, known as the probability flow ODE (PF-ODE, Song et al. 2021b). The only unknown term is the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, which is estimated using a neural network $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, t)$ trained through score matching (Song et al., 2021b) or an equivalent denoising loss (Karras et al., 2022). Subsequently, after training, one can sample $\mathbf{x}_0 \sim p_{\text{data}}(\mathbf{x}_0)$ by solving the SDE (or ODE if $\lambda = 0$) given in (2) starting from $\mathbf{x}_T \sim p_T(\mathbf{x}_T)$.

Instead of building path p_t from a data distribution, $q_{\text{data}}(\mathbf{x}, \mathbf{y})$, to a known prior, diffusion bridges (DBs) create path q_t such that $q_0(\mathbf{x}) = q_{\text{data}}(\mathbf{x})$ and $q_T(\mathbf{y}) = q_{\text{data}}(\mathbf{y})$. For $(\mathbf{x}, \mathbf{y}) \sim q_{\text{data}}(\mathbf{x}, \mathbf{y})$, DBs leverage the distribution p_t induced by (1) with $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{x}_T = \mathbf{y}$ to sample $\mathbf{x}_t$. In this way, $q_t(\mathbf{x}_t) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim q_{\text{data}}(\mathbf{x},\mathbf{y})}\big[p_t(\mathbf{x}_t|\mathbf{x}_0 = \mathbf{x}, \mathbf{x}_T = \mathbf{y})\big]$ and admits the forward SDE:

$$\mathrm{d}\mathbf{x}_t = [\mathbf{u}(\mathbf{x}_t, t) + g(t)^2 \mathbf{h}(\mathbf{x}, \mathbf{x}_T, t)]\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t, \tag{3}$$

given $\mathbf{x}_T = \mathbf{y}$ and $\mathbf{h}(\mathbf{x}, \mathbf{x}_T, t) = \nabla_{\mathbf{x}} \log p(\mathbf{x}_T|\mathbf{x})$, the gradient of the log transition kernel induced by (1). For $\mathbf{u}$ and g in Table 4, $p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)$ can be sampled efficiently. Thus, $\mathbf{x}_t \sim q_t(\mathbf{x}_t)$ can be obtained by first sampling $(\mathbf{x}, \mathbf{y})$ and then $\mathbf{x}_t$.

DBs can be broadly categorized into those that condition on $\mathbf{x}_T$ and those that do not. Bortoli et al. (2023) show that conditioning is necessary for effectively learning the coupling encoded in q_{data}. This is crucial for many practical tasks, such as image denoising, where the denoised image should match the blurry input. For the conditioned DBs (Zhou et al., 2024), the family of the backward SDE, conditioned on $\mathbf{x}_T = \mathbf{y}$, is

$$\begin{aligned}\mathrm{d}\mathbf{x}_t =& [\mathbf{u}(\mathbf{x}_t, t) + g(t)^2 \mathbf{h}(\mathbf{x}, \mathbf{x}_T, t) \\ &- \tfrac{1+\tau^2}{2} g(t)^2 \mathbf{s}(\mathbf{x}_t|\mathbf{x}_T, t)]\,\mathrm{d}t + \tau g(t)\,\mathrm{d}\mathsf{w}_t,\end{aligned} \tag{4}$$

for all $\tau \geq 0$, where $\mathbf{s}(\mathbf{x}_t|\mathbf{x}_T, t) = \nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t|\mathbf{x}_T)$ is the score of the DB's distribution q_t given that $\mathbf{x}_T = \mathbf{y}$.

Notably, besides the noisy sample $\mathbf{x}_t$, the drift terms in (3) and (4) also depend on $\mathbf{y} = \mathbf{x}_T$. This dependency leads us to consider a more general diffusion process:

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t, \quad \mathbf{x}_0 \sim p(\mathbf{x}_0|\mathbf{y}), \tag{5}$$

where $\mathbf{y}$ represents other factors affecting the process and does not need to have the same shape as $\mathbf{x}_t$. When $\mathbf{y}$ consists of zero entries, (5) simplifies to the standard diffusion process in (1). Currently, all structure-preserving diffusion frameworks are based on (1).

3.2 Group Invariance and Equivariance

A set of functions $\mathcal{G} = \{\kappa : \mathcal{X} \to \mathcal{X}\}$ equipped with an associative binary operation $\circ : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$, composition in this case, is called a *group* if (1) for any $\kappa_1, \kappa_2 \in \mathcal{G}$, $\kappa_1 \circ \kappa_2 \in \mathcal{G}$; (2) $\mathcal{G}$ has an identity operator $\mathbf{e}$ with $\mathbf{e} \circ \kappa = \kappa \circ \mathbf{e} = \kappa$; and (3) for any $\kappa \in \mathcal{G}$, there exists an inverse operator κ^{-1} such that $\kappa^{-1} \circ \kappa = \kappa \circ \kappa^{-1} = \mathbf{e}$. For instance, let f_x be the operator that flips images horizontally, then $\mathcal{G} = \{\mathsf{f}_x, \mathbf{e}\}$ is a group as $\mathsf{f}_x^{-1} = \mathsf{f}_x$.

In this paper, we limit our focus to κ that do not alter Wiener processes. From now on we assume $\mathcal{G}$ consists of isometries κ that fix zero. That is, κ is a distance-preserving transformation such that $\|\kappa\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ and $\|\kappa\mathbf{x} - \kappa\mathbf{y}\|_2 = \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$; see Appx B.1 for details. Consequently, any $\kappa \in \mathcal{G}$ can be expressed as an orthogonal matrix $A_\kappa \in \mathbb{R}^{d\times d}$, where $\kappa\mathbf{x} = A_\kappa\mathbf{x}$.

For a distribution with density p defined on $\mathbb{R}^d$, we say p is $\mathcal{G}$-*invariant* if $p(\mathbf{x}) = p(\kappa\mathbf{x})$ for all $\kappa \in \mathcal{G}$. Likewise, for a conditional distribution $p(\mathbf{x}|\mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, we say it is $\mathbf{G}$-*invariant* if $p(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = p(\mathbf{x}|\mathbf{y})$ for all $(\kappa_1, \kappa_2) \in \mathbf{G}$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$. As shorthand, we will write $\mathbf{G} = \{\boldsymbol{\kappa} = (\kappa_1, \kappa_2) | \kappa_1 : \mathbb{R}^m \to \mathbb{R}^m, \kappa_2 : \mathbb{R}^n \to \mathbb{R}^n\}$ defined in $\mathbb{R}^{m+n}$ such that $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = (\kappa_1\mathbf{x}, \kappa_2\mathbf{y}) = (A_{\kappa_1}\mathbf{x}, A_{\kappa_2}\mathbf{y})$ with orthogonal $A_{\kappa_1} \in \mathbb{R}^{m\times m}$ and $A_{\kappa_2} \in \mathbb{R}^{n\times n}$.

Since diffusion models learn a distribution by estimating its score, the following lemma demonstrates that if a distribution is $\mathcal{G}$-invariant, its score must be $\mathcal{G}$-*equivariant*. We also independently present a special case of this result, excluding the condition on $\mathbf{y}$, which aligns with the findings in existing structure-preserving diffusion frameworks (Hoogeboom et al., 2022; Mathieu et al., 2023). Proofs are provided in Appx B.

Lemma 1. *$p(\mathbf{x}|\mathbf{y})$ is $\mathbf{G}$-invariant if and only if $\mathbf{s}(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = \kappa_1 \circ \mathbf{s}(\mathbf{x}|\mathbf{y})$ for all $(\kappa_1, \kappa_2) \in \mathbf{G}$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$. Likewise, $p(\mathbf{x})$ is $\mathcal{G}$-invariant if and only if $\mathbf{s}(\kappa\mathbf{x}) = \kappa \circ \mathbf{s}(\mathbf{x})$ for all $\kappa \in \mathcal{G}$.*

We visualize the meaning of Lem 1 in Fig 2 by illustrating an equivariant (unconditional) distribution under diagonal flipping about the line $x = y$. As can be seen, symmetric points across the diagonal have symmetric scores, showing that flipping a point before or after score evaluation yields the same result.

4 STRUCTURE PRESERVING PROCESSES

In this section, we discuss diffusion processes that maintain the structure of the data distribution. Specifically, given a data distribution that is assumed to be group invariant, we explore the sufficient and necessary configurations of diffusion processes to preserve this invariance throughout the diffusion trajectory. These insights will serve as a foundation for the design of the equivariant neural networks shown in Sec 5. We present our main theoretical results in Prop 1:

Proposition 1. *Given a diffusion process in (5) with $\mathbf{G}$-invariant $p_0(\mathbf{x}_0|\mathbf{y})$, let $[\mathbf{0}]_{p_t}$ be the set of ODE drifts preserving the distribution p_t. Then $p_t(\mathbf{x}_t|\mathbf{y})$ is $\mathbf{G}$-invariant for all $t \geq 0$ if and only if*

$$\kappa_1^{-1} \circ \mathbf{f}(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}, t) - \mathbf{f}(\mathbf{x}, \mathbf{y}, t) \in [\mathbf{0}]_{p_t} \tag{6}$$

for all $t > 0$, $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\boldsymbol{\kappa} \in \mathbf{G}$.

While Prop 1 is presented based on the conditional distribution $p_t(\mathbf{x}_t|\mathbf{y})$, it also applies to the unconditional case by setting $n = 0$. In this scenario, $p_t(\mathbf{x}_t|\mathbf{y})$ reduces to $p_t(\mathbf{x}_t)$, $\mathbf{G}$ simplifies to $\mathcal{G}$ consisting of κ_1, and (6) becomes $\kappa_1^{-1} \circ \mathbf{f}(\kappa_1\mathbf{x}, t) - \mathbf{f}(\mathbf{x}, t) \in [\mathbf{0}]_{p_t}$. Existing structure-preserving diffusion models (Yim et al., 2023; Xu et al., 2022; Hoogeboom et al., 2022; Qiang et al., 2023; Martinkus et al., 2023) are based on the special case that $\kappa_1^{-1} \circ \mathbf{f}(\kappa_1\mathbf{x}, t) - \mathbf{f}(\mathbf{x}, t) = \mathbf{0}$. However, this zero-drift condition is not the only one that preserves p_t. For instance, a spherical Gaussian can be preserved by any circular vector field, where the drift is aligned with the boundary of the level set of the density function.

To gain an intuitive understanding of why (6) results in the $\mathbf{G}$-invariance of $p_t(\mathbf{x}|\mathbf{y})$, we illustrate how the proposition works in DB models for $\mathbf{G} = \{\boldsymbol{\kappa} = (\kappa, \kappa) | \kappa \in \mathcal{G}\}$. For simplicity, we assume $\mathbf{f}(\kappa\mathbf{x}, \kappa\mathbf{y}, t) - \kappa \circ \mathbf{f}(\mathbf{x}, \mathbf{y}, t) = \mathbf{0}$, which implies $\mathbf{f}$ is equivariant. As discussed in Sec 3.1, in DBs, $\mathbf{y}$ corresponds to the starting point $\mathbf{x}_T$ of the backward process, where $\mathbf{x}_T$ can be intuitively understood as a noisy image and $\mathbf{x}_0$ as the corresponding denoised version. When $\mathbf{f}$ is equivariant, it essentially says for an infinitesimal step δ, the transition probability induced by the SDE in (5) satisfies

$$p(\mathbf{a}|\mathbf{b}, \mathbf{x}_T) = p(\kappa\mathbf{a}|\kappa\mathbf{b}, \kappa\mathbf{x}_T). \tag{7}$$

As $p_T(\mathbf{x}_T|\mathbf{x}_T) = p_T(\kappa\mathbf{x}_T|\kappa\mathbf{x}_T)$, which is $\mathbf{G}$-invariant, applying this relationship recursively from $t = T$ to 0 implies $p(\mathbf{x}_t|\mathbf{x}_T) = p(\kappa\mathbf{x}_t|\kappa\mathbf{x}_T)$ for $t \in [0, T]$. (Since the SDE is solved reversely, the base case becomes the invariance of p_T instead of p_0) Intuitively, suppose κ denotes the image flipping operator. This basically says, when input blurry image $\mathbf{x}_T$ is flipped, so is the denoised image $\mathbf{x}_0$. In Fig 3 (**Left**), we visualize the evolution of conditional p_t when the conditioned end point $\mathbf{x}_T$ is flipped w.r.t. $x = 0$. As we can see when $\mathbf{x}_T$ is flipped, the trajectory from $\mathbf{x}_T$ to $\mathbf{x}_t$ is also flipped, which corresponds to the invariance of $p(\mathbf{x}_t|\mathbf{x}_T)$.

For completeness, in Fig 3 (**Right**), we also present an example of unconditional p_t in Prop 1 by setting $n = 0$. In this case, to ensure that p_t is invariant, it suffices that $\mathbf{f}(\kappa_1\mathbf{x}, t) - \kappa_1 \circ \mathbf{f}(\mathbf{x}, t) = \mathbf{0}$. Here, we visualize the evolution of p_t driven by two different diffusion processes with p_0 invariant to flipping with respect to $x = 0$ (that is, $\kappa_1(x) = -x$). In the upper plot, we have drift $\mathbf{f}(x, t) = \frac{1-x}{1-t}$ that pushes x to 1 and is not flip-equivariant for $t \geq 0$. As we observe, for all $t > 0$, p_t is no longer flip-invariant, which corroborates Prop 1. In contrast, the lower plot illustrates the VP-SDE (see Table 4) with $\alpha_t = 1 - t$ for $t \in (0, 1)$. The drift here, $\mathbf{f}(x, t) = -\frac{x}{2(1-t)}$ is is flip-equivariant as $\mathbf{f}(-x, t) = -\mathbf{f}(x, t)$. As shown, p_t has a symmetric

density for all $t \geq 0$, which is also aligned with Prop 1.

In summary, Prop 1 shows for conditional p_t, (6) ensures the coupling relationship between condition $\mathbf{y}$ and the noise sample $\mathbf{x}_t$ is $\mathbf{G}$-invariant. In contrast, for unconditional p_t, it ensures the sample $\mathbf{x} \sim p_t$ follows the same distribution when transformed by $\kappa \in \mathcal{G}$.

The following proposition generalize this result to all groups consisting of linear isometries and linear drifts of the form $\mathbf{u}(\mathbf{x}, t) = u(t)\mathbf{x}$.

Proposition 2. *Assume $\mathbf{u}(\mathbf{x}, t) = u(t)\mathbf{x}$ for some scalar function $u : \mathbb{R} \to \mathbb{R}$. Given any group $\mathcal{G}$ (or $\mathbf{G}$) composed of linear isometries, if the unconditional p_t induced by (1) is G-invariant at $t = 0$, then it is $\mathcal{G}$-invariant for all $t \geq 0$. Likewise, if the conditional $q_t(\mathbf{x}_t|\mathbf{x}_T)$ induced by (3) is $\mathbf{G}$-invariant at $t = 0$ then it is $\mathbf{G}$-invariant for all $t \geq 0$.*

Since the drift terms of both VP and VE-SDE in Table 4 (Appx A) take the form $u(t)\mathbf{x}$, Prop 2 indicates that the induced diffusion process and the corresponding diffusion bridges are structure-preserving for any group composed of linear isometries.

5 STRUCTURE PRESERVING MODELS

In this section, we explore applying the insights from Sec 4 to ensure the data generated by SPDM adheres to a $\mathbf{G}$-invariant distribution. As mentioned in Sec 3, sampling a diffusion model involves solving the SDE in (2) or (4) by estimating the score using a neural network $\mathbf{s}_{\boldsymbol{\theta}}$. We will discuss several effective methods to design and train $\mathbf{s}_{\boldsymbol{\theta}}$ so that it meets the properties outlined in Prop 1, achieving theoretically guaranteed structure-preserving sampling. A summary of these methods limitations can be found in Appx I.

5.1 Structure Preserving Sampling

Unconditioned Distribution Sampling. By Prop 1, if a diffusion process is structure preserving, p_t is $\mathcal{G}$-invariant for all $t \geq 0$. So given the prior distribution p_T is $\mathcal{G}$-invariant and by Lem 1, for all $t \geq 0$, the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is $\mathcal{G}$-equivariant. Thus, if the score estimator $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, t)$ perfectly learns the $\mathcal{G}$-equivariant property and satisfies (6), the drift of backward SDE (2):

$$\overleftarrow{\mathbf{f}}_{\boldsymbol{\theta},\lambda}(\mathbf{x}_t, t) = \mathbf{u}(\mathbf{x}_t, t) - \tfrac{1}{2}(1 + \lambda^2) g^2(t)\ \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, t) \quad (8)$$

also satisfies (6) as $[\mathbf{0}]_{p_t}$ is closed under addition (see Appx B.3). Applying Prop 1 with reversed t, we can then conclude that the generated samples must follow a $\mathcal{G}$-invariant distribution.

Equivariant Style-transfer Through Diffusion Bridges Conditioned on $\mathbf{x}_T$. When the drift $\mathbf{u}(\mathbf{x}, t)$ of original SDE in (1) satisfies (6), given stacked group $\mathbf{G} = \{(\kappa, \kappa) | \kappa \in \mathcal{G}\}$, we can show that the drift $\mathbf{u}(\mathbf{x}_t, t) + g(t)^2 \mathbf{h}(\mathbf{x}_t, \mathbf{x}_T, t)$ in (3) also satisfies (6) (see Lem 13 in Appx B.5 for the proof). As a result, by Prop 1, $p_t(\mathbf{x}_t|\mathbf{x}_T)$ is $\mathbf{G}$-invariant for all $t \in [0, T]$ and thus by Lem 1, its score is equivariant and thus satisfies (6). In this way, if the score estimator $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, \mathbf{x}_T, t)$ perfectly learns the equivariant property such that $\mathbf{s}_{\boldsymbol{\theta}}(\kappa \mathbf{x}_t, \kappa \mathbf{x}_T, t) = \kappa \circ \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, \mathbf{x}_T, t)$, the drift of reverse-time SDE (4) satisfies (6); therefore, the invariant coupling between $\mathbf{y} = \mathbf{x}_T$ and $\mathbf{x}_t$ is preserved for all $t \in [0, T]$ during the sampling process.

Building on this observation, to ensure the generated data preserves the necessary geometric structure, it suffices to train a group equivariant score estimator $\mathbf{s}_{\boldsymbol{\theta}}$. We present two theoretically guaranteed $\mathcal{G}$-equivariant implementations of $\mathbf{s}_{\boldsymbol{\theta}}$, SPDM+WT and SPDM+FA.

Weight Tying (SPDM+WT). Currently, most existing diffusion models are based on the U-Net backbone (Salimans et al., 2017; Ronneberger et al., 2015). As the only components that are not equivariant are CNNs, we replace them with group-equivariant CNNs (Cohen and Welling, 2016; Ravanbakhsh et al., 2017; Esteves et al., 2018; Kondor and Trivedi, 2018; Knigge et al., 2022; Yarotsky, 2022) to make the entire network equivariant.

In particular, as we only consider linear isometry groups, we can make CNNs equivariant by tying the weights of the convolution kernels k, which will also reduce the total number of parameters and improve the computation efficiency (Ravanbakhsh et al., 2017). (For more general groups, refer to Cohen and Welling (2016); Knigge et al. (2022) for methods to make CNNs $\mathcal{G}$-equivariant.) We provide more details on our selections of weight-tied kernels in Appx C.

Output Combining (SPDM+FA). When $\mathcal{G}$ contains finite elements, we can achieve $\mathcal{G}$-equivariance through frame averaging (FA) (Puny et al., 2022), leveraging the following fact: for any function $\mathbf{r} : \mathbb{R}^d \to \mathbb{R}^d$,

$$\tilde{\mathbf{r}}(\mathbf{x}, \mathbf{y}) = \tfrac{1}{|\mathcal{G}|} \textstyle\sum_{\kappa \in \mathcal{G}} \kappa^{-1}\ \mathbf{r}(\kappa \mathbf{x}, \kappa \mathbf{y}) \quad (9)$$

is $\mathcal{G}$-equivariant, where $|\mathcal{G}|$ denotes the number of elements in $\mathcal{G}$ and the second argument of $\mathbf{r}$ can be discarded for the approximation of the score not conditioned on $\mathbf{y}$. Based on this fact, we can obtain an equivariant estimator $\tilde{\mathbf{s}}_{\boldsymbol{\theta}}(\cdot, t)$ of the score by setting $\mathbf{r}(\cdot) = \mathbf{s}_{\boldsymbol{\theta}}(\cdot, t)$. Note that unlike other FA-based diffusion models (Martinkus et al., 2023; Duval et al., 2023), our method trains $\mathbf{s}_{\boldsymbol{\theta}}(\cdot, t)$ using regular score-matching and only adopts FA during inference time. This design significantly saves training costs. To see why FA is not necessary during training, we note that

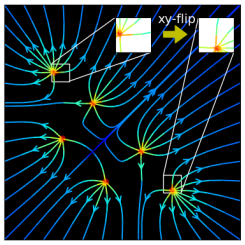


Figure 2: The vector fields of score functions that are equivariant under xy-flip.

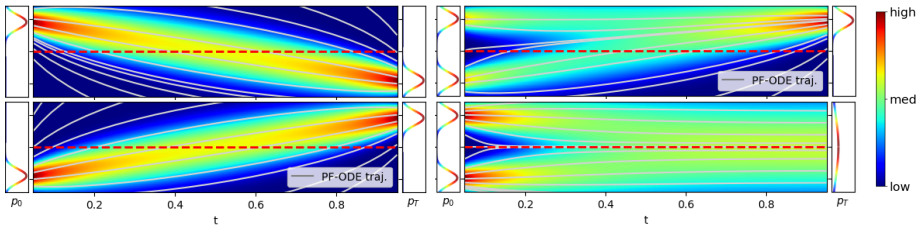


Figure 3: **Left:** The evolution of p_t driven by DB processes induced by the VE-SDE in Table 4 (Appx A) conditioned on the end point $\mathbf{x}_T = -1$ (upper) and $\mathbf{x}_T = 1$ (lower). **Right:** The upper plot has $f(x,t) = \frac{1-x}{1-t}$ and $g(t) = 1$. The lower is the VP-SDE in Table 4 (Appx A) with $\alpha_t = 1 - t$.

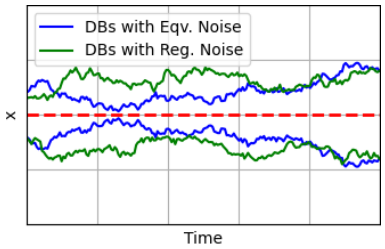


Figure 4: The trajectories of DBs with and without equivariant noise.

Table 1: Equivariance of DB outputs when enforcing equivariant drift by frame-averaging (FA) and adding equivariant noise (EN).

	LYSTO	CT-PET
Model	$\Delta\hat{\mathbf{x}}_0 \downarrow$	$\Delta\hat{\mathbf{x}}_0 \downarrow$
DDBM (Baseline)	52.44	164.91
DDBM+FA	44.50	153.39
DDBM+EN	31.93	70.52
DDBM+FA+EN (SPDM+FA)	**0.00**	**0.00**

our previous discussion has shown that the ground truth score function is equivariant; therefore, when $\mathbf{s}_{\boldsymbol{\theta}}$ is well trained, estimator $\kappa^{-1}\mathbf{s}_{\boldsymbol{\theta}}(\kappa\,\cdot, t)$ for $\kappa \in \mathcal{G}$ will produce very similar output, estimating the value of the score function at $\mathbf{x}_t$ (given $\mathbf{y}$). Thus, their average is also a valid estimator. To boost the model's performance, additional regularizers can be used to encourage $\kappa^{-1}\mathbf{s}_{\boldsymbol{\theta}}(\kappa\,\cdot, t)$ for $\kappa \in \mathcal{G}$ to have the same output; we discuss this technique with extra details in Appx D.

5.2 $\mathcal{G}$-equivariant Trajectory

The sampling of SPDM is governed by the SDE in (2) or (4), collectively written as

$$\mathrm{d}\mathbf{x}_t = \overleftarrow{\mathbf{f}}_{\boldsymbol{\theta},\lambda}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}t + \lambda g(t)\,\mathrm{d}\mathbf{w}_t \tag{10}$$

where $\mathbf{y}$ can be optionally discarded. In practice, the sampling process solves (10) iteratively through

$$\mathbf{x}_{i-1} \leftarrow \overleftarrow{\mathbf{f}}_{\boldsymbol{\theta},\lambda}(\mathbf{x}_i, \mathbf{y}, t_i)(t_{i-1} - t_i) + \lambda g(t_i)\sqrt{t_i - t_{i-1}}\boldsymbol{\epsilon}_i \tag{11}$$

with preset time steps $\{t_i\}_{i=1}^n$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, I)$.

While the techniques discussed in Sec 5.1 ensure the equivariance of $\overleftarrow{\mathbf{f}}_{\boldsymbol{\theta},\lambda}$, they do not guarantee the equivariance of the sampled noise sequence $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$, in a per-sample sense. As a result, the sample trajectory $\mathbf{x}_t$ may not be equivariant. This is visualized in Fig 4 with green curves, where the drifts of DBs are equivariant to the flip about $x = 0$ but the trajectory is not. Due to this asymmetry, the output of the SPDM will not be equivariant. One option to address this problem is to adopt ODE sampling by setting $\lambda = 0$. However, this method is not always preferred as SDE sampling can significantly improve image quality (Karras et al., 2022; Song et al., 2021b; Zhou et al., 2024). For $\lambda > 0$, we also need $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$ to be equivariant in the sense that for $\kappa \in \mathcal{G}$, if the starting point $\mathbf{x}_n$ is updated to $\kappa\mathbf{x}_n$, then $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$ is also updated to $\{\tilde{\boldsymbol{\epsilon}}_i\}_{i=1}^n$ with $\tilde{\boldsymbol{\epsilon}}_i = \kappa\boldsymbol{\epsilon}_i$. In this way, the trajectory becomes equivariant, as shown by the blue curves in Fig 4. Table 1 presents the effectiveness of sampling from an equivariant model with and without the noise satisfying the equivariant property. As shown, perfect equivariance is achieved only by combining equivariant noise (EN) and FA as seen in the SPDM+FA implementation.

In Appx E, we present a simple technique to achieve EN by fixing the random seed and matching some artificial features between $\mathbf{x}_n$ and $\boldsymbol{\epsilon}_n$. In our empirical study, we use this method to inject noise into $\mathbf{x}_t$ as shown in Fig 1 so that the rotation of the input results in a precise output rotation.

6 EMPIRICAL STUDY

In this section, we present experiments demonstrating the effectiveness of the methods from Sec 5. The results support our theoretical work in Sec 4 and offer additional insights.

Datasets. To evaluate the performance and equivariance capabilities of our models over image generation tasks, we adopt the rotated MNIST (Larochelle et al., 2007), LYSTO (Jiao et al., 2023), and ANHIR (Borovec et al., 2020) datasets. In order to validate equivariant trajectory sampling of our models we evaluate denoinsing LYSTO images, and style transfer from CT scan images to PET scan images of the same patient from the CT-PET dataset (Gatidis et al., 2022). See Appx F for a more detailed discussion on the datasets used.

Models. We implement regular diffusion models and bridge models (DDBM) (Zhou et al., 2024) based on VP-SDEs (Ho et al., 2020; Song et al., 2021a), which are structure-preserving with respect to C_4, D_4, and flipping, as per Prop 2. For generation tasks, we present the performance of the standard diffusion model, VP-SDE, as a baseline, along with SP-GAN (Birrell et al., 2022), the only GAN-based model with theoretical group invariance guarantees. We also report the mean performance of GE-GAN (Dey et al., 2021). The tested models and their invariance and equivariance properties are summarized in Table 5 (Appx G).

For style-transfer tasks, in addition to the original DDBM implementation (Zhou et al., 2024), we report the performance of the popular style-transfer method Pix2Pix (Isola et al., 2017) and the unconditional diffusion bridge model I^{2}SB (Liu et al., 2023). For the denoising task on LYSTO, all models use pixel-space implementations. For the CT-PET dataset, all models except Pix2Pix are trained in a latent space, with images first encoded by a fine-tuned pretrained VAE from Stable Diffusion (Rombach et al., 2022). FA was applied during fine-tuning and inference to ensure equivariance.

All models, except SPDM-WT, are trained with data augmentation using randomly selected operators from their respective groups. We apply both non-leaky augmentation as in EDM (Karras et al., 2022) and self-conditioning (Chen et al., 2023) to improve diffusion model sample quality. For sampling we make use of both the DDPM and DDIM samplers (Ho et al., 2020; Song et al., 2021a). Additional model configuration details can be found in Appx G with training resources and training times in Appx G.6.

6.1 Image Generation Tasks

To demonstrate our models are able to match or exceed expected performance over the listed image generation datasets, we report the FID score (Heusel et al., 2017) of each in Table 2. To ensure consistency, we reproduced the results of SP-GAN and GE-GAN[1] and computed the FID using the standard InceptionV3 model. Details on our FID calculation are provided in Appx H. Samples from each model are presented in Appx J. Table 2 shows that diffusion-based models consistently outperform SP-GAN across all datasets, with SPDM+WT and SPDM+FA performing on pair or better.

Table 2: Model Comparison on Rotated MNIST, LYSTO and ANHIR datasets.

Model	Rotated MNIST					
	FID↓				Inv-FID↓	$\Delta\hat{\mathbf{x}}_0\downarrow$
	1%	5%	10%	100%	100%	100%
VP-SDE	5.97	**3.05**	3.47	2.81	2.21	36.98
SPDM+WT	5.80	3.34	3.57	3.50	2.20	**0.00**
SPDM+FA	**5.42**	3.09	**2.83**	**2.64**	**2.07**	**0.00**
SP-GAN	149	99	88	81	–	–
SP-GAN (Reprod.)	16.59	11.28	9.02	10.95	19.92	–
GE-GAN	–	–	4.25	2.90	–	–
GE-GAN (Reprod.)	15.82	7.44	5.92	4.17	58.61	–

Model	LYSTO			ANHIR		
	FID↓	Inv-FID↓	$\Delta\hat{\mathbf{x}}_0\downarrow$	FID↓	Inv-FID↓	$\Delta\hat{\mathbf{x}}_0\downarrow$
VP-SDE	7.88	0.66	20.77	8.03	0.57	39.82
SPDM+WT	12.75	**0.59**	**0.00**	11.73	0.43	**0.00**
SPDM+FA	**5.31**	0.6	**0.00**	**7.57**	**0.31**	**0.00**
SP-GAN	192	–	–	90	–	–
SP-GAN (Reprod.)	16.29	0.66	–	17.12	0.28	–
GE-GAN	3.90	–	–	5.19	–	–
GE-GAN (Reprod.)	23.20	27.84	–	14.16	6.87	–

Table 3: Model Comparison on LYSTO denoising and CT-PET style transfer datasets.

Model	LYSTO				CT-PET			
	FID↓	$L_1\downarrow$	SSIM↑	$\Delta\hat{\mathbf{x}}_0\downarrow$	FID↓	$L_1\downarrow$	SSIM↑	$\Delta\hat{\mathbf{x}}_0\downarrow$
DDBM	17.28	0.076	0.696	52.44	18.13	**0.041**	0.861	164.91
SPDM+FA	**16.21**	**0.071**	0.721	**0.00**	**17.74**	0.042	0.860	**0.00**
Pix2Pix	78.43	0.087	0.654	113.63	20.26	0.043	**0.862**	172.11
I^{2}SB	20.45	0.073	**0.722**	105.83	27.51	0.051	0.832	96.83

Validating Equivariance. To quantify the degree of $\mathcal{G}$-invariance of the learned sampling distribution, we introduce a metric called *Inv-FID*. Given a set of sampled images $\mathcal{D}_s$ from $\hat{p}_0$, Inv-FID calculates the maximum FID between $\kappa_1(\mathcal{D}_s)$ and $\kappa_2(\mathcal{D}_s)$ for $\kappa_1, \kappa_2 \in \mathcal{G}$. If $\hat{p}_0$ is perfectly $\mathcal{G}$-invariant, applying any $\kappa \in \mathcal{G}$ to its outcomes will leave the resulting distribution unchanged. Thus, the closer the FID between $\kappa_1(\mathcal{D}_s)$ and $\kappa_2(\mathcal{D}_s)$ is to zero, the more $\mathcal{G}$-invariant $\hat{p}_0$ is. As shown in Table 2, diffusion models with theoretical guarantees, that is SPDM+WT and SPDM+FA, tend to achieve lower scores. Interestingly, the differences in Inv-FID scores across diffusion models are relatively

[1]We note that the reproduced FID of GE-GAN is significantly higher than reported by Dey et al., as their score is based on a customized InceptionV3 finetuned on LYSTO and ANHIR. While included in the table for reference, these scores are not comparable with other FIDs.

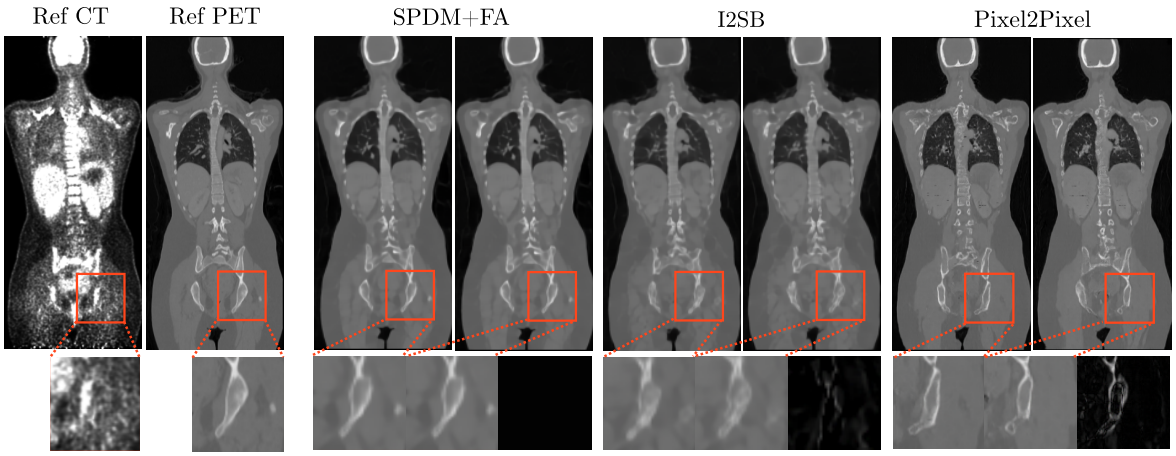


Figure 5: PET images generated by SPDM+FA, I2SB, and Pixel2Pixel for the CT-PET style transfer task. The leftmost group shows the input CT and ground-truth PET.

small, indicating that these models naturally learn invariant properties. Therefore, in situations where invariance in the sampling distribution is not crucial, standard diffusion models may be sufficient.

Validating Equivariant Sampling Trajectory. To empirically validate our methods' theoretical guarantees of equivariant sampling trajectories, we implemented an image-denoising task using SDEdit (Meng et al., 2022), as shown in Fig 1. Given a low-resolution (or corrupted) image $\tilde{\mathbf{x}}_0$, we add equivariant noise to obtain $\mathbf{x}_t$ using the technique from Sec 5.2. This technique is also applied when solving backward SDEs to obtain the denoised image $\mathrm{dn}(\tilde{\mathbf{x}}_0)$, where dn represents the denoising process. As discussed in Sec 5.2, if a diffusion model is $\mathcal{G}$-equivariant, we should have $\mathrm{dn}(\kappa\,\tilde{\mathbf{x}}_0) - \kappa\,\mathrm{dn}(\tilde{\mathbf{x}}_0) \approx \mathbf{0}$ for all $\kappa \in \mathcal{G}$. In Table 2, we report the average maximum pixel-wise distance $\Delta\mathbf{x}_0$ between $\mathrm{dn}(\kappa\,\tilde{\mathbf{x}}_0)$ and $\kappa\,\mathrm{dn}(\tilde{\mathbf{x}}_0)$ over 16 randomly sampled corrupted $\tilde{\mathbf{x}}_0$ images. κ is randomly picked for each $\tilde{\mathbf{x}}_0$. The results show that theoretically equivariant models consistently have nearly zero $\Delta\mathbf{x}_0$, while models without theoretical guarantees produce significantly different outputs, which could be problematic in applications like medical image analysis. Likewise, in Table 3, for a model $\mathbf{m}_\theta$, we adopt a similar idea to measure its equivariance $\Delta\hat{\mathbf{x}}_0$ by reporting the average maximum pixel-wise distance between $\mathbf{m}_\theta(\kappa\,\tilde{\mathbf{y}})$ and $\kappa\,\mathbf{m}_\theta(\tilde{\mathbf{y}})$, where $\mathbf{y}$ is the input.

Results. Among models with theoretically guaranteed structure-preserving properties, SPDM+WT struggles to achieve FID scores comparable to FA methods on complex datasets like LYSTO. This is likely due to the weight-tying technique limiting the model's expressiveness and optimization. In contrast, SPDM+FA maintains sample quality and achieves the best performance on most datasets. This result corroborates our discussion made in Sec 5.1, it being sufficient to train a score-based model using regular score-matching and combine the score-based model's outputs during the inference time to ensure equivariance without compromising the model's performance.

6.2 Equivariant Image Style Transfer Tasks

We compare the performance of Pix2Pix (Isola et al., 2017), I^2SB (Liu et al., 2023), and our SPDM+FA models on the tasks of LYSTO image denosing and style-transfer converting a CT to PET scan image, as described at the start of the section. The results are shown in Fig 5. In addition to FID for measuring sample quality and $\Delta\hat{\mathbf{x}}_0$ for measuring the model's equivariance, we report the L_1 loss between the output and ground truth for local structure similarity and SSIM (Wang et al., 2004) for global feature alignment.

Results. SPDM+FA achieves nearly perfect group equivariance for both tasks, the best scores under most measures, and close-to-the-best scores in the rest. A qualitative sample comparison is provided in Fig 5. Two images are generated from each model: the left conditioned on the original CT and the right an x-flipped CT that is x-flipped, reverted, again after generation. Perfect equivariance would result in two identical images. The bottom row zooms in on a selected area, showing the difference between the two samples. SPDM+FA's black patch indicates perfect equivariance, while the I2SB and Pixel2Pixel's, with white pixels, indicate imperfect equivariance.

Beyond its perfect equivariance and the high image quality reflected in SPDM+FA's low FID score (Table 3), the visualizations in Fig 5 further show that our method generates PET images that closely match the ground truth, particularly in preserving bone shape. In

contrast, other methods suffer from lower reconstruction accuracy and significant bone shape distortions when the input CT is flipped. Such issues could lead to misdiagnoses and unreliable clinical conclusions.

These findings validate the effectiveness of the techniques introduced in Sec 5 and reinforce our framework's applicability to diffusion bridges, guiding the development of equivariant bridge models.

7 DISCUSSION

In this paper, we investigated structure-preserving diffusion models (SPDM), an extended diffusion framework that accounts for invariants in the diffusion process. This extension allows us to effectively characterize the structure-preserving properties of a broader range of diffusion processes, including diffusion bridges used by DDBM (Zhou et al., 2024). We presented a characterization of the drift terms that achieve a structure-preserving process, complementing existing work that primarily focuses on sufficient conditions. Based on the developed theoretical insights, we discussed several effective techniques to ensure the invariant distributions of samples and the equivariant properties of diffusion bridges. Empirical results on image generation and style-transfer tasks support our theoretical claims and demonstrate the effectiveness of the proposed methods in achieving structure-preserving sampling while maintaining high image quality.

Acknowledgements

We would like to thank, Neel Dey, for providing the preprocessed ANHIR dataset used in Dey et al. (2021) and for clarifying some details around how the computation of FID was carried out within the forgoing paper. We also thank the reviewers and the area chair for the constructive comments. We gratefully acknowledge funding support from NSERC and the Canada CIFAR AI Chairs program. Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute.

References

Anderson, B. D. (1982). Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326.

Baranchuk, D., Voynov, A., Rubachev, I., Khrulkov, V., and Babenko, A. (2022). Label-efficient semantic segmentation with diffusion models. In *International Conference on Learning Representations*.

Birrell, J., Katsoulakis, M., Rey-Bellet, L., and Zhu, W. (2022). Structure-preserving GANs. In *Proceedings of the 39th International Conference on Machine Learning*, pages 1982–2020.

Borovec, J., Kybic, J., Arganda-Carreras, I., Sorokin, D. V., Bueno, G., Khvostikov, A. V., Bakas, S., Chang, E. I.-C., Heldmann, S., Kartasalo, K., Latonen, L., Lotz, J., Noga, M., Pati, S., Punithakumar, K., Ruusuvuori, P., Skalski, A., Tahmasebi, N., Valkonen, M., Venet, L., Wang, Y., Weiss, N., Wodzinski, M., Xiang, Y., Xu, Y., Yan, Y., Yushkevich, P., Zhao, S., and Muñoz-Barrutia, A. (2020). ANHIR: Automatic non-rigid histological image registration challenge. *IEEE Transactions on Medical Imaging*, 39(10):3042–3052.

Bortoli, V. D., Guan-Horng Liu, T. C., Theodorou, E. A., and Nie, W. (2023). Augmented bridge matching. arXiv:2311.06978.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*.

Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. (2023). Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*.

Cohen, T. and Welling, M. (2016). Group equivariant convolutional networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2990–2999.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. S. (2023). Diffdock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*.

De Bortoli, V., Mathieu, E., Hutchinson, M. J., Thornton, J., Teh, Y. W., and Doucet, A. (2022). Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*.

Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142.

Dey, N., Chen, A., and Ghafurian, S. (2021). Group equivariant generative adversarial networks. In *International Conference on Learning Representations*.

Duval, A. A., Schmidt, V., Hernández-García, A., Miret, S., Malliaros, F. D., Bengio, Y., and Rolnick, D. (2023). FAENet: Frame averaging equivariant GNN for materials modeling. In *Proceedings of the 40th International Conference on Machine Learning*, pages 9013–9033.

Ehrendorfer, M. (2006). *The Liouville equation and atmospheric predictability*, page 59–98. Cambridge University Press.

Elesedy, B. and Zaidi, S. (2021). Provably strict generalisation benefit for equivariant models. In *Proceedings of the 38th International Conference on Machine Learning*, pages 2959–2969.

Esteves, C., Allen-Blanchette, C., Makadia, A., and Daniilidis, K. (2018). Learning $SO(3)$ equivariant representations with spherical CNNs. In *ECCV*, pages 54–70.

Gao, L., Du, Y., Li, H., and Lin, G. (2022). RotEqNet: Rotation-equivariant network for fluid systems with symmetric high-order tensors. *Journal of Computational Physics*, 461:111205.

Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., and Rubin, D. (2022). A whole-body FDG-PET/CT dataset with manually annotated tumor lesions. *Scientific Data*, 9:601–608.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. (2022). Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, pages 8867–8887.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jiao, Y., van der Laak, J., Albarqouni, S., Li, Z., Tan, T., Bhalerao, A., Ma, J., Sun, J., Pocock, J., Pluim, J. P., Koohbanani, N. A., Bashir, R. M. S., Raza, S. E. A., Liu, S., Graham, S., Wetstein, S., Khurram, S. A., Watson, T., Rajpoot, N., Veta, M., and Ciompi, F. (2023). LYSTO: The lymphocyte assessment hackathon and benchmark dataset. arXiv:2301.06304.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. S. (2022). Torsional diffusion for molecular conformer generation. In *Advances in Neural Information Processing Systems*.

Karras, T., Aittala, M., Aila, T., and Laine, S. (2022). Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*.

Kim, D., Lai, C.-H., Liao, W.-H., Murata, N., Takida, Y., Uesaka, T., He, Y., Mitsufuji, Y., and Ermon, S. (2024). Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In *The Twelfth International Conference on Learning Representations*.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations*.

Knigge, D. M., Romero, D. W., and Bekkers, E. J. (2022). Exploiting redundancy: Separable group convolutional networks on Lie groups. In *Proceedings of the 39th International Conference on Machine Learning*, pages 11359–11386.

Köhler, J., Klein, L., and Noe, F. (2020). Equivariant flows: Exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, pages 5361–5370.

Kondor, R. and Trivedi, S. (2018). On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Proceedings of the 35th International Conference on Machine Learning*.

Kong, Z., Ping, W., Huang, J., Zhao, K., and Catanzaro, B. (2021). Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*.

Lafarge, M. W., Bekkers, E. J., Pluim, J. P., Duits, R., and Veta, M. (2021). Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Medical Image Analysis*, 68:101849.

Larochelle, H., Erhan, D., Courville, A., Bergstra, J., and Bengio, Y. (2007). An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th International Conference on Machine Learning*.

Lee, D., Lee, D., Bang, D., and Kim, S. (2024). Disco: Diffusion Schrödinger bridge for molecular conformer optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(12):13365–13373.

Lin, W., Xiao, J., and Micheli-Tzanakou, E. (1998). A computational intelligence system for cell classification. In *Proceedings of IEEE International Conference on Information Technology Applications in Biomedicine*, pages 105–109.

Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E., Nie, W., and Anandkumar, A. (2023). I^2SB: Image-to-image Schrödinger bridge. In *Proceedings of the*

40th International Conference on Machine Learning, pages 22042–22062.

Martinkus, K., Ludwiczak, J., LIANG, W.-C., Lafrance-Vanasse, J., Hotzel, I., Rajpal, A., Wu, Y., Cho, K., Bonneau, R., Gligorijevic, V., and Loukas, A. (2023). Abdiffuser: full-atom generation of in-vitro functioning antibodies. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Mathieu, E., Dutordoir, V., Hutchinson, M. J., De Bortoli, V., Teh, Y. W., and Turner, R. E. (2023). Geometric neural diffusion processes. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., and Ermon, S. (2022). SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*.

Nica, B. (2012). The Mazur–Ulam theorem. *Expositiones Mathematicae*, 30(4):397–398.

Oksendal, B. (2003). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 22(1).

Pohlman, S., Powell, K., Obuchowski, N., Chilcote, W., and Grundfest-Broniatowski, S. (1996). Quantitative classification of breast tumors in digitized mammograms. *Medical Physics*, 23(8):1337–1345.

Puny, O., Atzmon, M., Smith, E. J., Misra, I., Grover, A., Ben-Hamu, H., and Lipman, Y. (2022). Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*.

Qiang, B., Song, Y., Xu, M., Gong, J., Gao, B., Zhou, H., Ma, W., and Lan, Y. (2023). Coarse-to-Fine: a hierarchical diffusion model for molecule generation in 3D. In *Proceedings of the 40th International Conference on Machine Learning*.

Rangayyan, R., El-Faramawy, N., Desautels, J., and Alim, O. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810.

Ravanbakhsh, S., Schneider, J., and Póczos, B. (2017). Equivariance through parameter-sharing. In *Proceedings of the 34th International Conference on Machine Learning*.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241.

Salimans, T., Karpathy, A., Chen, X., and Kingma, D. P. (2017). PixelCNN++: Improving the pixelCNN with discretized logistic mixture likelihood and other modifications. In *International Conference on Learning Representations*.

Shao, H.-C., Li, Y., Wang, J., Jiang, S., and Zhang, Y. (2023). Real-time liver motion estimation via deep learning-based angle-agnostic x-ray imaging. *Medical Physics*, 50(11):6649–6662.

Shawe-Taylor, J. (1993). Symmetries and discriminability in feedforward network architectures. *IEEE Transactions on Neural Networks*, 4(5):816–826.

Shi, C., Luo, S., Xu, M., and Tang, J. (2021). Learning gradient fields for molecular conformation generation. In *International Conference on Machine Learning*.

Song, J., Meng, C., and Ermon, S. (2021a). Denoising diffusion implicit models. In *International Conference on Learning Representations*.

Song, Y., Dhariwal, P., Chen, M., and Sutskever, I. (2023). Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*.

Song, Y. and Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2021b). Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*.

Veeling, B. S., Linmans, J., Winkens, J., Cohen, T., and Welling, M. (2018). Rotation equivariant CNNs for digital pathology. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 210–218.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., and Cattin, P. C. (2022). Diffusion models for implicit image segmentation ensembles. In *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, pages 1336–1348.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. (2022). Geodiff: A geometric diffusion model for

molecular conformation generation. In *International Conference on Learning Representations*.

Yarotsky, D. (2022). Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55:407–474.

Yim, J., Trippe, B. L., De Bortoli, V., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. (2023). SE(3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, pages 40001–40039.

Zhang, Q. and Chen, Y. (2023). Fast sampling of diffusion models with exponential integrator. In *The Eleventh International Conference on Learning Representations*.

Zhou, L., Lou, A., Khanna, S., and Ermon, S. (2024). Denoising diffusion bridge models. In *The Twelfth International Conference on Learning Representations*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]. All assumptions for the given theorem are clearly stated if needed, with more detailed derivations provided in the paper appendix. All dataset are described in detail, and model training paramters are provided in the appendix.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [No] The primary motivation for the given experiments is in validating the theoretical guarantees from our primary proposition that characterizes diffusion model structure preserving for isometry groups.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No] We will provide anonymized source code if requested but not otherwise.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] Detailed proofs are provided in the appendix section.
   (b) Complete proofs of all theoretical results. [Yes] Detailed proofs are provided in the appendix section.
   (c) Clear explanations of any assumptions. [Yes] These are provided when necessary within the main body of the paper. All assumptions are stated explicitly in the appendix.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No] We will provide anonymized source code if requested but not otherwise. One of the medical imaging datasets used is under restricted licence due to privacy concerns, so we are not able to provide any data relating to this dataset such as processed data or model checkpoints.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] All key training parameters for the various models are provided in the appendix of the paper.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] All custom measures are defined within the main text or a citation is provided for clarity. Additional details for some of the measures used in benchmarking are outlined in the paper Appendix.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] We briefly mention the computing hardware used to train all the models mentioned in the paper in the introductory paragraph of the Empirical study section.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator if your work uses existing assets. [Yes] We cite the work of any authors we make use of (e.g., code from existing machine learning models, dataset curators, etc.)
   (b) The license information of the assets, if applicable. [Yes] One of the dataset used is under restricted licence due to primary concerns. This is mentioned in the main body of the text when the dataset is introduced.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes] All necessary information for replicating the dataset used in the experiments is either provided either

in the main paper or in the supplemental sections.

(d) Information about consent from data providers/curators. [Yes] In so far as we mention necessary licences where required.

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Yes] As mentioned above in the checklist, and in the main body of the paper, one of the dataset used is under restricted license due to patient privacy concerns.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A COMMON DIFFUSION PROCESS COEFFICIENTS

Here we provide a overview of some commonly used drift and diffusion coefficients, namely, those corresponding to he variance preserving (VP, Ho et al. 2020; Song et al. 2021a) and variance exploding (VE, Song et al. 2021b) SDEs.

SDE	$\mathbf{u}(\mathbf{x}, t)$	$g(t)^2$	$p(\mathbf{x}_t\|\mathbf{x}_0)$	$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T\|\mathbf{x}_t)$	$p_t(\mathbf{x}_t\|\mathbf{x}_0, \mathbf{x}_T)$
VP	$\frac{\mathrm{d}\log\alpha_t}{\mathrm{d}t}\,\mathbf{x}$	$\frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t} - \frac{\mathrm{d}\log\alpha_t^2}{\mathrm{d}t}\sigma_t^2$	$\mathcal{N}(\alpha_t\mathbf{x}_0, \sigma_t^2\mathbf{I})$	$\frac{(\alpha_t/\alpha_T)\mathbf{x}_T - \mathbf{x}_t}{\sigma_t^2(\eta_t/\eta_T - 1)}$	$\mathcal{N}\Big(\frac{\eta_T}{\eta_t}\frac{\alpha_t}{\alpha_T}\mathbf{x}_T + \alpha_t\mathbf{x}_0(1 - \frac{\eta_T}{\eta_t}), \sigma_t^2(1 - \frac{\eta_T}{\eta_t})\Big)$
VE	$\mathbf{0}$	$\frac{\mathrm{d}\sigma_t^2}{dt}$	$\mathcal{N}(\mathbf{x}_0, \sigma_t^2\mathbf{I})$	$\frac{\mathbf{x}_T - \mathbf{x}_t}{\sigma_T^2 - \sigma_t^2}$	$\mathcal{N}\Big(\frac{\sigma_t^2}{\sigma_T^2}\mathbf{x}_T + (1 - \frac{\sigma_t^2}{\sigma_T^2})\mathbf{x}_0, \sigma_t^2(1 - \frac{\sigma_t^2}{\sigma_T^2})\Big)$

Table 4: Choices of $\mathbf{u}(\mathbf{x}, t)$ and $g(\mathbf{x})$ where $\eta_t = \frac{\alpha_t^2}{\sigma_t^2}$ (Zhou et al., 2024).

B DERIVATION DETAILS OF THE THEORETICAL RESULTS

In this section, we provide detailed derivations of our theoretical results. For conciseness, we complete most of the proofs in measure theory notation and show their equivalence to those presented in the main text.

In Appx B.1, we demonstrate that the isometry assumption on group operators results in their linearity. In Appx B.2, we briefly review the Liouville equations, which play a crucial role in characterizing the distribution evolution of particles driven by an ODE drift. In Appx B.3, we discuss a special family of ODE drifts that preserve distributions, characterizing the equivalence of various drifts by inducing the same evolution of p_t. This result helps derive the equivalent conditions on drifts to achieve structure-preserving ODE and SDE processes. Appx B.4 discusses the structure-preserving conditions for ODE processes, and we extend the results to SDE processes in Appx B.5.

B.1 Isometries

The groups $\mathcal{G}$ involved in our discussions from Sec 3.2 are assumed to consist of isometries κ satisfying $\|\kappa\mathbf{x}\| = \|\mathbf{x}\|$. Then, for $\kappa \in \mathcal{G}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have $\|\kappa\mathbf{x} - \kappa\mathbf{y}\|_2 = \|\mathbf{x} - \mathbf{y}\|_2$. Since $\kappa \in \mathcal{G}$ is bijective, by the Mazur–Ulam theorem (Nica, 2012), κ is affine and thus can be written as

$$\kappa(\mathbf{x}) = A\mathbf{x} + \mathbf{b} \tag{12}$$

for some $A_\kappa \in \mathbb{R}^{d\times d}$ and $\mathbf{b}_\kappa \in \mathbb{R}^d$. Besides, A_κ is orthogonal:

Lemma 2. *If $\kappa(\mathbf{x}) = A_\kappa\mathbf{x} + \mathbf{b}_\kappa$ is an isometry, then $A_\kappa^\top A_\kappa = \mathbf{I}$.*

Proof. As κ is an isometry, then for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$\|A_\kappa\mathbf{x} - A_\kappa\mathbf{y}\| = \|\kappa\mathbf{x} - \kappa\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\|. \tag{13}$$

In addition,

$$\langle A_\kappa\mathbf{x}, A_\kappa\mathbf{y}\rangle = \frac{1}{4}\left[\|A_\kappa\mathbf{x} - A_\kappa(-\mathbf{y})\|^2 - \|A_\kappa\mathbf{x} - A_\kappa\mathbf{y}\|^2\right] = \langle\mathbf{x}, \mathbf{y}\rangle \tag{14}$$

That is $\left\langle A_\kappa^\top A_\kappa\mathbf{x}, \mathbf{y}\right\rangle = \langle\mathbf{x}, \mathbf{y}\rangle$, which implies $A_\kappa^\top A_\kappa = \mathbf{I}$. □

Remark 1. *Lem 2 suggests that $D\kappa(\mathbf{x}) = A_\kappa$ for all $\mathbf{x} \in \mathbb{R}^d$.*

In addition, since $\|\kappa\mathbf{x}\|_2 = \|\mathbf{x}\|_2$, we have $\|A_\kappa\mathbf{x} + \mathbf{b}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$. Setting $A_\kappa\mathbf{x} = -\mathbf{b}$ yields $\|\mathbf{b}\| = 0$, or equivalently, $\mathbf{b} = \mathbf{0}$.

Therefore, for all the group operators κ appearing in our discussion, we can write:

$$\kappa\mathbf{x} = A_\kappa\mathbf{x} \tag{15}$$

for some orthogonal $A_\kappa \in \mathbb{R}^{d\times d}$.

The following lemma can significantly simplify the discussion on the geometric properties of diffusion processes in Appx B.2 and Appx B.5:

Lemma 3. *Let $C_c^\infty(\mathbb{R}^d)$ be the set of compactly supported functions. Then for $\kappa \in \mathcal{G}$,*

$$\{\phi \circ \kappa | \phi \in C_c^\infty(\mathbb{R}^d)\} = C_c^\infty(\mathbb{R}^d) \tag{16}$$

Proof. If $\phi \in C_c^\infty(\mathbb{R}^d)$ has a compact support C, then $\phi \circ \kappa$ has a support $\kappa^{-1}C$, which is also compact because $\kappa^{-1} \in \mathcal{G}$ is also affine (thus continuous) and a continuous image of a compact set is compact. Moreover, since ϕ is infinitely differentiable, so is $\phi \circ \kappa$. Thus, $\{\phi \circ \kappa | \phi \in C_c^\infty(\mathbb{R}^d)\} \subseteq C_c^\infty(\mathbb{R}^d)$. In addition, for $\psi \in C_c^\infty(\mathbb{R}^d)$, we have $\phi = \psi \circ \kappa^{-1} \in C_c^\infty(\mathbb{R}^d)$ such that $\phi \circ \kappa = \psi$. Hence, $C_c^\infty(\mathbb{R}^d) \subseteq \{\phi \circ \kappa | \phi \in C_c^\infty(\mathbb{R}^d)\}$. □

B.2 Liouville Equation

Our proof relies on the Liouville equation in measure theory notation. We provide an intuitive and easy-to-follow proof here and show its equivalence to the popular version in the probability density notation in Remark 2.

Consider N non-interacting particles moving according to a deterministic ODE in $\mathbb{R}^d$:

$$\mathrm{d}\mathbf{x}_t = \mathbf{u}(\mathbf{x}_t, t)\,\mathrm{d}t. \tag{17}$$

Then their distribution is characterized by a measure $\mu_t^{(N)}$ such that for any compactly supported function $\phi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\int \phi(\mathbf{x})\,\mathrm{d}\mu_t^N(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^N \phi(\mathbf{x}_t^i). \tag{18}$$

Then

$$\frac{\partial}{\partial t}\int \phi(\mathbf{x})\,\mathrm{d}\mu_t^N(\mathbf{x}) = \frac{1}{N}\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^N \phi(\mathbf{x}_t^i) = \frac{1}{N}\sum_{i=1}^N \nabla\phi(\mathbf{x}_t^i)\cdot\mathbf{u}(\mathbf{x}_t^i, t) \tag{19}$$

$$= \int \nabla\phi(\mathbf{x})\cdot\mathbf{u}(\mathbf{x}, t)\,\mathrm{d}\mu_t^N(\mathbf{x}). \tag{20}$$

Then if we suppose the initial distribution

$$\mu_0^N(\mathbf{x}, 0) \to^\star \mu_0(\mathbf{x}) \text{ as } N \to \infty \tag{21}$$

in a sense that $\int \phi(\mathbf{x})\,\mathrm{d}\mu_0(\mathbf{x}) \to \int \phi(\mathbf{x})\,\mathrm{d}\mu_0(\mathbf{x})$ for any $\phi \in C_c^\infty(\mathbb{R}^d)$. Then we can establish the limit $\mu_t^N(\mathbf{x}) \to^\star \mu_t(\mathbf{x})$ and μ_t satisfies

$$\frac{\partial}{\partial t}\int \phi(\mathbf{x})\,\mathrm{d}\mu_t(\mathbf{x}) = \int \nabla\phi(\mathbf{x})\cdot\mathbf{u}(\mathbf{x}, t)\,\mathrm{d}\mu_t(\mathbf{x}). \tag{22}$$

Notably, the setting we consider in the main text assume that the drift could optionally depend on some additional (fixed) term $\mathbf{y}$ such that

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}t, \tag{23}$$

where $\mathbf{x}_t \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ with $m > 0$ and $n \geq 0$.[2] In this case, the process can be rewritten as

$$\mathrm{d}\begin{bmatrix}\mathbf{x}_t \\ \mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{f}(\mathbf{x}_t, \mathbf{y}, t) \\ \mathbf{0}\end{bmatrix}\mathrm{d}t = \mathbf{u}\big([\mathbf{x}_t, \mathbf{y}]^\top, t\big)\,\mathrm{d}t. \tag{24}$$

Applying (22), we obtain

$$\boxed{\frac{\partial}{\partial t}\int \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) = \int \nabla_1\phi(\mathbf{x}, \mathbf{y})\cdot\mathbf{f}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \qquad \textit{(Liouville eq. meas.)}} \tag{25}$$

where $\nabla_1\psi(\mathbf{x}, \mathbf{y}, \ldots) := \frac{\partial\psi(\mathbf{x},\mathbf{y},\ldots)}{\partial\mathbf{x}}$ denote the gradient with respect to the first argument.

[2] We use $n = 0$ to indicate the case that $\mathbf{f}$ does not depend on $\mathbf{y}$. Unless otherwise stated, we will continue to use this convention.

Remark 2. *Let λ denote the Lebesgue measure. When the probability measure $\mu_t(\mathbf{x}|\mathbf{y})$ has density $p_t(\mathbf{x}|\mathbf{y}) \in C^1(\mathbb{R}^m \times \mathbb{R}^n \times [0,T])$ with respect to $\mathbf{x}$, we have*

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_c \phi(\mathbf{x},\mathbf{y}) p_t(\mathbf{x}|\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) = \frac{\partial}{\partial t}\int_c \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \int_c \nabla_1\phi(\mathbf{x},\mathbf{y})\cdot\mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) = \int_c \nabla_1\phi(\mathbf{x},\mathbf{y})\cdot\mathbf{f}(\mathbf{x},\mathbf{y},t) p_t(\mathbf{x}|\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) \\
&= [p_t(\mathbf{x}|\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\cdot\phi(\mathbf{x},\mathbf{y})]_{\partial c} - \int_c \phi(\mathbf{x},\mathbf{y})\ \nabla_{\mathbf{x}}\cdot\big(p_t(\mathbf{x}|\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\big)\,\mathrm{d}\lambda(\mathbf{x}) \\
&= -\int_c \phi(\mathbf{x},\mathbf{y})\ \nabla_{\mathbf{x}}\cdot\big(p_t(\mathbf{x}|\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\big)\,\mathrm{d}\lambda(\mathbf{x}).
\end{aligned}$$

As this holds for all $\phi \in C_c^\infty(\mathbb{R}^{m+n})$, we obtain the regular Liouville equation (Oksendal, 2003; Ehrendorfer, 2006):

$$\boxed{\frac{\partial}{\partial t}p_t(\mathbf{x}|\mathbf{y}) = -\nabla_{\mathbf{x}}\cdot\big(p_t(\mathbf{x}|\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\big). \qquad \text{(Liouville eq. density)}} \tag{26}$$

B.3 Distribution-preserving Drifts.

While a zero drift implies $p_t = p_0$ for all $t > 0$, the converse is not necessarily true:

Example 1. *Let p_0 be the density of a spherical Gaussian $\mathcal{N}(\mathbf{0},\mathbf{I})$ in $\mathbb{R}^2$. For $\mathbf{f}(\mathbf{x},t) = [y,-x]^\top$, by the Liouville equation, at $t=0$*

$$\begin{aligned}
\frac{\partial}{\partial t}p_t(x,y) &= -\nabla\cdot\Big[\frac{1}{Z}\exp(-\frac{x^2+y^2}{2})\,[y,-x]^\top\Big] &(27)\\
&= \frac{\partial}{\partial x}\left[\frac{1}{Z}\exp(-\frac{x^2+y^2}{2})y\right] - \frac{\partial}{\partial y}\left[\frac{1}{Z}\exp(-\frac{x^2+y^2}{2})x\right] &(28)\\
&= \frac{1}{Z}\left[-\exp(-\frac{x^2+y^2}{2})xy + \exp(-\frac{x^2+y^2}{2})xy\right] = 0. &(29)
\end{aligned}$$

As a result, $\mathbf{f}$ does not change p_0, although it is not zero.

In general,

Lemma 4. *Given a measure μ, drift $\mathbf{f}$ does not change the distribution if for all $\phi \in C_c^\infty(\mathbb{R}^d)$*

$$0 = \int \nabla\phi(\mathbf{x})\cdot\mathbf{f}(\mathbf{x},t)\,\mathrm{d}\mu(\mathbf{x}) \tag{30}$$

for all $\mathbf{x}$ and t. We use $[\mathbf{0}]_\mu(\mathbf{x})$ to denote the set of drifts that do not alter distribution μ. That is, if $\mathbf{f}$ satisfies (30), we have $\mathbf{f} \in [\mathbf{0}]_\mu$.

Proof. This is an immedate result of (25) by setting the left-hand side zero. □

Remark 3. *For any μ, $\mathbf{0} \in [\mathbf{0}]_\mu$.*

Remark 4. *If $\mathbf{f},\mathbf{g} \in [\mathbf{0}]_\mu$, then $\alpha\mathbf{f} + \beta\mathbf{g} \in [\mathbf{0}]_\mu$, for $\alpha,\beta \in \mathbb{R}$,* i.e., *$[\mathbf{0}]_\mu$ is a vector space.*

Remark 5. *In the main text, we use the notation $[\mathbf{0}]_p$ instead of $[\mathbf{0}]_\mu$ to represent distribution-preserving drifts that maintain a distribution with density p, which corresponds to the distribution measure μ.*

B.4 Structural Preserving ODE Processes

In this section, we discuss the sufficient and necessary condition of structurally preserved ODE processes. Here, we consider ODE process:

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t,\mathbf{y},t)\,\mathrm{d}t \tag{31}$$

with $\mathbf{x}_t \in \mathbb{R}^m$, and $\mathbf{y} \in \mathbb{R}^n$ denote additional conditions of the process. Here, we assume $m > 0$ and $n \geq 0$, where $n = 0$ denote the case when $\mathbf{f}$ does not depend on $\mathbf{y}$. We note that for a similar setting with discrete time step and drift $\mathbf{f}$ that does not depend on $\mathbf{y}$, a sufficient condition on $\mathcal{G}$-invariance of μ_t for $t \geq 0$ has been discussed by Papamakarios et al. (2021) and Köhler et al. (2020).

Let $\mu_t(\mathbf{x}_t|\mathbf{y})$ be the probability measure of $\mathbf{x}_t$ induced by the ODE process (31) conditioned on $\mathbf{y}$. Let $\mathbf{G} = \{\boldsymbol{\kappa} = (\kappa_1, \kappa_2) | \kappa_1 : \mathbb{R}^m \to \mathbb{R}^m, \kappa_2 : \mathbb{R}^n \to \mathbb{R}^n\}$ be a group of isometries defined in $\mathbb{R}^{m+n}$ such that $\boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = (\kappa_1 \mathbf{x}, \kappa_2 \mathbf{y})$. It is easy to see that the sets of κ_1 and κ_2 are also groups of isometries. We will respectively denote them as $\mathcal{G}_1$ and $\mathcal{G}_2$. In addition, by Lem 2, we have

$$\boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} A_{\kappa_1} & \mathbf{0} \\ \mathbf{0} & A_{\kappa_2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{\kappa_1} \\ \mathbf{b}_{\kappa_2} \end{bmatrix}, \tag{32}$$

where $A_{\kappa_1} \in \mathbb{R}^{m \times m}$ and $A_{\kappa_2} \in \mathbb{R}^{n \times n}$ are orthogonal.

Furthermore, by Remark 1, we have

$$D\boldsymbol{\kappa} = A_{\boldsymbol{\kappa}} = \begin{bmatrix} A_{\kappa_1} & \mathbf{0} \\ \mathbf{0} & A_{\kappa_2} \end{bmatrix}. \tag{33}$$

We say $\mu_t(\mathbf{x}_t|\mathbf{y})$ is $\mathbf{G}$-*invariant* if for all $\boldsymbol{\kappa} \in \mathbf{G}$, $\mu_t(\kappa_1 \mathbf{x}_t | \kappa_2 \mathbf{y}) = \mu_t(\mathbf{x}_t|\mathbf{y})$. Lem 5 shows that this definition is equivalent to the one given in Sec 3.2.

Lemma 5. *Assume $\mu(\cdot|\cdot)$ has density $p(\mathbf{x}|\mathbf{y})$. Then $\mu(\cdot|\cdot)$ is $\mathbf{G}$-invariant if and only if the density $p(\mathbf{x}|\mathbf{y}) = p(\kappa_1 \mathbf{x}|\kappa_2 \mathbf{y})$ for all $\boldsymbol{\kappa} \in \mathbf{G}$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$.*

Proof. $\mu(\cdot|\cdot)$ is $\mathbf{G}$-invariant if and only if for all $\boldsymbol{\kappa} \in \mathbf{G}$, $\phi \in C_c^\infty(\mathbb{R}^{m+n})$,

$$\int \phi(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\boldsymbol{\mu}(\mathbf{x}|\mathbf{y}) = \int \phi(\boldsymbol{\kappa}^{-1}(\mathbf{x}, \mathbf{y})) \, \mathrm{d}\boldsymbol{\mu}(\mathbf{x}|\mathbf{y}). \tag{34}$$

That is,

$$\begin{aligned} &\int \phi(\mathbf{x}, \mathbf{y}) p(\mathbf{x}|\mathbf{y}) \, \mathrm{d}\lambda(\mathbf{x}) = \int \phi(\boldsymbol{\kappa}^{-1}(\mathbf{x}, \mathbf{y})) \, \mathrm{d}\boldsymbol{\mu}(\mathbf{x}|\mathbf{y}) = \int \phi(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mu(\kappa_1 \mathbf{x}|\kappa_2 \mathbf{y}) \\ = &\int \phi(\mathbf{x}, \mathbf{y}) p(\kappa_1 \mathbf{x}|\kappa_2 \mathbf{y}) \, \mathrm{d}\lambda(\kappa_1 \mathbf{x}) \overset{\text{(Lem 8)}}{=} \int \phi(\mathbf{x}, \mathbf{y}) p(\kappa_1 \mathbf{x}|\kappa_2 \mathbf{y}) \, \mathrm{d}\lambda(\mathbf{x}). \end{aligned}$$

Therefore, $p(\mathbf{x}|\mathbf{y}) = p(\kappa_1 \mathbf{x}|\kappa_2 \mathbf{y})$. Since every step is reversible, the proof is completed. □

Then we give the equivalent conditions on the drift terms to ensure the structure-preserving property of ODE flows.

Lemma 6. *Consider the ODE process in (31) with $\mathbf{G}$-invariant $\mu_0(\cdot|\cdot)$. Then, μ_t is $\mathbf{G}$-invariant for all $t \geq 0$ if and only if*

$$A_{\kappa_1}^\top \mathbf{f}(\kappa_1 \mathbf{x}, \kappa_2 \mathbf{y}, t) - \mathbf{f}(\mathbf{x}, \mathbf{y}, t) \in [\mathbf{0}]_{\mu_t}. \tag{35}$$

for all $t \geq 0$, $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2) \in \mathbf{G}$.

Proof. ($\Rightarrow$) Assume that μ_t is $\mathbf{G}$-invariant for all $t \geq 0$. For all $\phi \in C_c^\infty(\mathbb{R}^{m+n})$ and $\boldsymbol{\kappa} \in \mathbf{G}$, let $\psi = \phi \circ \boldsymbol{\kappa}$. We

note that by Lem 3, $\psi \in C_c^\infty(\mathbb{R}^{m+n})$. Then, for $t \geq 0$, we have

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\mu_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&\overset{(25)}{=} \int \nabla_1\phi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&\overset{(\mathbf{G}-\mathrm{inv})}{=} \int \nabla_1\phi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top D\kappa_1(\mathbf{x})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \Big(A_{\kappa_1}(\mathbf{x})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t) - \mathbf{f}(\mathbf{x},\mathbf{y},t)\Big)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}).
\end{aligned}$$

By Lem 3, ψ can be any functions in $C_c^\infty(\mathbb{R}^{m+n})$. Thus, (35) follows.

($\Leftarrow$) Assume (35) holds and $\boldsymbol{\mu}_0$ is $\mathbf{G}$-invariant. For all $\phi \in C_c^\infty(\mathbb{R}^{m+n})$ and $\boldsymbol{\kappa} \in \mathbf{G}$, let $\psi = \phi \circ \boldsymbol{\kappa}$. Then we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\mathbf{x}|\mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}t}\int \psi(\mathbf{x},\mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\mathbf{x}|\mathbf{y}) \\
\overset{(25)}{=}& \int (\nabla_1\psi)(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \overset{(\mathbf{A})}{=} \int (\nabla_1\phi)\,(\kappa_1\mathbf{x},\kappa_2\mathbf{y})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t)\,\mathrm{d}\boldsymbol{\mu}_t(\mathbf{x}|\mathbf{y}) \\
=& \int (\nabla_1\phi)\,(\mathbf{x},\mathbf{y})^\top\ \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\boldsymbol{\mu}_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}),
\end{aligned} \tag{36}$$

where ($\mathbf{A}$) is due to:

$$\begin{aligned}
0 &\overset{(35)}{=} \int \nabla_1\psi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})^\top \Big(A_{\kappa_1}(\mathbf{x})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t) - \mathbf{f}(\mathbf{x},\mathbf{y},t)\Big)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \int \nabla_1\psi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})^\top D\kappa_1(\mathbf{x})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \int \nabla_1\phi(\kappa_1\mathbf{x},\kappa_2\mathbf{y})^\top \mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) - \int \nabla_1\psi(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y})
\end{aligned}$$

Besides, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \phi_1(\mathbf{x},\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \overset{(25)}{=} \int \nabla\phi_1(\mathbf{x},\mathbf{y})^\top \mathbf{f}(\mathbf{x},\mathbf{y},t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}). \tag{37}$$

As $\mu_0(\mathbf{x}|\mathbf{y}) = \mu_0(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{x})$, (36) and (37) together suggest that $\mu_t(\mathbf{x}|\mathbf{y})$ and $\mu_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})$ share the same Liouville's equation. Therefore, $\mu_t(\mathbf{x}|\mathbf{y}) = \mu_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})$ for all $t \geq 0$. □

B.5 Structural Preserving SDE Processes

In this section, we assume all the measures involved have densities. We first show that Lebesgue measure is $\mathcal{G}$-invariant, where $\mathcal{G}$ is a group of isometries.

Lemma 7. *For all $\kappa \in \mathcal{G}$, $\det D\kappa(\mathbf{x}) = \det A_\kappa = 1$ or -1 for all $\mathbf{x} \in \mathbb{R}^d$.*

Proof. For $\kappa \in \mathcal{G}$, by Lem 2 and Remark 1, we have $D\kappa(\mathbf{x})^\top D\kappa(\mathbf{x}) = A_\kappa^\top A_\kappa = I$. Then $\det(D\kappa(\mathbf{x}))^2 = (\det A_\kappa)^2 = 1$, which implies $\det D\kappa(\mathbf{x}) = \det A_\kappa = \pm 1$ □

Lemma 8. *The Lebesgue measure λ is $\mathcal{G}$-invariant.*

Proof. For all $\phi \in C_c^\infty(\mathbb{R}^d)$ and $\kappa \in \mathcal{G}$, we have

$$\begin{aligned}\int \phi(\mathbf{x})\,\mathrm{d}\lambda(\mathbf{x}) &= \int \phi(\kappa\mathbf{x})\,\mathrm{d}\lambda(\kappa\mathbf{x}) \overset{\text{(Lem 8)}}{=} \int \phi(\kappa\mathbf{x})\,|\det D\kappa(\mathbf{x})|\,\mathrm{d}\lambda(\mathbf{x}) \\ &= \int \phi(\kappa\mathbf{x})\,\mathrm{d}\lambda(\mathbf{x}) = \int \phi(\mathbf{x})\,\mathrm{d}\lambda(\kappa^{-1}\mathbf{x})\end{aligned}$$

Therefore, $\lambda = \kappa^{\#}\lambda$. □

To deal with the invariance property associated with the diffusion term, we prove a lemma similar to Lem F.4 of (Yim et al., 2023). The lemma basically says Laplacian is invariant with respect to isometries:

Lemma 9. *For $\boldsymbol{\kappa} \in \mathbf{G}$ and $v : \mathbb{R}^{m+n} \to \mathbb{R}$, we have*

$$\Delta_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y}) = (\Delta_1 v) \circ \boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}), \tag{38}$$

where

$$(\Delta_1 u)(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{m} \frac{\partial^2}{\partial x_k^2} u(\mathbf{x}, \mathbf{y}) = (\nabla_1 \cdot \nabla_1 u)(\mathbf{x}, \mathbf{y}), \tag{39}$$

$$(\nabla_1 u)(\mathbf{x}, \mathbf{y}) = \frac{\partial u}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{y}). \tag{40}$$

Proof. Let

$$M = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{41}$$

where $\mathbf{0}$ denotes a zero matrix of a proper size. Then it is easy to see

$$\Delta_1 v(\mathbf{x}, \mathbf{y}) = \nabla \cdot \Big(M\ \nabla v(\mathbf{x}, \mathbf{y}) \Big). \tag{42}$$

As a result, for all $\phi \in C_c^\infty(\mathbb{R}^{m+n})$, we have

$$\begin{aligned}&\int \phi(\mathbf{x}, \mathbf{y})(\Delta_1 v) \circ \boldsymbol{\kappa}(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) = \int \phi(\kappa_1^{-1}\mathbf{x}, \kappa_2^{-1}\mathbf{y})\ (\Delta_1 v)(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= \int \phi(\kappa_1^{-1}\mathbf{x}, \kappa_2^{-1}\mathbf{y})\ \nabla \cdot \Big(M\ \nabla v(\mathbf{x}, \mathbf{y}) \Big)\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= \big[\phi(\kappa_1^{-1}\mathbf{x}, \kappa_2^{-1}\mathbf{y})\ M \nabla v(\kappa_1^{-1}\mathbf{x}, \kappa_2^{-1}\mathbf{y}) \big]_{\partial c} - \int M \nabla v(\mathbf{x}, \mathbf{y}) \cdot \nabla(\phi \circ \boldsymbol{\kappa}^{-1})(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int M \nabla v(\mathbf{x}, \mathbf{y}) \cdot \Big((\nabla \phi)(\boldsymbol{\kappa}^{-1}(\mathbf{x}, \mathbf{y}))^\top D\boldsymbol{\kappa}^{-1}(\mathbf{x}, \mathbf{y}) \Big)\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int M \nabla v(\mathbf{x}, \mathbf{y}) \cdot \Big(\nabla \phi(\boldsymbol{\kappa}^{-1}(\mathbf{x}, \mathbf{y}))^\top A_{\boldsymbol{\kappa}}^\top \Big)\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int M \nabla v(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}) \cdot \Big(\nabla \phi(\mathbf{x}, \mathbf{y})^\top A_{\boldsymbol{\kappa}}^\top \Big)\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &\overset{\text{(Lem 8)}}{=} -\int \nabla v(\kappa_1\mathbf{x}, \kappa_2\mathbf{y})^\top M^\top A_{\boldsymbol{\kappa}} \nabla \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int \nabla v(\kappa_1\mathbf{x}, \kappa_2\mathbf{y})^\top \begin{bmatrix} A_{\kappa_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \nabla \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int \begin{bmatrix} \nabla_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})^\top & \mathbf{0} \end{bmatrix} \nabla \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) \\ &= -\int \begin{bmatrix} \nabla_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})^\top & \mathbf{0} \end{bmatrix} \nabla \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) + \Big[\phi(\mathbf{x}, \mathbf{y})\ \begin{bmatrix} \nabla_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})^\top & \mathbf{0} \end{bmatrix} \Big]_{\partial c} \\ &= \int \phi(\mathbf{x}, \mathbf{y})\ \ \nabla \cdot \begin{bmatrix} \nabla_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})^\top & \mathbf{0} \end{bmatrix} \mathrm{d}\lambda(\mathbf{x}, \mathbf{y}) = \int \phi(\mathbf{x}, \mathbf{y})\ \Delta_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}, \mathbf{y}).\end{aligned}$$

Thus, $(\Delta_1 v) \circ \boldsymbol{\kappa}(\mathbf{x}, \mathbf{y}) = \Delta_1(v \circ \boldsymbol{\kappa})(\mathbf{x}, \mathbf{y})$. □

Lemma 10. *$\mu(\cdot|\cdot)$ is $\mathbf{G}$-invariant if and only if*

$$\mathbf{s}(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = A_{\kappa_1}\ \mathbf{s}(\mathbf{x}|\mathbf{y}) \tag{43}$$

for all $\boldsymbol{\kappa} \in \mathbf{G}$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$, where $\mathbf{s}(\mathbf{x}|\mathbf{y})$ denotes the score function $\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})$.

Proof. ($\Rightarrow$) By Lem 5, if μ is $\mathbf{G}$-invariant, its density $p(\mathbf{x}|\mathbf{y}) = p(\kappa_1\mathbf{x}|\kappa_2\mathbf{y})$. Taking log on both sides, followed by taking the derivative with respect to $\mathbf{x}$ yields

$$A_{\kappa_1}^\top\ \mathbf{s}(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = \mathbf{s}(\mathbf{x}|\mathbf{y}), \tag{44}$$

as $D\kappa_1(\mathbf{x}) = A_{\kappa_1}$.

($\Leftarrow$) Conversely, (43) yields

$$p(\mathbf{x}|\mathbf{y}) = p(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) + C, \tag{45}$$

where C must be zero so that $p(\mathbf{x}|\mathbf{y})$ and $p(\kappa_1\mathbf{x}|\kappa_2\mathbf{y})$ are valid densities. □

Then we prove the following Lemma presented in the Sec 3.2.

Lemma 1. *$p(\mathbf{x}|\mathbf{y})$ is $\mathbf{G}$-invariant if and only if $\mathbf{s}(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = \kappa_1 \circ \mathbf{s}(\mathbf{x}|\mathbf{y})$ for all $(\kappa_1, \kappa_2) \in \mathbf{G}$, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$. Likewise, $p(\mathbf{x})$ is $\mathcal{G}$-invariant if and only if $\mathbf{s}(\kappa\mathbf{x}) = \kappa \circ \mathbf{s}(\mathbf{x})$ for all $\kappa \in \mathcal{G}$.*

Proof. The conditional density case is immediate given Lem 5 and Lem 10 while the unconditional one is the special case that $n = 0$. □

Lemma 11 (Song et al. (2021b))**.** *Let p_t be the marginal distribution of $\mathbf{x}_t$ that satisfies SDE:*

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t, \quad \mathbf{x}_0 \sim p_0(\mathbf{x}_0|\mathbf{y}). \tag{46}$$

Besides, let $\mathbf{s}_t(\cdot|\mathbf{y}) = \nabla \log p_t(\cdot|\mathbf{y})$. Then, the ODE

$$\mathrm{d}\mathbf{x} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}t \tag{47}$$

with

$$\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{y}, t) = \mathbf{f}(\mathbf{x}, \mathbf{y}, t) - \frac{1}{2}g(t)^2\ \mathbf{s}_t(\mathbf{x}|\mathbf{y}) \tag{48}$$

also has the same marginal distribution p_t for all $t \geq 0$.

Proof. The marginal distribution $p_t(\mathbf{x}|\mathbf{y})$ evolution is characterized by the Fokker-Planck equation (Oksendal, 2003):

$$\frac{\partial p_t(\mathbf{x}|\mathbf{y})}{\partial t} = -\nabla \cdot \big(\mathbf{f}(\mathbf{x}, \mathbf{y}, t)p_t(\mathbf{x}|\mathbf{y})\big) + \frac{1}{2}\nabla \cdot \nabla\big(g(t)^2 p_t(\mathbf{x}|\mathbf{y})\big) \tag{49}$$

$$= -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}[f_i(\mathbf{x}, \mathbf{y}, t)p_t(\mathbf{x}|\mathbf{y})] + \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}[g(t)^2\ p_t(\mathbf{x}|\mathbf{y})] \tag{50}$$

$$= -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left\{[f_i(\mathbf{x}, \mathbf{y}, t)p_t(\mathbf{x}|\mathbf{y})] - \frac{g(t)^2}{2}\big[p_t(\mathbf{x}|\mathbf{y})\frac{\partial}{\partial x_i}\log p_t(\mathbf{x}|\mathbf{y})\big]\right\} \tag{51}$$

$$= -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left[f_i(\mathbf{x}, \mathbf{y}, t) - \frac{g(t)^2}{2}\frac{\partial}{\partial x_i}\log p_t(\mathbf{x}|\mathbf{y})\right] p_t(\mathbf{x}|\mathbf{y}), \tag{52}$$

where the last line is the Fokker-Planck equation of

$$\mathrm{d}\mathbf{x} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}t \tag{53}$$

with $\tilde{\mathbf{f}}(\mathbf{x}, \mathbf{y}, t)$ given in (48). □

Now we are ready to give the if and only if statement on the structurally preserving property of the distributions induced by

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{y}, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}_t \tag{54}$$

Notably, a sufficient condition given by (55) with the left-hand side equal to zero is firstly discussed by Yim et al. (2023).

Then we give the equivalent conditions on the drift terms to ensure the structure-preserving property of SDE flows and its equivalence to the Prop 1 presented in the main text.

Proposition 3. *Given a diffusion process in* (54) *with* $\mathbf{G}$*-invariant* $\mu_0(\cdot|\cdot)$*,* $\mu_t(\cdot|\cdot)$ *is* $\mathbf{G}$*-invariant for all* $t \geq 0$ *if and only if*

$$A_{\kappa_1}^\top \mathbf{f}(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}, t) - \mathbf{f}(\mathbf{x}, \mathbf{y}, t) \in [\mathbf{0}]_{\mu_t}. \tag{55}$$

for all $t > 0$*,* $\mathbf{x} \in \mathbb{R}^m$*,* $\mathbf{y} \in \mathbb{R}^n$ *and* $\boldsymbol{\kappa} \in \mathbf{G}$*.*

Proof. ($\Rightarrow$) Let $\tilde{\mathbf{f}}$ denote the corresponding ODE drift shown in (48). Assume $\mu_t(\cdot|\cdot)$ is $\mathbf{G}$-invariant for all $t \geq 0$. Then by Lem 6, for all $\boldsymbol{\kappa} \in \mathbf{G}$, the ODE drift $\tilde{\mathbf{f}}$ satisfies

$$A_{\kappa_1}^\top \mathbf{f}(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}, t) - \mathbf{f}(\mathbf{x}, \mathbf{y}, t) \in [\mathbf{0}]_{\mu_t}. \tag{56}$$

That is,

$$A_{\kappa_1}^\top \Big(\mathbf{f}(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}, t) - \frac{1}{2}g(t)^2\mathbf{s}_t(\kappa_1\mathbf{x}|\kappa_2\mathbf{y})\Big) - \Big(\mathbf{f}(\mathbf{x}, \mathbf{y}, t) - \frac{1}{2}g(t)^2\mathbf{s}_t(\mathbf{x}|\mathbf{y})\Big) \in [\mathbf{0}]_{\mu_t}. \tag{57}$$

By Lem 10, $\mathbf{G}$-invariance of μ_t implies that $A_{\kappa_1}^\top \mathbf{s}_t(\kappa_1\mathbf{x}|\kappa_2\mathbf{y}) = \mathbf{s}_t(\mathbf{x}|\mathbf{y})$. Thus, (55) follows.

($\Leftarrow$) Assume (55) holds. For $\phi \in C_c^\infty(\mathbb{R}^{m+n})$ and $\boldsymbol{\kappa} \in \mathbf{G}$, let $\psi = \phi \circ \boldsymbol{\kappa}$, and then we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}) = \frac{\mathrm{d}}{\mathrm{d}t}\int \psi(\mathbf{x}, \mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\mathbf{x}, \mathbf{y}) \\
&\overset{(25)}{=} \int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) = \int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top \Big(\mathbf{f}(\mathbf{x}, \mathbf{y}, t) - \frac{1}{2}g^2(t)\mathbf{s}_t(\mathbf{x}|\mathbf{y})\Big)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) \\
&= \underbrace{\int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top \mathbf{f}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y})}_{\mathbf{I}} - \frac{1}{2}g^2(t)\underbrace{\int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top \mathbf{s}_t(\mathbf{x}|\mathbf{y})\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y})}_{\mathbf{II}}.
\end{aligned} \tag{58}$$

By (55) and applying the same argument to derive ($\mathbf{A}$) in the proof of Lem 6. We have

$$\int \nabla_1\phi(\kappa_1\mathbf{x}, \kappa_2\mathbf{y})^\top \mathbf{f}(\kappa_1\mathbf{x}, \kappa_2\mathbf{y}, t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) = \int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top \mathbf{f}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}\mu_t(\mathbf{x}|\mathbf{y}) = \mathbf{I}$$

Then,

$$\begin{aligned}
\mathbf{I} &= \int \nabla_1\phi(\mathbf{x}, \mathbf{y})^\top \mathbf{f}(\mathbf{x}, \mathbf{y}, t)\,\mathrm{d}\mu_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}) \\
&= -\int \phi(\mathbf{x}, \mathbf{y})\ \nabla_{\mathbf{x}} \cdot \big(p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})\mathbf{f}(\mathbf{x}, \mathbf{y}, t)\big)\,\mathrm{d}\lambda(\mathbf{x}).
\end{aligned}$$

In addition,

$$\begin{aligned}
\mathbf{II} &= \int \nabla_1\psi(\mathbf{x}, \mathbf{y})^\top p_t(\mathbf{x}|\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) = -\int \psi(\mathbf{x}, \mathbf{y})\Delta_1 p_t(\mathbf{x}|\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) \\
&= -\int \phi(\mathbf{x}, \mathbf{y})\Delta_1 p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) \overset{(\text{Lem } 9)}{=} -\int \phi(\mathbf{x}, \mathbf{y})\ \Delta_1(p_t \circ \boldsymbol{\kappa}^{-1})(\mathbf{x}|\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x})
\end{aligned}$$

As a result, by Eq (58), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_c \phi(\mathbf{x},\mathbf{y})p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})\,\mathrm{d}\lambda(\mathbf{x}) = \frac{\mathrm{d}}{\mathrm{d}t}\int \phi(\mathbf{x},\mathbf{y})\,\mathrm{d}\boldsymbol{\mu}_t(\kappa_1^{-1}\mathbf{x},\kappa_2^{-1}\mathbf{y})$$
$$= -\int_c \phi(\mathbf{x},\mathbf{y})\left[\nabla_{\mathbf{x}}\cdot\big(p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\big) - \frac{1}{2}g^2(t)\Delta_1(p_t\circ\boldsymbol{\kappa}^{-1})(\mathbf{x},\mathbf{y})\right]\mathrm{d}\lambda(\mathbf{x})$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y}) = -\nabla_{\mathbf{x}}\cdot\big(p_t(\kappa_1^{-1}\mathbf{x}|\kappa_2^{-1}\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)\big) - \frac{1}{2}g^2(t)\Delta_1(p_t\circ\boldsymbol{\kappa}^{-1})(\mathbf{x},\mathbf{y}). \tag{59}$$

By the Fokker-Planck equation, we also have

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(\mathbf{x}|\mathbf{y}) = -\nabla_{\mathbf{x}}\cdot(p_t(\mathbf{x}|\mathbf{y})\mathbf{f}(\mathbf{x},\mathbf{y},t)) + \frac{1}{2}g^2(t)\ (\Delta_1 p_t)(\mathbf{x}|\mathbf{y}). \tag{60}$$

Therefore, $p_t = p_t\circ\boldsymbol{\kappa}^{-1}$, which, by Lem 5, implies $\boldsymbol{\mu}_t$ is $\mathbf{G}$-invariant. □

Proposition 1. *Given a diffusion process in* (5) *with* $\mathbf{G}$*-invariant* $p_0(\mathbf{x}_0|\mathbf{y})$*, let* $[\mathbf{0}]_{p_t}$ *be the set of ODE drifts preserving the distribution* p_t*. Then* $p_t(\mathbf{x}_t|\mathbf{y})$ *is* $\mathbf{G}$*-invariant for all* $t\geq 0$ *if and only if*

$$\kappa_1^{-1}\circ\mathbf{f}(\kappa_1\mathbf{x},\kappa_2\mathbf{y},t) - \mathbf{f}(\mathbf{x},\mathbf{y},t) \in [\mathbf{0}]_{p_t} \tag{6}$$

for all $t>0$*,* $\mathbf{x}\in\mathbb{R}^m$*,* $\mathbf{y}\in\mathbb{R}^n$ *and* $\boldsymbol{\kappa}\in\mathbf{G}$*.*

Proof. Prop 1 is equivalent to Prop 3 but in probability density notations by Lem 5. □

Finally, we show how our theoretical results can be applied to characterize the structure-preserving properties of the diffusion bridges. The main results are given in Lem 13 and are collectively presented with the counterparts for the regular diffusion processes in Prop 2.

Lemma 12. *Let* p_t *denote the distribution of* $\mathbf{x}_t$ *generated by SDE:*

$$\mathrm{d}\mathbf{x}_t = \mathbf{u}(\mathbf{x}_t,t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t. \tag{61}$$

Then $p(\kappa\mathbf{x}_T|\kappa\mathbf{x}_t) = p(\mathbf{x}_T|\mathbf{x}_t)$ *for all* $\kappa\in\mathcal{G}_r$ *and* $T\geq t$ *if*

$$A_\kappa^\top\mathbf{u}(\kappa\mathbf{x},t) - \mathbf{u}(\mathbf{x},t) \in [\mathbf{0}]_{\mu_t}, \tag{62}$$

for all $\kappa\in\mathcal{G}$*.*

Proof. Without loss of generality, it is sufficient to show

$$p(\kappa\mathbf{x}_t|\kappa\mathbf{x}_0) = p(\mathbf{x}_t|\mathbf{x}_0) \tag{63}$$

for all $\kappa\in\mathcal{G}$ and $t\geq 0$. Then let $\mathbf{y}=\mathbf{x}_0$, $\mathbf{f}(\mathbf{x}_t,\mathbf{y},t)=\mathbf{u}(\mathbf{x},t)$ and $\mathbf{G}=\{(\kappa,\kappa)|\kappa\in\mathcal{G}_r\}$. As (62) implies $A_\kappa^\top\mathbf{u}(\kappa\mathbf{x},t)-\mathbf{u}(\mathbf{x},t)\in[\mathbf{0}]_{\mu_t}$. Then, combined with (62), Prop 3 shows $\mu_t(\mathbf{x}_t|\mathbf{y})$ is $\mathbf{G}$-invariant. By Lem 5, we have $p(\kappa\mathbf{x}_t|\kappa\mathbf{x}_0)=p(\mathbf{x}_t|\mathbf{x}_0)$, which completes the proof. □

Lemma 13. *Assume the two ends* $(\mathbf{x}_0,\mathbf{x}_T)\in\mathbb{R}^d\times\mathbb{R}^d$ *of the diffusion bridges follow a* $\mathbf{G}$*-invariant conditional distribution* $\boldsymbol{\mu}_{0|T}(\mathbf{x}_0|\mathbf{x}_T)$*, where* $\mathbf{G}=\{(\kappa,\kappa)|\kappa\in\mathcal{G}\}$*. Let* $\boldsymbol{\mu}_{t|T}$ *denote the measure of* $(\mathbf{x}_t,\mathbf{x}_T)$ *induced by diffusion bridge:*

$$\mathrm{d}\mathbf{x}_t = \big(\mathbf{u}(\mathbf{x}_t,t) + g(t)^2\mathbf{h}(\mathbf{x}_t,t,\mathbf{x}_T,T)\big)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t, \tag{64}$$

where $\mathbf{h}(\mathbf{x}_t,t,\mathbf{x}_T,T)=\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T|\mathbf{x}_t)$ *is the gradient of the log transition kernel from* t *to* T *generated by the original SDE:*

$$\mathrm{d}\mathbf{x}_t = \mathbf{u}(\mathbf{x}_t,t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathsf{w}_t. \tag{65}$$

If

$$A_\kappa^\top\mathbf{u}(\kappa\mathbf{x},t) - \mathbf{u}(\mathbf{x},t) = \mathbf{0} \tag{66}$$

for all $\kappa\in\mathcal{G}$*, then* $\mathbf{u}(\mathbf{x},t)+g(t)^2\mathbf{h}(\mathbf{x},t,\mathbf{y},T)$ *satisfies (55) and* $\mu_{t|T}$ *is* $\mathbf{G}$*-invariant for all* $t\in[0,T]$*.*

Proof. By Lem 12, we have $p(\kappa \mathbf{x}_T|\kappa \mathbf{x}_t) = p(\mathbf{x}_T|\mathbf{x}_t)$ for all $\kappa \in \mathcal{G}$. As a result,

$$A_\kappa \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T|\mathbf{x}_t) = \nabla_{\kappa \mathbf{x}_t} \log p(\kappa \mathbf{x}_T|\kappa \mathbf{x}_t). \tag{67}$$

Or equivalently,

$$\mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T) = A_\kappa^\top \mathbf{h}(\kappa \mathbf{x}_t, t, \kappa \mathbf{x}_T, T). \tag{68}$$

As a result,

$$\mathbf{f}(\mathbf{x}, \mathbf{y}, t) = \mathbf{u}(\mathbf{x}, t) + g(t)^2 \mathbf{h}(\mathbf{x}, t, \mathbf{y}, T). \tag{69}$$

satisfies (55), and thus Prop 3 implies $\boldsymbol{\mu}_{t|T}$ is $\mathbf{G}$-invariant. □

Proposition 2. *Assume $\mathbf{u}(\mathbf{x}, t) = u(t)\mathbf{x}$ for some scalar function $u : \mathbb{R} \to \mathbb{R}$. Given any group $\mathcal{G}$ (or $\mathbf{G}$) composed of linear isometries, if the unconditional p_t induced by (1) is G-invariant at $t = 0$, then it is $\mathcal{G}$-invariant for all $t \geq 0$. Likewise, if the conditional $q_t(\mathbf{x}_t|\mathbf{x}_T)$ induced by (3) is $\mathbf{G}$-invariant at $t = 0$ then it is $\mathbf{G}$-invariant for all $t \geq 0$.*

Proof. The first part of the proposition regarding the unconditional p_t follows Prop 3 while the second part basically restates Lem 13 in density notation where their equivalence can be seen by Lem 5. □

C GROUP INVARIANT WEIGHT TIED CONVOLUTIONAL KERNELS

As stated in Sec 5, we currently limit our attention to linear groups $\mathcal{G}_\mathcal{L}$. In this setting, we can directly impose $\mathcal{G}_\mathcal{L}$-equivariance into the diffusion model by constructing specific CNN kernels.

In particular, for a given linear group $\mathcal{G}_\mathcal{L}$ we can construct a group equivariant convolutional kernel $\mathsf{k} \in \mathbb{R}^{d \times d}$, of the form

$$\mathsf{k} = \begin{array}{|c|c|c|c|} \hline k_{1,1} & k_{1,2} & \cdots & k_{1,d} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ \hline k_{d-1,1} & k_{d-1,2} & \cdots & k_{d-1,d} \\ \hline k_{d,1} & k_{d,2} & \cdots & k_{d,d} \\ \hline \end{array}, \tag{70}$$

such that

$$\mathsf{h}(\mathsf{k} * \mathbf{x}) = \mathsf{k} * \mathsf{h}(\mathbf{x})$$

for any $h \in \mathcal{G}_\mathcal{L}$ and $\mathbf{x} \sim p_{data}$ by constraining the individual kernel values to obey a system of equalities set by the group invariance condition

$$\mathsf{h}(\mathsf{k}) = \mathsf{k}. \tag{71}$$

Example: Vertical Flipping. A concrete example, which was discussed in Sec 3.2, is to consider the group $\mathcal{G} = \{\mathsf{f}_x, \mathbf{e}\}$ where f_x is a vertical flipping operation with $\mathsf{f}_x^{-1} = \mathsf{f}_x$. A convolutional kernel $\mathsf{k} \in R^{3 \times 3}$ constrained to be equivariant to actions from this group would take the form:

$$\mathsf{k} = \begin{array}{|c|c|c|} \hline d & a & d \\ \hline e & b & e \\ \hline f & c & f \\ \hline \end{array} \tag{72}$$

It should be clear given the form of k that

$$\mathsf{f}_x(\mathsf{k} * \mathbf{x}) = \mathsf{k} * \mathsf{f}_x(\mathbf{x})$$

and consequently also for f_x^{-1}, as desired. Likewise, we also present the weight-tied kernels for C4 and D4.

Example: The C_4 Cyclic and D_4 Dihedral Group. Recall that the C_4 cyclic group is composed of planar 90 deg rotations about the origin, and can be denoted as $C_4 = \{\mathsf{e}, \mathsf{r}_1, \mathsf{r}_2, \mathsf{r}_3\}$ where r_i represents a rotation of $i \times 90$ deg. Taking a convolutional kernel $\mathsf{k} \in \mathbb{R}^{5\times 5}$ and constraining it to be C_4-equivariant results in k being of the form:

$$\mathsf{k} = \begin{array}{|c|c|c|c|c|} \hline a & b & c & d & a \\ \hline d & e & f & e & b \\ \hline c & f & g & f & c \\ \hline b & e & f & e & d \\ \hline a & d & c & b & a \\ \hline \end{array}. \tag{73}$$

The D_4 dihedral group can then be "constructed" from C_4 by adding the vertial flipping operation from the past example; that is, $D_4 = \{\mathsf{e}, \mathsf{r}_1, \mathsf{r}_2, \mathsf{r}_3, \mathsf{f}_x, \mathsf{f}_x \circ \mathsf{r}_1, \mathsf{f}_x \circ \mathsf{r}_2, \mathsf{f}_x \circ \mathsf{r}_3\}$. This requires further constraints to k so that

$$\mathsf{k} = \begin{array}{|c|c|c|c|c|} \hline a & b & c & b & a \\ \hline b & e & f & e & b \\ \hline c & f & g & f & c \\ \hline b & e & f & e & b \\ \hline a & b & c & b & a \\ \hline \end{array}. \tag{74}$$

Naturally, constraining convolutional kernels in this fashion has the advantage of reducing the number of model parameters – with a possible loss in expressiveness when the kernel size is relatively small in comparison to the size of the group and structure of the data. For a more general discussion on $\mathcal{G}$-equivariant convolutional kernels in the context of CNNs we refer the reader to Cohen and Welling (2016) and Knigge et al. (2022).

D EQUIVARIANCE REGULARIZATION

Instead of achieving $\mathcal{G}$-equivalence by adopting specific model architectures, as described in Sec 5, or by frame averaging Puny et al. (2022), we can also directly add a regularizer to the score-matching loss to inject this preference. Specifically, according to Lem 1, the estimated score $\mathbf{s}_{\boldsymbol{\theta}}(\cdot, t)$ is equivariant if

$$\mathbf{s}_{\boldsymbol{\theta}}(\kappa\mathbf{x}, \kappa\mathbf{y}, t) = \kappa\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}, t), \tag{75}$$

for all $\kappa \in \mathcal{G}$. (For the unconditional distribution, similar techniques can be applied by omitting the second argument of $\mathbf{s}_{\boldsymbol{\theta}}$.) Thus, we propose the following regularizer to encourage the two terms to match for all $\mathbf{x}$ and t:

$$\mathcal{R}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) = \mathbb{E}\left[\frac{1}{|\mathcal{G}|}\sum_{\kappa\in\mathcal{G}} \left\|\mathbf{s}_{\boldsymbol{\theta}}(\kappa\mathbf{x}, \kappa\mathbf{y}, t) - \kappa\mathbf{s}_{\bar{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{y}, t)\right\|^2\right] \tag{76}$$

where the expectation is taken over the same variables in the regular score-matching loss and $\bar{\boldsymbol{\theta}}$ denotes the exponential moving average (EMA) of the model weights

$$\bar{\boldsymbol{\theta}} \leftarrow \text{stopgrad}(\mu\bar{\boldsymbol{\theta}} + (1-\mu)\boldsymbol{\theta}) \quad \text{with} \quad \mu \in [0, 1), \tag{77}$$

which helps improve training stability. In practice, iterating over all elements in $\mathcal{G}$ may be intractable. Thus, for each optimization step, $\mathcal{R}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ is one-sample approximated by:

$$\mathcal{R}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \approx \mathbb{E}\left[\left\|\mathbf{s}_{\boldsymbol{\theta}}(\kappa\mathbf{x}, \kappa\mathbf{y}, t) - \kappa\mathbf{s}_{\bar{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{y}, t)\right\|^2\right], \tag{78}$$

with randomly picked $\kappa \in \mathcal{G}$.

E CONSTRUCTIONS OF EQUIVARIANT NOISY SEQUENCE

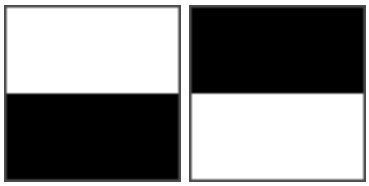
In this section, we present a method to construct an equivariant noisy sequence $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$ with respect to some $\mathbf{x}_n \sim q(\mathbf{x})$ without knowing the "true" orientation of $\mathbf{x}_n$.

Let q denote the distribution of $\mathbf{x}_n$. Construct a function $\phi : \mathbb{R}^d \to \mathbb{R}^d$ such that: (1) for all $\kappa \in \mathcal{G}$, $\mathbf{x} \sim q$ or $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, I)$, $\phi(\kappa\mathbf{x}) = \kappa\phi(\mathbf{x})$ almost surely; (2) for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, there exists a unique $\kappa \in \mathcal{G}$ such that $\phi(\mathbf{x}) = \kappa\phi(\mathbf{y})$. For example, for $\mathbb{R}^2$ with $\mathcal{G}$ consisting of element-swapping operators, ϕ can be the function that outputs one-hot vector indicating the max element of the input. We will present some selections of ϕ for common $\mathcal{G}$ below.

Given starting point $\mathbf{x}_n$ and a noise sequence $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$, choose $\kappa \in \mathcal{G}$ such that $\phi(\mathbf{x}_n) = \kappa\phi(\boldsymbol{\epsilon}_n)$. Then we use the noise sequence $\tilde{\boldsymbol{\epsilon}}_i = \kappa\boldsymbol{\epsilon}_i$ for the evaluation of (11). To see why this approach works, assume that $\mathbf{x}_n$ is updated to $\mathsf{r}\mathbf{x}_n$ for some $\mathsf{r} \in \mathcal{G}$. Then, $\phi(\mathsf{r}\mathbf{x}_n) = \mathsf{r}\phi(\mathbf{x}_n) = (\mathsf{r} \circ \kappa)\phi(\boldsymbol{\epsilon}_n)$, and thus the sequence becomes $\{\mathsf{r} \circ \kappa\, \boldsymbol{\epsilon}_i\}_{i=1}^n = \{\mathsf{r}\tilde{\boldsymbol{\epsilon}}_i\}_{i=1}^n$. Note that this is a general method to create an equivariant noise sequence with respect to any input.

Below, we present some choices of ϕ for some common linear operator groups for 2D images.

Example: Vertical Flipping. The function ϕ_v can be chosen to output either of two images:

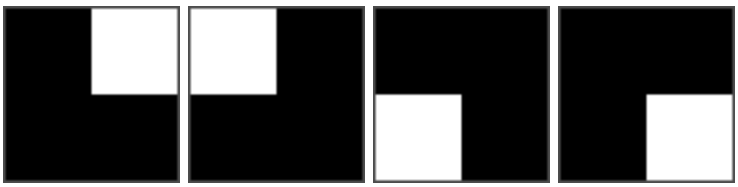


Specifically, if the input image $\mathbf{x}$ has the max value on the upper half of the image, ϕ returns the left plot; otherwise, the right one. (Here, we assume that it is almost surely that the max value cannot appear in both halves.)

It is obvious that if the input is flipped vertically, the output will be flipped in the same way. Therefore, the first condition is satisfied. For the second, if $\phi_v(\mathbf{x})$ and $\phi_v(\mathbf{y})$ have the same output, κ is the identity operator; otherwise, κ is the vertical flipping. For multichannel input, ϕ can be applied independently to each channel.

Applying the same idea, we can derive the corresponding ϕ_h for horizontal flipping.

Example: C_4 Cyclic Group. We can use a similar idea to derive ϕ_{C4} for C_4 cyclic that is composed of planar 90 deg rotations about the origin. In this case, ϕ_{C4} has four possible outputs



such that ϕ_{C4} assigns the quadrant white if the input has the max value in that quadrant. (Here, we assume it is almost surely that the max value cannot appear in multiple quadrants.) Then, it is straightforward to see that ϕ_{C4} satisfies the two conditions of ϕ.

Example: D_4 Dihedral Group. As we have mentioned in Appx C, the D_4 dihedral group can be "constructed" from C_4 by adding the vertial flipping operation. As a result, we can combine ϕ_v and ϕ_{C4} to construct the corresponding ϕ_{D4} for D_4. Assume that ϕ_v assigns one to the elements corresponding to the white pixels and zero to the ones associated with the black. Likewise, let ϕ_{C4} assign two to the elements corresponding to the white pixels and zero to the rest. Then we define $\phi_{D4} = \phi_v + \phi_{C4}$. It is easy to check that both ϕ_v and ϕ_{D4} satisfy the first condition of ϕ for all $\kappa \in D_4$ (i.e., vertical flipping, rotation, and their composition). We also note that the range ϕ_{D4} contains eight distinct elements. Starting from one element, we get all the elements by applying one of the eight operators in $D_4 = \{\mathsf{e}, \mathsf{r}_1, \mathsf{r}_2, \mathsf{r}_3, \mathsf{f}_x, \mathsf{f}_x \circ \mathsf{r}_1, \mathsf{f}_x \circ \mathsf{r}_2, \mathsf{f}_x \circ \mathsf{r}_3\}$, which suggests ϕ_{D4} satisfies the second condition.

F DATASET DETAILS

This section contains detailed discussion on the contents and preprocessing of each dataset mentioned in Sec 6.

F.1 Rotated MNIST

Rotated MNIST dataset (Larochelle et al., 2007) contains random 90° rotations of MNIST images (Deng, 2012), resulting in a C_4-invariant distribution. This dataset was generated following the description in Knigge et al. (2022), and has been commonly used to evaluate group-invariant CNN models, as seen in (Dey et al., 2021; Birrell et al., 2022), with experiments on 1% (600), 5% (3000), and 10% (6000) of the dataset.

F.2 LYSTO

The LYSTO dataset (Jiao et al., 2023) consists of 20,000 labeled image patches at a resolution of 299x299x3 extracted at 40X magnification from breast, colon, and prostate cancer samples stained with CD3 or CD8 dyes. This data was preprocessed by first scaling all the images to 128x128x3 before randomly sampling 64x64x3 image patches from each image to generate the final dataset. The data was first scaled to increase the feature density within each randomly sampled patch. The resulting data exhibits D_4 invariance due to natural rotational and mirror invariance.

F.3 LYSTO Denosing

To construct the LYSTO denosing dataset, we take the LYSTO 64x64x3 patches, generated as described above, and downscale them to 1/4 the resolution and then upscaled back using LANCZOS interpolation to form conditional training pairs.

F.4 ANHIR

The ANHIR dataset (Borovec et al., 2020) provides whole-slide images of lesions, lung-lobes, and mammary-glands at a variety of different resolutions, from 15kx15k to 50kx50k. We only make use of the lung images. This data was processed following the method outlined in Dey et al. (2021) from which random 64x64x3 image patches are extracted.

F.5 CT-PET

The CT-PET dataset (Gatidis et al., 2022) includes 1014 (501 positives and 512 controls) annotated whole-body paired FDG-PET/CT scans, comprised of 3D voxels, of patients with malignant lymphoma, melanoma, and non-small cell lung cancer. We extract 40 slices per voxel to construct the training datasets. This data is restricted under TCIA restricted lience. Formal access must be filed for and granted before the data can be made available to practitioners from the Cancer Imaging Archive (CIA). Scripts for processing this data, such as into slices, are provided by CIA on the dataset page.

The style-transfer dataset was constructed by first slicing the 3D voxel volumes of the patients into 2D images of the middle of the patients. These images were then cropped to just contain the torso and head and then scaled to 256x256x3. A patient's CT scan slice was then paired with the matching PET scan image slice to form the final dataset.

G MODEL DETAILS

The implementation details and hyperparamters used while training the models presented in Sec 6 over the listed datasets are given below. Table 5 provides a summary of all the models discussed within the paper and their theoretical invariance (equivariance) guarantees.

We refer the reader to the reader to Appx.G. of Birrell et al. (2022) for a more in-depth discussion of the implementation details and training parameters used to produce the results of SP-GAN reported in Sec 6. We will only summarize the training parameters we changed from the defaults given in the forgoing reference.

Model	Arch.	$\mathcal{G}_{\mathcal{L}}$-inv Smpl	Eqv Traj
VP-SDE	U-Net	✗	✗
SPDM+WT	U-Net (WT)	✓	✓
SPDM+FA	U-Net	✓	✓
DDBM	U-NET	✗	✗
SPDM+FA (Bridge)	U-NET	✓	✓
SP-GAN	CNN	✓	–
GE-GAN	CNN	✗	–
Pix2Pix	U-NET & CNN	✗	–
I^2SB	U-NET	✗	✗

Table 5: Model summary. VP-SDE denotes the regular diffusion model with variance preservation configuration in Table 4. ✓: Theo. guaranteed ✗: Not theo. guaranteed –: Not Applicable

All Diffusion models (VP-SDE, SPDM+WT, SPDM+FA, SPDM+FA(Bridge)) are trained using the Adam optimizer Kingma and Ba (2015) with learning rate $\eta = 0.0001, 0.0002$, and weight decay rate of $\gamma = 0.0$; separately, we make use of the exponential moving average (EMA) of the model weights (77) with $\mu = 0.999, 0.9999, 0.9999432189950708$, which are the values commonly used when training this style of diffusion model.

G.1 Rotated MNIST

For the Rotated MNIST datasets, the diffusion models are configured using the following model parameters: dropout rate $\mathsf{d} = 0.1$.

Model	Loss	Batch	Cond.	Aug	Attn. res.	Num. Ch.	Num. Heads	Ch. Scal.	Scale Shift
SPDM	L_2	32	True	True	8	128	16	1,2,2	True
SPDM+WT	L_2	32	True	False	8	128	16	1,2,2	True
SPDM+FA	$L_2 + FA$	32	True	True	8	128	16	1,2,2	True

The GAN based methods were trained using the scripts provided by Dey et al. (2021) and Birrell et al. (2022) with the following settings:

Model	Loss	Batch	Cond.	Aug	latent dim	gp weight	lr	alpha
SP-GAN	D_2^L	64	True	True	64	10.0	$1e{-}4$	2
GE-GAN	RA	64	True	True	64	10.0	$1e{-}4$	–

G.2 LYSTO

For the LYSTO dataset, the diffusion models are configured using the following model parameters: dropout rate $\mathsf{d} = 0.1$.

Model	Loss	Batch	Cond.	Aug	Attn. res.	Num. Ch.	Num. Heads	Ch. Scal.	Scale Shift
VP-SDE	L_2	32	True	True	32,16,8	128	64	1,2,2,2	True
SPDM+WT	L_2	32	True	False	32,16,8	128	64	1,2,2,2	True
SPDM+FA	$L_2 + FA$	32	True	True	32,16,8	128	64	1,2,2,2	True

The GAN based methods were trained using the scripts provided by Dey et al. (2021) and Birrell et al. (2022) with the following settings:

Model	Loss	Batch	Cond.	Aug	latent dim	gp weight	lr	alpha
SP-GAN	D_2^L	32	True	True	128	10.0	$1e{-}4$	2
GE-GAN	RA	32	True	True	128	10.0	$1e{-}4$	–

G.3 ANHIR

For the ANHIR dataset, the diffusion models are configured using the following model parameters: dropout rate $\mathsf{d} = 0.1$.

Model	Loss	Batch	Cond.	Aug	Attn. res.	Num. Ch.	Num. Heads	Ch. Scal.	Scale Shift
VP-SDE	L_2	32	True	True	32,16,8	128	64	1,2,2,2	True
SPDM+WT	L_2	32	True	False	32,16,8	128	64	1,2,2,2	True
SPDM+FA	$L_2 + FA$	32	True	True	32,16,8	128	64	1,2,2,2	True

The GAN based methods were trained using the scripts provided by Dey et al. (2021) and Birrell et al. (2022) with the following settings:

Model	Loss	Batch	Cond.	Aug	latent dim	gp weight	lr	alpha
SP-GAN	D_2^L	32	True	True	128	10.0	$1e{-}4$	2
GE-GAN	RA	32	True	True	128	10.0	$1e{-}4$	–

G.4 LYSTO Denoising Task

For the LYSTO denoising dataset, the diffusion models are configured using the following model parameters:

Model	Loss	Batch	Cond.	Aug	Attn. res.	Num. Ch.	Num. Heads	Ch. Scal.	Scale Shift
VP-SDE	L_2	32	True	True	32,16,8	128	64	1,2,2,2	True
SPDM+FA	$L_2 + FA$	32	True	True	32,16,8	128	64	1,2,2,2	True
Pix2Pix	GAN	32	True	True	–	–	–	–	–
I^2SB	L_2	64	True	True	–	–	–	–	–

The entries in the above table are left empty for both Pix2Pix and I^2SB as their architectures differ from the other models. In particular, Pix2Pix being a GAN makes use of a U-NET for the generator and custom discriminator architecture. The model comes with several predefined U-NET architecture configurations that can be selected; however, it is lacking a configuration for 64x64 images. We defined a suitable configuration by modifying that provided for 128x128 by reducing the number of down-sampling layers from 7 to 4. All other settings are left as default. The I^2SB model by default it make use of a preconfigured U-NET architecture which it downloads from openai. For this task we simply make use of the default configuration settings without any modification.

G.5 CT-PET Style Transfer Task

For the CT-PET dataset, all models except Pix2Pix are trained in latent space, where the original images are first encoded by a fine-tuned pretrained VAE from stable diffusion (Rombach et al., 2022). FA was applied during the fine-tuning and inference to ensure equivariance. The VAE takes the 256x256x3 images and encodes them into a 32x32x4 latent space representation.

Models are configured using the following model parameters:

The entries in the above table are left empty for both Pix2Pix as the architecture differs from the other models. Pix2Pix makes use of a U-NET for the generator and custom discriminator architecture. The model comes with a predefined U-NET architecture for 256x256x3 images which we use. All other settings are left as default. As discussed above the I^2SB model by default it make use of a preconfigured U-NET architecture. We can't make

Model	Loss	Batch	Cond.	Aug	Attn. res.	Num. Ch.	Num. Heads	Ch. Scal.	Scale Shift
DDBM	L_2	32	True	True	32,16,8	128	64	1,2,2,2	True
SPDM+FA	$L_2 + FA$	32	True	True	32,16,8	128	64	1,2,2,2	True
Pix2Pix	GAN	32	True	True	–	–	–	–	–
I^2SB	L_2	64	True	True	32,16,8	128	64	1,2,2,2	True

use of the default architecture for this task as it is incompatible with the latent space embedding produced by the VAE. Instead we configure the U-NET architecture in an identical fashion to SPBM.

G.6 Computational Resources

All the model results reported within Sec 6 were trained using NVIDIA A40 or L40S equivalent. Training times for each model are reported in Table 6.

Table 6: Model training times for experiments discussed in Sec 6.

Training Times for LYSTO & ANHIR				**Training Times for LYSTO & CT-PET**			
Model	GPUs	LYSTO	ANHIR	Model	GPUs	LYSTO	CT-PET
VP-SDE	2	5 days	5 days	DDBM	4	2 days	2 days
SPDM+WT	2	2 weeks	2 weeks	SPDM+FA	4	2 days	2 days
SPDM+FA	2	5 days	5 days	Pix2Pix	1	2 hours	1 day
SP-GAN	1	2 days	2 days	I^2SB	2	3 days	3 days
GE-GAN	1	2 days	2 days				

H FID COMPUTATION

In Sec 6.1 we report the Fréchet intercept distance (FID) (Heusel et al., 2017) score of the various models on the datasets described in Sec 6 and Appx F, respectively under C_4 and D_4 groups. In order to make the FID score robust to changes in image orientation, meaning the features the underlying InceptionV3 model extracts from the reference dataset can be compared to those extracted from the generated samples, we average the reference statistics over all actions within the group considered. In particular, suppose $\mathcal{D}_{ref}$ is a reference dataset (e.g., all the images within the rotated MNIST dataset) and $\mathcal{D}_s$ is a collection of samples generated from the model being evaluated with respect to group $\mathcal{G}$. Let $\mathrm{T}(\cdot)$ denote the operation that returns the mean and covariance statistics of the features extracted from a dataset; i.e., $T(\mathcal{D}_f) = (\mu_s, \Sigma_s)$. Moreover, recall that FID is computed using the expression:

$$\mathrm{FID} = d^2(\mathrm{T}(\mathcal{D}_{ref}, \mathrm{T}(\mathcal{D}_s)) = \|\mu_{ref} - \mu_s\|_2^2 + \mathrm{Tr}(\Sigma_{ref} + \Sigma_s - 2(\Sigma_{ref}\Sigma_s)^{1/2}). \tag{79}$$

Then the calculation of the FID with respect to the group $\mathcal{G}$ is done by first computing

$$\mathrm{T}_{\mathcal{G}}(\mathcal{D}) = \frac{1}{|\mathcal{D}|}\sum_{\mathsf{h}\in\mathcal{G}} T(A_{\mathsf{h}}\mathcal{D}) = (\hat{\mu}, \hat{\Sigma}), \tag{80}$$

where $A_{\mathsf{h}}\mathcal{D} = \{A_{\mathsf{h}}\mathsf{x} \mid \mathsf{x} \in \mathcal{D}\}$. Then we compute the final FID score as

$$\mathrm{FID}_{\mathcal{G}} = d^2(\mathrm{T}_{\mathcal{G}}(\mathcal{D}_{ref}), \mathrm{T}(\mathcal{D}_s)). \tag{81}$$

This formulation ensures that the reference statistics used in computing the FID score of a model conditioned to be equivalent are not biased. All FID values reported in Table 2 and Table 3, potentially excluding those reported by other authors, were calculated in this fashion.

I LIMITATIONS

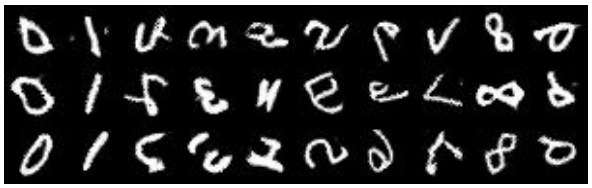
Here we provide a summary discussion of some limitations of the proposed method. Note that although our theory can handle arbitrary groups of linear isometries, our implemented methods are restrictioned to groups

with finitely many elements, as the techniques +WT and +FA proposed cannot be immediately generalized to infinite groups without necessary approximation. With that said, we would like to note that when a specific group is chosen it may be feasible to design specialized models to achieve perfect equivariance even if the group contains infinite elements. For example, when not considering the structure of the drift's attribute, existing work typically employs GNN to achieve SO(3) or SE(3) equivariance. We believe this approach can also be adapted to fit our more general setting. As our work concentrates on general groups, we have decided to reserve the study of specific groups for future research.

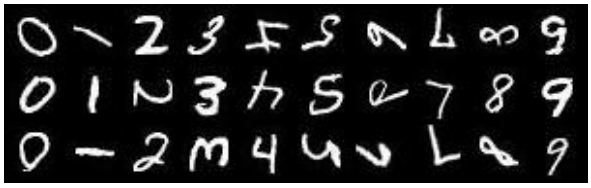
J SAMPLE IMAGES

Here, we include a collection of generated image samples from the models discussed within the paper across the various datasets in Section 6.

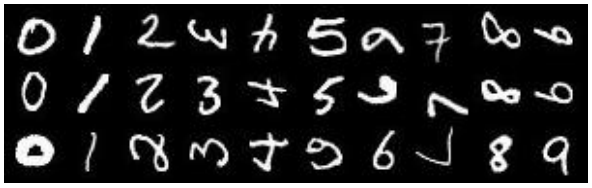
(a) Reference C_4 rotated MNIST images.



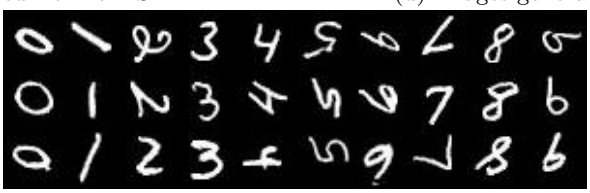
(b) Images generated from SP-GAN.



(c) Images generated from VP-SDE.

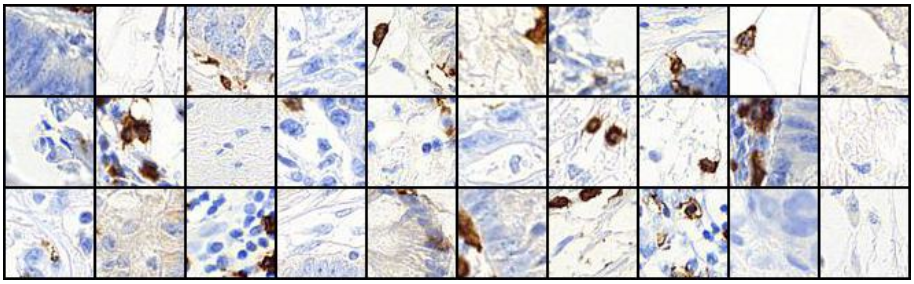


(d) Images generated from SPDM+WT

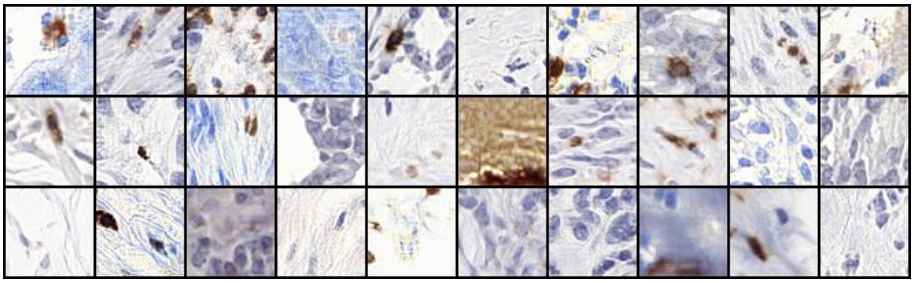


(e) Images generated from SPDM+FA

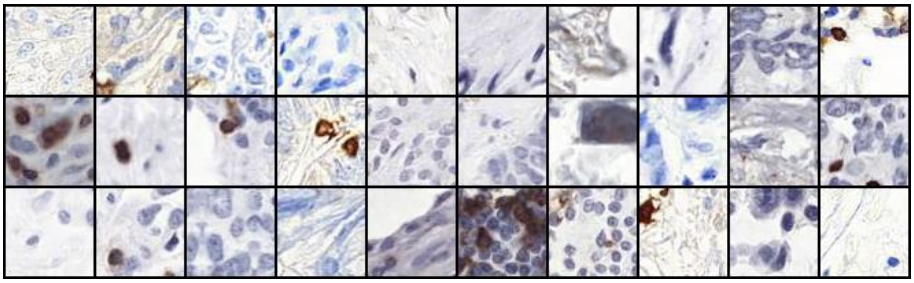
Figure 6: Sample comparison between models trained on the Rotated MNIST 28x28x1 dataset as described in Sec 6.



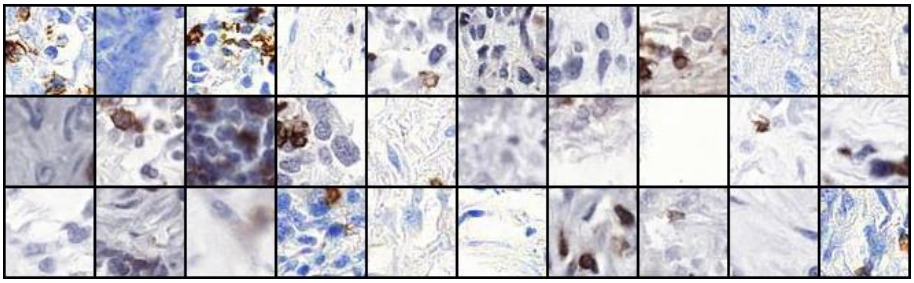
(a) Reference images.



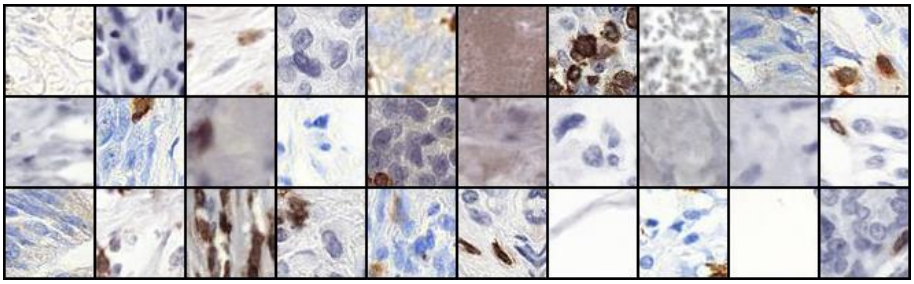
(b) Images generated from SP-GAN.



(c) Images generated from VP-SDE

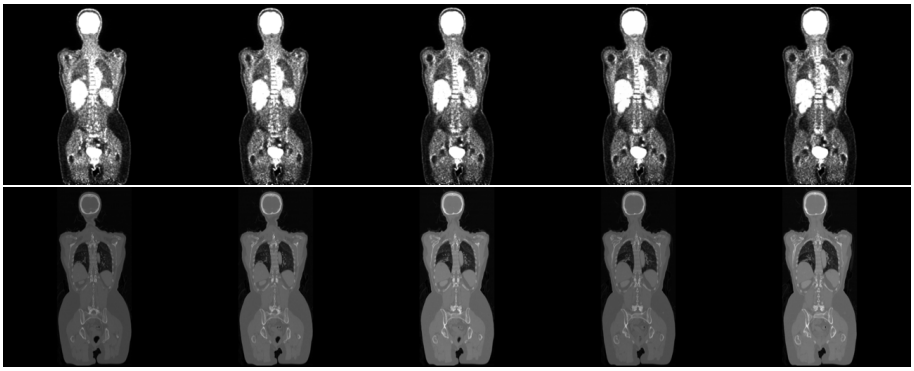


(d) Images generated from SPDM+WT.

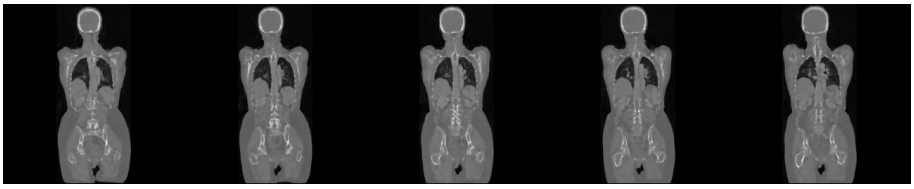


(e) Images generated from SPDM+FA

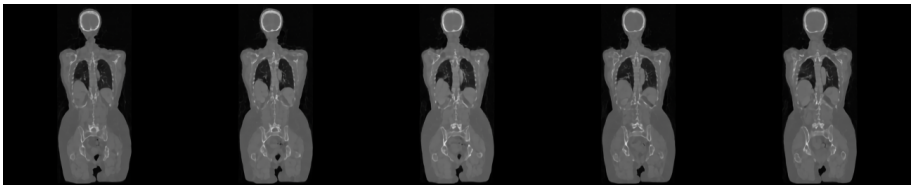
Figure 7: Sample comparison between models trained on the LYSTO 64x64x3 dataset from Sec 6.



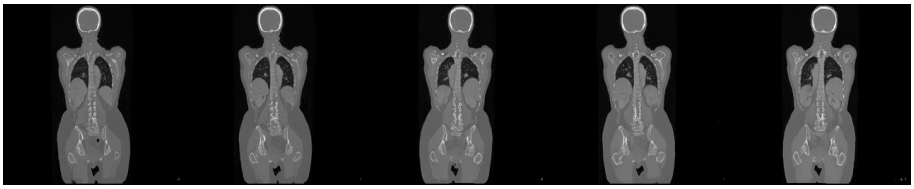
(a) Reference images of CT and PET images.



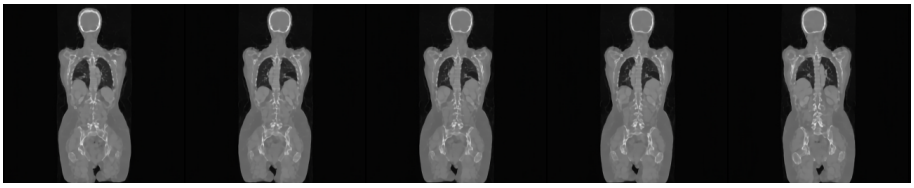
(b) Images generated from DDBM.



(c) Images generated from SPDM+FA



(d) Images generated from Pix2Pix.



(e) Images generated from I^2SB

Figure 8: Sample comparison between models trained on the CT-PET 256x256x3 dataset from Sec 6 and Appx F.