
Diffusion Generative Modeling on Lie Group Representations

Marco Bertolini*, Tuan Le* & Djork-Arné Clevert

Machine Learning Research

Pfizer Worldwide Research and Development

Friedrichstraße 110, 10117 Berlin, Germany

{marco.bertolini,tuan.le,djork-arne.clevert}@pfizer.com

Abstract

We introduce a novel class of score-based diffusion processes that operate directly in the representation space of Lie groups. Leveraging the framework of Generalized Score Matching, we derive a class of Langevin dynamics that decomposes as a direct sum of Lie algebra representations, enabling the modeling of any target distribution on any (non-Abelian) Lie group. Standard score-matching emerges as a special case of our framework when the Lie group is the translation group. We prove that our generalized generative processes arise as solutions to a new class of paired stochastic differential equations (SDEs), introduced here for the first time. We validate our approach through experiments on diverse data types, demonstrating its effectiveness in real-world applications such as $SO(3)$ -guided molecular conformer generation and modeling ligand-specific global $SE(3)$ transformations for molecular docking, showing improvement in comparison to Riemannian diffusion on the group itself. We show that an appropriate choice of Lie group enhances learning efficiency by reducing the effective dimensionality of the trajectory space and enables the modeling of transitions between complex data distributions.

1 Introduction

Deep probabilistic generative modeling amounts to creating data from a known tractable prior distribution. Score-based models (Hyvärinen & Dayan, 2005; Sohl-Dickstein et al., 2015; Ho et al., 2020; Huang et al., 2021; Song et al., 2021, 2020b) achieve this by learning to reverse a corruption process of the data. Most algorithms assume an Euclidean data space X , yet many scientific applications (Brehmer & Cranmer, 2020; Zhang et al., 2024; Klimovskaia et al., 2020; Karpadne et al., 2018) involve distributions on curved manifolds $\mathcal{M}$. While significant progress has been made in developing the theory of diffusion in curved spaces (De Bortoli et al., 2022; Huang et al., 2022), key challenges remain: parametrizing vector fields on general $\mathcal{M}$ is unsolved, and Langevin updates require projection to preserve the manifold structure. Even when $\mathcal{M} = G$ is a Lie group, denoising score-matching remains a challenge for general non-Abelian groups, thus necessitating explicit trajectory simulation. Recent findings (Abramson et al., 2024) highlight this complexity, as diffusion was performed in raw Cartesian coordinates rather than explicitly modeling the torsion space, given its representational difficulty and lack of performance gain.

An appropriate representation that leverages the symmetry property of the data should, however, enable models to better capture the underlying physical laws. The limited performance of manifold-based diffusion must thus stem from technical and computational difficulties rather than fundamental principles. This work seeks to reconcile this expectation with the empirical findings by addressing the question: *Given a Lie group G acting on Euclidean space X through a map (representation)*

*Shared first authorship.

$\rho_X : G \rightarrow GL(X)$, can we construct a **generative process on X that models any distribution on G** , thus retaining the advantages of flat-space diffusion while capturing non-trivial manifold structures? We address this question by constructing the diffusion process directly in the representation space, defined as the image of the group action map, $\text{Im}(\rho_X) \subseteq GL(X)$. This yields a matrix-valued diffusion process in $GL(X)$ which, when applied to elements of X , induces a stochastic flow corresponding to infinitesimal Lie group transformations, i.e., Lie algebra elements. In this way, the process preserves the geometric inductive bias of the (curved) Lie group while remaining entirely within the flat vector space X . Our construction builds on the framework of *Generalized Score Matching* (GSM) (Lyu, 2009; Lin et al., 2016), which estimates probability densities via the generalized score function $\mathcal{L} \log p(\mathbf{x})$ for a suitable linear operator $\mathcal{L}$. We show that the G -induced generative process satisfies a continuous-time stochastic differential equation (SDE) involving this generalized score. As illustrated in Figure 1, our approach differs to diffusion processes directly on the Lie group: rather than mapping data to the group and back via the representation map, we remain in X throughout, using the differential of the representation $d\rho_X : TG \rightarrow TX$ to guide the Langevin dynamics.

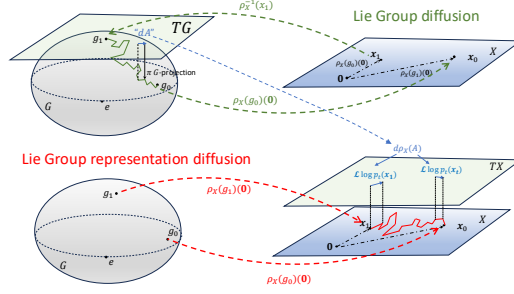


Figure 1: Comparison of strategies between Lie group (top) diffusion and our proposed Lie group representation diffusion (bottom).

In short, we propose an exact SDE-based diffusion framework that enables general generative modeling on Lie group representations, thus combining the advantages of curved dynamics with the theoretical and practical effectiveness of Euclidean diffusion. Our method realizes **simulation-free** training of Lie group-like diffusion models, and it provides a novel approach to denoising score-matching for general non-Abelian groups. Our main contributions can be summarized as follows:

Generalized score matching via Lie algebras: We extend generalized score matching on X to estimate the score of any distribution on a Lie group G acting on X . We elucidate the conditions for a suitable G (valid for any differentiable manifold X). We recover standard score-matching as a specific case of our framework, corresponding to the group $G = T(n)$ of translations on $X = \mathbb{R}^n$.

Lie group representation diffusion processes as exact solution of a novel class of SDEs: We introduce a new class of solvable SDEs that govern Lie group diffusion via Euclidean coordinates, significantly expanding the range of processes that can be addressed using score-based modeling techniques. We also show that our approach extends naturally to flow matching (Appendix F).

Dimensionality reduction, bridging non-trivial distributions and trajectory disentanglement: Through extensive experiments, we demonstrate that: (1) our approach can estimate, regardless of the choice of G , any probability density (Sections 5 (2,3,4d distributions) and (QM9)); (2) by appropriately selecting G to align with the data structure, the learning process is significantly simplified, effectively reducing its dimensionality (Section 5(MNIST)) (3) our framework enables solutions to processes that are challenging or unfeasible with standard score matching, such as bridging between complex data-driven distributions (Section 5 (MNIST) and (CrossDocked)).

2 Diffusion dynamics through Lie algebras

We start this section by setting up notation and review the connection between vector fields and Lie algebra actions on manifolds. A Lie group G is a group that is also a finite-dimensional differentiable manifold, such that the group operations of multiplication $\cdot : G \times G \rightarrow G$ and inversion are C^∞ -functions[†]. A Lie algebra $\mathfrak{g}$ is a vector space equipped with an operation, the Lie bracket, $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$, satisfying the Jacobi identity. Every Lie group gives rise to a Lie algebra as its tangent space at the identity, $\mathfrak{g} = T_e G$, and the Lie bracket is the commutator of tangent vectors,

[†]We restrict ourselves to real Lie groups. It would be interesting to extend our analysis to the complex case (Le et al., 2021).

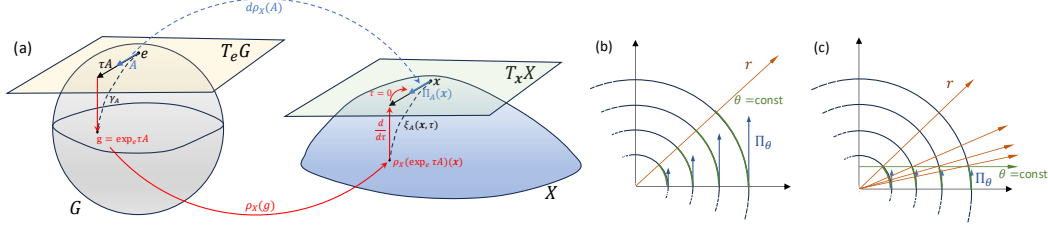


Figure 2: (a) Depiction of the fundamental vector field definition (1). Flow coordinates for a pair of commuting (b) and not-commuting ones vector fields (c).

$[A, B] = AB - BA$. In this work, we are interested in how Lie groups and Lie algebras act on spaces. Given a manifold X , a (left) **group action** of G on X is an associative map[†] $\rho_X : G \times X \rightarrow X$ such that $\rho_X(e, x) = x, \forall x \in X$. Fundamental concepts associated with a group action are the ones of orbits and stabilizers. The **orbit** of $\mathbf{x} \in X$ is the set of elements in X which can be reached from $\mathbf{x}$ through the action of G , i.e., $G \cdot \mathbf{x} = \{\rho_X(g)(\mathbf{x}), g \in G\}$. The **stabilizer** subgroup of G with respect to $\mathbf{x}$ is the set of group elements that fix $\mathbf{x}$, $G_{\mathbf{x}} = \{g \in G | \rho_X(g)(\mathbf{x}) = \mathbf{x}\}$. The **action of a Lie algebra** on X , $\mathfrak{A} : \mathfrak{g} \rightarrow \text{Vect}(X)$ is a Lie algebra homomorphism and maps elements of $\mathfrak{g}$ to vector fields on X such that the map $\mathfrak{g} \times X \rightarrow TX, (A, \mathbf{x}) \mapsto \mathfrak{A}(A)(\mathbf{x})$ is smooth. Given $A \in \mathfrak{g}$ and a group action ρ_X , the **flow** on X induced by ρ_X is given by $\xi_A : X \times \mathbb{R} \rightarrow X, (\mathbf{x}, \tau) \mapsto \rho_X(\exp(\tau A))(\mathbf{x})$, where the map $\exp : \mathfrak{g} \rightarrow G$ is defined by $\exp(A) = \gamma_A(1)$, where $\gamma_A : \mathbb{R} \rightarrow G$ is the unique one-parameter subgroup of G whose tangent vector at the identity is A . The **infinitesimal action** of $\mathfrak{g}$ on X , $d\rho_X : \mathfrak{g} \rightarrow \text{Vect}(X)$, is defined as the differential of the map ρ_X , that is,

$$d\rho_X : A \mapsto \left. \frac{d}{d\tau} \right|_{\tau=0} \rho_X(\exp(\tau A))(\mathbf{x}) \equiv \Pi_A(\mathbf{x}). \quad (1)$$

Π_A is called the **fundamental vector field** corresponding to $A \in \mathfrak{g}$. Given a fixed point $\mathbf{x}_0 \in X$, we denote $\tau = \xi_A(\mathbf{x}_0)^{-1}(\mathbf{x})$ the **fundamental flow coordinate**, which is the parameter such that applying the flow to $\mathbf{x}_0$ gives $\mathbf{x}$. Central to our discussion is the fact that any smooth vector field $V : X \rightarrow TX$ on X can be interpreted as a differential operator acting on smooth functions $f : X \rightarrow \mathbb{R}$. The operator $V(f)$ represents the directional derivative of f at $\mathbf{x} \in X$ in the direction of $V(\mathbf{x})$. We denote $\mathcal{L}_A = \Pi_A \cdot \nabla$ the differential operator corresponding to Π_A . In the following we will use both Π_τ and Π_A interchangeably, when no potential confusion arises. When $\dim \mathfrak{g} > 1$ we indicate as $\mathbf{\Pi}(\mathbf{x}) = (\Pi_{A_1} \quad \Pi_{A_2} \quad \dots)$ the matrix of the collection of fundamental vector fields.

Let us work out the example for $X = \mathbb{R}^2$ and $G = \text{SO}(2)$, the group of rotations in the plane.

The Lie algebra $\mathfrak{so}(2)$ consists of all matrices of the form $A_\alpha = \begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix}$, where $\alpha \in \mathbb{R}$, and the Lie bracket is identically zero. The flow on X induced by ρ_X is given by the exponential map $\rho_{\mathbb{R}^2}(\exp(\tau A_\alpha))(\mathbf{x}) = \begin{pmatrix} \cos(\alpha\tau) & -\sin(\alpha\tau) \\ \sin(\alpha\tau) & \cos(\alpha\tau) \end{pmatrix} \mathbf{x}$, and without loss of generality we can set $\alpha = 1$. The infinitesimal action is computed as

$$d\rho_{\mathbb{R}^2}(A) = \left. \frac{d}{d\tau} \right|_{\tau=0} \begin{pmatrix} \cos \tau & -\sin \tau \\ \sin \tau & \cos \tau \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}.$$

and thus the fundamental vector field defines the derivation $\mathcal{L}_A(\mathbf{x}) = -x_2 \frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial x_2}$. Let $\mathbf{x}_0 \in \mathbb{R}^2$ be a fixed point, then the flow equation $\mathbf{x}(\tau) \equiv \xi_A(\mathbf{x}_0, \tau) = \rho_{\mathbb{R}^2}(\exp(\tau A), \mathbf{x}_0)$ gives a system of two equations, which we can solve to find the expression of the fundamental flow coordinate

$$\begin{cases} \mathbf{x} \cdot \mathbf{x}_0 &= |\mathbf{x}_0|^2 \cos \tau, \\ \mathbf{x} \times \mathbf{x}_0 &= |\mathbf{x}_0|^2 \sin \tau, \end{cases} \Rightarrow \tau = \arctan \frac{\mathbf{x} \times \mathbf{x}_0}{\mathbf{x} \cdot \mathbf{x}_0}, \quad \text{where } \mathbf{x} \times \mathbf{y} = y_2 x_1 - x_1 y_2. \quad (2)$$

Note that $\frac{\partial}{\partial \tau} = \frac{\partial x_1}{\partial \tau} \frac{\partial}{\partial x_1} + \frac{\partial x_2}{\partial \tau} \frac{\partial}{\partial x_2} = -x_2 \frac{\partial}{\partial x_1} + x_1 \frac{\partial}{\partial x_2} = \Pi_A(\mathbf{x})^\top \nabla = \mathcal{L}_A$.

[†]In the manuscript we adopt both notations $\rho_X(g)(\mathbf{x})$, derived from defining $\rho_X : G \rightarrow \text{GL}(X)$, and $\rho_X(g, \mathbf{x})$, derived from the definition $\rho_X : G \times X \rightarrow X$, which are obviously equivalent.

2.1 Intuition behind Lie group-induced generalized score matching

Score matching aims at estimating a (log) probability density $p(\mathbf{x})$ by learning to match its score function, i.e., its gradient in data space. Generalized score matching replaces the gradient operator with a general linear operator $\mathcal{L}$ (Lyu, 2009). The learning objective is given by minimizing the generalized Fisher divergence

$$D_{\mathcal{L}}(p||q_{\theta}) = \int_X p(\mathbf{x}) |\mathcal{L} \log p(\mathbf{x}) - \mathbf{s}_{\theta}(\mathbf{x})|^2 d\mathbf{x}, \quad (3)$$

where $\mathbf{s}_{\theta} = \mathcal{L} \log q_{\theta}$. The requirement on the choice of $\mathcal{L}$ is that it preserves all the information about the original density. Formally, we require $\mathcal{L}$ to be *complete*, that is, given two densities $p(\mathbf{x})$ and $q(\mathbf{x})$, $\mathcal{L}p(\mathbf{x}) = \mathcal{L}q(\mathbf{x})$ (almost everywhere [‡]) implies that $p(\mathbf{x}) = q(\mathbf{x})$ (almost everywhere).

Given a Lie group G acting on X , the collection of fundamental fields Π corresponding to a choice of basis $\mathbf{A} = (A_1, A_2, \dots)$ of $\mathfrak{g}$ is a linear operator, thus potentially suitable for score-matching. It is then natural to set $\mathcal{L}$ to the derivation associated with the fundamental fields Π , i.e., $\mathcal{L} = \Pi(\mathbf{x})^{\top} \nabla$. It then follows that $\mathcal{L} \log p(\mathbf{x})$ computes the directional derivatives of $\log p(\mathbf{x})$ with respect to the fundamental flow coordinates τ , and provided that Π meets some consistency conditions (which we will address in the next section), we can employ $\mathcal{L} \log p(\mathbf{x})$ to sample from $p(\mathbf{x})$ using Langevin dynamics:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \mathbf{s}_{\theta}(\mathbf{x}_t) d\rho_X(\exp(\tau \mathbf{A}))(\mathbf{x}_t) = \mathbf{x}_t - \sum_i \underbrace{\mathcal{L}_i \log p_t(\mathbf{x}_t)}_{\text{generalized scores}} \underbrace{\Pi_{A_i}(\mathbf{x}_t)}_{A_i \text{ directions}} \Delta t, \quad (4)$$

where Δt is the step size and we have temporarily set aside stochasticity and denoising aspects. This process mirrors the intuition depicted in Figure 1: each infinitesimal step of the dynamics corresponds to infinitesimal transformations along the flow on X induces by the G -action, and each component of the generalized score is learned through maximum likelihood over the orbits ξ_{A_i} of the corresponding transformations.

2.2 Sufficient conditions for Lie group-induced generalized score matching

We now address the properties our setup $(X, G, \mathfrak{g}, \Pi)$ must satisfy to meet the sufficient conditions for score-matching and Langevin dynamics. We note that these results hold for any differentiable manifold X . Proofs for these results can be found in Appendix C.

Condition 1: Completeness of Π . We start by establishing an algebraic-geometric condition for Π 's completeness:

Proposition 2.1. *The linear operator $\Pi(\mathbf{x})$ is complete if Π is the local frame of a vector bundle E over X whose rank is $n \geq \dim X$ almost everywhere. If $\text{rank } E = n$ everywhere, then $E = TX$, the tangent bundle of X .*

The following result specifies which Lie groups yield operators Π satisfying the above proposition:

Proposition 2.2. *The operator Π induced by $\mathfrak{g}$ is complete if and only if the subspace $U \subseteq X$ such that $\dim \frac{G}{G_{\mathbf{x}}} < n$ for $\mathbf{x} \in U$, where $n = \dim X$, has measure zero in X .*

As an example, consider standard score-matching on mass-centered point clouds. Here $X = \mathbb{R}^{3N-3}$, since the points' coordinates satisfy $\sum_{i=1}^N \mathbf{x}_i = 0$. Without loss of generality, X can be parametrized by $\mathbf{x}_1, \dots, \mathbf{x}_{N-1}$, with $\mathbf{x}_N$ determined by the center of mass condition. The group $G = T(3N)$ acts transitively on X , with a 3-dimensional stabilizer subgroup $G_X = \{(0, \dots, 0, \mathbf{a})^{\top} \in \mathbb{R}^{3N}\}$ fixing the space. Thus, $\dim G/G_X = n$ for all $\mathbf{x} \in X$, satisfying Proposition 2.2.

Condition 2: Homogeneity of X . While the completeness of the operators is necessary for estimating the target density, it is not sufficient to ensure that the Langevin dynamics (4) will behave appropriately, as the following example illustrates. Let $X = \mathbb{R}$, and $G = \mathbb{R}_+^*$, the multiplicative group of non-zero positive real numbers. The orbits under the action $\rho_X(a, x) = ax$ are $\mathcal{O}_+ = (0, \infty)$,

[‡]Almost everywhere means everywhere except for a set of measure zero, where we assume the standard Lebesgue measure.

$\mathcal{O}_- = (-\infty, 0)$, and $\mathcal{O}_0 = \{0\}$. If the dynamics begins within $\mathcal{O}_+$, it will be never be able to reach values in $\mathcal{O}_-$, as G -transformations cannot move the system outside its initial orbit. We therefore ask that each pair of points of X is connected through the G action. This amounts to require that X is **homogeneous** for G , that is, $\forall \mathbf{x}, \mathbf{y} \in X$ there exists a $g \in G$ such that $\rho_X(g)\mathbf{x} = \mathbf{y}$. We note that this condition solely ensures that the generation outcome is independent of the initial sampling condition, that is, that Langevin dynamics can generate any point of the target distribution from any point of the prior. Beyond this, the formalism remains fully applicable in the non-homogeneous case, where the dynamic is restricted to orbits of the group, effectively partitioning the distribution. Though the formalism still applies within each orbit, global generation across X would not be supported without homogeneity.

Condition 3: Commutativity of Π . The final requirement is that Π forms a (locally) **commuting** frame of vector fields, $[\mathcal{L}_A, \mathcal{L}_B]f(\mathbf{x}) = 0 \ \forall A, B$ and $\forall f \in C^\infty(X)$. In this case, the coordinates τ_i 's are orthogonal, and their flows commute, meaning the orbits parametrized by τ_i correspond to $\{\tau_j = 0\}_{j \neq i}$. For non-commuting flows this is not the case, as Figure 2b-c illustrates: (b) $V_1 = x_1 \partial_{x_1} + x_2 \partial_{x_2}$, $V_2 = x_1 \partial_{x_2} - x_2 \partial_{x_1}$ satisfy $[V_1, V_2] = 0$, and the orbits parametrized by $\tau_1 = r$ correspond to subspaces with constant $\tau_2 = \theta$; (c) $W_{1,2} = V_{1,2}/|\mathbf{x}|$ do not commute, and the loci $\theta = \text{const}$ no longer coincide with the r -orbits, causing θ to vary along these, despite the fact that r, θ are still orthogonal at each point. This last condition ensures that the updates governed by the different elements A_i of $\mathfrak{g}$ in (4) remain independent of one another. Notably, this does not exclude non-Abelian groups; even if $A_{1,2} \in \mathfrak{g}$ do not commute in the Lie algebra, their flows on X can, as shown in the $\mathfrak{g} = \mathfrak{so}(3)$ example in Appendix B.3.

3 Lie algebra score-based generative modeling via SDEs

In this section, we formalize the framework we developed above from the point of view of SDEs. Namely, we show that there exists a class of SDEs, which, when reversed, can generate data according to dynamics similar to (4), guided by the generalized score of the fundamental vector fields of the Lie algebra $\mathfrak{g}$. Our main result is the following.

Theorem 3.1. *Let G be a Lie group acting on X satisfying the conditions of Section 2.2, and let $\mathfrak{g}$ be its Lie algebra. The pair of SDEs*

$$d\mathbf{x} = \left[\beta(t)\Pi(\mathbf{x})\mathbf{f}(\mathbf{x}) + \frac{\gamma(t)^2}{2}\rho_X(\Omega) \right] dt + \gamma(t)\Pi(\mathbf{x})d\mathbf{W}, \quad (5)$$

$$d\mathbf{x} = \left[\beta(t)\Pi(\mathbf{x})\mathbf{f}(\mathbf{x}) - \frac{\gamma^2(t)}{2}\rho_X(\Omega) - \gamma^2(t)\Pi(\mathbf{x})\nabla^\top \cdot \Pi(\mathbf{x}) - \gamma(t)^2\Pi(\mathbf{x})\mathcal{L} \log p_t(\mathbf{x}) \right] dt + \gamma(t)\Pi(\mathbf{x})d\mathbf{W}, \quad (6)$$

where $\beta, \gamma : \mathbb{R} \rightarrow \mathbb{R}$ are time-dependent functions, $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ the fundamental vector fields, $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ the drift, $\Omega = \sum_i A_i^2$ is known as the quadratic Casimir element of $\mathfrak{g}$, and $\mathcal{L} = \Pi(\mathbf{x})^\top \nabla$, is such that

1. The forward-time SDE (5) is exactly solvable:

$$\mathbf{x}(t) = \left(\prod_i O_i(\tau_i(t)) \right) \mathbf{x}(0) = \left(\prod_{i=1}^n e^{\tau_i(t)A_i} \right) \mathbf{x}(0), \quad (7)$$

where $O_i = e^{\tau_i(t)A_i}$ is the finite group action and $\tau(t)$ is the solution to the SDE

$$d\tau(\mathbf{x}) = \beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}. \quad (8)$$

2. The SDE (6) is the reverse-time process of (5).

3. The Langevin dynamic of the above SDEs decomposes as a direct sum of $\mathfrak{g}$ infinitesimal actions (1), defining an infinitesimal transformation along the flows ξ_τ .

We refer to Appendix D for the full proof of the above result. Here we limit ourselves to a few comments regarding the extra terms that appear in the paired SDEs.

The appearance of the Casimir element (we assume the identity as bilinear form on $\mathfrak{g}$ (Kac & Kac, 1983)) compensates for the deviation of the tangent vector from the orbit due to the curvature of the flow coordinates. This can be seen in the example of $SO(2)$ acting on $\mathbb{R}^2$ (which will be discussed thoroughly below). An infinitesimal transformation along the θ direction, represented by Π_θ , moves any point $\mathbf{x}$ along a vector tangent to its $SO(2)$ orbit, a circle of radius $r = \sqrt{x_1^2 + x_2^2}$. Due to the orbit’s non-zero curvature, this movement would shift the point to an orbit of radius $r' > r$. The term $\rho_X(\Omega)$ compensates for this displacement, ensuring the final point remains close to the original orbit. This is illustrated in Figure 3. With this result at hand we can formulate our procedure for our Lie group-induced score-based generative modeling with SDEs.

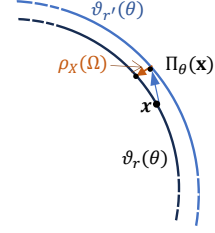


Figure 3: Quadratic Casimir for $SO(2)$.

Perturbing data through the SDE. The forward-time SDE (5) defines a noising diffusion process respecting the decomposition of the Lie algebra $\mathfrak{g}$ infinitesimal actions on X . In fact, given a data sample $\mathbf{x}(0) \sim p_0$, the solution (7) takes the form of a product of finite group element actions O_i on $\mathbf{x}(0)$, where the specific order is irrelevant since the Lie algebra generators commute. For each factor, we first determine $\boldsymbol{\tau}(0) = \boldsymbol{\tau}(\mathbf{x}(0))$, and employ these as initial conditions for the forward SDE (8). By choosing appropriately the drift terms $f_i^t s$, for instance, to be affine in the flow coordinates τ_i , we can solve for $\boldsymbol{\tau}(t)$ with standard techniques (Särkkä & Solin, 2019), as this will follow a Gaussian distribution. Alternatively, we can sample from $\boldsymbol{\tau}(t)$ by first simulating (8), then performing sliced score matching Song et al. (2020a); Pang et al. (2020) to sample from $p_t(\mathbf{x}(\boldsymbol{\tau}(t))|\mathbf{x}(0))$.

Generating samples through the reverse SDE. The time-reverse SDE (6) guides the generation of samples $\mathbf{x}(0) \sim p_0(\mathbf{x})$ starting from samples $\mathbf{x}(T) \sim p_T(\mathbf{x})$, provided we can estimate the generalized score $\mathcal{L} \log p_t(\mathbf{x})$ of each marginal distribution. To sample from p_T , we use the fact that the distribution in the flow coordinates $\boldsymbol{\tau}$ is tractable (with an appropriate choice of the drift terms and time-dependent functions β, γ in (5)), and that (since $p_t(\mathbf{x})d\mathbf{x} = p_t(\boldsymbol{\tau})d\boldsymbol{\tau}$)

$$p_t(\mathbf{x}) = p_t(\boldsymbol{\tau}) \left| \frac{\partial \boldsymbol{\tau}}{\partial \mathbf{x}} \right| = p_t(\boldsymbol{\tau}) |\boldsymbol{\Pi}^{-1}(\mathbf{x})|, \quad (9)$$

where the extra term corresponds to the determinant of the Jacobian of the coordinate transformation induced by the fundamental flow coordinates. In particular, when $\mathbf{f}(\boldsymbol{\tau})$ is affine in $\boldsymbol{\tau}$, it follows that $p_T(\boldsymbol{\tau}) = \mathcal{N}(\boldsymbol{\tau} | \mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Thus, we can sample $\boldsymbol{\tau}(T) \sim p_T(\boldsymbol{\tau})$ simply as a collection of independent Gaussian random variables, and use the flow map to obtain $\mathbf{x}(T) = \boldsymbol{\xi}_A(\boldsymbol{\tau}(T), \mathbf{x}_0)$, which will follow the distribution (9) for $t = T$. We describe training and sampling procedures in Algorithms 1 and 2 in Appendix E.

Estimating the generalized score. Analogously to standard score matching, we train a time-dependent neural network $\mathbf{s}_\theta(\mathbf{x}(t), t) : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ to estimate the generalized score $\mathcal{L} \log p_t(\mathbf{x}(t)|\mathbf{x}(0))$ at any time point, that is, we minimize the objective

$$\mathbb{E}_t \left\{ w(t) \mathbb{E}_{\mathbf{x}(0) \sim p_0(\mathbf{x})} \mathbb{E}_{\mathbf{x}(t) \sim p_t(\mathbf{x}|\mathbf{x}(0))} \left[|\mathbf{s}_\theta(\mathbf{x}(t), t) - \mathcal{L} \log p_t(\mathbf{x}(t)|\mathbf{x}(0))|^2 \right] \right\}, \quad (10)$$

where $w : [0, T] \rightarrow \mathbb{R}_+$ is a time-weighting function. Now, from Condition 3 above and the property that $\mathcal{L}_{A_i}$ computes the direction derivative along the flow of $\Pi_{A_i}(\mathbf{x})$, it follows that $\mathcal{L} \log p_t(\mathbf{x}(t)|\mathbf{x}(0)) = \nabla_{\boldsymbol{\tau}(t)} \log p_t(\mathbf{x}(\boldsymbol{\tau}(t))|\mathbf{x}(\boldsymbol{\tau}(0)))$. Under the above assumptions, $p_t(\boldsymbol{\tau}) = \mathcal{N}(\boldsymbol{\tau} | \boldsymbol{\mu}(\mathbf{x}(0), t), \boldsymbol{\Sigma}(t))$, where the form of the mean and the variance depends on the explicit form of (8). Using the parametrization $\boldsymbol{\tau}(t) = \boldsymbol{\mu}(\mathbf{x}(0), t) + \sqrt{\boldsymbol{\Sigma}(t)} \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we obtain

$$\mathcal{L} \log p_t(\mathbf{x}(t)|\mathbf{x}(0)) = -\boldsymbol{\Sigma}^{-1}(\boldsymbol{\tau}(t) - \boldsymbol{\mu}(\mathbf{x}(0), t)) = -\sqrt{\boldsymbol{\Sigma}(t)}^{-1} \boldsymbol{\eta}_t. \quad (11)$$

3.1 Examples

Standard Score Matching. Standard score matching can be recovered as a special case of our formalism by choosing $X = \mathbb{R}^n$ and $G = T(n)$. As we show explicitly in Appendix B.1, we have $\mathcal{L} = \nabla$ and the Lie algebra action $\boldsymbol{\Pi}(\mathbf{x}) = \mathbf{I}$, the identity on X . Since $\boldsymbol{\Pi}$ is $\mathbf{x}$ -independent, its divergence vanishes, as well as the quadratic Casimir ($T(N)$ is Abelian), so that the SDEs (5) take the known form

$$d\mathbf{x} = \beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}, \quad d\mathbf{x} = [\beta(t)\mathbf{f}(\mathbf{x}) - \gamma(t)^2 \nabla \log p_t(\mathbf{x})]dt + \gamma(t)d\mathbf{W}. \quad (12)$$

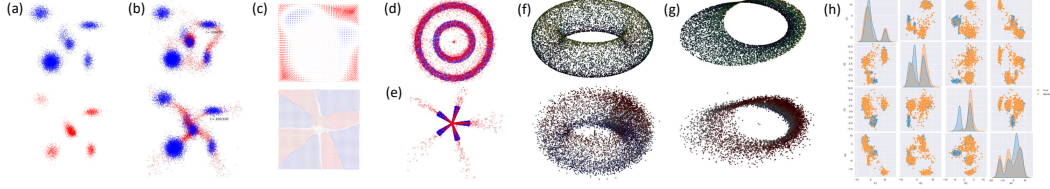


Figure 5: (a) 2d mixture of Gaussians (top: ground truth, bottom: generated); (b) generating process using single scores for the subgroups $\text{SO}(2), \mathbb{R}_+$ with the corresponding scores (c); (d,e) one-dimensional learning for a symmetric distributions; 3d-distributions: torus (f) and Möbius strip (g) (top: ground truth, bottom: generated); (h) 4d mixture of Gaussian for the group $G = \text{SO}(4) \times \mathbb{R}_+$.

$G = \text{SO}(2) \times \mathbb{R}_+$. A simple but non-trivial case is given by $G = \text{SO}(2) \times \mathbb{R}_+$ describing rotations and dilations acting on $X = \mathbb{R}^2$. A basis for $\mathfrak{g} = \mathfrak{so}(2) \oplus \mathbb{R}$ is given by $A_r = \mathbf{I}$ and $A_\theta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, yielding $\Pi(\mathbf{x}) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$, which satisfies all the conditions of section 2.2.

Following our discussion above and in Appendix B.1 we have (since $\rho(\Omega) = A_r^2 + A_\theta^2 = \mathbf{I} - \mathbf{I} = \mathbf{0}$)

$$d\mathbf{x} = \beta(t) (f_r(r) A_r \mathbf{x} + f_\theta(\theta) A_\theta \mathbf{x}) dt + \gamma(t) (dW_r A_r \mathbf{x} + dW_\theta A_\theta \mathbf{x}), \quad (13)$$

and we see that the SDE splits into contributions from the two Lie algebra summands. To find an explicit solution, let $\gamma(t) = \sqrt{\beta(t)}$ and $f_r = -\frac{1}{4} \log(x^2 + y^2)$, $f_\theta = -\frac{1}{2} \arctan \frac{y}{x}$. This choice corresponds, in the flow coordinates system, to a 2d Ornstein-Uhlenbeck system (Gardiner, 1985)

which has a Gaussian solution with mean $\begin{pmatrix} r(0) \\ \theta(0) \end{pmatrix} e^{-\int_0^t \beta(s) ds}$ and variance $(1 - e^{-\int_0^t \beta(s) ds}) \mathbf{I}$. Let

us define $\sigma(t) = \sqrt{1 - e^{-\int_0^t \beta(s) ds}}$, such that $r(t) = r(0) + \lambda(t) = r(0) - r(0)\sigma(t)^2 + \sigma(t)\eta_r$ and similarly $\theta(t) = \theta(0) + \varphi(t) = \theta(0) - \theta(0)\sigma(t)^2 + \sigma(t)\eta_\theta$, where $\eta_r, \eta_\theta \in \mathcal{N}(0, 1)$, then it is an easy calculation to show that

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = e^{\lambda(t)} \begin{pmatrix} \cos \varphi(t) & -\sin \varphi(t) \\ \sin \varphi(t) & \cos \varphi(t) \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}. \quad (14)$$

We can look at the asymptotic behavior of the solution. Assuming that $\beta(t)$ is a monotonous increasing function, that is, $\beta(t + \epsilon) > \beta(t)$ for $\epsilon > 0$, then $\lim_{t \rightarrow \infty} \sigma(t) = 1$ and hence

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = e^{-r_0 + \eta_r} \begin{pmatrix} \cos \theta_0 & \sin \theta_0 \\ -\sin \theta_0 & \cos \theta_0 \end{pmatrix} \begin{pmatrix} \cos \eta_\theta & \sin \eta_\theta \\ -\sin \eta_\theta & \cos \eta_\theta \end{pmatrix} \begin{pmatrix} e^{r_0} \cos \theta_0 \\ e^{r_0} \sin \theta_0 \end{pmatrix} = \begin{pmatrix} e^{\eta_r} \cos \eta_\theta \\ e^{\eta_r} \sin \eta_\theta \end{pmatrix},$$

where $\theta_0 = \theta(0)$, $r_0 = r(0)$. Note that, even if (15) is not Gaussian, we can still easily draw samples from it by sampling the two Gaussian variables η_r, η_θ .

Dihedral and bond angles. The above formalism can be applied to obtain transformations of physically meaningful quantities, as bond and torsion angles for molecules' conformations. Let γ_i be the dihedral angle between the planes identified by the points $\{\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}\}$ and $\{\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_{i+2}\}$, respectively (Figure 4a). The Lie algebra element corresponding to an infinitesimal change in γ_i is given by a $3N \times 3N$ -dimensional 3×3 -block diagonal matrix, whose $j = 1, \dots, N$ block is given by $H(j - (i + 1))\hat{\mathbf{x}}_{i+1,i} \cdot \mathbf{A}$, where $\mathbf{A} = (A_x, A_y, A_z)$ is the vector of the Lie algebra basis for $\mathfrak{so}(3)$, $\hat{\mathbf{x}}_{i+1,i} = (\mathbf{x}_{i+1} - \mathbf{x}_i)/|\mathbf{x}_{i+1} - \mathbf{x}_i|$ and $H(i) = 1$ if $i > 0$ and 0 otherwise is the Heaviside step function. For bond angles β_i (Figure 4c) we construct the corresponding $\mathfrak{so}(2) \in \mathfrak{so}(3)$ algebra element blocks as $H(j - i)(\mathbf{x}_{i+1,i} \times \mathbf{x}_{i-1,i}) \cdot \mathbf{A}$. Examples of the dynamics generated by these operators are presented in Figure 4(b,d).

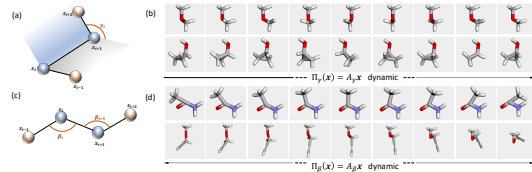


Figure 4: Lie algebra $\mathfrak{so}(2) \subset \mathfrak{so}(3)$ dynamics for torsion (a,b) and bond angles (c,d) in molecular conformers.

4 Related Work

Representation theory applied to neural networks has been studied both theoretically (Esteves, 2020; Chughtai et al., 2023; Puny et al., 2021; Smidt, 2021) and applied to a variety of groups, architectures

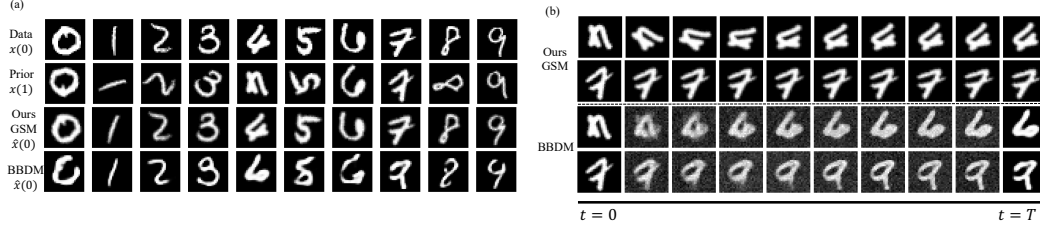


Figure 6: (a) Original and rotated MNIST samples with generated samples from our model and BBDM. (b) Reverse diffusion trajectories of our model against BBDM. Intermediate samples from BBDM resemble interpolation of mixed digits. For the first BBDM case, the 4-digit transitions into a 6-digit.

and data type: CNNs (Cohen & Welling, 2016; Romero et al., 2020; Liao & Liu, 2023; Finzi et al., 2020; Weiler & Cesa, 2019; Weiler et al., 2018), Graph Neural Networks (Satorras et al., 2021), Transformers, (Geiger & Smidt, 2022; Romero & Cordonnier, 2020; Hutchinson et al., 2021), point clouds (Thomas et al., 2018), chemistry (Schütt et al., 2021; Le et al., 2022a). On the topic of disentanglement of group action and symmetry learning, Pfau et al. (2020) factorize a Lie group from the orbits in data space, while Winter et al. (2022) learn through an autoencoder architecture invariant and equivariant representations of any group acting on the data. Fumero et al. (2021) learns disentangled representations solely from data pairs. Dehmamy et al. (2021) propose an architecture based on Lie algebras that can automatically discover symmetries from data. Xu et al. (2022) predict molecular conformations from molecular graphs in an roto-translation invariant fashion with equivariant Markov kernels.

Related to our study is the field of diffusion on Riemannian manifolds. De Bortoli et al. (2022) propose diffusion in a product space, a condition which is not a necessary in our framework, defined by the flow coordinates in the respective Riemannian sub-manifolds. When the Riemannian manifold is a Lie group, their method yields dynamics similar to ours, as illustrated in an example in Section 3.1. In fact, our formalism could be integrated with their approach to create a unified framework for diffusion processes on the broader class of Riemannian manifolds admitting a Lie group action. These techniques has been applied in a variety of use cases (Corso et al., 2023; Ketata et al., 2023; Yim et al., 2023; Jing et al., 2022) for protein docking, ligand and protein generation. The works Zhu et al. (2024); Kong & Tao (2024) leverage trivialized momentum to perform diffusion on the Lie algebra (isomorphic to $\mathbb{R}^n$) instead of the Lie group, thereby eliminating curvature terms, although their approach is to date only feasible for Abelian groups. An interesting connection with our work is the work of Kim et al. (2022): the authors propose a bijection to map a non-linear problem to a linear one, to approximate a bridge between two non-trivial distributions. Our case can be seen as a bijection between the (curved) Lie group manifold and the (flat) Euclidean data space.

5 Experiments

2d, 3d and 4d distributions. In Figure 5 we illustrate the framework for a variety of $d = 2, 3$ -dimensional distributions. In all cases we take $G = \text{SO}(d) \times \mathbb{R}_+$. Figure 5(a,b,c) displays a mixture of Gaussians: in (a) (bottom) we see that our generalized score-matching can learn any distribution, regardless of its inherent symmetry; (b) shows the output of the generation process using only one score (top $\mathbf{g} = \mathbf{so}(2)$, bottom $\mathbf{g} = \mathbf{r}_+$), while (c) shows the vector fields corresponding to the scores, where we color-coded the field directions. Figures 5(d,e) depicts radial and angular distributions, where the score is learned using the respective Lie algebra elements. This reflects the ability to leverage the symmetry properties of the data and perform diffusion in a lower-dimensional space. We also show in Figure 5h ($G = \text{SO}(4) \times \mathbb{R}_+$) that our method can be applied to higher dimensional Lie groups. We list quantitative comparisons in terms of W2-distances for our generalized score model against standard diffusion model in Appendix E.1.

Rotated MNIST. In this experiment we show that our framework can be applied to effectively learn a bridge between two non-trivial distributions, adopting however only techniques from score-matching and DDPM. Let $p_T(\mathbf{x})$ be the rotated MNIST dataset and $p_0(\mathbf{x})$ the original (non-rotated) MNIST dataset. We can learn to sample from p_0 starting from element of p_T by simply modeling a $\text{SO}(2)$ dynamic. Some examples of our results are shown in Figure 6. Notice that our formalism

allows us to reduce the learning to a 1-dimensional score $\mathcal{L}_\theta = x_1 \partial_{x_1} - x_2 \partial_{x_2}$, which reflects the true dimensionality of the problem. We trained the model with $T = 100$ time-steps, but for sampling it suffices to set $T = 10$. As it can be seen in the example trajectories 6b, the model starts converging already at $t/T \sim 0.5$. We employ a CNN which processes input images $\mathbf{x}(t)$, and the resulting feature map is flattened and concatenated with a scalar input t , then passed through fully connected layers to produce the final output.

We compare our approach to the Brownian Bridge Diffusion Model (BBDM) (Li et al., 2023). Unlike our method, BBDM operates unconstrained in the full MNIST pixel space ($\mathbb{R}^{28 \times 28}$), where intermediate states represent latent digits. As shown in Figure 6a, this can result in incorrect transitions, such as adding extraneous pixels or altering the original digit, even generating entirely different digits (Figure 6b).

We further evaluate both methods on the classification accuracy as well as FID scores of generated MNIST digits. Since the task is to correctly rotate a MNIST digit into the correct orientation aligning with the ground-truth data distribution, we observe that our GSM model achieves superior classification accuracies (0.96 vs 0.80) and FID scores (85.77 vs 133.4) as shown in Table 1.

Table 1: FID and Accuracy scores comparing GSM against BBDM.

Model	Avg Acc ($\uparrow$)	Avg FID ($\downarrow$)
GSM	0.96 $\pm$ 0.02	85.8 $\pm$ 15.7
BBDM	0.80 $\pm$ 0.10	133.4 $\pm$ 19.0

Further details can be found in Appendix E.2.

QM9. We use our framework to train a generative model $p_\theta(X|M)$ for conformer sampling of small molecules M from the QM9 dataset (Ramakrishnan et al., 2014). We only keep the lowest energy conformer as provided in the original dataset, that is, for each molecule only one 3D conformer is maintained. Here $X = \mathbb{R}^{3N}$ and we choose $G = (\text{SO}(3) \times \mathbb{R}_+)^N$, where each factor acts on the space $\mathbb{R}^3$ spanned by the Cartesian coordinates of the molecule’s atoms, respectively.

As Figure 7a shows, our generative process yields conformers that are energetically very similar to the ground truth conformers, while showing some variability, as it can be seen in the last example where the torsion angle is differently optimized. We train another model $p_\gamma(X|M)$ via standard Fisher denoising score-matching, i.e., choosing $G = T(3)^N$ as in Sec. 3.1, and generate 5 conformers per molecule for both models p_θ, p_γ . We then compute the UFF energy (Rappe et al., 1992) implemented in the RDKit for all generated conformers and extract the lowest energy geometry as generated sample. To compare against the reference geometry, we compute the energy difference $\Delta = U_{\text{true}} - U_{\text{gen}}$ for both models. Figure 7b shows that both diffusion models tend to generate conformers that have lower energies than the ground true conformer according to the UFF parametrization, while the diffusion model that implements the dynamics according to $G = (\text{SO}(3) \times \mathbb{R}_+)^N$ (colored in blue) achieves slightly lower energy conformers, mean $\Delta_\theta = -0.2159$ against mean $\Delta_\gamma = -0.2144$ for the standard diffusion model (colored in orange).

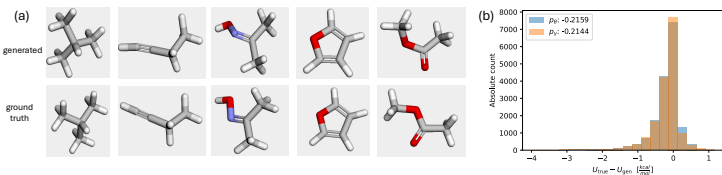


Figure 7: (a) Generated 3D conformer from the QM9 validation set (top) and ground truth conformer (bottom). (b) Energy difference distribution between diffusion models (p_θ, p_γ) and ground-truth energy. Both models generate a similar Δ energy distribution.

CrossDocked2020: Global E(3) and Protein-Ligand Complexes. In this final experiment, we train a generative model for global SE(3) transformations acting on small molecules. Specifically, given a pair consisting of a compound and a protein pocket, our goal is to generate the trajectory by which the ligand best fits into the pocket. Importantly, the internal structure of the compound remains fixed, which presents a challenge with standard diffusion processes. Thus, while the SE(3) transformations are global with respect to the ligand, they do not represent global symmetries of the overall system. We derive in appendix B.4 the relevant operators that guide the dynamics (6). Figure 8a shows examples of docked molecules using SE(3)-guided score-matching diffusion. The true and generated molecules at different generation steps are visualized as point clouds, showing a

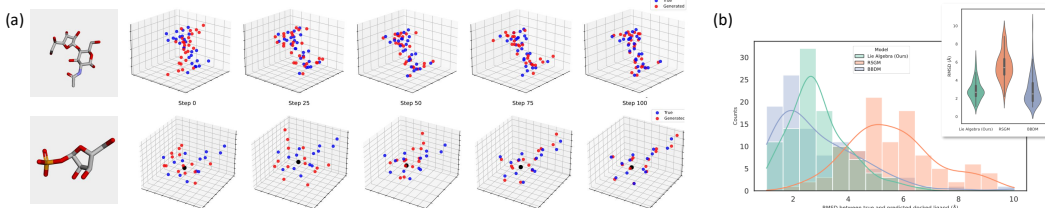


Figure 8: (a) SE(3) trajectories for molecular docking; (b) Comparison with RSGM.

good agreement. Figure 8b shows that our model achieves a lower RMSD (2.9 ± 1.0 Å vs 5.6 ± 1.2 Å) for the docked ligands than the RSGM method (De Bortoli et al., 2022; Corso et al., 2023) (for details, we refer to Appendix E.3.1). We also compare our method against the Brownian Bridge Diffusion Model (BBDM) which operates on the $T(3)^N$ group, as a standard (Euclidean) diffusion baseline with the constraint to start and end at valid rigidly transformed molecules during training. We use the same network architecture as in the GSM and RSGM experiments to learn the correct SO(3) rotation. Unlike existing experiments (our method and RSGM), the Euclidean BBDM in this setting attempts to learn only global SO(3) rotation, neglecting translation. Since the problem is implicitly 3-dimensional but the equivariant score network predicts all $3N$ ligand atom coordinates, final samples with implausible coordinate trajectories tend to have higher energies due to unphysical poses including bond stretching, non-planar aromatic rings, and deformed rings. In terms of mean/std RMSD on the CrossDocked2020 test set, our method (Lie algebra: 2.91 ± 1.0 Å) is comparable with BBDM (2.92 ± 1.57 Å). However, since BBDM models all atomic coordinates, the overall dynamics do not follow a global SO(3) rotation, achieving $\text{MAE}(D(x_0, x_0), D(\hat{x}_0, \hat{x}_0)) = 0.43 \pm 0.21$, while RSGM and our method achieve 0.0 by design. This indicates that Lie algebra induced diffusion offers a clear advantage over standard diffusion models in this particular bridging problem.

6 Conclusions and Outlook

We presented a method for generative modeling on *any* Lie group G representation on a space X through generalized score matching. **Our framework generates a curved Lie group diffusion dynamics in flat Euclidean space**, without the need to transform the data and of performing group projections. Specifically, we introduced a new class of exactly-solvable SDEs that guide the corruption and generation processes. Thus, our framework does not merely complement existing methods, but *expands* the space of exactly solvable diffusion processes. Our framework is particularly relevant given recent findings (Abramson et al., 2024) showing that unconstrained models outperform equivariant ones: with our framework there is no need of a tradeoff, as we retain the expressivity of unconstrained models on raw Cartesian coordinates with the benefits of group inductive bias. Moreover, our techniques descend quite straightforwardly to flow matching (Lipman et al., 2022). We spell out the connection in appendix F and we plan to expand on this in future work.

Code Availability

Our source code will be made available on <https://github.com/pfizer-opensource/symmetry-induced-score-matching>.

References

- Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Babrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.
- Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.
- LE Blumenson. A derivation of n-dimensional spherical coordinates. *The American Mathematical Monthly*, 67(1):63–66, 1960.
- Johann Brehmer and Kyle Cranmer. Flows for simultaneous manifold learning and density estimation. *Advances in neural information processing systems*, 33:442–453, 2020.
- Bilal Chughtai, Lawrence Chan, and Neel Nanda. A toy model of universality: Reverse engineering how networks learn group operations. In *International Conference on Machine Learning*, pp. 6243–6267. PMLR, 2023.
- Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.
- Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi S. Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kKF8_K-mBbS.
- Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.
- Nima Dehmamy, Robin Walters, Yanchen Liu, Dashun Wang, and Rose Yu. Automatic symmetry discovery with lie algebra convolutional network. *Advances in Neural Information Processing Systems*, 34:2503–2515, 2021.
- Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL <https://openreview.net/forum?id=AAWuCvzaVt>.
- Carlos Esteves. Theoretical aspects of group equivariant neural networks. *arXiv preprint arXiv:2004.05154*, 2020.
- Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *International Conference on Machine Learning*, pp. 3165–3176. PMLR, 2020.
- Marco Fumero, Luca Cosmo, Simone Melzi, and Emanuele Rodolà. Learning disentangled representations via product manifold projection. In *International conference on machine learning*, pp. 3530–3540. PMLR, 2021.
- Crispin W Gardiner. Handbook of stochastic methods for physics, chemistry and the natural sciences. *Springer series in synergetics*, 1985.
- Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.
- René Haas, Inbar Huberman-Spiegelglas, Rotem Mulayoff, Stella Graßhof, Sami S Brandt, and Tomer Michaeli. Discovering interpretable directions in the semantic latent space of diffusion models. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–9. IEEE, 2024.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

- Chin-Wei Huang, Jae Hyun Lim, and Aaron C Courville. A variational perspective on diffusion-based generative models and score matching. *Advances in Neural Information Processing Systems*, 34: 22863–22876, 2021.
- Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron C Courville. Riemannian diffusion models. *Advances in Neural Information Processing Systems*, 35:2750–2761, 2022.
- Michael J Hutchinson, Charline Le Lan, Sheheryar Zaidi, Emilien Dupont, Yee Whye Teh, and Hyunjik Kim. Lietransformer: Equivariant self-attention for lie groups. In *International Conference on Machine Learning*, pp. 4533–4543. PMLR, 2021.
- Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.
- Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35:24240–24253, 2022.
- Victor G Kac and Victor G Kac. The invariant bilinear form and the generalized casimir operator. *Infinite Dimensional Lie Algebras: An Introduction*, pp. 14–24, 1983.
- Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554, 2018.
- Mohamed Amine Ketata, Cedrik Laue, Ruslan Mammadov, Hannes Stärk, Menghua Wu, Gabriele Corso, Céline Marquet, Regina Barzilay, and Tommi S Jaakkola. Diffdock-pp: Rigid protein-protein docking with diffusion models. *arXiv preprint arXiv:2304.03889*, 2023.
- Dongjun Kim, Byeonghu Na, Se Jung Kwon, Dongsoo Lee, Wanmo Kang, and Il-Chul Moon. Maximum likelihood training of implicit nonlinear diffusion model. *Advances in neural information processing systems*, 35:32270–32284, 2022.
- Anna Klimovskaia, David Lopez-Paz, Léon Bottou, and Maximilian Nickel. Poincaré maps for analyzing complex hierarchies in single-cell data. *Nature communications*, 11(1):2966, 2020.
- Lingkai Kong and Molei Tao. Convergence of kinetic langevin monte carlo on lie groups. *arXiv preprint arXiv:2403.12012*, 2024.
- Tuan Le, Marco Bertolini, Frank Noé, and Djork-Arné Clevert. Parameterized hypercomplex graph neural networks for graph classification. In *International Conference on Artificial Neural Networks*, pp. 204–216. Springer, 2021.
- Tuan Le, Frank Noé, and Djork-Arné Clevert. Equivariant graph attention networks for molecular property prediction. *arXiv preprint arXiv:2202.09891*, 2022a.
- Tuan Le, Frank Noe, and Djork-Arné Clevert. Representation learning on biomolecular structures using equivariant graph attention. In *The First Learning on Graphs Conference*, 2022b. URL <https://openreview.net/forum?id=kv4xUo5Pu6>.
- Adam Leach, Sebastian M Schmon, Matteo T. Degiacomi, and Chris G. Willcocks. Denoising diffusion probabilistic models on $SO(3)$ for rotational alignment. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022. URL <https://openreview.net/forum?id=BY88eBbkpe5>.
- Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdlm: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1952–1961, June 2023.
- Dengfeng Liao and Guangzhong Liu. Lie group equivariant convolutional neural network based on laplace distribution. *Remote Sensing*, 15(15):3758, 2023.
- Lina Lin, Mathias Drton, and Ali Shojaie. Estimation of high-dimensional graphical models using regularized score matching. *Electronic journal of statistics*, 10(1):806, 2016.

- Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.
- Siwei Lyu. Interpretation and generalization of score matching. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 359–366, 2009.
- Tianyu Pang, Kun Xu, Chongxuan Li, Yang Song, Stefano Ermon, and Jun Zhu. Efficient learning of generative models via finite-difference score matching. *Advances in Neural Information Processing Systems*, 33:19175–19188, 2020.
- Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, and Youngjung Uh. Understanding the latent space of diffusion models through the lens of riemannian geometry. *Advances in Neural Information Processing Systems*, 36:24129–24142, 2023.
- David Pfau, Irina Higgins, Alex Botev, and Sébastien Racanière. Disentangling by subspace diffusion. *Advances in Neural Information Processing Systems*, 33:17403–17415, 2020.
- Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. *arXiv preprint arXiv:2110.03336*, 2021.
- Joshua A Rackers, Lucas Tecot, Mario Geiger, and Tess E Smidt. A recipe for cracking the quantum scaling limit with machine learned electron densities. *Machine Learning: Science and Technology*, 4(1):015027, feb 2023. doi: 10.1088/2632-2153/acb314. URL <https://dx.doi.org/10.1088/2632-2153/acb314>.
- Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, Aug 2014. ISSN 2052-4463. doi: 10.1038/sdata.2014.22. URL <https://doi.org/10.1038/sdata.2014.22>.
- A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. III Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992. doi: 10.1021/ja00051a040. URL <https://doi.org/10.1021/ja00051a040>.
- David Romero, Erik Bekkers, Jakub Tomczak, and Mark Hoogendoorn. Attentive group equivariant convolutional networks. In *International Conference on Machine Learning*, pp. 8188–8199. PMLR, 2020.
- David W Romero and Jean-Baptiste Cordonnier. Group equivariant stand-alone self-attention for vision. *arXiv preprint arXiv:2010.00977*, 2020.
- Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019.
- Victor Garcia Satorras, Emiel Hoogetboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.
- Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.
- Tess E Smidt. Euclidean symmetry and equivariance in machine learning. *Trends in Chemistry*, 3(2): 82–85, 2021.
- Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.
- Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In *Uncertainty in Artificial Intelligence*, pp. 574–584. PMLR, 2020a.

- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020b.
- Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in neural information processing systems*, 34:1415–1428, 2021.
- Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.
- Yingheng Wang, Yair Schiff, Aaron Gokaslan, Weishen Pan, Fei Wang, Christopher De Sa, and Volodymyr Kuleshov. Infodiffusion: Representation learning using information maximizing diffusion models. In *International Conference on Machine Learning*, pp. 36336–36354. PMLR, 2023.
- Maurice Weiler and Gabriele Cesa. General $e(2)$ -equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019.
- Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018.
- Robin Winter, Marco Bertolini, Tuan Le, Frank Noé, and Djork-Arné Clevert. Unsupervised learning of group invariant and equivariant representations. *Advances in Neural Information Processing Systems*, 35:31942–31956, 2022.
- Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.
- Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40001–40039. PMLR, 23–29 Jul 2023. URL <https://proceedings.mlr.press/v202/yim23a.html>.
- Yangtian Zhang, Zuobai Zhang, Bozita Zhong, Sanchit Misra, and Jian Tang. Diffpack: A torsional diffusion model for autoregressive protein side-chain packing. *Advances in Neural Information Processing Systems*, 36, 2024.
- Yuchen Zhu, Tianrong Chen, Ling kai Kong, Evangelos A Theodorou, and Molei Tao. Trivialized momentum facilitates diffusion generative modeling on lie groups. *arXiv preprint arXiv:2405.16381*, 2024.

A Appendix A: Summary of Notation and Intuition

Symbol	Name	Definition	Intuition
G	Lie group		A continuous symmetry group, e.g., rotations ($SO(3)$), translations, scalings. Encodes the structure of transformations acting on the data.
e	Identity element of G	$e \cdot g = g \forall g \in G$	The identity transformation leaving everything unchanged.
$\mathfrak{g}$	Lie algebra of G	$T_e G$	Tangent space at the identity; represents infinitesimal group transformations.
X	Data manifold		The space where the data lives, often $\mathbb{R}^n$, but can be more general or even discrete (e.g., graph for molecules, grid for images, etc.).
ρ_X	Group action	$\rho_X : G \times X \rightarrow X$	Specifies how each abstract group element $g \in G$ transforms data points in X via matrices.
$G \cdot \mathbf{x}$	Orbit of $\mathbf{x}$ under G	$\{\rho_X(g)(\mathbf{x}), g \in G\}$	The set of all points reachable from x via group actions. Captures the "symmetry class" of x .
$G_{\mathbf{x}}$	Stabilizer subgroup at $\mathbf{x}$	$\{g \in G \rho_X(g)(\mathbf{x}) = \mathbf{x}\}$	Subgroup of G that leaves $\mathbf{x}$ unchanged. Describes residual symmetries at that point.
$d\rho_X$	Infinitesimal action	$d\rho_X : \mathfrak{g} \rightarrow \text{Vect}(X)$	Maps infinitesimal transformations to vector fields on X ; captures how a tiny "step" in G moves a point in X .
$\exp$	Exponential map	$\exp(A) = \gamma_A(1)$, where $\gamma_A : \mathbb{R} \rightarrow G$	Geodesic path on G determined by the direction given by the vector $A \in \mathfrak{g}$.
ξ_A	Flow on X induced by ρ_X	$\rho_X(\exp(\tau A))(\mathbf{x})$	Path on X corresponding to a geodesic path on G determined by A .
$\Pi_A(\mathbf{x})$	Fundamental vector field from $A \in \mathfrak{g}$	$\left. \frac{d}{d\tau} \right _{\tau=0} \rho_X(\exp(\tau A))(\mathbf{x})$	A vector field on X generated by a direction A in the Lie algebra; describes how x moves under an infinitesimal group transformation.

Table 2: Summary of Lie group/Lie algebra related quantities with their notation, definition and intuitive meaning.

B Examples of Lie groups and Lie algebra actions

In this appendix we list some important Lie groups and Lie algebra actions, their corresponding fundamental vector fields as well as the fundamental flow coordinates. These will be useful in the main text.

B.1 $T(N)$

Let $X = \mathbb{R}^N$ and $G = T(N)$, the group of translations in N -dimensional space. Element of $T(N)$ are represented by a vector $\mathbf{v} = (v_1, v_2, \dots, v_N)^\top \in \mathbb{R}^N$, where v_i are the translation components along the x_i axes for $i = 1, \dots, N$, thus $T(N) \simeq \mathbb{R}^N$. Explicitly, for a $\mathbf{x} \in X$ its action is given by $\rho_{\mathbb{R}^N}(\mathbf{v}, \mathbf{x}) = \mathbf{x} + \mathbf{v}$.

The corresponding Lie algebra $\mathfrak{t}(N)$ is also isomorphic to $\mathbb{R}^N$, and it consists of vectors $\mathbf{a} = (a_1, a_2, \dots, a_N)^\top \in \mathbb{R}^N$. The Lie bracket of any two elements in $\mathfrak{t}(N)$ vanishes, as $T(N)$ is Abelian.

To derive the infinitesimal action, we first note that the exponential map is trivial, $\exp(\tau \mathbf{A}) = \tau \mathbf{A}$. Hence, we have

$$\Pi_A(\mathbf{x}) = \left. \frac{d}{d\tau} \right|_{\tau=0} \rho_{\mathbb{R}^N}(\tau \mathbf{A}, \mathbf{x}) = \left. \frac{d}{d\tau} \right|_{\tau=0} (\mathbf{x} + \tau \mathbf{A}) = \mathbf{A} . \quad (15)$$

Thus, the fundamental vector field Π_A corresponding to $\mathbf{A} \in \mathfrak{t}(N)$ is the constant vector field:

$$\Pi_A = a_1 \frac{\partial}{\partial x_1} + a_2 \frac{\partial}{\partial x_2} + \dots + a_N \frac{\partial}{\partial x_N} = \mathbf{A} \cdot \nabla .$$

B.2 $X = \mathbb{R}^N, G = \mathbb{R}_+^*$ (group of dilations)

Let us consider $X = \mathbb{R}^N$ and $G = \mathbb{R}_+^*$, the group of dilations in N -dimensional space. The group $\mathbb{R}_+^*$ consists of all positive scaling factors. Each element of $G = \mathbb{R}_+^*$ can be represented by a scalar $\lambda > 0$ that scales all vectors in $\mathbb{R}^N$ by this factor.

The action of $G = \mathbb{R}_+^*$ on $\mathbb{R}^N$ is a dilation, meaning that every vector $\mathbf{x} = (x_1, x_2, \dots, x_N)^\top \in \mathbb{R}^N$ is scaled by the factor λ . Explicitly, the group action is given by

$$\rho_{\mathbb{R}^N}(\lambda, \mathbf{x}) = \lambda \mathbf{x} . \quad (16)$$

The Lie algebra $\mathfrak{g} = \mathbb{R}$ corresponding to the dilation group $G = \mathbb{R}_+^*$ consists of real numbers representing the logarithm of the scaling factor. Specifically, an element $A \in \mathfrak{g}$ corresponds to a generator of the dilation, and the exponential map $\exp : \mathfrak{g} \rightarrow G$ is given by: $\exp(\tau A) = e^{\tau A}$, where τ is a real parameter.

The infinitesimal action corresponds to taking the derivative at $\tau = 0$. For a vector $\mathbf{x} \in \mathbb{R}^N$ and $A \in \mathfrak{g}$, the fundamental vector field Π_A is computed as:

$$d\rho_{\mathbb{R}^N}(A) = \left. \frac{d}{d\tau} \right|_{\tau=0} \rho_{\mathbb{R}^N}(e^{\tau A}, \mathbf{x}) = \left. \frac{d}{d\tau} \right|_{\tau=0} (e^{\tau A} \mathbf{x}) = A \mathbf{x} , \quad (17)$$

and

$$\mathcal{L}_A(\mathbf{x}) = A \mathbf{x} \cdot \nabla .$$

Now, solving the equation

$$\mathbf{x} = e^{\tau A} \mathbf{x}_0 \quad (18)$$

in terms of τ we obtain

$$\tau = \frac{1}{A} \log \frac{|\mathbf{x}|^2}{\mathbf{x} \cdot \mathbf{x}_0} = \frac{1}{A} \log \frac{|\mathbf{x}|^2}{|\mathbf{x}| |\mathbf{x}_0|} = \frac{1}{A} \log \frac{|\mathbf{x}|}{|\mathbf{x}_0|} = \frac{1}{2A} \log \frac{|\mathbf{x}|^2}{|\mathbf{x}_0|^2} . \quad (19)$$

In the usual case of $A = 1$ (generator of the Lie algebra), $\mathbf{x}_0 = \frac{1}{\sqrt{N}}(1, 1, \dots, 1)^\top$ to be the unit vector we obtain the usual expression as flow coordinate

$$\tau = \frac{1}{2} \log(x_1^2 + x_2^2 + \dots + x_N^2) . \quad (20)$$

B.3 $X = \mathbb{R}^3, G = \mathbf{SO}(3) \times \mathbb{R}_+^*$

The dilation part is solved in the previous section, so we actually just focus on the action of $\mathbf{SO}(3)$ on $\mathbb{R}^3$. The orbits are given by spheres centered at the origin, and we can decompose the action of $\mathbf{SO}(3)$

by varying the azimuthal or the polar angle defined by a vector $\mathbf{x}$. Namely, we have the two actions

$$\begin{aligned}\rho_{\mathbb{R}^3}(\varphi, \mathbf{x}) &= \begin{pmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \\ \rho_{\mathbb{R}^3}(\theta, \mathbf{x}) &= \left[I + \sin \theta \begin{pmatrix} 0 & 0 & \cos \varphi \\ 0 & 0 & \sin \varphi \\ -\cos \varphi & -\sin \varphi & 0 \end{pmatrix} \right. \\ &\quad \left. + (1 - \cos \theta) \begin{pmatrix} -\cos^2 \varphi & -\cos \varphi \sin \varphi & 0 \\ -\cos \varphi \sin \varphi & -\sin^2 \varphi & 0 \\ 0 & 0 & -1 \end{pmatrix} \right] \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (21)\end{aligned}$$

If we take the differentials

$$\begin{aligned}d\rho_{\mathbb{R}^3}(\varphi, \mathbf{x})|_{\varphi=0} &= \begin{pmatrix} -\sin \varphi & -\cos \varphi & 0 \\ \cos \varphi & -\sin \varphi & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \Big|_{\varphi=0} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A_z \mathbf{x}, \\ d\rho_{\mathbb{R}^3}(\theta, \mathbf{x}) &= \left[\cos \theta \begin{pmatrix} 0 & 0 & \cos \varphi \\ 0 & 0 & \sin \varphi \\ -\cos \varphi & -\sin \varphi & 0 \end{pmatrix} \right. \\ &\quad \left. - \sin \theta \begin{pmatrix} \cos^2 \varphi & \cos \varphi \sin \varphi & 0 \\ \cos \varphi \sin \varphi & \sin^2 \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \right]_{\theta=0} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & \cos \varphi \\ 0 & 0 & \sin \varphi \\ -\cos \varphi & -\sin \varphi & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (\cos \varphi A_y - \sin \varphi A_x) \mathbf{x}, \quad (22)\end{aligned}$$

where

$$A_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad A_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad A_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (23)$$

form a basis for $\mathfrak{so}(3)$. The corresponding differential operators are

$$\mathcal{L}_\varphi = x\partial_y - y\partial_x, \quad \mathcal{L}_\theta = \frac{1}{\sqrt{x^2 + y^2}} [zx\partial_x + zy\partial_y - (x^2 + y^2)\partial_z], \quad (24)$$

and it is an easy calculation to show that they commute $[\mathcal{L}_\varphi, \mathcal{L}_\theta] = 0$. The attentive reader might have noticed that the commutation does not hold at the matrices level. While this is expected, since there is no 2-dimensional commuting subalgebra in $\mathfrak{so}(3)$, it is nonetheless quite puzzling since everything works out at the level of differential operators. This reflect the fact that the commutation properties are necessary at the level of the action of $\mathfrak{g}$ on X , and not necessarily at the Lie algebra level. In this case, however, we can elegantly resolve the puzzle, we found a matrix representation for the action $d\rho_{\mathbb{R}^3}(\theta)\mathbf{x}$ which does commute with the φ action. To do this we note that we can rewrite

$$\mathcal{L}_\theta = \frac{\cos \theta}{\sin \theta} x\partial_x + \frac{\cos \theta}{\sin \theta} y\partial_y - \frac{\sin \theta}{\cos \theta} z\partial_z, \quad (25)$$

which corresponds to simultaneous dilations, with different coefficient, in the z axis and x, y -plane. The finite action takes the form

$$\tilde{\rho}_{\mathbb{R}^3}(\theta, \mathbf{x}) = \exp \left[\log \sin \theta \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \log \cos \theta \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right], \quad (26)$$

and computing the first order term we obtain

$$d\tilde{\rho}_{\mathbb{R}^3}(\theta, \mathbf{x}) = \begin{pmatrix} \frac{\cos \theta}{\sin \theta} & 0 & 0 \\ 0 & \frac{\cos \theta}{\sin \theta} & 0 \\ 0 & 0 & -\frac{\sin \theta}{\cos \theta} \end{pmatrix} \mathbf{x}. \quad (27)$$

This matrix is diagonal and it trivially commutes with A_z . The price we had to pay to realize a system of commuting matrices is that in $\tilde{\rho}$ the flow parameter θ appear non-linearly, thus we traded-off commutativity at the level of the Lie algebra matrices for the linearity of the flow parameters at the group level. We remark that both give rise to the same differential operator on X , which is the relevant object for our purposes.

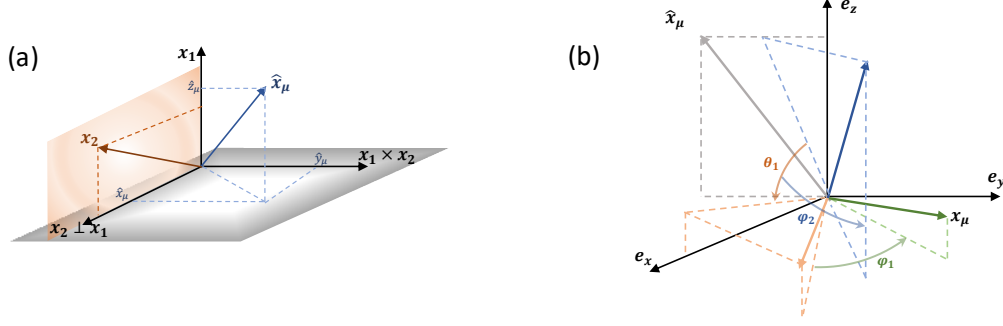


Figure 9: (a) The coordinates $\hat{\mathbf{x}}_\mu$ are the coordinates in the coordinate system defined by $\mathbf{x}_1$, the orthogonal projection of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$. $\mathbf{x}_2 \perp \mathbf{x}_1 = \mathbf{x}_2 - \mathbf{x}_1 \cdot \mathbf{x}_2$, and $\mathbf{x}_2 \times \mathbf{x}_1$. (b) Graphical depiction of the global symmetry transformations parametrized by the three angles $\varphi_2, \theta_1, \varphi_1$.

B.4 $X = \mathbb{R}^{3N}$ and global $\text{SO}(3)$

Let $X = \mathbb{R}^{3N}$ be parametrized by $\mathbf{x}_{i=1,\dots,N}$. We can describe a global $\text{SO}(3)$ action as follows

$$\begin{aligned} \mathbf{x}_1 &= R_{\mathbf{e}_z}(\varphi_1) R_{\mathbf{e}_y}(\theta_1) \begin{pmatrix} 0 \\ 0 \\ \hat{z}_1 \end{pmatrix}, \\ \mathbf{x}_2 &= R_{\mathbf{e}_z}(\varphi_1) R_{\mathbf{e}_y}(\theta_1) R_{\mathbf{e}_z}(\varphi_2) \begin{pmatrix} \hat{x}_2 \\ 0 \\ \hat{z}_2 \end{pmatrix}, \\ \mathbf{x}_{\mu=3,\dots,N} &= R_{\mathbf{e}_z}(\varphi_1) R_{\mathbf{e}_y}(\theta_1) R_{\mathbf{e}_z}(\varphi_2) \hat{\mathbf{x}}_\mu, \end{aligned} \quad (28)$$

where $R_{\mathbf{a}}(\omega)$ represents a rotation of an angle ω around the axis $\mathbf{a}$. We can then derive the operator $\Pi \in \mathbb{R}^{3N \times 3N}$ as follows. Let $R'(\omega)$ be the matrix where we take the partial derivative with respect to ω of all elements of R . Then

$$\begin{aligned} \Pi_{\varphi_1} &= ([A_z \mathbf{x}_1]^\top \quad [A_z \mathbf{x}_2]^\top \quad \dots \quad [A_z \mathbf{x}_N]^\top)^\top \\ \Pi_{\theta_1} &= \begin{pmatrix} (\cos \varphi_1 A_y - \sin \varphi_1 A_x) \mathbf{x}_1 \\ (\cos \varphi_1 A_y - \sin \varphi_1 A_x) \mathbf{x}_2 \\ \vdots \\ (\cos \varphi_1 A_y - \sin \varphi_1 A_x) \mathbf{x}_N \end{pmatrix} \\ \Pi_{\varphi_2} &= \begin{pmatrix} \mathbf{0} \\ (\sin \theta_1 \cos \varphi_1 A_x + \sin \theta_1 \sin \varphi_1 A_y + \cos \theta_1 A_z) \mathbf{x}_2 \\ \vdots \\ (\sin \theta_1 \cos \varphi_1 A_x + \sin \theta_1 \sin \varphi_1 A_y + \cos \theta_1 A_z) \mathbf{x}_N \end{pmatrix} \end{aligned} \quad (29)$$

Notice that these do represent global rotations since it is easy to see that $(\sin \theta_1 \cos \varphi_1 A_x + \sin \theta_1 \sin \varphi_1 A_y + \cos \theta_1 A_z) \mathbf{x}_1 = \mathbf{0}$. Formally, the true Lie algebra elements are 3×3 matrices of the form

$$A_\varphi = \begin{pmatrix} A_z & 0 & 0 & \dots & 0 \\ 0 & A_z & 0 & \dots & 0 \\ 0 & 0 & A_z & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_z \end{pmatrix} \quad (30)$$

and similarly for the other operators. Now, for the inverse relations we have

$$\begin{aligned}\theta_1 &= \arccos \frac{z_1}{(x_1^2 + y_1^2 + z_1^2)^{1/2}}, \\ \varphi_1 &= \text{sgn}(y_1) \arccos \frac{x_1}{(x_1^2 + y_1^2)^{1/2}}, \\ \varphi_2 &= \arctan \frac{\tilde{y}_2}{\tilde{x}_2},\end{aligned}\tag{31}$$

where $\tilde{\mathbf{x}}_2 = R_{\mathbf{e}_y}(\theta_1)^{-1} R_{\mathbf{e}_z}(\varphi_1)^{-1} \mathbf{x}_2 = R_{\mathbf{e}_y}(-\theta_1) R_{\mathbf{e}_z}(-\varphi_1) \mathbf{x}_2$.

B.5 $X = \mathbb{R}^4, G = \mathbf{SO}(4) \times \mathbb{R}_+$

Now we look at the case of a higher dimensional Lie group, namely $G = \mathbf{SO}(4) \times \mathbb{R}_+$. The parametrization is given by

$$\begin{aligned}x_1 &= e^r \cos \varphi_1, \\ x_2 &= e^r \sin \varphi_1 \cos \varphi_2, \\ x_3 &= e^r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3, \\ x_4 &= e^r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3.\end{aligned}\tag{32}$$

The Lie algebra elements corresponding to the $\mathbf{SO}(4)$ flow coordinates are

$$\begin{aligned}A_{\varphi_1} &= \begin{pmatrix} 0 & -\cos \varphi_2 & -\sin \varphi_2 \cos \varphi_3 & -\sin \varphi_2 \sin \varphi_3 \\ \cos \varphi_2 & 0 & 0 & 0 \\ \sin \varphi_2 \cos \varphi_3 & 0 & 0 & 0 \\ \sin \varphi_2 \sin \varphi_3 & 0 & 0 & 0 \end{pmatrix}, \\ A_{\varphi_2} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\cos \varphi_3 & -\sin \varphi_3 \\ 0 & \cos \varphi_3 & 0 & 0 \\ 0 & \sin \varphi_3 & 0 & 0 \end{pmatrix}, \\ A_{\varphi_3} &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.\end{aligned}\tag{33}$$

Next, we compute the three non-trivial commutators (note that an operators with itself always commute). First, we list the differential operators

$$\begin{aligned}\mathcal{L}_{\varphi_1} &= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_2 \partial_2 + x_1 x_3 \partial_3 + x_1 x_4 \partial_4 - (x_2^2 + x_3^2 + x_4^2) \partial_1], \\ \mathcal{L}_{\varphi_2} &= \frac{1}{\sqrt{x_3^2 + x_4^2}} [x_2 x_3 \partial_3 + x_2 x_4 \partial_4 - (x_3^2 + x_4^2) \partial_2], \\ \mathcal{L}_{\varphi_3} &= x_3 \partial_4 - x_4 \partial_3.\end{aligned}\tag{34}$$

where we used the notation $\partial_i = \partial_{x_i}$. These follow directly from (33) together with $L_{\varphi_i} = A_{\varphi_i} \mathbf{x} \cdot \nabla$, and using the relations

$$\sin \varphi_3 = \frac{x_4}{\sqrt{x_3^2 + x_4^2}}, \quad \cos \varphi_3 = \frac{x_3}{\sqrt{x_3^2 + x_4^2}}, \quad \cos \varphi_2 = \frac{x_2}{\sqrt{x_2^2 + x_3^2 + x_4^2}}, \quad \sin \varphi_2 = \frac{\sqrt{x_3^2 + x_4^2}}{\sqrt{x_2^2 + x_3^2 + x_4^2}}.\tag{35}$$

Specifically, we have

$$\begin{aligned}[\mathcal{L}_{\varphi_2}, \mathcal{L}_{\varphi_3}] &= \frac{1}{\sqrt{x_3^2 + x_4^2}} [-x_2 x_4 \partial_3 + x_2 x_3 \partial_4] - \frac{-x_3 x_4 + x_4 x_3}{(x_3^2 + x_4^2)^{1/2}} [x_2 x_3 \partial_3 + x_2 x_4 \partial_4 - (x_3^2 + x_4^2) \partial_2] \\ &\quad - \frac{1}{\sqrt{x_3^2 + x_4^2}} [x_2 x_3 \partial_3 - 2x_3 x_4 \partial_2 - x_2 x_4 \partial_4 + 2x_3 x_4 \partial_1] \\ &= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - x_1 x_4 \partial_3] - \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - 2x_3 x_4^2 \partial_1 - x_1 x_4 \partial_3 + 2x_3 x_4 \partial_1] \\ &= 0.\end{aligned}\tag{36}$$

$$\begin{aligned}
[\mathcal{L}_{\varphi_1}, \mathcal{L}_{\varphi_3}] &= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - x_1 x_4 \partial_3] - \frac{-x_3 x_4 + x_4 x_3}{(x_2^2 + x_3^2 + x_4^2)^{3/2}} [x_1 x_2 \partial_2 + x_1 x_3 \partial_3 + x_1 x_4 \partial_4 - (x_2^2 + x_3^2 + x_4^2) \partial_1] \\
&\quad - \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - 2x_3 x_4 \partial_1 - x_1 x_4 \partial_3 + 2x_3 x_4 \partial_1] \\
&= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - x_1 x_4 \partial_3] - \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_3 \partial_4 - 2x_3 x_4 \partial_1 - x_1 x_4 \partial_3 + 2x_3 x_4 \partial_1] \\
&= 0,
\end{aligned} \tag{37}$$

$$\begin{aligned}
[\mathcal{L}_{\varphi_1}, \mathcal{L}_{\varphi_2}] &= \frac{x_3 \partial_3 + x_4 \partial_4}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{x_1 x_2}{\sqrt{x_3^2 + x_4^2}} + \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{-x_1 x_3^2 - x_1 x_4^2}{(x_3^2 + x_4^2)^{3/2}} [x_2 x_3 \partial_3 + x_2 x_4 \partial_4 - (x_3^2 + x_4^2) \partial_2] \\
&\quad + \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{1}{(x_3^2 + x_4^2)^{1/2}} [x_1 x_2 x_3 \partial_3 - 2x_1 x_3^2 \partial_2 - 2x_1 x_4^2 \partial_2 + x_1 x_2 x_4 \partial_4] \\
&\quad - \frac{1}{(x_3^2 + x_4^2)^{1/2}} \frac{-x_2 x_4^2 - x_2 x_3^2 + (x_3^2 + x_4^2) x_2}{(x_2^2 + x_3^2 + x_4^2)^{3/2}} [x_1 x_2 \partial_2 + x_1 x_3 \partial_3 + x_1 x_4 \partial_4 - (x_2^2 + x_3^2 + x_4^2) \partial_1] \\
&\quad - \frac{1}{(x_3^2 + x_4^2)^{1/2}} \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [x_1 x_2 x_4 \partial_4 - 2x_2 x_4^2 \partial_1 + x_1 x_2 x_3 \partial_3 - 2x_2 x_3^2 \partial_1 - (x_3^2 + x_4^2)(x_1 \partial_2 - 2x_2 \partial_1)] \\
&= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{-x_1 (x_3^2 + x_4^2)}{(x_3^2 + x_4^2)^{3/2}} [x_2 x_3 \partial_3 + x_2 x_4 \partial_4 - (x_3^2 + x_4^2) \partial_2] \\
&\quad + \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{1}{(x_3^2 + x_4^2)^{1/2}} [x_1 x_2 x_3 \partial_3 - 2x_1 x_3^2 \partial_2 - 2x_1 x_4^2 \partial_2 + x_1 x_2 x_4 \partial_4] \\
&\quad - \frac{1}{(x_3^2 + x_4^2)^{1/2}} \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [-2x_2 x_4^2 \partial_1 - 2x_2 x_3^2 \partial_1 - (x_3^2 + x_4^2)(x_1 \partial_2 - 2x_2 \partial_1)] \\
&= \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{-x_1}{(x_3^2 + x_4^2)^{1/2}} [-(x_3^2 + x_4^2) \partial_2] + \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} \frac{1}{(x_3^2 + x_4^2)^{1/2}} [-2x_1 x_3^2 \partial_2 - 2x_1 x_4^2 \partial_2] \\
&\quad - \frac{1}{(x_3^2 + x_4^2)^{1/2}} \frac{1}{\sqrt{x_2^2 + x_3^2 + x_4^2}} [-(x_3^2 + x_4^2) x_1 \partial_2] \\
&= 0.
\end{aligned} \tag{38}$$

B.6 $G = \text{SO}(N)$

We present here the formalism for the $G = \text{SO}(N)$ for any $N \geq 4$. The parametrization is given by (Blumenson, 1960)

$$\begin{aligned}
x_1 &= e^r \cos \varphi_1, \\
x_2 &= e^r \sin \varphi_1 \cos \varphi_2, \\
x_3 &= e^r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3, \\
&\vdots \\
x_j &= e^r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \cdots \sin \varphi_{j-1} \cos \varphi_j, \\
&\vdots \\
x_{n-1} &= e^r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \cdots \sin \varphi_{n-2} \cos \varphi_{n-1}, \\
x_n &= e^r \sin \varphi_1 \sin \varphi_2 \sin \varphi_3 \cdots \sin \varphi_{n-2} \sin \varphi_{n-1}.
\end{aligned} \tag{39}$$

The corresponding Lie algebra elements are given by

$$\begin{aligned}
A_{\varphi_{n-1}} &= \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 \\ 0 & \cdots & 1 & 0 \end{pmatrix}, A_{\varphi_{n-2}} = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 \\ 0 & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -\cos \varphi_{n-1} & -\sin \varphi_{n-1} \\ 0 & \cdots & \cos \varphi_{n-1} & 0 & 0 \\ 0 & \cdots & \sin \varphi_{n-1} & 0 & 0 \end{pmatrix}, \\
A_{\varphi_{n-3}} &= \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -\cos \varphi_{n-2} & -\sin \varphi_{n-2} \cos \varphi_{n-1} & -\sin \varphi_{n-2} \sin \varphi_{n-1} \\ 0 & \cdots & \cos \varphi_{n-2} & 0 & 0 & 0 \\ 0 & \cdots & \sin \varphi_{n-2} \cos \varphi_{n-1} & 0 & 0 & 0 \\ 0 & \cdots & \sin \varphi_{n-2} \sin \varphi_{n-1} & 0 & 0 & 0 \end{pmatrix}, \\
&\vdots \\
A_{\varphi_j} &= \frac{1}{x_j} \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -x_{j+1} & -x_{j+2} & \cdots & -x_n \\ 0 & \cdots & x_{j+1} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & x_{j+2} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & x_n & 0 & 0 & \cdots & 0 \end{pmatrix}, \\
&\vdots \\
A_{\varphi_1} &= \frac{1}{x_1} \begin{pmatrix} 0 & -x_2 & -x_3 & -x_4 & \cdots & -x_n \\ x_2 & \ddots & \vdots & \vdots & \vdots & \vdots \\ x_3 & \cdots & 0 & 0 & 0 & 0 \\ x_4 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \cdots & 0 & 0 & 0 & 0 \\ x_n & \cdots & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{40}
\end{aligned}$$

C Proofs of condition for suitable Lie group

Here we provide the statements with proofs of the results in Section 2.2.

Proposition C.1. *The linear operator induced by Π is complete if Π is the local frame of a vector bundle E over X whose rank is $n \geq \dim X$ almost everywhere. If $\text{rank } E = n$ everywhere, then $E = TX$, the tangent bundle of X .*

Proof. We start by noting that, given the expression of the fundamental fields as derivations, we can write $\mathcal{L}(\mathbf{x}) = \Pi(\mathbf{x})^\top \nabla$. Let $\pi : E \rightarrow X$ be the projection map, then $\text{rank } \pi^{-1}(\mathbf{x}) = \min(\text{rank } \Pi(\mathbf{x}), n)$, since $\text{rank } \nabla = n$. Now, consider $\mathcal{L} \log p(\mathbf{x}) = \mathcal{L} \log q(\mathbf{x})$, which implies $\mathcal{L} \log \frac{p(\mathbf{x})}{q(\mathbf{x})} = 0$. Let $U \subseteq X$ such that $\text{rank } \Pi \geq n \forall \mathbf{x} \in U$, and by assumption $X \setminus U$ has measure zero. Then the above holds if and only if $\nabla \log \frac{p(\mathbf{x})}{q(\mathbf{x})} = 0$, which implies $\frac{p(\mathbf{x})}{q(\mathbf{x})} = c$, constant $\forall \mathbf{x} \in U$. Now, $p(\mathbf{x})$ and $q(\mathbf{x})$ are probability densities by assumption, thus $c = 1$, which proves the claim. $\square$

Proposition C.2. *The operator Π induced by $\mathfrak{g}$ is complete if and only if the subspace $U \subseteq X$ such that $\dim \frac{G}{G_{\mathbf{x}}} < n$ for $\mathbf{x} \in U$, where $n = \dim X$, has measure zero in X .*

Proof. First, we recall that the dimension of an orbit $\mathcal{O}_{\mathbf{x}}$ of $\mathbf{x} \in X$ equals the dimension of the image of the map $d\rho_{\mathbf{x}} : \mathfrak{g} \rightarrow T_{\mathbf{x}}X : \mathbf{A} \mapsto \Pi(\mathbf{x})$. Suppose first that Π is complete. Then, from Proposition C.1 the rank of $\Pi(\mathbf{x})$ is $\geq n$ almost everywhere, and therefore $\dim G/G_{\mathbf{x}} \geq n$ almost everywhere, which implies one direction of the claim. The reverse is quite straightforward. Assume that the rank

of $\Pi(\mathbf{x})$ is $\geq n$ almost everywhere. As Π represent the action of the infinitesimal transformations of G , it means that locally G cannot fix points in X , thus proving the claim. $\square$

D Proof of main theorem

Here we provide the full proof of Theorem 3.1:

Theorem D.1. *Let G be a Lie group acting on X satisfying the conditions of Section 2.2, and let $\mathfrak{g}$ be its Lie algebra. The pair of SDEs*

$$d\mathbf{x} = \left[\beta(t)\Pi(\mathbf{x})\mathbf{f}(\mathbf{x}) + \frac{\gamma(t)^2}{2}\rho_X(\Omega) \right] dt + \gamma(t)\Pi(\mathbf{x})d\mathbf{W}, \quad (41)$$

$$d\mathbf{x} = \left[\beta(t)\Pi(\mathbf{x})\mathbf{f}(\mathbf{x}) - \frac{\gamma^2(t)}{2}\rho_X(\Omega) - \gamma^2(t)\Pi(\mathbf{x})\nabla^\top \cdot \Pi(\mathbf{x}) - \gamma(t)^2\Pi(\mathbf{x})\mathcal{L} \log p_t(\mathbf{x}) \right] dt + \gamma(t)\Pi(\mathbf{x})d\mathbf{W}, \quad (42)$$

where $\beta, \gamma : \mathbb{R} \rightarrow \mathbb{R}$ are time-dependent functions, $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ the fundamental vector fields, $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ the drift, $\Omega = \sum_i A_i^2$ is the quadratic Casimir element of $\mathfrak{g}$, and $\mathcal{L} = \Pi(\mathbf{x})^\top \nabla$ is such that

1. The forward-time SDE (41) is exactly solvable, with solution

$$\mathbf{x}(t) = \left(\prod_i O_i(\tau_i(t)) \right) \mathbf{x}(0) = \left(\prod_{i=1}^n e^{\tau_i(t)A_i} \right) \mathbf{x}(0), \quad (43)$$

where $O_i = e^{\tau_i(t)A_i}$ is the finite group action and $\tau(t)$ is the solution to the SDE

$$d\tau(\mathbf{x}) = \beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}. \quad (44)$$

2. The SDE (6) is the reverse-time process of (5).

3. The Langevin dynamic of the above SDEs decomposes as a direct sum of $\mathfrak{g}$ infinitesimal actions (1), each defining an infinitesimal transformation along the flows ξ_τ .

Proof. We start by proving 3. We start by rewriting (41) in terms of the fundamental flow coordinates $\tau_i = \xi_{A_i}^{-1}(\mathbf{x}_0)(\mathbf{x}) : X \rightarrow \mathbb{R}$. For this we employ Itô's Lemma for the multivariate case: given the SDE (41) and a transformation $\tau(\mathbf{x})$, it is given by

$$\begin{aligned} d\tau(\mathbf{x}) &= (\nabla_{\mathbf{x}}\tau)^\top \left[\beta(t)\Pi(\mathbf{x})\mathbf{f}(\mathbf{x}) + \frac{\gamma^2(t)}{2}\rho_X(\Omega) \right] dt + \frac{\gamma^2(t)}{2} \text{Tr} [\Pi(\mathbf{x})^\top (H_{\mathbf{x}}\tau) \Pi(\mathbf{x})] dt \\ &\quad + \gamma(t)(\nabla_{\mathbf{x}}\tau)^\top \Pi(\mathbf{x})d\mathbf{W} \\ &= \beta(t)\mathbf{f}(\mathbf{x}) + \frac{\gamma^2(t)}{2} [(\nabla_{\mathbf{x}}\tau)^\top \Delta_{\tau}\mathbf{x} + \text{Tr} [\Pi(\mathbf{x})^\top (H_{\mathbf{x}}\tau) \Pi(\mathbf{x})]] dt + \gamma(t)d\mathbf{W} \end{aligned} \quad (45)$$

since $\nabla_{\mathbf{x}}\tau = \Pi^{-1}(\mathbf{x})$ as matrices. Now, the second term can be rewritten in components as

$$\begin{aligned} &\{(\nabla_{\mathbf{x}}\tau)^\top \Delta_{\tau}\mathbf{x} + \text{Tr} [\Pi(\mathbf{x})^\top (H_{\mathbf{x}}\tau) \Pi(\mathbf{x})]\}_{k,l} \\ &= \sum_i \sum_j \frac{\partial x_j}{\partial \tau_k} \left(\frac{\partial}{\partial x_j} \frac{\partial x_i}{\partial \tau_l} \right) \frac{\partial \tau}{\partial x_i} + \sum_i \sum_j \frac{\partial x_j}{\partial \tau_k} \frac{\partial x_i}{\partial \tau_l} \frac{\partial^2 \tau}{\partial x_i \partial x_j} \\ &= \sum_j \frac{\partial x_j}{\partial \tau_k} \frac{\partial}{\partial x_j} \left(\sum_i \frac{\partial x_i}{\partial \tau_l} \frac{\partial \tau}{\partial x_i} \right) \\ &= \frac{\partial}{\partial \tau_k} \left(\sum_i \frac{\partial x_i}{\partial \tau_l} \frac{\partial \tau}{\partial x_i} \right) \\ &= \frac{\partial}{\partial \tau_k} \frac{\partial \tau}{\partial \tau_l} \\ &= H_{\tau}\tau, \end{aligned} \quad (46)$$

which vanishes. Thus we proved that

$$d\boldsymbol{\tau}(\mathbf{x}) = \beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}, \quad (47)$$

and provided that τ_i is chosen so that $f_i(\mathbf{x}(\boldsymbol{\tau})) = f_i(\tau_i)$, this corresponds to a system of independent SDEs, as claimed.

Now, to prove 1, let $\boldsymbol{\tau}(t)$ be a solution to (47) and $\mathbf{x}(t)$ as in (7). Then a Taylor expansion yields

$$\mathbf{x}(t) = \left[I + \sum_i \tau_i(t)A_i + \frac{1}{2} \left(\sum_i \tau_i(t)A_i \right)^2 + \mathcal{O}(\tau_i^3) \right] \mathbf{x}(0) \quad (48)$$

since $[A_i, A_j] = 0$ and where $\mathcal{O}(\tau_i^3)$ represents terms of third order in τ_i 's. Then taking the differential and dropping higher order terms

$$\begin{aligned} d\mathbf{x}(t) &= \left[\sum_i d\tau_i(t)A_i + \frac{1}{2} \left(\sum_i d\tau_i(t)A_i \right)^2 \right] \mathbf{x}(0) \\ &= \left[\sum_i [\beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}] A_i + \frac{1}{2} \left(\sum_i [\beta(t)\mathbf{f}(\mathbf{x})dt + \gamma(t)d\mathbf{W}] A_i \right)^2 \right] \mathbf{x}(0) \\ &= [\beta(t)\boldsymbol{\Pi}(\mathbf{x})\mathbf{f}(\mathbf{x})dt + \gamma(t)\boldsymbol{\Pi}(t)d\mathbf{W}] + \frac{1}{2} \left(\sum_i \gamma(t)d\mathbf{W} A_i \right)^2 \mathbf{x}(0) \\ &= [\beta(t)\boldsymbol{\Pi}(\mathbf{x})\mathbf{f}(\mathbf{x})dt + \gamma(t)\boldsymbol{\Pi}(t)d\mathbf{W}] + \frac{\gamma(t)^2}{2} \left(\sum_i A_i^2 dt \right) \mathbf{x}(0) \\ &= \left[\beta(t)\boldsymbol{\Pi}(\mathbf{x})\mathbf{f}(\mathbf{x}) + \frac{\gamma(t)^2}{2} \rho_X(\Omega) \right] dt + \gamma(t)\boldsymbol{\Pi}(\mathbf{x})d\mathbf{W}, \end{aligned} \quad (49)$$

which in the forward SDE (5), proving our claim, where we used the relations $dW_i^2 = dt$ and $dW_i dW_j = 0$ for $j \neq i$.

Finally, we prove 2. To do this it suffices to apply Anderson's result (Anderson, 1982)

$$\begin{aligned} d\mathbf{x} &= \left[\beta(t)\boldsymbol{\Pi}_i(\mathbf{x})\mathbf{f}(\mathbf{x}) + \frac{\gamma^2(t)}{2} \rho_X(\Omega) - \gamma^2(t) \nabla \cdot (\boldsymbol{\Pi}(\mathbf{x})\boldsymbol{\Pi}(\mathbf{x})^\top) \right. \\ &\quad \left. - \gamma(t)^2 \boldsymbol{\Pi}(\mathbf{x})\boldsymbol{\Pi}(\mathbf{x})^\top \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right] dt + \gamma(t)\boldsymbol{\Pi}(\mathbf{x})d\mathbf{W}_i, \end{aligned} \quad (50)$$

and note that $\boldsymbol{\Pi}(\mathbf{x})^\top \nabla_{\mathbf{x}} = \boldsymbol{\mathcal{L}}$, the generalized score, and

$$\begin{aligned} [\nabla_{\mathbf{x}} \cdot (\boldsymbol{\Pi}(\mathbf{x})\boldsymbol{\Pi}(\mathbf{x})^\top)]_i &= \frac{\partial}{\partial x_k} (\Pi_{ij}\Pi_{kj}) \\ &= \frac{\partial}{\partial x_k} (\Pi_{ij}) \Pi_{kj} + \Pi_{ij} \frac{\partial}{\partial x_k} \Pi_{kj} \\ &= \frac{\partial x_k}{\partial \tau_j} \frac{\partial}{\partial x_k} \left(\frac{\partial x_i}{\partial \tau_j} \right) + \Pi_{ij} [\nabla^\top \cdot \boldsymbol{\Pi}(\mathbf{x})]_j \\ &= \frac{\partial}{\partial \tau_j} \left(\frac{\partial x_i}{\partial \tau_j} \right) + \Pi_{ij} [\nabla^\top \cdot \boldsymbol{\Pi}(\mathbf{x})]_j \\ &= [\text{Tr } H_{\boldsymbol{\tau}}(\mathbf{x})]_i + \Pi_{ij} [\nabla^\top \cdot \boldsymbol{\Pi}(\mathbf{x})]_j \end{aligned} \quad (51)$$

where we recall that the divergence of a matrix is a vector whose components are the divergence of its rows. Recalling the relationship between the trace of the Hessian and the Laplacian we can write in operator form

$$\nabla_{\mathbf{x}} \cdot (\boldsymbol{\Pi}(\mathbf{x})\boldsymbol{\Pi}(\mathbf{x})^\top) = \boldsymbol{\Pi}(\mathbf{x})\nabla^\top \cdot \boldsymbol{\Pi}(\mathbf{x}) + \rho_X(\Omega), \quad (52)$$

Plugging this back in into the previous expression we obtain our claim

$$\begin{aligned} d\mathbf{x} &= \left[\beta(t)\boldsymbol{\Pi}(\mathbf{x})\mathbf{f}(\mathbf{x}) - \frac{\gamma^2(t)}{2} \rho_X(\Omega) - \gamma^2(t)\boldsymbol{\Pi}(\mathbf{x})\nabla^\top \cdot \boldsymbol{\Pi}(\mathbf{x}) \right. \\ &\quad \left. - \gamma(t)^2 \boldsymbol{\Pi}(\mathbf{x})\boldsymbol{\mathcal{L}} \log p_t(\mathbf{x}) \right] dt + \gamma(t)\boldsymbol{\Pi}(\mathbf{x})d\mathbf{W}. \end{aligned} \quad (53)$$

□

Algorithm 1 Training with variance-preserving scheduler	Algorithm 2 Sampling with variance-preserving scheduler
1: repeat 2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ 3: $t \sim \text{Uniform}(\{1, \dots, T\})$ 4: $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 5: $\boldsymbol{\tau}_0 = M_G(\mathbf{x}_0)$ $\triangleright$ Flow coordinates. M_G is group-dependent 6: $\boldsymbol{\tau}_t = \alpha_t \boldsymbol{\tau}_0 + \sigma_t \boldsymbol{\eta}$ $\triangleright$ Sample from $p(\boldsymbol{\tau}_t \boldsymbol{\tau}(\mathbf{x}_0))$ 7: $\mathbf{x}_t = M_G^{-1}(\boldsymbol{\tau}_t)$ $\triangleright$ Cartesian coordinates 8: Take gradient descent step on $\left\ \nabla_{\theta} \left[\mathbf{s}_{\theta}(\mathbf{x}_t, t) + \frac{\boldsymbol{\eta}}{\sigma_t} \right] \right\ ^2$ 9: until converged	1: $\boldsymbol{\tau}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 2: $\mathbf{x}_T = M_G^{-1}(\boldsymbol{\tau}_T)$ $\triangleright$ Cartesian coordinates 3: for $t = T, \dots, 1$ do 4: $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\boldsymbol{\eta} = \mathbf{0}$ 5: $\mathbf{v}_{s,t} = \underbrace{\left(\sum_i \left(\frac{1}{2} \boldsymbol{\tau}_{t,i} + \mathbf{s}_{\theta}(\mathbf{x}_t, t)_i \right) A_i \right)}_{\in \mathbb{R}} \mathbf{x}_t$ $\triangleright$ Dynamics induced by drift and generalized scores 6: $\rho_X(\Omega) = \mathbf{v}_{c,t} = \left(\sum_i A_i^2 \right) \mathbf{x}_t$ $\triangleright$ Dynamics induced by quadratic Casimir elements 7: $\mathbf{v}_{d,t} = \sum_i \underbrace{\nabla \cdot [A_i \mathbf{x}_t]}_{\in \mathbb{R}} (A_i \mathbf{x}_t)$ $\triangleright$ Dynamics induced by divergences 8: $\mathbf{v}_t = \mathbf{v}_{s,t} + \frac{1}{2} \mathbf{v}_{c,t} + \mathbf{v}_{d,t}$ 9: $\tilde{\mathbf{x}}_{t-1} = \mathbf{x}_t + \beta_t \mathbf{v}_t$ $\triangleright$ Update state based on velocity 10: $\mathbf{x}_{t-1} = \tilde{\mathbf{x}}_{t-1} + \sqrt{\beta_t} \sum_i \eta_i A_i \mathbf{x}_t$ $\triangleright$ Stochastic dynamics 11: $\boldsymbol{\tau}_{t-1} = M_G(\mathbf{x}_{t-1})$ 12: end for 13: return $\mathbf{x}_0$

E Experiments

Practical implementation. In this section we list practical implementations for training and inference of our proposed Algorithm 1 and Algorithm 2 assuming a variance-preserving SDE for the flow-coordinates, see Eq. (55), because we know that this standard SDE is exactly solvable and related to the forward SDE in Cartesian space as stated in the main Theorem 3.1.

The implementation showcase the examples for $G_0 = (\text{SO}(2) \times \mathbb{R}_+)$ (see the second paragraph in 3.1) for data living in $\mathbf{x} \in \mathbb{R}^2$ and $G_1 = (\text{SO}(3) \times \mathbb{R}_+)$ for $\mathbf{x} \in \mathbb{R}^3$ from Appendix B.3.

The flow-maps for G_0 and G_1 can be computed by leveraging the bijection from Cartesian to polar $\boldsymbol{\tau} = (r, \theta)$ for G_0 and spherical $\boldsymbol{\tau} = (r, \theta, \phi)$ for G_1 , respectively. As stated in the main text and Appendix, we obtain

$$M_{G_0}(\mathbf{x}) = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \arctan(\frac{y}{x}) \end{pmatrix}, \quad M_{G_0}^{-1}(\boldsymbol{\tau}) = \begin{pmatrix} r \cos(\theta) \\ r \sin(\theta) \end{pmatrix}.$$

As mentioned in (13), the Lie algebra basis are $A_r = \mathbf{I}$ and $A_{\theta} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, yielding a quadratic Casimir operator $A_r^2 + A_{\theta}^2 = \mathbf{0}$, such that the dynamics induced by the Casimir elements in line 6 in Alg. 2 vanishes, i.e. $\mathbf{v}_c = \mathbf{0}$. The dynamics induced by the divergences (line 7 in Alg. 2) returns $\nabla \cdot A_r \mathbf{x} = \nabla \cdot \mathbf{x} = \sum_{i=1}^2 \frac{\partial}{\partial x_i} x_i = \sum_{i=1}^2 1 = 2$ and $\nabla \cdot A_{\theta} \mathbf{x} = \nabla \cdot (-x_2, x_1)^{\top} = \frac{\partial}{\partial x_1} (-x_2) + \frac{\partial}{\partial x_2} x_1 = 0$. Therefore, the divergence dynamics returns the velocity component $\mathbf{v}_d = 2A_r \mathbf{x} + 0A_{\theta} \mathbf{x} = 2\mathbf{x}$.

For G_1 the bijection to flow- and Cartesian coordinates is well-known as

$$M_{G_1}(\mathbf{x}) = \begin{pmatrix} \sqrt{x^2 + y^2 + z^2} \\ \arctan(\frac{\sqrt{x^2 + y^2}}{z}) \\ \arctan(\frac{y}{x}) \end{pmatrix}, \quad M_{G_1}^{-1}(\boldsymbol{\tau}) = \begin{pmatrix} r \sin(\theta) \cos(\phi) \\ r \sin(\theta) \sin(\phi) \\ r \cos(\theta) \end{pmatrix}.$$

With $A_r = \mathbf{I}$, $A_{\theta} = (\cos \phi A_y - \sin \phi A_x)$ and $A_{\phi} = A_z$ as defined in (22)-(23), the quadratic Casimir elements A_i^2 are left multiplied with the vector representation $\mathbf{x}$, we can distinguish each

group component as follows

$$\begin{aligned} A_r^2 \mathbf{x} &= \mathbf{I}^2 \mathbf{x} = \mathbf{I} \mathbf{x} = \mathbf{x} = (x_1, x_2, x_3)^\top \\ A_\theta^2 \mathbf{x} &= (\cos^2(\phi) A_y^2 + \sin^2(\phi) A_x^2 - \cos \phi \sin \phi A_y A_x - \cos \phi \sin \phi A_x A_y) \mathbf{x} = -\mathbf{x} = -(x_1, x_2, x_3)^\top \\ A_\phi^2 \mathbf{x} &= A_z^2 \mathbf{x} = -(x_1, x_2, 0)^\top, \end{aligned}$$

defining the Casimir dynamics in line 6 in Algorithm 2.

The dynamics induced by the divergences are computed in the same manner as shown in the SO(2) examples. Specifically, we obtain the (scalar) divergences

$$\begin{aligned} \nabla \cdot A_r \mathbf{x} &= \nabla \cdot \mathbf{x} = 3 \\ \nabla \cdot A_\phi \mathbf{x} &= \nabla \cdot A_z \mathbf{x} = \nabla \cdot (-x_2, x_1, 0) = \frac{\partial}{\partial x_1}(-x_2) + \frac{\partial}{\partial x_2} x_1 + \frac{\partial}{\partial x_3} 0 = 0 \\ \nabla \cdot A_\theta \mathbf{x} &= \frac{x_3}{\sqrt{x_1^2 + x_2^2}}, \end{aligned}$$

where the last divergence is point dependent.

In practice, it suffices to compute the quadratic Casimir elements directly using GPU-accelerated frameworks when these are point dependent as in A_θ , or pre-compute them should they be constant matrices. The divergences can be computed using automatic differentiation libraries from modern deep learning frameworks.

Experimental details. In this final section we present some further details regarding our experiment in Section 5. We provide the code to replicate our experiments in the Supplementary Information (SI). Following publication we will open-source our code.

E.1 2D and 3D toy datasets

Table 3: Comparison of GSM and Fisher Score matching on 2D and 3D synthetic datasets. Best results are in bold. When numbers are two close we consider them on par.

Dataset	Group	W2
MoG (2D)	$\text{SO}(2) \times \mathbb{R}^+$	0.34
MoG (2D)	$T(2)$	0.15
Concentric Circles (2D)	$\text{SO}(2) \times \mathbb{R}^+$	0.19
Concentric Circles (2D)	$T(2)$	0.17
Line (2D)	$\text{SO}(2) \times \mathbb{R}^+$	0.33
Line (2D)	$T(2)$	0.56
MoG (3D)	$\text{SO}(2) \times \mathbb{R}^+$	0.40
MoG (3D)	$T(3)$	0.44
Torus (3D)	$\text{SO}(3) \times \mathbb{R}^+$	0.14
Torus (3D)	$T(3)$	0.35
Möbius Strip (3D)	$\text{SO}(3) \times \mathbb{R}^+$	0.06
Möbius Strip (3D)	$T(3)$	0.16

We perform a quantitative evaluation using the Wasserstein-2 (W2) distance on synthetic 2D and 3D datasets, comparing standard (Fisher) score matching ($G = T(2)$, $G = T(3)$) with our proposed approach based on Lie groups ($G = \text{SO}(2) \times \mathbb{R}^+$) and ($G = \text{SO}(3) \times \mathbb{R}^+$). A strong bias in such experiments arises from the similarity between the prior and target distributions. The considered toy datasets are often symmetric with respect to the origin in $\mathbb{R}^{2,3}$, as in the standard Gaussian prior in Fisher score matching. The similarity of the prior distribution to the target one affects decisively the performance of the generating process.

To account for this, we report a normalized W2 metric, dividing the W2 distance between samples and target by the W2 distance between target and the corresponding priors. We observe that generalized

score matching (GSM, ours) performs on par or better in most datasets, particularly where symmetry provides a clear inductive bias as indicated in Table 3. In the MoG datasets, standard (Fisher) score matching ($G = T(N)$) outperforms the Lie group model ($G = \text{SO}(N) \times \mathbb{R}^N$), which is expected since no rotational symmetry is present, while translation symmetry effectively helps locate the Gaussian modes. The performance gap becomes even more pronounced in 3D, where GSM shows stronger advantages. We hypothesize that in higher dimensions, memorizing the target distribution becomes more difficult, and models that incorporate symmetry more explicitly benefit increasingly from this inductive bias.

E.2 MNIST

We parametrize the noising process through the SDE

$$d\tau = \sqrt{\beta(t)}d\mathbf{W}, \quad (54)$$

where we set the drift term to zero. Notice that this choice is consistent with a 2d-rotation of a function over the grid $\mathbf{x}_{i,j}$, given by $f(\mathbf{x}_{i,j}) = f_{i,j}$, denoting the value of the pixel of image f at the location i, j . We train a convolutional neural network (CNN) with three convolutional layers followed by fully connected layers that outputs a single value, being the score for the flow coordinate τ . For the specific details of the implementation we refer to the code-base in the SI. In sampling, we apply a smoothing function to compensate interpolation artifacts due to rotations on a discretized grid. We choose $T = 100$ time-steps in training but only need $T = 10$ time-steps during sampling.

E.2.1 BBDM

We implement the Brownian Bridge Diffusion Model (BBDM) (Li et al., 2023) and train it on the rotated MNIST dataset. The BBDM operates on the full pixel space $\mathbb{R}^{784}$ of the 28×28 MNIST digits and indicates a continuous time stochastic process conditioned on the starting $\mathbf{x}(0)$ and end point $\mathbf{x}(T)$ which are pinned together as paired data. In this case, we assume $\mathbf{x}(T) \sim p(\mathbf{x}_T)$ to be a randomly augmented MNIST digit obtained from an original MNIST digit $\mathbf{x}(0)$. During training, we sample an intermediate point $\mathbf{x}(t) \sim N(x_t | \mu_t(x(0), x(T)), \Sigma_t)$ where the mean function $\mu_t(t)(\mathbf{x}(0), \mathbf{x}(T))$ is a linear interpolation between the endpoints $(\mathbf{x}(0), \mathbf{x}(T))$ and use the score-network to predict the original data point $\hat{\mathbf{x}}(0) = s_\theta(x_t, t, x_T)$ as opposed to the noise or difference parameterization proposed in the original BBDM paper. We noticed that predicting the original data point led to better sampling quality including the inductive bias that MNIST digits are represented as binary tokens. Furthermore, we observe that the sampling quality is also better when the prior image x_T is input as context into the score network, enforcing a stronger signal throughout the trajectory. As opposed to our model, we trained the BBDM on $T = 1000$ diffusion timesteps using the sin-scheduler from BBDM.

To evaluate the quality of generated MNIST samples, we train a convolutional neural network (CNN) classifier to predict digit labels. This classifier provides a reliable metric for assessing the reconstruction accuracy by comparing the predicted labels of original (unrotated) and generated (rotated) images. The architecture consists of two convolutional blocks, followed by fully connected layers to predict the 10 MNIST digit labels. All convolutional layers use 3×3 kernels with padding 1, and max pooling uses 2×2 kernels. The model is trained using Adam optimizer (lr=0.001), cross-entropy loss, batch size 64, for 10 epochs on the standard MNIST training set (60,000 samples). The trained classifier achieves greater than 99% accuracy on the MNIST test set, providing a reliable metric for evaluating reconstruction quality.

To calculate the FID scores, we extract the embedding after the second convolutional block.

E.3 QM9 & CrossDocked2020

QM9. The conformer generation tasks is about learning a conditional probabilistic map $\mathbf{x} \sim p_\theta(X|M)$, where $\mathbf{x} \in \mathbb{R}^{3N}$ for a molecule with N atoms. We implement a variant of EQGAT (Le et al., 2022b) as neural network architecture where input features for the nodes consist of atom types and atomic coordinates, while edge features are encoded to indicate the existence of a single-, double, triple or aromatic bond based on the adjacency matrix. We use $L = 5$ message passing layers with $s_{\text{dim}} = 128$, $v_{\text{dim}} = 64$ scalar and vector features, respectively. To predict the scores for each atom, we concatenate the hidden scalar and vector embeddings $s \in \mathbb{R}^{128}$, $v \in \mathbb{R}^{3 \times 64}$ into one output

embedding $o = \mathbb{R}^{128+3*64}$ which is further processed by a 2-layer MLP with three output units. Notice that the predicted scores per atom are neither invariant nor equivariant since the scalar and vector features are transformed with an MLP.

We choose the drift f with its scaling β and the diffusion coefficients γ in such way that the forward SDE for the flow coordinates τ in (8) has the expression

$$d\tau = -\frac{1}{2}\beta(t)\tau dt + \sqrt{\beta(t)}d\mathbf{W}, \quad (55)$$

where for clarity we have omitted the dependency between the flow coordinates and the original data in Cartesian coordinates, i.e. $\tau(\mathbf{x})$, since the coordinate transformations with Lie algebra representation are described in B.3. The forward SDE in (55) is commonly known as *variance-preserving* SDE (Song et al., 2020b). We use the cosine scheduler proposed by Dhariwal & Nichol (2021) and $T = 100$ diffusion timesteps.

CrossDocked. For this experiment we adopt again an SDE of the form (55) for the three SO(3) flow coordinates $\theta_1, \varphi_1, \varphi_2$ and the three $T(3)$ center of mass Cartesian flow coordinates. The SO(3) flow coordinates are always computed and applied in the ligand center of mass. In this way there is no ambiguity regarding the non-commutativity of SE(3), as rotation around the origin commute with translations of the system. We train a variant of EQGAT as in the QM9 case, but now including also node and edge features of the protein pocket. Specifically, the adjacency matrix for the GNN is computed dynamically at each time step, according to the relative distance between ligand and protein. For this, we choose a cut-off of 5 Å. We also use in this experiment a cosine scheduler and $T = 100$ diffusion timesteps. Since this learning problem is 6-dimensional, we aggregate the last layer’s node embeddings from the ligand atoms into a global representation through summation. This embedding is fed as input into a 2-layer MLP to predict the six scores.

E.3.1 RSGM and BBDM n CrossDocked

RSGM We utilized the framework of Riemannian Score-Based Generative Models (RSGM) by (De Bortoli et al., 2022) to model rigid-body motions on $G = (\text{SO}(3) \times T(3))$, in similar fashion to (Corso et al., 2023; Yim et al., 2023) by choosing a variance exploding SDE for the rotation dynamics and variance preserving SDE for the global translations. The terminal distribution for the rotation is designed to converge to an isotropic Gaussian distribution on SO(3) (Leach et al., 2022), while the terminal distribution for the translation component converges to an isotropic Gaussian in $\mathbb{R}^3$. To obtain the tractable scores for rotation and translation, we use the code by the authors from DiffDock and SE(3)-Diffusion for Protein Backbone Modeling in <https://github.com/gcorso/DiffDock/blob/main/utils/so3.py> and https://github.com/jasonkyuyim/se3_diffusion/blob/master/data/se3_diffuser.py and make sure that the score outputs for rotation and translation are SO(3) equivariant using the same EQGAT model architecture. The (variance-preserving) scheduler for the translation dynamics is chosen in similar fashion to our experiment using the cosine scheduler, while the (variance-exploding) scheduler for the rotation dynamics is implemented as an linear increasing sequence in $\log_{10}$ space with $\sigma_{min} = 0.001$ and $\sigma_{max} = 2.0$ and $T = 100$ discretized diffusion steps as $\sigma(t) = 10^t$ for $t \in (\log_{10}(\sigma_{min}), \log_{10}(\sigma_{max}))$.

BBDM In similar fashion to the MNIST experiment, we train and evaluate a standard Euclidean diffusion model on CrossDocked2020 for rigid docking. We sample a rotated ligand endpoint x_T using the Riemannian Score-Based Generative Models (RSGM) scheduler, with the original ligand x_0 , and sample intermediates as $x_t = m_t x_0 + (1 - m_t) x_T + \sigma_t \epsilon$, where $\epsilon \sim N(0, I)$ and $m_t = \frac{t}{T}$, $\sigma_t = 2(m_t - m_t^2)$ using $T = 100$ diffusion steps. As the Euclidean baseline, we train an equivariant Fisher score network with $3N$ degrees of freedom to predict the ground-truth pose $\hat{x}_0$. In this setting intermediate perturb ligand coordinates x_t do not resemble ligands due to the linear interpolation and addition of Gaussian noise, while the learning task is to predict the ground-truth ligand coordinate, given the static protein pocket. The output prediction head in BBDM is $3N$, compared to to GSM and RSGM which model 6 dimensions accounting for global rotation and translation.

To compare all modeling approaches with respect to the dynamics using the same network architecture, we perform 5 dockings per protein-ligand complex in the CrossDocked test dataset comprising 100 complexes and compute the mean RMSD between ground-truth coordinates and predicted coordinates.

F Lie group-induced flow matching modeling

We briefly summarize the formalism of flow matching. Given a target distribution $p_0(\mathbf{x})$ and a vector field u_t generating the distribution $p_t(\mathbf{x})$, i.e., if it satisfies $p_t(\mathbf{x}) = [u_t]_* p_0(\mathbf{x})$ where $[u_t]_*$ is the push-forward map, the flow matching objective is defined as

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t, \mathbf{x}_t \sim p_t(\mathbf{x})} |v_{t;\theta}(\mathbf{x}_t) - u_t(\mathbf{x}_t)|^2. \quad (56)$$

Marginalizing over samples $\mathbf{x}_0 \sim p_0(\mathbf{x})$ we obtain the conditional flow matching objective

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0 \sim p_0(\mathbf{x}), \mathbf{x}_t \sim p_t(\mathbf{x}|\mathbf{x}_0)} |v_{t;\theta}(\mathbf{x}_t) - u_t(\mathbf{x}_t|\mathbf{x}_0)|^2. \quad (57)$$

Now, under the assumptions for learning the generalized score through the objective 10 we have that $p_t(\boldsymbol{\tau}(\mathbf{x})) = \mathcal{N}(\boldsymbol{\tau}|\boldsymbol{\mu}(\boldsymbol{\tau}(0), t), \boldsymbol{\Sigma}(t))$, where $\boldsymbol{\tau}(0) = \mathbf{x}(\boldsymbol{\tau})(0)$. Then the solution of the SDE from Theorem 3.1

$$\mathbf{x}(t) = \left(\prod_{i=1}^n e^{\tau_i(t) A_i} \right) \mathbf{x}(0), \quad (58)$$

is a flow inducing the distribution $p_t(\boldsymbol{\tau}(\mathbf{x}))$. Thus, the vector field that generates the conditional probability path is obtained by differentiating the path above with respect to t , yielding

$$\begin{aligned} u_t(\mathbf{x}(t)|\mathbf{x}(0)) &= \frac{d}{dt} \mathbf{x}(t) = \sum_i \frac{\partial \mathbf{x}(t)}{\partial \tau_i} \frac{\partial \tau_i}{\partial t} \\ &= \sum_i A_i \mathbf{x}(t) (\mu'_{t,i}(\tau_i(0)) + \sigma'_t(\tau_i(0)) \eta_i) \\ &= \Pi(\mathbf{x}(t)) \left[\boldsymbol{\mu}'_t(\boldsymbol{\tau}(0)) + \frac{\sigma'_t(\boldsymbol{\tau}(0))}{\sigma_t(\boldsymbol{\tau}(0))} (\boldsymbol{\tau}(t) - \boldsymbol{\mu}_t(\boldsymbol{\tau}(0))) \right], \end{aligned} \quad (59)$$

where we used the fact that

$$\boldsymbol{\tau}(t) = \boldsymbol{\mu}_t(\boldsymbol{\tau}(0)) + \sigma_t(\boldsymbol{\tau}(0)) \boldsymbol{\eta}, \quad (60)$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$. Thus, we see that the unique vector field that defines the flow (7) is again proportional to the fundamental vector field $\Pi(\mathbf{x})$ of the Lie algebra $\mathfrak{g}$ of G . In figure 10 we illustrate the flow generated by our formalism in the case of $\text{SO}(2)$ in comparison with the traditional flow matching of $T(2)$. The orange path depicts the linear (in Euclidean metric) displacement given by the traditional flow matching, assuming $G = T(2)$. In green and blue we depicted the orbits trajectories resulting from generalized flow matching with $G = \text{SO}(2) \times \mathbb{R}_+$. Although the start and end points are the same, the path is decomposed into transformations along the orbits of the two group factors. This is particularly useful when these correspond to meaningful degrees of freedom in the system. For example, when flowing between conformers of the same molecule, the intermediate states produced by traditional flow matching are often unphysical, as they involve linear interpolation between the Cartesian coordinates of the atoms. However, generalized score matching, following the degrees of freedom given by bond and torsion angles as described in Section 3.1, would not only yield efficient learning but also produce chemically meaningful intermediate states, as the path is broken down into updates of chemically relevant degrees of freedom.

G Further Related Work

In the context of interpreting the latent space of diffusion models, Park et al. (2023) explores the local structure of the latent space (trajectory) of diffusion models using Riemannian geometry. Similarly, Haas et al. (2024) propose a method to uncover semantically meaningful directions in the semantic latent space (h -space) (Wang et al., 2023) of denoising diffusion models (DDMs) by PCA. Wang et al. (2023) propose a method to learn disentangled and interpretable latent representations of diffusion models in an unsupervised way. We note that the aforementioned works aim to extract meaningful latent factors in traditional DDMs, often restricting to human-interpretable semantic features and focusing on image generation.

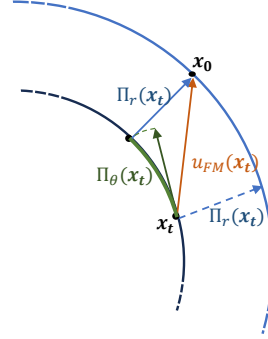


Figure 10: $\mathfrak{so}(2)$ (green and blue) vs. $\mathfrak{t}(2)$ (orange) induced flows.

H Further Outlook

In the context of generative chemistry, particularly for modeling interactions within protein pockets, our methods could be employed to decouple the intrinsic generation of ligands from the global transformations required to fit the ligand into the pocket. This approach can also be extended beyond 3D coordinates, for instance, by working with higher-order representations, such as modeling electron density (Rackers et al., 2023).

Moreover, for more complex problems, it is feasible that an optimal generation process can be achieved by combining different choices of G along the trajectory. In the context of ligand generation, we propose a **time-dependent group action** $G_t = tT(3N) + (1-t)(\text{SO}(3) \times \mathbb{R}_+)^N$: at the beginning of the diffusion process, when the point cloud is still far from forming a recognizable conformer, we can leverage the properties of a true Gaussian prior. As the point cloud is gradually optimized to “resemble a molecule”, we progressively transition to a generalized score-guided process. This shift allows us to fine-tune chemically relevant properties, such as bonds and torsion angles, ensuring that the intermediate and final conformers are chemically valid and accurate. This will be the focus of our forthcoming work.

A potential limitation of our work is that it currently does not extend to representations of finite groups. While finite groups also admit a rich representation theory, it remains an open question how to adapt our framework to those settings. Another limitation lies in our assumption that X is a vector space, whereas Lie groups can act on more general manifolds. Although the conditions discussed in Section 2.2 hold for arbitrary manifolds, our main theoretical results are restricted to actions on vector spaces. Extending the full analysis to curved spaces would be a compelling direction for future work, potentially enabling a general theory of diffusion on Riemannian manifolds via Lie group representations.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: We stated our contributions clearly in the Introduction Section 1.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: In the Appendix H we discuss how our work can be further improved.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: All the assumptions can be found in Section 2.2, while the proofs can be found in the Appendix C and D.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: We also share the code to reproduce our results in the Supplementary Material of this submission.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We describe the experimental details in the Appendix E and we submitted the code to reproduce our results in the Supplementary Material of this submission.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: All the details are presented in the Appendix and in addition we share our code to reproduce the results. Pseudo-code for training and inference are provided in Algorithms 1 and 2 in the Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: This can be seen in the description of the results in the Experiment Section 5.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We provide all the details in the Appendix and in addition we share the code to reproduce all our results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: we reviewed the Code Of Ethics and confirm that the research in the paper conform to it.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All the dataset we used are publicly available and properly cited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We provide our code as in the Supplementary Material of this submission.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.