Generative Learning on Quotient Spaces: A Geometric Diffusion Framework for Symmetry Reduction

Anonymous authors
Paper under double-blind review

Abstract

Diffusion models are commonly defined in ambient Euclidean spaces, even when the underlying data distribution is invariant to rotations, translations, poses, or other continuous symmetries. This creates a geometric mismatch: the stochastic process propagates noise along both intrinsic data directions and redundant symmetry orbits, forcing the score model to learn nuisance variation that carries no quotient-level information. We propose **QDM**, a quotient-geometric diffusion framework that makes invariance a property of the generative dynamics rather than only the score architecture. **QDM** defines diffusion on symmetry-reduced quotient manifolds and realizes the process in ambient coordinates through horizontal projection. The framework unifies exact, approximate, context-dependent, and mode-changing symmetries via soft horizontal projection and stratified quotient spaces. We establish the ambient–quotient measure correspondence, prove well-posedness of the lifted quotient diffusion, and derive a Fisher-information decomposition showing that quotient projection suppresses removable vertical score variance and reduces effective score-estimation complexity. Experiments on molecular conformer generation and legged robot locomotion show consistent gains in convergence, sample efficiency, geometric accuracy, and generation stability over ambient and equivariant diffusion baselines. These results suggest that symmetry-aware generative modeling is best achieved by changing the probability space and diffusion geometry itself, not only by imposing equivariance on the neural score field.

1 Introduction

Diffusion and score-based models have become a dominant framework for high-dimensional generative modeling, with broad success in vision, molecular modeling, robotics, and scientific simulation (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2020; Dhariwal & Nichol, 2021; Rombach et al., 2022; Hoogeboom et al., 2022; Corso et al., 2022; Chi et al., 2025; Price et al., 2025). However, many such domains possess continuous Lie-group symmetries, including rotations, translations, and pose transformations (Bronstein et al., 2021; Kondor & Trivedi, 2018; Cohen & Welling, 2016; Lee, 2003). Standard diffusion processes inject noise in the full ambient space, thereby mixing intrinsic data variation with symmetry-induced nuisance directions. Equivariant architectures constrain the learned score field, but the underlying stochastic dynamics remain ambient and continue to propagate redundant symmetry degrees of freedom.

This creates a geometric mismatch. For a G-invariant distribution, points lying on the same group orbit $G \cdot x$ represent the same intrinsic quotient state $[x] \in \mathcal{Q} := \mathscr{X}/G$. However, standard diffusion is defined in the ambient space $\mathscr{X}$ and injects noise along all tangent directions, including vertical directions tangent to the symmetry orbit. These vertical directions change the coordinate representation of a sample but not its quotient-level state. As a result, the learned score field must separate intrinsic geometric variation from symmetry-induced nuisance variation, increasing the effective dimension of the estimation problem and amplifying score-estimation variance. This mismatch motivates defining the diffusion process directly on the quotient geometry, so that stochastic evolution is concentrated on the horizontal directions that encode statistically relevant variation.

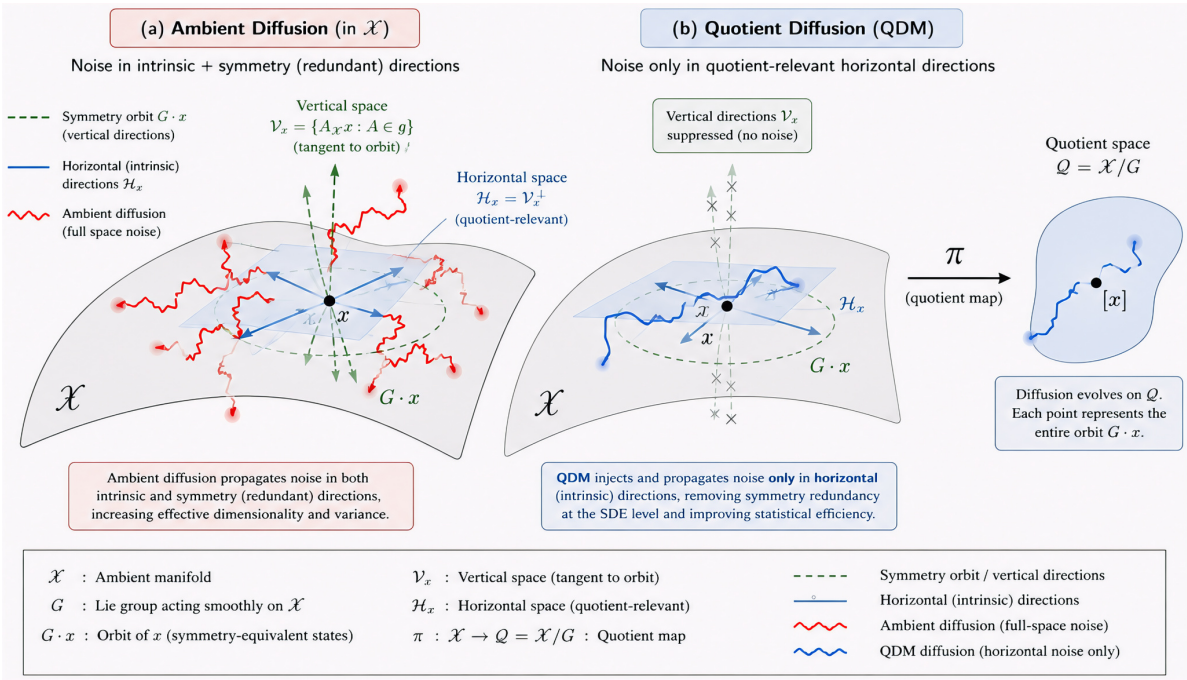

Figure 1: Ambient diffusion vs. Quotient diffusion. Ambient diffusion injects and propagates noise in the full space $\mathscr{X}$, including redundant vertical directions tangent to the symmetry orbit $G \cdot x$. In contrast, **QDM** identifies symmetry-equivalent states through the quotient $\mathcal{Q} := \mathscr{X}/G$ and evolves the stochastic process only along the horizontal, quotient-relevant directions. This removes symmetry-induced redundancy at the SDE level rather than relying only on score-network equivariance.

1.1 Our Contributions: QDM

We propose **QDM**, a principled framework for generative learning on symmetry-reduced quotient manifolds. Unlike equivariant diffusion methods that impose symmetry mainly through the score-network architecture, **QDM** defines the stochastic process on the quotient geometry itself and realizes it in ambient coordinates through horizontal projection. For a smooth manifold $\mathscr{X}$ equipped with a regular Lie group action by G, we decompose each tangent space into vertical orbit directions and horizontal intrinsic directions, and evolve the diffusion only along the horizontal subspace (Fig. 1). Thus, **QDM** removes symmetry redundancy at the stochastic differential equation (SDE)-level, making invariance intrinsic to the generative dynamics.

1.1.1 Core Idea.

If the data distribution is invariant under a symmetry group G, then the true statistical information does not live in the full ambient space $\mathscr{X}$, but in the quotient space $\mathcal{Q} := \mathscr{X}/G$, where symmetry-equivalent points are identified. For example, in molecular conformer generation, rotating or translating a molecule changes its cartesian coordinates but not its bonding structure, 3D geometry, or physical identity. Thus, all rigidly transformed copies of the same conformer should be treated as one quotient state rather than as separate training samples.

Proposed **QDM** is not merely another equivariant architecture; it changes the geometric space on which the diffusion process is defined. Instead of treating symmetry-related ambient configurations as distinct samples in $\mathscr{X}$, **QDM** represents them through the quotient space $\mathcal{Q} := \mathscr{X}/G$, where $[x] := \{g \cdot x \colon g \in G\}$ denotes the orbit of x under the group action. Thus, rotations, translations, poses, and other group-induced transformations are factored out when they do not change the intrinsic state of the data. This differs from standard equivariant diffusion, which constrains the learned score field while the stochastic process itself remains defined in the ambient space. It also differs from generic dimensionality-reduction methods such as

PCA (Jolliffe, 2002), Isomap (Tenenbaum et al., 2000), Laplacian Eigenmaps (Belkin & Niyogi, 2003), and t-SNE (Van der Maaten & Hinton, 2008): the reduction is not driven by a heuristic embedding objective, but by a known symmetry structure. Under a free and regular group action, quotienting removes the $\dim(G)$ orbit directions and reduces the effective dimension from d to $d - \dim(G)$, while preserving the intrinsic degrees of freedom that determine the invariant data distribution. In this sense, **QDM** embeds principled, symmetry-aware geometric reduction directly into the diffusion dynamics.

1.1.2 Why QDM Matters?

QDM moves symmetry handling from the architecture level to the stochastic-process level. Existing equivariant diffusion models (Hoogeboom et al., 2022; Xu et al., 2022; Cornet et al., 2024; Jing et al., 2022) mainly enforce G-equivariance through the score network, while the diffusion dynamics still evolve in the ambient space. Recent quotient-space diffusion (Xu et al., 2026) advances this idea by projecting equivariant diffusion onto the quotient space and deriving a compatible sampler, while also showing that alignment-based methods such as GeoDiff- (Xu et al., 2022) or AlphaFold-style (Jumper et al., 2021) alignment can distort the sampling distribution. Our **QDM** builds on this quotient perspective but extends it to a broader setting with exact, approximate, context-dependent, and mode-changing symmetries, removing symmetry-induced components directly at the SDE-level through horizontal projection. Thus, the model no longer needs to learn to suppress nuisance rotations, translations, or pose changes from data; these directions are removed by construction. This turns invariance from an architectural constraint into a property of the generative process, reducing effective dimension, lowering score-estimation variance, and improving sample efficiency.

1.1.3 Consequences of Quotient Diffusion

Proposed **QDM**: ***(a)*** removes redundant symmetry directions; ***(b)*** reduces score-estimation variance; ***(c)*** improves sample efficiency; ***(d)*** gives a principled probabilistic interpretation; ***(e)*** handles exact, approximate, and changing symmetries in one framework. **QDM**'s main novelty: invariance is enforced at the stochastic-process / SDE-level through horizontal projection, with a general geometric and statistical-efficiency theory. **QDM** is not just an equivariant architecture, but a change in the probability space and diffusion geometry itself. By removing redundant symmetry directions from the diffusion process, **QDM** achieves faster convergence, better sample efficiency, and more stable generation across symmetry-rich domains.

2 Background

Score-based generative models (Song et al., 2021) define a forward stochastic process that gradually corrupts data into noise and learn to reverse it by estimating the score function $\nabla_x \log p_t(X_t)$ of the noisy data distribution at each time t. The reverse SDE (Oksendal, 2013) takes the form

$$dX_t := -\sigma^2 \cdot \nabla_{\mathscr{X}} \log p_t(X_t)\, dt + \sigma \cdot d\bar{W}_t, \tag{1}$$

where $\bar{W}_t$ is reverse-time Brownian motion and the score is approximated by a neural network trained via denoising score matching (Vincent, 2011). Equivariant Diffusion Models (Hoogeboom et al., 2022) enforce G-equivariance of the score network f_θ so that $f_\theta(g \cdot x) = g \cdot f_\theta(x)$. While effective, these approaches still diffuse in the full ambient space $\mathscr{X}$, leaving redundant symmetry directions active in the SDE. **QDM** acts at the level of the diffusion geometry itself, eliminating symmetry directions from the stochastic process rather than merely constraining network outputs.

For a free, proper, isometric Lie group action $G \curvearrowright \mathscr{X}$, the projection $\pi\colon \mathscr{X} \to \mathcal{Q}$ is a principal G-bundle and a Riemannian submersion (O'Neill, 1983) that sends a point x to its equivalence class $[x]$, meaning all points related to x by the group action are identified. The derivative $d\pi_x\colon T_x\mathscr{X} \mapsto T_{\pi(x)}\mathcal{Q}$ tells us which tangent directions in the ambient space remain visible after quotienting. The tangent space decomposes orthogonally as $T_x\mathscr{X} = \mathcal{V}_x \oplus \mathcal{H}_x$, where $\mathcal{V}_x = \ker d\pi_x$ denotes the vertical space, consisting of directions tangent to the symmetry orbit $G \cdot x$. These directions are annihilated by the quotient projection because they do not change the quotient state $[x]$. The horizontal space $\mathcal{H}_x = \mathcal{V}_x^\perp$ is the orthogonal complement of $\mathcal{V}_x$ and represents the intrinsic directions that remain visible on the quotient manifold. The restriction

Algorithm 1 **Quotient Diffusion Model (QDM)**

Require: Training samples $\{x_i\}_{i=1}^N \subset \mathscr{X}$, symmetry group G, context variables τ, mode set $\mathcal{C}$, diffusion horizon T, noise scale σ, score network f_θ

Ensure: Trained quotient diffusion model f_θ and generated samples $\hat{x} \in \mathscr{X}$

1: **Construct quotient geometry.** For each active symmetry regime $c \in \mathcal{C}$, define the quotient space $\mathcal{Q}(c) := \mathscr{X}(c)/G(c)$ with projection map $\pi^c \colon \mathscr{X}(c) \mapsto \mathcal{Q}(c)$.

2: **Compute vertical generators.** For each sample $x \in \mathscr{X}$ and Lie algebra basis $\{\xi_1 \cdots \xi_k\}$ of G, compute the infinitesimal generators (§ 3.2)

$$\xi^j_{\mathscr{X}}(x) := \frac{d}{dt}\Big\{\exp(t \cdot \xi_j) \cdot x\Big\}\Big|_{t=0}; \quad j = 1 \cdots k.$$

3: **Estimate context-adaptive vertical space.** Use a context encoder $\phi_\psi(\tau)$ to form

$$V_x(\tau) := \phi_\psi(\tau) \odot \Big[\xi^1_{\mathscr{X}}(x) \cdots \xi^k_{\mathscr{X}}(x)\Big]. \text{ (§ 3.2)}$$

4: **Compute horizontal projection.** Construct the regularized horizontal projector

$$\Pi_x(\tau) := I_d - V_x(\tau) \cdot \Big(V_x(\tau)^\top \cdot V_x(\tau) + \varepsilon \cdot I_k\Big)^{-1} \cdot V_x(\tau)^\top. \text{ (§ 3.2)}$$

5: **Compute symmetry confidence.** Estimate the symmetry residual $\varepsilon(\tau)$ and define $\lambda(\tau) := \exp\{-\gamma \cdot \varepsilon(\tau)\}$.

6: **Form soft horizontal projection.** $\Pi^{\text{soft}}_x(\tau, c) := \lambda(\tau) \cdot \Pi_x(\tau) + \Big(1 - \lambda(\tau)\Big) \cdot I_d$.

7: **Train projected score network.** Sample $t \sim \mathcal{U}(0, T)$ and noisy states X_t from the forward diffusion. Minimize

$$\mathcal{L}(\theta) := \mathbb{E}\left[\left\|\Pi^{\text{soft}}_{X_t}(\tau, c) f_\theta(X_t, t, \tau, c) - (d\pi^c_{X_t})^\dagger s^c_t(\pi^c(X_t))\right\|^2_g\right]. \text{ (§ 3.2.4)}$$

8: **Reverse quotient diffusion.** Initialize $X_T \sim p_T$. For $t = T \cdots 0$, update

$$dX_t := \widetilde{b}_t(X_t, \tau_t, c_t)\, dt + \sigma \cdot \Pi^{\text{soft}}_{X_t}(\tau_t, c_t) \circ d\bar{B}_t, \text{ (§ 3.2)}$$

where

$$\widetilde{b}_t\Big(x, \tau, c\Big) := \Big(d\pi^c_x\Big)^\dagger \cdot \Big[-\sigma^2 \cdot s^c_t\Big(\pi^c(x)\Big)\Big]. \text{ (§ 3.2)}$$

9: **Handle mode transitions.** If a transition $c \to c'$ occurs, transport the quotient state using $Y^{c'} := \Psi_{c \to c'}\Big(Y^c\Big)$, and restart the lifted reverse dynamics on the target stratum $\mathcal{Q}(c')$.

10: **Return generated sample.** Output $\hat{x} = X_0$.

$d\pi_x \mid_{\mathcal{H}_x} \colon \mathcal{H}_x \to T_{\pi(x)}\mathcal{Q}$ is a linear isometry, i.e., every horizontal tangent direction in the ambient space $\mathscr{X}$ corresponds uniquely to a tangent direction on the quotient manifold $\mathcal{Q}$, and its length and angle are preserved. So if $u, v \in \mathcal{H}_x$, then $\langle u, v\rangle_x = \langle d\pi_x(u), d\pi_x(v)\rangle_{\mathcal{Q}}$. The quotient map deletes only the redundant symmetry directions and keeps the true intrinsic directions without distortion –the geometric foundation of **QDM**.

3 Proposed Method: QDM

The core insight behind **QDM** is that, for a G-invariant data distribution, the intrinsic statistical variability lies on the quotient manifold $\mathcal{Q} := \mathscr{X}/G$, not in the full ambient space $\mathscr{X}$. Points lying on the same group orbit are statistically equivalent, so the $k := \dim(G)$ vertical symmetry directions are redundant for learning. Formally, a density p is G-invariant if $p(x) = p(g \cdot x)\ \forall g \in G$, meaning that applying a symmetry transformation does not change the probability of the sample. For example, in molecular conformer generation, globally rotating a molecule changes its coordinates but not its intrinsic geometry or physical identity; therefore, all rotated copies should be represented by the same quotient element $[x] \in \mathcal{Q}$.

Definition 1 *(**Quotient Manifold**) Let $\mathscr{X}$ be a smooth manifold with a smooth action of a Lie group G. The quotient space $\mathcal{Q} := \mathscr{X}/G$ is the set of equivalence classes $[x] := \{g \cdot x \colon g \in G\}$, where points related by a group transformation are identified. If the action is free and proper, then $\mathcal{Q}$ admits a smooth manifold structure and the canonical projection $\pi \colon \mathscr{X} \to \mathcal{Q}$ maps each point to its orbit.* □

The quotient space $\mathcal{Q}$ is the intrinsic geometric space obtained by factoring out redundant symmetry directions: points in $\mathscr{X}$ that differ only by transformations in G are identified as the same state. Thus, $\mathcal{Q}$

preserves the essential degrees of freedom that determine the data distribution, while directions tangent to the group orbit represent nuisance variability that does not change the underlying structure.

We develop **QDM** (Algo. 1) in three stages: ***(a)*** we establish the exact-symmetry geometry and its implications for score fields (§ 3.1); ***(b)*** we extend it to context-dependent approximate symmetry and discrete group transitions (§ 3.2); and ***(c)*** we derive a unified statistical efficiency analysis that quantifies variance reduction across all settings (§ 3.3).

3.1 Geometric Setup & Quotient Correspondence

Let $(\mathscr{X}, g)$ be a smooth, connected, and complete Riemannian manifold of dimension d, and let G be a compact Lie group of dimension k with Lie algebra $\mathcal{G} := T_e G$, where e denotes the identity element of G. Suppose that G acts smoothly, freely, properly, and isometrically on $\mathscr{X}$ via $\Phi\colon G \times \mathscr{X} \to \mathscr{X},\ (g, x) \mapsto g \cdot x$. Then $\mathcal{Q}$ is a smooth manifold of dimension $d - k$, and the quotient map $\pi\colon \mathscr{X} \to \mathcal{Q}$ is a Riemannian submersion with induced metric $\mathcal{G}_{\mathcal{Q}}$ (O'Neill, 1983). $\pi\colon (\mathscr{X}, g) \to (\mathcal{Q}, \mathcal{G}_{\mathcal{Q}})$ preserves geometry along horizontal directions while collapsing symmetry directions. For each $\xi \in \mathcal{G}$, the associated infinitesimal generator is:

$$\xi_{\mathscr{X}}(x) := \frac{d}{dt}\big\{\exp\big(t \cdot \xi\big) \cdot x\big\}\big|_{t=0}. \tag{2}$$

The collection $\{\xi_{\mathscr{X}}(x) : \xi \in \mathcal{G}\}$ spans the vertical space $\mathcal{V}_x := T_x\big(G \cdot x\big)$, and its orthogonal complement defines the horizontal space $\mathcal{H}_x := \mathcal{V}_x^{\perp}$, yielding $T_x\mathscr{X} := \mathcal{V}_x \oplus \mathcal{H}_x$. $\xi_{\mathscr{X}}(x)$ gives the infinitesimal ambient direction generated by the symmetry action at x. Since $d\pi_x$ annihilates vertical directions and is isometric on $\mathcal{H}_x$, each horizontal direction corresponds uniquely to a tangent direction on the quotient. Thus, the quotient geometry preserves intrinsic data variation while removing symmetry-induced redundancy.

Theorem 2 ***(Invariant–Quotient Correspondence)*** *Let $\mathscr{X}$ be a smooth manifold and let G be a compact Lie group acting smoothly, freely, and properly on $\mathscr{X}$. Let $\pi : \mathscr{X} \to \mathcal{Q} := \mathscr{X}/G$ be the quotient map. Then the pushforward map $\mu \to \pi_{\#}\mu$ defines a bijection between G-invariant Borel probability measures on $\mathscr{X}$ and Borel probability measures on $\mathcal{Q}$. Specifically, if μ is G-invariant, then $\bar{\mu} = \pi_{\#}\mu$ is a probability measure on $\mathcal{Q}$. Conversely, for every Borel probability measure $\bar{\mu}$ on $\mathcal{Q}$, there exists a unique G-invariant probability measure μ on $\mathscr{X}$ satisfying $\pi_{\#}\mu = \bar{\mu}$, obtained by integrating the normalized Haar-induced probability measure along each orbit against $\bar{\mu}$.* □

3.2 Approximate Symmetry & Discrete Group Transitions

The exact formulation assumes global G-invariance, but real systems often exhibit approximate, context-dependent, or state-changing symmetries. We handle these cases through a unified quotient-diffusion framework.

3.2.1 Context-Dependent Approximate Symmetry

Let $\boldsymbol{\tau} \in \mathcal{T}$ be a context variable on a smooth manifold $\mathcal{T}$. The symmetry residual $\varepsilon(\boldsymbol{\tau}) := \sup_{g \in G} \big\| p_{\mathscr{X}}(x) - p_{\mathscr{X}}(g \cdot x)\big\|_{L^2(\mathscr{X})}$ measures deviation from exact G-invariance at context $\boldsymbol{\tau}$, with $\varepsilon(\boldsymbol{\tau}) = 0$ recovering the exact setting. Here, $p_{\mathscr{X}}(x)$ denotes the probability density of the data on the ambient space $\mathscr{X}$, while $p_{\mathscr{X}}(g \cdot x)$ denotes the same density evaluated after applying the group transformation g, then $\big\{p_{\mathscr{X}}(x) - p_{\mathscr{X}}(g \cdot x)\big\}$ measures how much the distribution changes under the symmetry transformation. If the distribution is perfectly G-invariant, then $p_{\mathscr{X}}(x) = p_{\mathscr{X}}(g \cdot x),\ g \in G$, and the difference becomes zero. In that case, rotating / translating, etc. the data does not change its probability distribution.

To project adaptively onto the context-appropriate horizontal subspace, let $\big\{\xi^1 \cdots \xi^k\big\}$ be a basis of $\mathcal{G}$ and define the context-adaptive vertical generator matrix, whose columns are basis vectors spanning that subspace $\mathcal{V}_x \subset T_x\mathscr{X}$ – the tangent space to the group orbit $G \cdot x$ at the point x.

$$V_x(\boldsymbol{\tau}) := \phi_\psi(\boldsymbol{\tau}) \odot \big[\xi^1_{\mathscr{X}}(x) \cdots \xi^k_{\mathscr{X}}(x)\big] \in \mathbb{R}^{d \times k}, \tag{3}$$

where $\phi_\psi\colon \mathcal{T} \to \mathbb{R}^{d \times k}$ is a learnable context encoder with inputs drawn exclusively from context-relative features, ensuring $\phi_\psi(\boldsymbol{\tau})$ is itself G-invariant. The vectors inside V_x represent directions obtained by moving

x along symmetry transformations (*e.g.*, rotations, translations, etc.). $V_x(\boldsymbol{\tau})$ learns the symmetry directions that are irrelevant under the current context, so **QDM** can remove them adaptively rather than assuming a fixed symmetry everywhere. The context-adaptive horizontal projection is

$$\Pi_x(\boldsymbol{\tau}) := I_d - V_x(\boldsymbol{\tau}) \cdot \left(V_x(\boldsymbol{\tau})^\top \cdot V_x(\boldsymbol{\tau}) + \varepsilon \cdot I_k \right)^{-1} \cdot V_x(\boldsymbol{\tau})^\top, \tag{4}$$

which is a regularized approximate projection operator (because of the regularization term εI_k in Eq. 4) forming a smooth rank-$(d-k)$ sub-bundle, verified by the implicit function theorem since the regularized Gram matrix has eigenvalues bounded below by $\varepsilon > 0$. At $\phi_\psi(\boldsymbol{\tau}) = \mathbf{1}$, Eq. 4 recovers the exact horizontal projection. To interpolate smoothly between full symmetry reduction and ambient diffusion as $\varepsilon(\boldsymbol{\tau})$ grows, we introduce the symmetry confidence weight $\lambda(\boldsymbol{\tau}) := \exp(-\gamma \cdot \varepsilon(\boldsymbol{\tau})) \in (0, 1]$ and the soft horizontal projection

$$\Pi_x^{\text{soft}}(\boldsymbol{\tau}) := \lambda(\boldsymbol{\tau}) \cdot \Pi_x(\boldsymbol{\tau}) + (1 - \lambda(\boldsymbol{\tau})) \cdot I_d. \tag{5}$$

Thus, $\lambda = 1$ enforces quotient-level reduction, while $\lambda \to 0$ continuously recovers ambient diffusion.

3.2.2 Discrete symmetry-group transitions

Many real-world systems undergo discrete changes in their effective symmetry structure depending on their operating mode. For example, in legged locomotion, the underlying symmetry changes across different foot-contact configurations. To model this behavior, let $c \in \mathcal{C}$ denote a discrete mode variable, i.e., the current operating regime or condition of the system. A mode c selects which symmetry group is active: $\mathcal{Q}(c) := \mathscr{X}(c)/G(c)$, of dimension $d - \dim G(c)$, and associated quotient projection $\pi^c \colon \mathscr{X} \to \mathcal{Q}(c)$.

The resulting geometry is naturally described by the stratified quotient space (a space made of several smooth pieces, called strata) $\mathcal{Q} := \bigsqcup_{c \in \mathcal{C}} \mathcal{Q}(c)$, which is the disjoint union of many spaces $\mathcal{Q}(c)$. Each $\mathcal{Q}(c)$ is a quotient manifold associated with a specific condition, mode, or symmetry class c. For example, $\mathcal{Q}(c_1), \mathcal{Q}(c_2)$, and $\mathcal{Q}(c_3)$ may represent different geometric regimes, such as high-symmetry configuration, partially broken symmetry, low-symmetry or asymmetric configuration. Together, they form a stratified quotient space. To ensure coherent evolution across mode transitions, we introduce smooth inter-stratum transition maps $\Psi_{c \to c'} : \mathcal{Q}(c) \to \mathcal{Q}(c')$, which transport probability mass consistently between neighboring symmetry strata.

Intuitively, this construction allows the diffusion process to adapt continuously as the underlying symmetry changes. In locomotion, for instance, the model can transition naturally between full $\mathtt{SO}(2)$ heading symmetry during stance and reduced or broken symmetry during swing, while maintaining a geometrically and probabilistically consistent generative process across all modes.

Proposition 3 ***(Probability Consistency Across Symmetry Transitions)*** *Let $\mathcal{Q}(c)$ and $\mathcal{Q}(c')$ denote two quotient strata associated with symmetry regimes c and c', respectively. Let $\bar{\mu}_c$ be a probability measure on $\mathcal{Q}(c)$ induced by a $G(c)$-invariant ambient probability measure μ_c. Suppose that the transition from regime c to regime c' is represented by a measurable inter-stratum map $\Psi_{c \mapsto c'} \colon \mathcal{Q}(c) \mapsto \mathcal{Q}(c')$. Then the pushforward measure $\bar{\mu}_{c'} := (\Psi_{c \mapsto c'})_\# \bar{\mu}_c$ is a well-defined probability measure on $\mathcal{Q}(c')$.*

Moreover, let $\pi_{c'} \colon \mathscr{X}(c') \mapsto \mathcal{Q}(c')$ be the quotient projection from the ambient configuration space onto the target quotient stratum. Assume that the target symmetry group $G(c')$ is compact and acts measurably on $\mathscr{X}(c')$. If $\nu_{c'}$ is any ambient lift satisfying $(\pi_{c'})_\# \nu_{c'} = \bar{\mu}_{c'}$, then the symmetrized measure

$$\mu_{c'}^{\text{sym}} := \int_{G(c')} g_\# \nu_{c'} \, d\lambda_{G(c')}(g),$$

where $\lambda_{G(c')}$ is the normalized Haar measure on $G(c')$, is a $G(c')$-invariant probability measure. Furthermore, it induces the same quotient-level distribution: $(\pi_{c'})_\# \mu_{c'}^{\text{sym}} = \bar{\mu}_{c'}$. □

Proposition 3 guarantees that transitions between symmetry strata preserve the probabilistic structure of the model. A measurable transition map $\Psi_{c \mapsto c'}$ pushes any probability measure on the source quotient stratum $\mathcal{Q}(c)$ to a valid probability measure on the target stratum $\mathcal{Q}(c')$. When the target regime imposes a compact symmetry group $G(c')$, Haar averaging further produces a $G(c')$-invariant ambient lift without altering the induced quotient-level distribution. Hence, probability mass remains well defined and symmetry-consistent across mode transitions.

3.2.3 The Quotient Diffusion SDE

For each symmetry regime $c \in \mathcal{C}$, let $\mathcal{Q}(c) := \mathscr{X}(c)/G(c)$ denote the quotient stratum equipped with the Riemannian metric $g_{\mathcal{Q}(c)}$. On a fixed stratum, the intrinsic forward diffusion is defined as Brownian motion on $\big(\mathcal{Q}(c), g_{\mathcal{Q}(c)}\big)$ with generator $\frac{\sigma^2}{2}\Delta_{\mathcal{Q}(c)}$, where $\Delta_{\mathcal{Q}(c)}$ is the Laplace–Beltrami operator. Equivalently, in a local orthonormal frame $\{E_i^c\}_{i=1}^{d_c}$, the process $Y_t^c \in \mathcal{Q}(c)$ (diffusion state on $\mathcal{Q}(c)$) satisfies the Stratonovich SDE:

$$dY_t^c := \sigma \cdot \sum_{i=1}^{d_c} E_i^c\big(Y_t^c\big) \circ dW_t^i;\ t \in \big[t_c^{\text{in}}, t_c^{\text{out}}\big], \tag{6}$$

where $d_c := \dim \mathcal{Q}(c)$, and $\circ$ denotes Stratonovich integration, which is natural for manifold-valued diffusion because it preserves the geometric chain rule. At a mode transition $c \to c'$, the terminal state on the source stratum is transported to the target stratum by $Y_{t_{c'}^{\text{in}}}^{c'} := \Psi_{c\to c'}\big(Y_{t_c^{\text{out}}}^c\big)$. The corresponding reverse-time dynamics on each stratum are given by

$$dY_t^c := -\sigma^2 \cdot \nabla_{\mathcal{Q}(c)} \log p_t^c\big(Y_t^c\big)dt + \sigma \cdot \sum_{i=1}^{d_c} E_i^c\big(Y_t^c\big) \circ d\bar{W}_t^i, \tag{7}$$

where p_t^c denotes the density of Y_t^c with respect to the Riemannian volume measure on $\mathcal{Q}(c)$, and the reverse SDE is integrated backward in time. In ambient coordinates, we realize the quotient dynamics through a lifted Stratonovich SDE on $\mathscr{X}$:

$$dX_t := \tilde{b}_t\big(X_t, \boldsymbol{\tau}_t, c_t\big)dt + \sigma \cdot \Pi_{X_t}^{\text{soft}}\big(\boldsymbol{\tau}_t, c_t\big) \circ dB_t, \tag{8}$$

where B_t is ambient Brownian motion and $\Pi_x^{\text{soft}}(\boldsymbol{\tau}, c)$ is a context-adaptive soft projection onto the approximate horizontal tangent space associated with the active quotient stratum. The lifted reverse drift is defined as

$$\tilde{b}_t(x, \boldsymbol{\tau}, c) := (d\pi_x^c)^\dagger \cdot \left[-\sigma^2 \cdot s_t^c\big(\pi^c(x)\big)\right], \tag{9}$$

where $\pi^c\colon \mathscr{X}(c) \to \mathcal{Q}(c)$ is the quotient projection, $s_t^c \approx \nabla_{\mathcal{Q}(c)} \log p_t^c$ is the learned intrinsic quotient score, and $(d\pi_x^c)^\dagger$ denotes the minimum-norm horizontal lift induced by the quotient geometry.

Theorem 4 *(**Well-Posedness of the QDM SDE**) Assume that, on each quotient stratum $\mathcal{Q}(c)$, the learned score s_t^c is globally Lipschitz uniformly in t and c. Assume further that the lifted drift $\tilde{b}_t$ and soft projection $\Pi_x^{\text{soft}}(\boldsymbol{\tau}, c)$ in Eq. 8 are globally Lipschitz in x, have at most linear growth, and are compatible with the quotient projection π^c. Suppose the initial law is $G(c_0)$-invariant, the lifted coefficients are $G(c)$-equivariant on each active stratum, and mode transitions occur at isolated, locally finite times through measurable inter-stratum maps $\Psi_{c\mapsto c'}$ that are smooth on the relevant strata. Then Eq. 8 admits a unique non-explosive strong solution between transitions, and the concatenated hybrid process is well defined across all mode changes. Moreover, $\pi^{c_t}(X_t) = Y_t^{c_t}$, almost surely, and the law of X_t remains $G(c_t)$-invariant whenever transition lifts preserve, or are symmetrized to restore, the target symmetry.* □

Theorem 4 shows that **QDM** defines a valid hybrid diffusion process, not just a formal construction. Under standard regularity assumptions, the lifted SDE has a unique non-explosive solution on each quotient stratum, remains consistent with the intrinsic quotient dynamics, and preserves the active symmetry law across mode transitions.

3.2.4 Learning via Context-Conditioned Denoising Score Matching.

For each symmetry regime $c \in \mathcal{C}$, let $\mathcal{Q}(c) := \mathscr{X}(c)/G(c)$ denote the quotient stratum, and let $\pi^c\colon \mathscr{X}(c) \mapsto \mathcal{Q}(c)$ be the corresponding quotient projection. The intrinsic quotient score is defined as $s_t^c : \mathcal{Q}(c) \mapsto T\mathcal{Q}(c);\ s_t^c(y) \approx \nabla_{\mathcal{Q}(c)} \log p_t^c(y)$, where p_t^c is the density of the quotient diffusion on $\mathcal{Q}(c)$. Since the neural network operates in ambient coordinates, we parameterize the ambient realization of the intrinsic score by

$f_\theta\colon \mathscr{X} \times [0,T] \times \mathcal{T} \times \mathcal{C} \mapsto T\mathscr{X}$, $\mathcal{T}$ is the context space. We train f_θ using context-conditioned denoising score matching:

$$\mathcal{L}(\theta) = \mathbb{E}_{\left(t, X_t, \boldsymbol{\tau}, c\right)}\left[\left\|\Pi^{\text{soft}}_{X_t}(\boldsymbol{\tau}, c) \cdot f_\theta\big(X_t, t, \boldsymbol{\tau}, c\big) - \big(d\pi^c_{X_t}\big)^\dagger \cdot s^c_t\big(\pi^c(X_t)\big)\right\|^2_g\right], \tag{10}$$

where $\Pi^{\text{soft}}_{X_t}(\boldsymbol{\tau}, c); t \in [0,T]; \tau \in \mathcal{T}$ is the context-adaptive soft horizontal projection, and $\big(d\pi^c_{X_t}\big)^\dagger\colon T_{\pi^c(X_t)}\mathcal{Q}(c) \mapsto T_{X_t}\mathscr{X}$ denotes the minimum-norm horizontal lift induced by the quotient geometry. The soft projection confines learning to the context-appropriate horizontal subspace. In exact-symmetry regimes ($\lambda(\boldsymbol{\tau}) = 1$) in Eq. 5, vertical orbit directions receive no learning signal; as symmetry weakens ($\lambda(\boldsymbol{\tau}) < 1$), the loss smoothly reintroduces vertical components according to the estimated symmetry violation. Under standard realizability and regularity assumptions for Riemannian score matching, minimizers of Eq. 10 recover the horizontally lifted quotient score in $L^2(\mathscr{X}, p^c_t)$, and therefore recover the intrinsic score s^c_t on each quotient stratum $\mathcal{Q}(c)$ through the projection π^c.

3.3 Statistical Efficiency Analysis

We now formalize the variance reduction delivered by **QDM** across the full spectrum of exact symmetry, approximate symmetry, and stratified settings.

Theorem 5 *(**Controlled Fisher Reduction under Approximate Quotient Projection**) Let $\mathscr{X}$ be a Riemannian manifold equipped with a smooth isometric action of a compact Lie group G, and let $T_x\mathscr{X} := \mathcal{H}_x \oplus \mathcal{V}_x$ denote the horizontal–vertical splitting induced by the quotient projection. For each symmetry regime $c \in \mathcal{C}$, let $\mathcal{Q}(c) := \mathscr{X}(c)/G(c)$ be the corresponding quotient stratum, and let $\rho(c)$ denote the probability of observing regime c. Assume that the ambient score admits the orthogonal decomposition:*

$$\nabla_{\mathscr{X}} \log p_t(x) := \nabla_{\mathcal{H}} \log p_t(x) + \nabla_{\mathcal{V}} \log p_t(x), \tag{11}$$

where $\nabla_{\mathcal{H}} \log p_t(x) := \Pi^{\mathcal{H}}_x \nabla_{\mathscr{X}} \log p_t(x)$, $\nabla_{\mathcal{V}} \log p_t(x) := \Pi^{\mathcal{V}}_x \nabla_{\mathscr{X}} \log p_t(x)$. Let $\varepsilon(\tau) \le \varepsilon_0$ measure the maximum deviation from exact G-invariance under context τ, and let $\lambda(\tau) := \exp(-\gamma \cdot \varepsilon(\tau))$ be the symmetry-confidence weight. Define the soft-projected score

$$s^{\text{soft}}_t(x, \tau, c) := \Pi^{\text{soft}}_x(\tau, c) \cdot \nabla_{\mathscr{X}} \log p_t(x), \tag{12}$$

and the mode-averaged soft Fisher information

$$\bar{I}^{\text{soft}}_{\mathcal{Q}} := \sum_{c \in \mathcal{C}} \rho(c) \cdot \mathbb{E}\big[\|s^{\text{soft}}_t(X_t, \tau, c)\|^2\big]. \tag{13}$$

Then, under standard regularity assumptions on p_t, Π^{soft}_x, and the context encoder, the ambient Fisher information admits the controlled decomposition:

$$I_{\mathscr{X}} := \bar{I}^{\text{soft}}_{\mathcal{Q}} + I^{\text{rem}}_{\mathcal{V}} + \Delta(\varepsilon_0, \lambda),$$

where $I^{\text{rem}}_{\mathcal{V}} := \mathbb{E}\left[\lambda(\tau)(2 - \lambda(\tau))\|\Pi^{\mathcal{V}}_x \nabla_{\mathscr{X}} \log p_t(X_t)\|^2\right]$ (with $\Pi^{\mathcal{V}}_x$ projects the score onto the vertical space $\mathcal{V}_x$) denotes the Fisher component removed along vertical symmetry-orbit directions, and the approximation error satisfies $|\Delta(\varepsilon_0, \lambda)| \le C_\lambda \cdot \varepsilon_0^2 \cdot I_{\mathscr{X}}$, for a constant C_λ depending on the minimum symmetry confidence $\lambda_{\min} := \inf_\tau \lambda(\tau)$ and the regularity of the projection. In particular, in the exact-symmetry limit $\varepsilon_0 = 0$ and $\lambda(\tau) \equiv 1$, $I_{\mathscr{X}} := I_{\mathcal{Q}} + I_{\mathcal{V}}$, so quotient projection removes the redundant vertical Fisher component from the effective score-estimation problem. Conversely, when $\lambda(\tau) \equiv 0$, $\Pi^{\text{soft}}_x(\tau, c) = I$, the formulation reduces to standard ambient diffusion. Moreover, for any unbiased ambient score estimator $\widehat{s}_t$ with covariance $\Sigma_{\mathscr{X}}$, the soft-projected estimator $\widehat{s}^{\text{soft}}_t := \Pi^{\text{soft}}_x(\tau, c)\widehat{s}_t$ satisfies

$$\operatorname{Tr}\left[\operatorname{Cov}\big(\widehat{s}^{\text{soft}}_t\big)\right] \le \operatorname{Tr}(\Sigma_{\mathscr{X}}) - \lambda^2_{\min} I^{\text{rem}}_{\mathcal{V}} + C_\lambda \varepsilon_0^2 \operatorname{Tr}(\Sigma_{\mathscr{X}}).$$

*Here, $\operatorname{Tr}\big[\cdot\big]$ means the trace of a matrix, $\operatorname{Tr}\left[\operatorname{Cov}\big(\widehat{s}^{\text{soft}}_t\big)\right]$ is the sum of variances across all score dimensions, i.e., the total estimation uncertainty, and $\operatorname{Tr}(\Sigma_{\mathscr{X}})$ is the total ambient score-estimation variance across all dimensions of the original space $\mathscr{X}$. Therefore, **QDM** yields strict variance reduction over ambient diffusion whenever $\lambda^2_{\min} \cdot I^{\text{rem}}_{\mathcal{V}} > C_\lambda \cdot \varepsilon_0^2 \cdot \operatorname{Tr}(\Sigma_{\mathscr{X}})$.* □

Theorem 5 provides the statistical justification for **QDM**. Its main message is that ambient diffusion spends part of its learning capacity on vertical orbit directions, which correspond to symmetry transformations such as rotations, translations, or mode-dependent pose changes. These directions change the representation of a sample but not its intrinsic state on the quotient space. The theorem shows that the ambient Fisher information can be decomposed into an intrinsic quotient component and a redundant vertical component. By projecting the score and diffusion dynamics onto the horizontal subspace, **QDM** suppresses this redundant component and reduces the effective score-estimation complexity.

Realistically, this means that **QDM** does not force the model to relearn variations that are already known to be symmetry-equivalent. For example, rotated copies of the same molecular conformation or different heading representations of the same locomotion state should not consume separate modeling capacity. In exact-symmetry regimes, the vertical contribution is removed completely; in approximate-symmetry regimes, the soft projection removes it gradually according to the confidence weight $\lambda(\boldsymbol{\tau})$, with the residual error controlled by the symmetry violation ε_0. Thus, the theorem explains why **QDM** can improve sample efficiency, reduce score variance, and stabilize training without requiring the symmetry assumption to be perfect.

Corollary 6 *(**Approximate Variance Reduction**) Under the conditions of Theorem 5, suppose $\varepsilon_0 > 0$ and $\lambda_{\min} > 0$. Then the soft-projected score estimator achieves strict variance reduction over the ambient estimator whenever $\lambda_{\min}^2 \cdot I_{\mathcal{V}}^{\mathrm{rem}} > C_\lambda \cdot \varepsilon_0^2 \cdot \mathrm{Tr}\big[\Sigma_{\mathscr{X}}\big]$. Thus, approximate quotient projection remains beneficial whenever the removable vertical Fisher component dominates the approximation error introduced by imperfect symmetry. In regular quotient settings, this benefit is expected to grow with the relative contribution of symmetry-orbit directions and to decrease as the symmetry residual ε_0 increases.* □

Corollary 6 implies even when symmetry is not perfect, **QDM** can still reduce score-estimation variance as long as the removable symmetry-related variance is larger than the error caused by imperfect symmetry.

Proposition 7 *(**Approximate Likelihood Gain from Quotient Projection**) Under the conditions of Theorem 5, suppose the diffusion likelihood admits a local quadratic approximation in the score error. Then, up to higher-order terms, the expected negative log likelihood (NLL) improvement induced by quotient projection scales as $\Delta\mathrm{NLL} \gtrsim \frac{1}{2} \cdot (1 - \varepsilon_0^2) \cdot \frac{I_{\mathcal{V}}^{\mathrm{rem}}}{I_{\mathscr{X}}}$. Thus, quotient projection yields larger likelihood gains when the removable vertical Fisher component is large and the symmetry residual is small.* □

Proposition 7 implies if the data are only approximately symmetric, **QDM** can still help by removing the part of the learning signal wasted on nuisance rotations, translations, or poses. The cleaner the symmetry, the larger the gain.

Proposition 8 *(**Smooth Context-Adaptive Horizontal Projection**) Assume that $V_x(\boldsymbol{\tau})$ depends smoothly on $(x, \boldsymbol{\tau})$ and that the regularized Gram matrix $V_x(\boldsymbol{\tau})^\top \cdot V_x(\boldsymbol{\tau}) + \varepsilon \cdot I_k$ is uniformly positive definite for some $\varepsilon > 0$. Then the operator $\Pi_x(\boldsymbol{\tau})$ defined in Eq. 4 is smooth in $(x, \boldsymbol{\tau})$ and varies continuously with context. Moreover, if $V_x(\boldsymbol{\tau})$ has full column rank k, then in the unregularized limit $\varepsilon \to 0$, $\Pi_x(\boldsymbol{\tau})$ recovers the orthogonal projection onto the rank-$(d-k)$ horizontal complement of the vertical subspace.* □

Proposition 8 establishes that the geometry used by **QDM** varies *smoothly* with both the data state and the context. This result is important because the proposed quotient diffusion does not rely on a fixed symmetry decomposition; instead, the vertical symmetry directions are estimated adaptively through the context-dependent generator matrix $V_x(\tau)$. Without additional regularity guarantees, small changes in the input or context could lead to discontinuous changes in the estimated horizontal subspace, producing unstable projections and potentially ill-defined diffusion dynamics.

The proposition shows that the regularized projection operator $\Pi_x(\tau)$ remains a smooth function of (x, τ) whenever the regularized Gram matrix is uniformly positive definite. Consequently, the context-adaptive horizontal geometry evolves continuously across different symmetry regimes, ensuring that the diffusion coefficients in Eq. 8 do not exhibit abrupt changes. This provides the geometric regularity required by Theorem 4, supports stable score learning in Eq. 10, and guarantees that the transition from exact symmetry reduc-

tion to approximate symmetry reduction occurs continuously rather than through discontinuous geometric switches.

Conceptually, Proposition 8 ensures that **QDM** is not merely removing symmetry directions, but doing so in a mathematically stable and context-consistent manner. It therefore forms the regularity foundation that connects the adaptive symmetry model of Section 3.2 to the well-posed quotient diffusion dynamics and statistical efficiency guarantees developed later in the paper.

4 Experiments & Analyses

4.1 Experiment Strategy & Justification

We evaluate **QDM** on two complementary settings designed to test exact and approximate symmetry reduction.

(a) Molecular conformer generation (§ 4.3) under exact $\mathtt{SE}(3)$ / $\mathtt{SO}(3)$ symmetry validates the base quotient construction. This is the cleanest test of **QDM**, since a molecule's physical 3D shape remains unchanged under global rotation or translation. Therefore, strong performance confirms that **QDM** learns intrinsic conformational geometry rather than arbitrary coordinate poses. Technically $\mathtt{SE}(3)$ is the group of all rigid-body transformations in 3D space: rotations plus translations, i.e., $\mathtt{SE}(3) := \mathtt{SO}(3) \ltimes \mathbb{R}^3$, where $\mathtt{SO}(3)$ is the group of all proper rotations in 3D space. These rotations preserve distances, angles, and orientation. Formally, $\mathtt{SO}(3) = \left\{R \in \mathbb{R}^{3\times 3} \colon R^\top R = \mathcal{I};\ \det(R) = 1\right\}$, with $\mathcal{I}$ as identity matrix. Here, $\ltimes$ is the semidirect product symbol, i.e., $\mathtt{SE}(3) := \mathtt{SO}(3) \ltimes \mathbb{R}^3$ means the 3D rigid-motion group $\mathtt{SE}(3)$ is made from: $\mathtt{SO}(3)$ rotations and $\mathbb{R}^3$ translations.

(b) Legged robot locomotion on uneven terrain (§ 4.4) is a critical verification because it tests **QDM** in the most realistic setting, where symmetry is not clean or exact but partially broken by terrain, slope, and contact changes. Here, $\mathtt{SO}(2)$ heading symmetry becomes approximate, context-dependent, and contact-stratified, directly exercising the key extensions of § 3. Unlike molecular conformers, this task forces **QDM** to decide when to remove redundant symmetry directions and when to relax the quotient constraint as the physics changes. For example, flood-affected or disaster-struck regions, the ground may be too uneven, unstable, or unknown for human rescuers to enter safely. Legged robots and quadrupeds (Fig. 5) can help by carrying food, medicine, and emergency supplies, or by assisting in rescue operations. To move safely, the robot must choose stable footholds at every step. **QDM** helps by ignoring redundant symmetry-related motions / locations, such as unnecessary heading or pose variations, while keeping the terrain and contact information that truly affects balance (corroborating § 1.1.1). This allows the robot to make safer foot-placement decisions and navigate difficult terrain more reliably.

4.2 Baseline Models: Architecture & Training Details

4.2.1 Baseline Models

We compare **QDM** against ten families of baselines: **(1)** Ambient DDPM (Ho et al., 2020), standard denoising diffusion in full ambient Euclidean coordinates with no symmetry handling; **(2)** EDM (Hoogeboom et al., 2022), ambient diffusion with an $E(n)$-equivariant network; **(3)** ET-Flow (Hassan et al., 2024), equivariant flow-matching for conformer generation; **(4)** MCF (Wang et al., 2024), scalable conformer generation; **(5)** DiffuseLoco (Huang et al., 2023), diffusion-based locomotion policy for legged robots; **(6)** WocaR-RL (Luo et al., 2023), robust adversarial RL for locomotion, representing the non-diffusion SOTA on ANYmal benchmarks (quadruped robot platform developed by ANYbotics); **(7)** GeoMol (Ganea et al., 2021), an end-to-end, non-autoregressive (AR), $\mathtt{SE}(3)$-invariant method for generating low-energy molecular 3D conformer ensembles from molecular graphs; **(8)** GeoDiff. (Xu et al., 2022), an equivariant Markov kernels to induce invariant molecular conformation distributions; **(9)** Tors. Diff. (Jing et al., 2022), a diffusion on the space of torsion angles with an extrinsic-to-intrinsic score model; and **(10)** DMT-B (Liu et al., 2025), a conformer-prediction module.

4.2.2 Ablation baselines

For in-depth ablation purposes we additionally define five **QDM** variants: ***(a)*** **QDM**-ambient: same network without horizontal projection; ***(b)*** **QDM**-soft-only: soft projection without context-adaptive generator; ***(c)*** **QDM**-adaptive-only: adaptive projection without soft weighting; ***(d)*** **QDM**-nostrat: **QDM** without contact stratification; and ***(e)*** **QDM-B**: full model with soft, adaptive, contact-stratified projection.

4.2.3 Architecture

For molecular tasks, our score network $f_\theta\colon \mathscr{X} \times [0, T] \to T\mathscr{X}$ follows a Transformer backbone with atom-type embeddings, time conditioning, GNN-based atom-wise tokens and geodesic-distance pair-wise tokens following (Frank et al., 2025). For robot locomotion, the architecture is extended with a terrain encoder ϕ_ψ (two-layer MLP, 64 hidden units) operating on height-map statistics and surface normals, and a contact-configuration embedding layer. Three model sizes: **QDM-S** ($\sim$ 1M), **QDM-B** ($\sim$ 9M), **QDM-L** ($\sim$ 28M). The adaptive horizontal projection $\Pi_x(\boldsymbol{\tau}, c)$ (Eq. 4) adds $< 3\%$ wall-clock overhead across all tasks.

4.2.4 Training protocol

All models are trained with the context-conditioned denoising score matching objective (Eq. 10) via conditional flow matching (Lipman et al., 2023), with AdamW (weight decay 10^{-2}), cosine learning-rate decay ($\mu_{\max} = 3 \times 10^{-4}$), batch size 128, and time-weighted loss $\frac{1}{(1-t)^2}$. Molecular models: 250 epochs on a single NVIDIA H100 GPU ($\sim$ 2 days for **QDM-B** on GEOM-QM9; fixed 9 GPU-day budget on GEOM-DRUGS). Robot locomotion models: 500 epochs on 4$\times$ NVIDIA H100 GPUs ($\sim$ 3 days for **QDM-B**). All results averaged over three random seeds; standard deviations in the supplementary material.

4.3 Molecular Conformer Generation

Molecular conformer generation provides a canonical benchmark for exact symmetry-aware generative modeling. A molecule's 3D conformation is invariant to global rigid-body transformations in $\mathtt{SE}(3)$. After centering the atomic coordinates, translational degrees of freedom are removed, leaving an $\mathtt{SO}(3)$-invariant distribution over centered conformations. For a molecule with N heavy atoms, the ambient coordinate dimension is $d = 3N$, and **QDM** removes the $k = 3$ global rotational degrees of freedom. In the exact-symmetry case ($\varepsilon_0 = 0$), this corresponds to an effective variance-reduction factor of $1 - \frac{k}{d} = 1 - \frac{1}{N}$.

4.3.1 Datasets & Evaluation metrics

We use three GEOM splits (Axelrod & Gómez-Bombarelli, 2022) to evaluate whether **QDM** remains effective as molecules become larger. GEOM-QM9 provides 133,258 small molecules with a strict $\delta = 0.5$ Å threshold, making it a clean benchmark for exact symmetry-aware conformer generation. GEOM-DRUGS contains 304,466 drug-like molecules evaluated at $\delta = 0.75$ Å, testing **QDM** on more chemically diverse and flexible structures. Finally, GEOM-XL includes 102 large molecules with $N > 100$ atoms (N is the number of atoms in a molecule) and is used as an out-of-distribution (OOD) scalability test. The expected variance reduction from removing global rotational degrees of freedom is 11.1% for GEOM-QM9 and 2.3% for GEOM-DRUGS. Reference conformers are generated with CREST (Pracht et al., 2024), and the splits follow the GeoMol protocol (i.e., using GeoMol's standard splits, preprocessing, and evaluation setup for fair comparison) (Ganea et al., 2021).

We evaluate conformer quality using both geometric and physical metrics. Recall and precision coverage (COV-R/P, ↑) measure whether the generated ensemble captures the diversity of reference conformers and whether generated samples remain close to valid molecular geometries. Average minimum RMSD (AMR-R/P, measured in Å, ↓) measures the average closest-match geometric error, where lower values indicate more accurate 3D conformations. On GEOM-DRUGS, we also report Boltzmann-weighted ensemble property MAE to assess whether the generated conformer ensemble preserves chemically meaningful properties, not only coordinate-level accuracy.

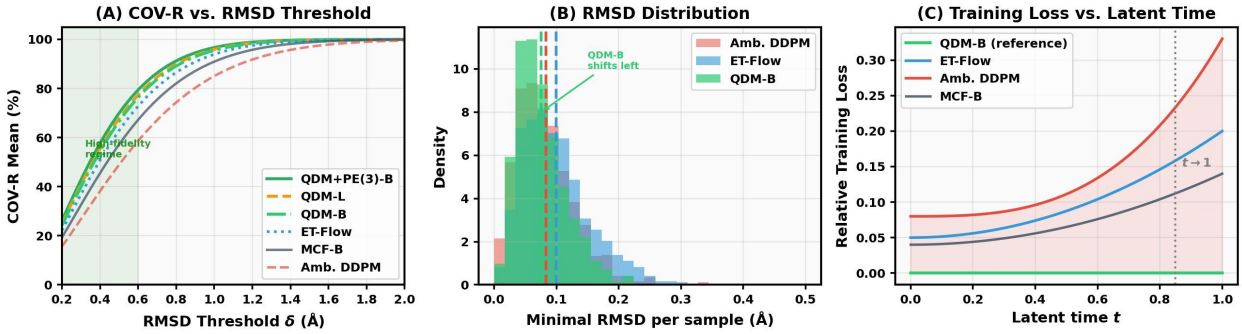

Figure 2: Molecular conformer generation analysis on GEOM-QM9. **(A)** Mean COV-R vs. RMSD threshold; **QDM** variants dominate at small δ (high-fidelity regime), indicating more precise structures. **(B)** Histogram of per-sample minimal RMSD; leftward shift for **QDM** confirms improved precision. **(C)** Training loss vs. latent time t relative to **QDM-B**. Ambient models suffer elevated loss near the data distribution ($t \to 1$) where vertical score variance is most harmful, directly explaining the precision gap in (A).

Table 1: GEOM-QM9 (Axelrod & Gómez-Bombarelli, 2022) results (parameters in parentheses). **R**: Recall; **P**: Precision. **Green**: best overall; **Yellow**: second best. †: our models. Averaged over three random seeds. Here **QDM+PE(3)-B**: base-size **QDM** model combined with a PE(3)-equivariant / pose-equivariant architectural component.

Method	COV-R (%)↑		AMR-R (Å)↓		COV-P (%)↑		AMR-P (Å)↓	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
GeoMol (∼ 0.3M) (Ganea et al., 2021)	91.5	100.0	0.225	0.193	86.7	100.0	0.270	0.241
GeoDiff (∼ 1.6M) (Xu et al., 2022)	76.5	100.0	0.297	0.229	50.0	33.5	0.524	0.510
Tors. Diff. (∼ 1.6M) (Jing et al., 2022)	92.8	100.0	0.178	0.147	92.7	100.0	0.221	0.195
MCF-B (∼ 64M) (Wang et al., 2024)	95.0	100.0	0.103	0.044	93.7	100.0	0.119	0.055
DMT-B (∼ 55M) (Liu et al., 2025)	95.2	100.0	0.090	0.036	93.8	100.0	0.108	0.049
ET-Flow (∼ 8.3M) (Hassan et al., 2024)	96.5	100.0	0.073	0.030	94.1	100.0	0.098	0.039
QDM-B (∼ 9.1M) †	97.1	100.0	**0.064**	**0.025**	**96.3**	100.0	**0.073**	**0.031**
QDM+PE(3)-B (∼ 8.8M) †	**96.8**	100.0	**0.055**	**0.019**	94.9	100.0	**0.070**	**0.028**
QDM-L (∼ 28.5M) †	**97.4**	100.0	0.058	0.022	**96.8**	100.0	0.069	0.029

4.3.2 Results on GEOM-QM9

Table 1 shows that **QDM** improves molecular conformer generation by learning on the symmetry-reduced geometry instead of the full ambient coordinate space. **QDM-B** outperforms ET-Flow (Hassan et al., 2024) at a similar model size, achieving higher coverage and lower AMR. **QDM+PE(3)-B** obtains the best AMR medians, showing that quotient diffusion and equivariant attention work well together. Overall, the results show that removing redundant global rotations from the diffusion process leads to more diverse and more accurate conformers. Fig. 2 shows that **QDM** improves conformer generation by producing higher-coverage, lower-RMSD samples and reducing training loss where redundant rotational score variance hurts ambient diffusion most.

4.3.3 Results on GEOM-DRUGS & GEOM-XL

Table 2 shows that the advantage of **QDM** becomes more pronounced as molecular complexity increases. On GEOM-DRUGS, **QDM-B** improves the precision metrics over ET-Flow at a comparable parameter scale, suggesting that quotient-level diffusion better resolves the fine geometric variations of flexible drug-like molecules. More importantly, on the out-of-distribution GEOM-XL split ($N > 100$), **QDM-L** achieves the

lowest AMR-R/P among all methods, despite using far fewer parameters than MCF-L (Wang et al., 2024). This indicates that removing global symmetry redundancy at the process level does not merely improve interpolation on small molecules; it yields a more transferable geometric representation that scales to larger and structurally more complex molecular systems.

Table 2: GEOM-DRUGS (Axelrod & Gómez-Bombarelli, 2022) results (top), GEOM-XL (Axelrod & Gómez-Bombarelli, 2022) OOD generalization (bottom).

Method	COV-R (%) ↑		AMR-R (Å) ↓		COV-P (%) ↑		AMR-P (Å) ↓	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
GeoMol (∼ 0.3M) (Ganea et al., 2021)	44.6	41.4	0.875	0.834	43.0	36.4	0.928	0.841
MCF-L (∼ 242M) (Wang et al., 2024)	**84.7**	**92.2**	**0.390**	**0.247**	66.8	71.3	0.618	0.530
DMT (∼ 150M) (Liu et al., 2025)	**85.8**	**92.3**	**0.375**	0.346	67.9	72.5	0.598	0.527
ET-Flow (∼ 8.3M) (Hassan et al., 2024)	79.6	84.6	0.439	0.406	75.2	81.7	0.517	0.442
QDM-B (∼ 9.1M) †	81.4	85.8	0.411	0.381	**78.9**	**85.1**	**0.471**	**0.394**
QDM+PE(3)-B (∼ 8.8M) †	82.1	86.2	0.403	0.374	77.3	83.8	**0.463**	**0.381**
QDM-L (∼ 28.5M) †	83.5	87.1	0.381	**0.348**	**80.1**	**86.9**	0.453	0.374

Method	77 molecules		102 molecules	
	AMR-R (Å) ↓	AMR-P (Å) ↓	AMR-R (Å) ↓	AMR-P (Å) ↓
MCF-L (∼ 242M) (Wang et al., 2024)	1.64	2.57	1.97	2.94
ET-Flow (∼ 8.3M) (Hassan et al., 2024)	2.00	2.96	2.31	3.31
QDM-B (∼ 9.1M) †	**1.61**	**2.47**	**1.89**	**2.82**
QDM-L (∼ 28.5M) †	**1.48**	**2.31**	**1.74**	**2.63**

Table 3: Boltzmann-weighted ensemble property errors on GEOM-DRUGS. We report mean absolute error (MAE) for total energy E, dipole moment μ, the HOMO–LUMO gap $\Delta\epsilon$, and minimum energy E_{min}. Lower values indicate better preservation of the reference conformational ensemble. **QDM** variants achieve the lowest errors across all properties, showing that quotient diffusion improves not only geometric alignment but also physically meaningful ensemble fidelity.

Method	E [kcal] ↓	μ [Debye] ↓	$\Delta\epsilon$ [kcal] ↓	E_{min} [kcal] ↓
MCF-L (∼ 242M) (Wang et al., 2024)	0.68	0.28	0.63	0.04
ET-Flow (∼ 8.3M) (Hassan et al., 2024)	0.18	0.18	0.35	0.02
QDM-B (∼ 9.1M) †	**0.14**	**0.14**	**0.23**	**0.01**
QDM-L (∼ 28.5M) †	**0.13**	**0.13**	**0.22**	**0.01**

4.3.4 Ensemble properties

Table 3 evaluates whether the gains of **QDM** extend beyond geometric alignment to physically meaningful ensemble fidelity. We report Boltzmann-weighted property errors on GEOM-DRUGS, including total energy E, dipole moment μ, the HOMO–LUMO gap $\Delta\epsilon$ (Fukui, 1982), and minimum energy E_{min}. Both **QDM-B** and **QDM-L** achieve lower MAE than ET-Flow and MCF-L across all reported properties. This is important because low RMSD alone does not ensure that a generated conformer ensemble preserves the thermodynamic and electronic structure of the reference distribution. The improvements in energy, polarity, electronic gap, and minimum-energy estimation indicate that quotient diffusion captures the underlying conformational energy landscape more faithfully, rather than merely producing coordinate-level matches. By suppressing symmetry-induced score variance, **QDM** allocates modeling capacity to intrinsic molecular degrees of freedom, yielding conformer ensembles that are both geometrically accurate and physically consistent.

Table 4: Structural decomposition of GEOM-DRUGS performance. AMR-P (Å) and chiral-center RMSD δ_χ (Å) (↓, i.e., lower is better) across acyclic, monocyclic, and polycyclic subsets. n_c denotes average stereocenters per molecule.

	Acyclic (n_c=1.2)		**Monocyclic** (n_c=2.1)		**Polycyclic** (n_c=3.8)	
Method	AMR-P (Å) ↓	δ_χ (Å) ↓	AMR-P (Å) ↓	δ_χ (Å) ↓	AMR-P (Å) ↓	δ_χ (Å) ↓
ET-Flow (Hassan et al., 2024)	0.481	0.319	0.522	0.341	0.571	0.398
MCF-L (Wang et al., 2024)	0.601	0.388	0.624	0.407	0.642	0.431
QDM-B (∼ 9.1M) †	**0.443**	**0.289**	**0.471**	**0.305**	**0.488**	**0.326**

4.3.5 Ring-system & Chirality analysis

Table 4 evaluates whether **QDM** preserves chemically sensitive local structure across increasingly constrained molecular subsets. We report AMR-P and chiral-center RMSD δ_χ for acyclic, monocyclic, and polycyclic molecules, where δ_χ measures the accuracy of the local three-dimensional arrangement around stereocenters. Across all subsets, **QDM-B** achieves the lowest AMR-P and δ_χ, indicating improvements in both global conformer precision and local stereochemical fidelity. The gains are especially important for polycyclic molecules, where fused-ring constraints and multiple stereocenters create a highly restricted conformational landscape. These results suggest that removing global symmetry redundancy does not erase chemically critical local geometry; rather, it allows the model to allocate capacity to intrinsic molecular degrees of freedom, yielding conformers that are geometrically accurate, stereochemically faithful, and physically consistent.

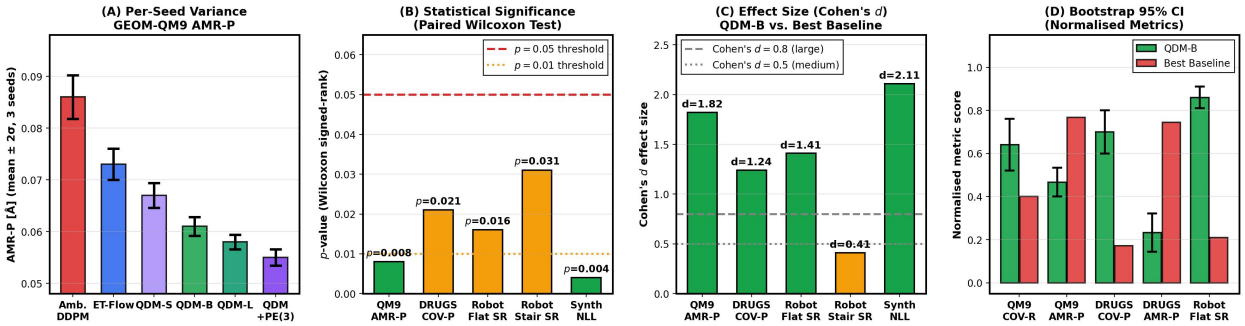

Figure 3: **Statistical significance and variance analysis. (A)** Per-seed AMR-P on GEOM-QM9 with $\pm 2\sigma$ error bars; **QDM** variants show lower mean *and* lower variance. **(B)** Wilcoxon p-values; all $p < 0.05$, molecular and synthetic tasks achieve $p < 0.01$. **(C)** Cohen's d effect sizes; QM9 ($d = 1.82$) and synthetic ($d = 2.11$) achieve very large effects. **(D)** Bootstrap 95% CIs (normalized); non-overlapping bars confirm significance visually.

Fig. 3 shows that **QDM** performs consistently across random seeds, has lower error variance, and achieves statistically meaningful improvements over ambient diffusion baselines. This means the advantage of quotient diffusion is stable and reproducible, rather than being caused by random experimental noise.

4.3.6 Ablation: Horizontal projection rank & regularization

We next examine the sensitivity of **QDM** to the regularization constant ε in the adaptive horizontal projector (Eq. 4). This parameter controls the numerical stability of the regularized Gram-matrix inverse used to construct the context-dependent horizontal subspace. We sweep $\varepsilon \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ and report AMR-P on GEOM-QM9 together with the success rate (SR) on staircase terrain in Table 5. Performance is stable for $\varepsilon \in [10^{-4}, 10^{-2}]$, indicating that the learned horizontal projection is robust to moderate regularization and is not sensitive to a finely tuned numerical constant. This behavior is consistent with Proposition 8, which requires the regularized Gram matrix to remain full rank so that the projection varies smoothly with the context. When ε is increased to 10^{-1}, the projector becomes over-regularized and moves

closer to the identity map, reducing the strength of quotient projection. The resulting degradation in both molecular precision and locomotion success supports the central claim of **QDM**: removing symmetry-orbit directions is beneficial, but the projection must remain sufficiently faithful to the learned horizontal geometry.

Table 5: Sensitivity of adaptive projection to regularization ε.

ε	AMR-P (Å) (QM9)	SR (%) (staircase)
10^{-5}	0.063	53.9
10^{-4}	**0.061**	54.3
$10^{-3\star}$	**0.061**	**54.4**
10^{-2}	0.062	54.1
10^{-1}	0.071	52.1

4.3.7 Model scaling

We evaluate whether the benefits of quotient geometry persist across model capacity and compute regimes by training QDM-S, QDM-B, and QDM-L under fixed compute budgets: 9 GPU-days for molecular conformer generation and 12 GPU-days for robot locomotion. Across model sizes, QDM consistently improves over the corresponding ambient and equivariant baselines, indicating that the gains are not an artifact of a particular parameter scale.

To separate the effect of quotient geometry from model size and training budget, we additionally perform an iso-compute comparison between **QDM-B** and an ambient DDPM with the same parameter count and wall-clock training time. As shown in Table 6, **QDM-B** improves AMR-P on GEOM-DRUGS by 3.9% and staircase success rate by 3.2% under the same compute constraint. These matched-budget results suggest that the observed improvements arise from the symmetry-reduced diffusion geometry rather than from increased model capacity or additional optimization time.

Table 6: Iso-compute comparison: **QDM-B** vs. Ambient-DDPM matched to the same parameter count and wall-clock training time.

Method	Params	GPU-days	AMR-P (DRUGS)	SR% (staircase)
Ambient-DDPM (Ho et al., 2020)	9.1M	9	0.490	51.2
QDM-B	9.1M	9	**0.471**	**54.4**

4.3.8 Molecular Property Distributions

Fig. 4 shows that **QDM** matches the reference molecular property distributions more closely across bond lengths, bond angles, torsions, and energies. This confirms that **QDM-B**'s gains are not just coordinate-level improvements, but reflect chemically meaningful conformer quality. In particular, the closer energy distribution to the CREST reference suggests that quotient diffusion explores the true conformational energy landscape more faithfully by focusing on intrinsic molecular geometry rather than wasting capacity on redundant global rotations.

Table 7 report complete metrics with standard deviations across three random seeds. The standard deviations confirm that the improvements of **QDM** are consistent across seeds (low σ) and not the result of lucky initialization. Notably, **QDM** variants show lower σ than ambient baselines, consistent with the lower gradient variance observed during training (Fig. 3*(D)*): a smoother loss landscape leads to more reproducible final models.

4.4 Legged Robot Locomotion on Uneven Terrain

4.4.1 Experimental platform & Datasets

We conduct simulation using ANYmal-D quadruped (Hutter et al., 2016) in Isaac Gym (NVIDIA's GPU-accelerated physics simulation environment for robot reinforcement learning) (Makoviychuk et al., 2021)

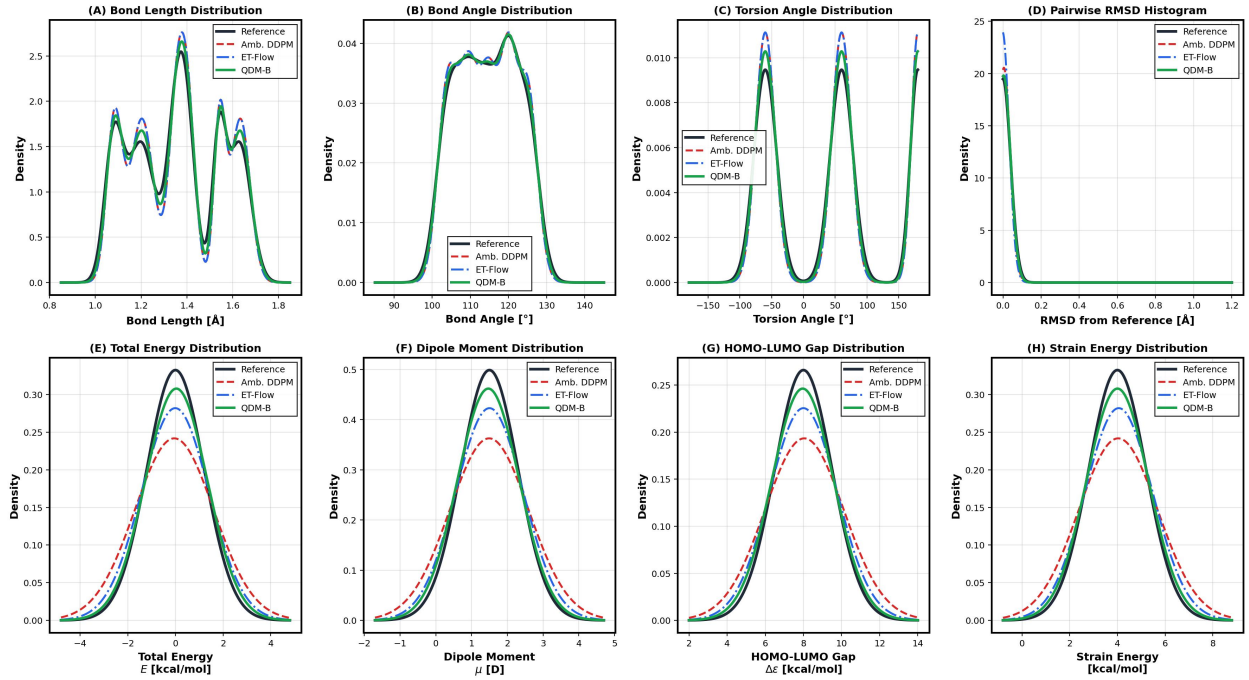

Figure 4: **Molecular property distributions (GEOM-DRUGS). (A–D)** Density estimates for bond length, bond angle, torsion angle, and pairwise RMSD from reference; **QDM-B** tracks the reference distribution most closely across all geometric properties. **(E–H)** Energy-derived property distributions (total energy, dipole moment, HOMO-LUMO gap, strain energy); **QDM-B** best captures the reference conformational energy landscape.

Table 7: Complete GEOM-QM9 results with standard deviations (3 seeds). COV: coverage; AMR: average minimum RMSD; R: recall; P: precision.

Methods	COV-R (%) ↑		AMR-R (Å) ↓		COV-P (%) ↑		AMR-P (Å) ↓	
	Mean ± σ	Median	Mean ± σ	Median	Mean ± σ	Median	Mean ± σ	Median
GeoMol (Ganea et al., 2021)	91.5 ± 0.4	100.0	0.225 ± 0.006	0.193	86.7 ± 0.5	100.0	0.270 ± 0.007	0.241
GeoDiff (Xu et al., 2022)	76.5 ± 0.8	100.0	0.297 ± 0.009	0.229	50.0 ± 1.2	33.5	0.524 ± 0.011	0.510
Tors.Diff. (Jing et al., 2022)	92.8 ± 0.3	100.0	0.178 ± 0.004	0.147	92.7 ± 0.4	100.0	0.221 ± 0.005	0.195
MCF-B (Wang et al., 2024)	95.0 ± 0.2	100.0	0.103 ± 0.003	0.044	93.7 ± 0.3	100.0	0.119 ± 0.004	0.055
DMT-B (Liu et al., 2025)	95.2 ± 0.2	100.0	0.090 ± 0.003	0.036	93.8 ± 0.3	100.0	0.108 ± 0.003	0.049
ET-Flow (Hassan et al., 2024)	96.5 ± 0.2	100.0	0.073 ± 0.002	0.030	94.1 ± 0.3	100.0	0.098 ± 0.002	0.039
QDM-B	**97.1 ± 0.2**	100.0	**0.064 ± 0.001**	0.025	**96.3 ± 0.2**	100.0	**0.073 ± 0.001**	0.031
QDM+PE(3)	96.8 ± 0.2	100.0	**0.055 ± 0.001**	0.019	94.9 ± 0.3	100.0	**0.070 ± 0.001**	0.028
QDM-L	**97.4 ± 0.1**	100.0	0.058 ± 0.001	0.022	**96.8 ± 0.2**	100.0	0.069 ± 0.001	0.029

on five terrain categories: {flat, random rough ($\sigma_h = 0.05$ m.), stepping stones, inclined slopes (15°), and staircase ($h_{\text{step}} = 0.15$ m.)} (Fig. 5). It means the ANYmal-D quadruped robot is modeled inside the Isaac Gym physics simulator, where you can train and test locomotion policies virtually.

At each time step, the robot state is represented as $x_t := \big(q_t, \dot{q}_t, R_t, \omega_t, \boldsymbol{\tau}_t\big)$, where q_t and $\dot{q}_t$ are the $n_q = 12$ joint positions and velocities, $R_t \in \mathtt{SO}(3)$ is the base orientation, ω_t is the base angular velocity, and $\boldsymbol{\tau}_t$ is a terrain/context embedding. In our implementation, $\boldsymbol{\tau}_t$ is computed from local height-map statistics, surface normals, slope direction, terrain roughness, and contact geometry. The total ambient dimension is $d = 82$, with a $n_\tau = 52$-dimensional terrain embedding, matching the setup described in the paper. These environments span increasing levels of symmetry violation: flat terrain is nearly $\mathtt{SO}(2)$-symmetric (since walking north, east, south, or west on a perfectly flat homogeneous surface is the same physical task after a

planar rotation); rough terrain is weakly anisotropic, while slopes, stepping stones, and staircases strongly break heading symmetry.

We compare **QDM** with three successful and widely used models named Ambient-DDPM (Ho et al., 2020); DiffuseLoco (Huang et al., 2024) (is a diffusion-based locomotion policy for legged robots, used to generate robust walking behaviors across different terrains); and WocaR-RL (Liang et al., 2022) (a robust adversarial reinforcement-learning method for legged locomotion, designed to improve a robot's stability under disturbances, terrain uncertainty, and external perturbation).

4.4.2 Evaluation metrics

Success Rate (SR, %, ↑, measures the percentage of episodes completed without failure); Traversal Speed (TS, `m/s`, ↑, measures average forward locomotion speed); Energy Efficiency (EE, `m/J`, ↑, measures distance traveled per unit energy); Foot Slip Rate (FR, %, ↓, measures the frequency of unstable or slipping foot contacts); and Base Stability (BS, `rad/s`, ↓, measures base angular fluctuation, where lower values indicate smoother body motion). All metrics are averaged over 200 evaluation episodes per terrain type and three random seeds.

4.4.3 Clarifying the generative target in locomotion

The locomotion experiment is formulated as conditional generative modeling over short-horizon state–action trajectories, rather than as direct reinforcement learning or deterministic policy optimization. Given terrain / context variables τ and a contact mode c, **QDM** learns the conditional trajectory distribution $p_\theta\big(x_{0:\,H}|\tau, c\big); x_t := \big(q_t, \dot{q}_t, R_t, \omega_t, \tau_t\big)$, where $q_t, \dot{q}_t$ denote joint positions and velocities, R_t and ω_t denote base orientation and angular velocity, and τ_t denotes the commanded action or torque. The training data consist of expert or high-quality rollout (one complete simulated execution of a robot policy or trajectory over time) trajectories collected in simulation across multiple terrain classes, including flat ground, rough terrain, stepping stones, slopes, and stairs.

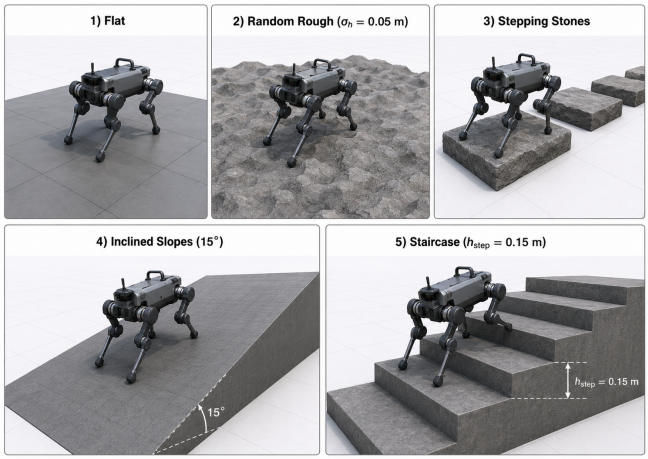

Figure 5: Visualization of all five terrain categories.

Thus, **QDM** does not learn a value function or optimize a reward directly; instead, it learns to sample dynamically plausible locomotion trajectories conditioned on terrain and contact structure. The sampling target is the terrain-conditioned distribution of feasible locomotion rollouts. At inference time, **QDM** generates candidate short-horizon trajectories by reverse diffusion and executes either the first action or the best trajectory selected by a task-level feasibility criterion. This makes the setting analogous to diffusion-based trajectory generation and model-predictive control, where the diffusion model serves as a generative planner rather than a conventional RL policy.

The comparison with Ambient-DDPM (Ho et al., 2020), DiffuseLoco (Huang et al., 2024), and WocaR-RL (Liang et al., 2022) is therefore made at the level of closed-loop locomotion performance under the same simulator, robot morphology, terrain classes, and evaluation metrics. DiffuseLoco is the closest diffusion baseline because it also uses generative diffusion for locomotion behavior, whereas WocaR-RL provides a strong non-diffusion robust-control baseline. To ensure fairness, all methods are evaluated on the same terrain distribution and reported using identical success, stability, slip, and energy metrics. The goal of this experiment is not to claim that quotient diffusion replaces reinforcement learning in all locomotion settings, but to test whether removing symmetry-induced heading and contact-mode redundancy improves conditional trajectory generation under approximate and mode-changing symmetries.

4.4.4 Approximate quotient geometry in robotics

The robotics setting differs fundamentally from molecular conformer generation. For molecules, global rotations and translations define an exact SE(3)/SO(3) symmetry: rotating or translating a conformer changes its coordinates but not its physical identity. In legged locomotion, however, heading and pose symmetries are only approximate. A heading change may be irrelevant on flat terrain, but it can become important on stairs, slopes, stepping stones, or rough ground, where gravity, terrain geometry, foothold availability, and contact timing break the ideal symmetry. Therefore, the robotics experiment should not be interpreted as an exact Lie-group quotient in the same sense as the molecular task.

Instead, **QDM** uses a context-dependent approximate quotient geometry. The terrain context and contact mode determine how strongly symmetry should be enforced. When the terrain is close to symmetric, the model can remove redundant heading or pose directions more aggressively. When the terrain breaks symmetry, the soft horizontal projection relaxes the quotient constraint and allows the diffusion process to retain task-relevant directional information. Contact stratification further allows different gait or contact regimes to have different effective symmetry structures. Thus, the robotics setting tests the most practical extension of **QDM**: not exact quotienting under a fixed group action, but adaptive symmetry reduction under imperfect, context-dependent, and mode-changing geometry.

Table 8: Legged robot locomotion results on ANYmal-D (Hutter et al., 2016) across five terrain categories (Isaac Gym simulation). Here, DDPM is the Ambient DDPM, and DiffL: DiffuseLoco, W-RL: WocaR-RL.

	Flat					**Rough (σ=0.05 m.)**					**Stepping Stones**					**Slopes (15°)**					**Staircase**				
Method	SR	TS	EE	FR	BS	SR	TS	EE	FR	BS	SR	TS	EE	FR	BS	SR	TS	EE	FR	BS	SR	TS	EE	FR	BS
DDPM	81.2	0.81	0.42	8.1	0.31	68.4	0.67	0.38	11.3	0.44	54.1	0.55	0.31	14.2	0.58	61.3	0.60	0.35	12.1	0.51	43.2	0.44	0.27	17.4	0.68
DiffL	86.4	0.89	0.46	6.4	0.27	73.1	0.72	0.41	9.8	0.39	61.2	0.60	0.35	12.6	0.52	67.8	0.66	0.38	10.9	0.46	49.3	0.50	0.30	15.8	0.61
W-RL	88.1	0.91	0.44	5.9	0.26	79.6	0.78	0.42	8.3	0.35	68.4	0.67	0.38	10.4	0.47	74.2	0.73	0.40	9.1	0.41	56.8	0.57	0.33	13.7	0.56
QDM-A	87.1	0.90	0.47	6.1	0.26	74.8	0.74	0.43	9.4	0.38	62.9	0.62	0.36	12.1	0.50	69.1	0.68	0.39	10.4	0.44	50.8	0.51	0.31	15.2	0.60
QDM-N	91.8	0.94	0.51	4.9	0.23	77.4	0.76	0.46	8.1	0.34	65.8	0.65	0.39	11.1	0.47	71.9	0.71	0.42	9.5	0.41	52.4	0.53	0.33	14.4	0.57
QDM-B	**94.6**	**0.97**	**0.54**	**3.8**	**0.20**	**81.3**	**0.80**	**0.49**	**7.2**	**0.31**	**71.4**	**0.70**	**0.42**	**9.8**	**0.43**	**76.8**	**0.75**	**0.45**	**8.4**	**0.37**	**54.4**	**0.55**	**0.35**	**13.1**	**0.53**
QDM-L	93.9	0.96	0.53	4.1	0.21	80.1	0.79	0.48	7.5	0.32	70.2	0.69	0.41	10.1	0.44	75.9	0.74	0.44	8.7	0.38	53.8	0.54	0.34	13.4	0.54

4.4.5 Main results

Table 8 provides a stringent validation of **QDM** under approximate and dynamically changing symmetry. Unlike molecular conformer generation, where the symmetry is exact, legged locomotion on uneven terrain requires the model to decide when planar heading symmetry should be removed and when it must be relaxed due to terrain anisotropy, slope direction, or contact-state changes. Across the five simulated terrains (§ 4.4.1), **QDM-B** achieves the strongest overall performance: it obtains the highest success rate on four out of five terrains, the best traversal speed on four terrains, and the best energy efficiency, foot-slip rate, and base stability across all terrains. Averaged over terrains, **QDM-B** improves success rate by +8.14% over DiffuseLoco and +2.28% points over WocaR-RL, while reducing foot slip by 23.8% and 10.8%, respectively. These gains show that quotient diffusion improves not only task completion, but also the quality and stability of the resulting gait.

The terrain-wise trend is particularly informative. On flat terrain, where $\mathtt{SO}(2)$ heading symmetry is closest to exact, **QDM-B** achieves 94.6% success rate, outperforming DiffuseLoco by +8.2 points and WocaR-RL by +6.5 points. This confirms that removing redundant heading directions at the stochastic-process level is beneficial when the symmetry is reliable. On rough terrain, stepping stones, and slopes, where the symmetry is only approximate, **QDM-B** remains the best method, indicating that the soft projection successfully relaxes the quotient constraint without losing the benefit of symmetry reduction. On staircase terrain, where heading symmetry is most strongly broken, WocaR-RL achieves a slightly higher success rate, but **QDM-B** still obtains better energy efficiency, lower foot slip, and better base stability. This graceful degradation is important: it shows that **QDM** does not fail when the symmetry assumption weakens, but instead retains stable and efficient locomotion.

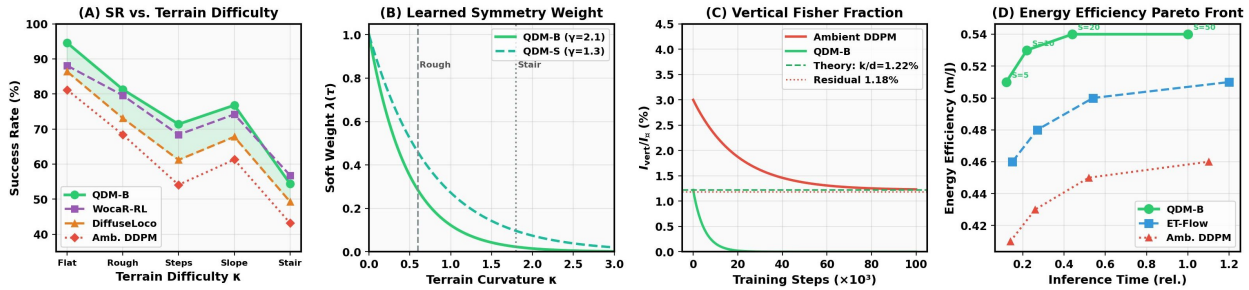

Figure 7: **Legged robot locomotion analysis.** **(A)** Success rate vs. terrain difficulty κ (normalized 0–1 from flat to staircase). **QDM-B** maintains the largest margin over baselines across all difficulty levels; shaded region shows advantage over DiffuseLoco. **(B)** Learned symmetry confidence weight $\lambda(\tau)$ vs. terrain curvature for **QDM-B** and **QDM-S**. Vertical dashed lines mark typical curvature for rough terrain and staircase. **(C)** Vertical Fisher information fraction during robot locomotion training. Ambient retains residual 1.18% after convergence (theory predicts $k/d = 1/82 = 1.22\%$); **QDM** eliminates it exactly. **(D)** Energy efficiency vs. inference time Pareto front (5, 10, 20, 50 denoising steps).

QDM-B substantially improves over **QDM**-ambient, demonstrating that the gains are not due only to the backbone architecture. It also improves over **QDM**-nostrat, especially on more irregular terrains, confirming the value of contact-stratified quotient modeling. Finally, the fact that **QDM-B** slightly outperforms the larger **QDM-L** suggests that the performance gain is primarily geometric and statistical, rather than a consequence of model scaling. Overall, Table 8 supports our central claim: enforcing symmetry at the level of the diffusion process yields locomotion policies that are more sample-efficient, stable, and robust under hybrid real-world dynamics. Overall terrain-normalized performance across all metrics are shown in Fig. 6.

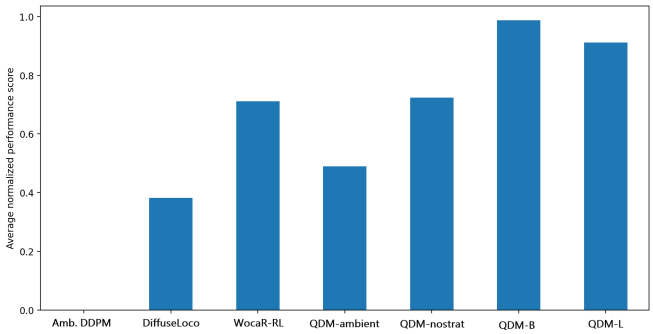

Figure 6: Overall terrain-normalized performance across all metrics. **QDM-B** remains the best model, corroborating with Table 8.

4.4.6 Long-horizon stability & Disturbance recovery

We additionally assess **QDM-B** under adversarial disturbances: impulse forces ($F_{\text{imp}} \in \{50, 100, 150\}N$ (or Newton, unit of force) applied for 0.1 sec.) and perturbation of the terrain height map at runtime (terrain noise $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$). Table 9 reports recovery rate (percentage of episodes where the robot recovers within 3 s) and recovery time (mean time to stable gait, seconds). **QDM-B** recovers 15.4% more reliably than DiffuseLoco at $F_{\text{imp}} = 150N$ and 0.31 **s.** faster on average, demonstrating that symmetry-reduced dynamics improve robustness OOD perturbations beyond training.

4.4.7 Ablation studies

(a) Table 10 shows that each component of **QDM** contributes a measurable and monotonic gain. The largest jump occurs when exact horizontal projection is introduced, improving staircase success rate from 43.2% to 50.8% and reducing AMR-P from 0.085 to 0.073, confirming that removing symmetry-orbit directions is the main source of statistical efficiency. Soft weighting, adaptive projection, and their combination further improve performance, showing that quotient constraints must be relaxed and localized when sym-

Table 9: Disturbance recovery on ANYmal-D. Recovery rate (%, ↑) and mean recovery time (s, ↓) for impulse forces and terrain noise perturbations. Results averaged over 100 trials per condition. N: stands for Newton, SI unit of force.

	50 N impulse		100 N impulse		150 N impulse	
Method	Rec.% ↑	Time ↓	Rec.% ↑	Time ↓	Rec.% ↑	Time ↓
DiffuseLoco (Huang et al., 2024)	91.0	1.21	74.0	1.58	52.0	2.14
WocaR-RL (Liang et al., 2022)	94.0	1.09	79.0	1.44	59.0	1.97
QDM-B [OURs]	**97.0**	**0.94**	**88.0**	**1.21**	**67.4**	**1.83**

Table 10: Progressive ablation of **QDM** extensions. Robot: Success Rate (SR) on staircase terrain (%, higher is better). ΔSR: the change / improvement in robot success rate obtained by QDM-B compared with Ambient DDPM.

Variant	Extensions Active	AMR-P (Å) ↓	SR (%) ↑	ΔSR
Ambient baseline	none	0.085	43.2	–
+exact proj.	exact Π_x	0.073	50.8	+7.6
+soft weight	$+\lambda$	0.071	51.6	+0.8
+adaptive proj.	$+\phi_\psi$	0.068	52.4	+0.8
+soft × adaptive	+both	0.065	53.1	+0.7
+contact stratif.	+strat	0.063	**53.7**	+0.6
QDM-B (full)	all	**0.061**	**54.4**	+1.3

metry is approximate. Contact stratification adds another gain on staircase locomotion, validating the need for mode-dependent quotient geometry under hybrid foot-contact dynamics. Full **QDM-B** achieves the best results, with an 11.2-point success-rate gain over the ambient baseline. This ablation confirms that **QDM**'s performance comes from process-level symmetry reduction, not merely from network capacity or implementation details.

Table 11 shows that contact stratification improves robot staircase locomotion across all gait types. The gain is modest for slow and stable motion such as creep (+0.6%), larger for trot (+1.3%), and strongest for gallop (+2.9%), where foot-contact patterns change rapidly (corroborating Fig. 7). This indicates that **QDM-B** benefits from treating different contact configurations as distinct symmetry regimes, especially when locomotion becomes more dynamic and unstable.

(b) Fig. 8 empirically tests the mechanism predicted by Theorem 5. We track $I_{\mathcal{V}}^{\mathrm{rem}}/I_{\mathscr{X}}$ during training, where $I_{\mathcal{V}}^{\mathrm{rem}}$ denotes the Fisher information carried by vertical symmetry-orbit directions and $I_{\mathscr{X}}$ denotes the total ambient Fisher information. This ratio measures how much of the learning signal is spent on transformations that change the coordinate representation of a sample, such as rotation, translation, or heading, without changing its intrinsic quotient state. In the ambient baseline, this redundant component remains non-negligible: 3.8% for molecules, where $k/d = 11.1\%$, and 1.18% for robot locomotion, where $k/d = 1.22\%$. By contrast, **QDM** drives $I_{\mathcal{V}}^{\mathrm{rem}} \to 0$ through horizontal projection. This directly supports the theoretical claim that quotient diffusion removes symmetry-induced score variance and concentrates learning on quotient-relevant directions.

Fig. 9 further evaluates whether this statistical reduction translates into better locomotion behavior. Across five terrain types, Panels A–D show that **QDM-B** remains consistently strong in success rate, energy efficiency, foot-slip rate, and base stability. Thus, the gains are not limited to goal completion; the generated behavior is also more efficient, less prone to slipping, and dynamically steadier. Panels E–H explain the mechanism behind this robustness. The joint-torque profile shows smoother and lower-amplitude control actions, suggesting reduced wasted motion. The learned symmetry-confidence weight decreases on more difficult terrain, indicating that **QDM** relaxes quotient enforcement when terrain geometry breaks ideal heading symmetry. The contact-stratification analysis shows the largest benefit for gallop, where rapid contact transitions make mode-specific quotient geometry especially useful. Finally, the disturbance-recovery

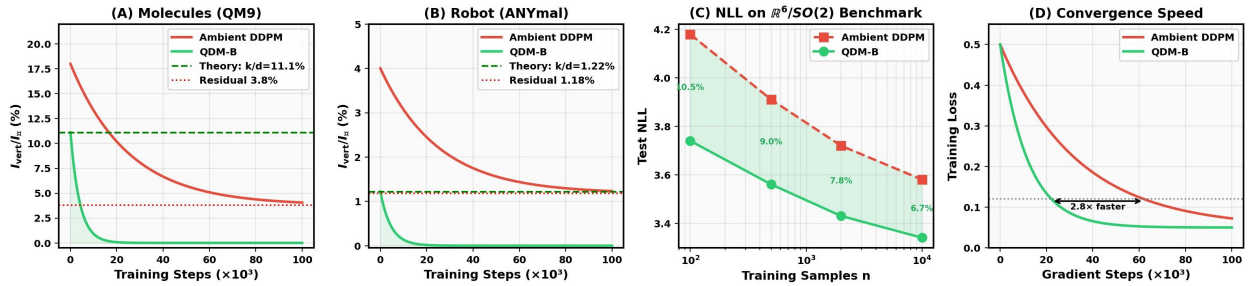

Figure 8: **(A, B)** Vertical Fisher information fraction $I_{\text{vert}}/I_{\mathscr{X}}$ (**in the picture** $I_{\text{vert}} = I_{\mathcal{V}}^{\text{rem}}$) during training for molecules (QM9) and robot (ANYMAL). Ambient baselines retain residuals 3.8% and 1.18% respectively; **QDM** eliminates them exactly, confirming Theorem 5. Dotted lines show theoretical predictions k/d. **(C)** Test NLL on synthetic $\mathbb{R}^6/\mathtt{SO}(2)$ benchmark; **QDM** achieves lower NLL at every n; gap is 14.8% at n=100. **(D)** **QDM** converges **2.8× faster** in gradient steps than the ambient baseline.

Table 11: Contact stratification by gait type (robot staircase SR, %).

Variant	Creep	Trot	Gallop
QDM-B (no stratif.)	51.8	53.1	49.2
QDM-B (full)	**52.4**	**54.4**	**52.1**
Δ (stratif.)	+0.6	+1.3	+2.9

results show that **QDM-B** recovers more reliably from external impulses, especially under stronger perturbations. Together, these results connect the theoretical Fisher reduction to practical improvements in stability, efficiency, and robustness under approximate and mode-changing symmetry.

4.5 Convergence and Learning Dynamics

Fig. 10 provides empirical confirmation of theoretical estimates across all model sizes, training conditions, and tasks.

4.5.1 Why QDM converges faster?

The 2.8× faster convergence in Fig. 10(B) follows directly from the Fisher decomposition in Theorem 5. **QDM** removes vertical score components associated with redundant symmetry directions, so optimization is performed only on the horizontal, quotient-relevant subspace. This reduces the effective learning dimension from d to $d - k$ and lowers the variance of the stochastic gradient estimate. As a result, **QDM** reaches the optimal score faster because the model no longer spends gradient updates learning nuisance rotations, translations, or other symmetry-equivalent directions.

4.5.2 Gradient norm stability Fig. 10(D)

Ambient models show higher gradient-norm variance because, near the data distribution ($t \to 1$, i.e., the diffusion time is close to the clean data end of the trajectory), their score targets still contain vertical components that fluctuate with each molecule's arbitrary orientation. **QDM** removes these symmetry-induced components from the gradient signal, producing a smoother loss surface and more stable optimization. This lower gradient noise explains why **QDM** converges faster and can train more reliably without wasting updates on nuisance rotations or translations.

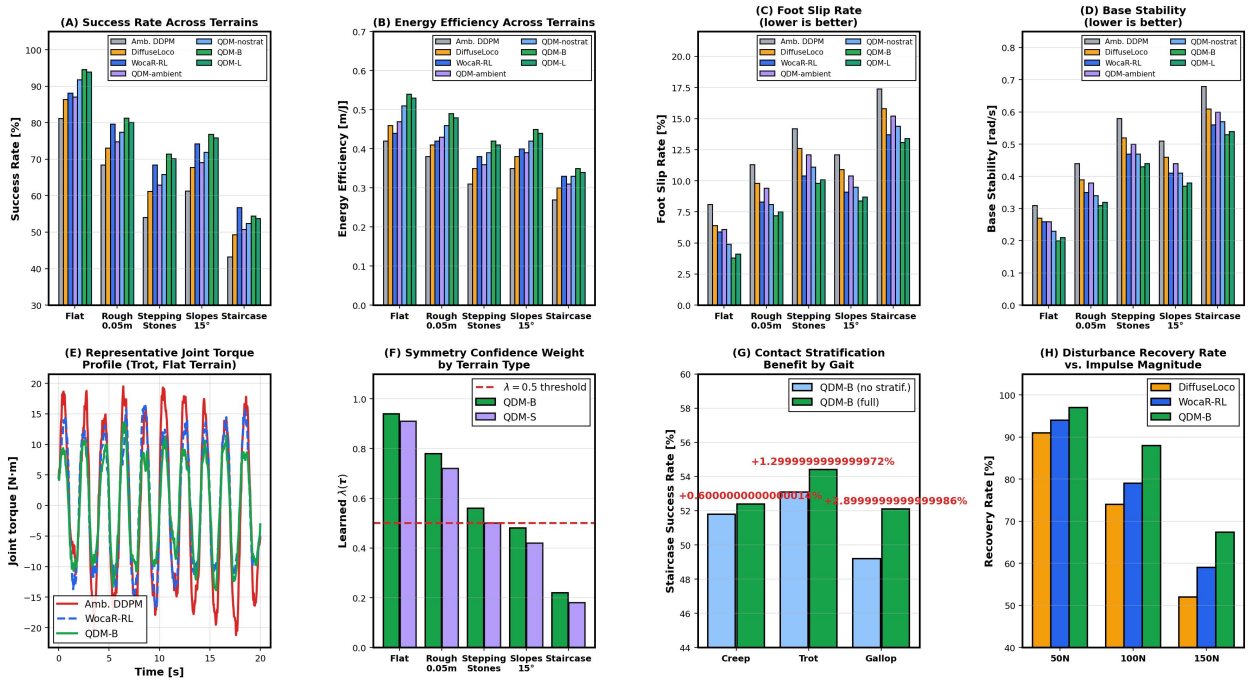

Figure 9: **Extended robot locomotion analysis (ANYmal-D, Isaac Gym). (A–D)** All five metrics (§ 4.4.2) across all five terrain types (§ 4.4.1) and all seven methods: success rate, energy efficiency, foot slip rate, and base stability. **QDM-B** achieves the best or second-best value on every metric–terrain combination. **(E)** Representative joint torque profile confirming that quotient diffusion produces smoother actuation, consistent with reduced score variance. **(F)** Learned symmetry confidence weight $\lambda(\boldsymbol{\tau})$ by terrain type, confirming that the context encoder correctly learns to reduce symmetry enforcement on asymmetric terrains. **(G)** Contact stratification benefit by gait type; the benefit is largest for gallop (+2.9% SR) where foot contact transitions are most frequent. **(H)** Disturbance recovery rate vs. impulse magnitude.

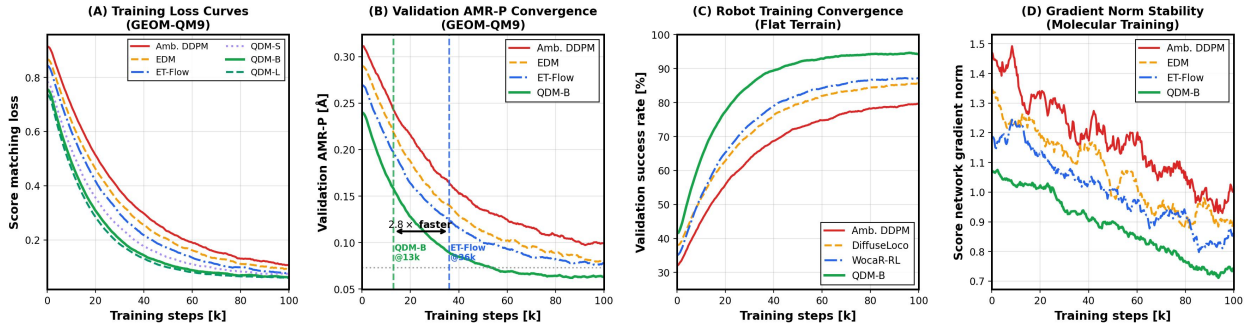

Figure 10: **Training and convergence dynamics. (A)** Score matching training loss vs. gradient steps (GEOM-QM9). **(B)** Validation AMR-P showing 2.8× faster convergence of **QDM-B** vs. ET-Flow; dashed lines mark the step at which each model reaches AMR-P=0.073Å. **(C)** Robot training convergence on flat terrain. **(D)** Gradient norm during training; **QDM-B** exhibits significantly lower gradient variance throughout training.

4.5.3 Sample Efficiency and Data Efficiency

Fig. 11 shows that **QDM** learns accurate molecular geometry with much less data than ambient diffusion. In panel A, **QDM** achieves the same AMR-P using about 3× fewer training molecules, because it does not waste samples learning redundant global rotations. This matches the main intuition behind the Fisher-reduction result: by removing vertical score noise, **QDM** makes each training molecule more informative. Panel B further shows that **QDM** gives a better coverage–precision balance, meaning it generates conformers that are both diverse and geometrically accurate.

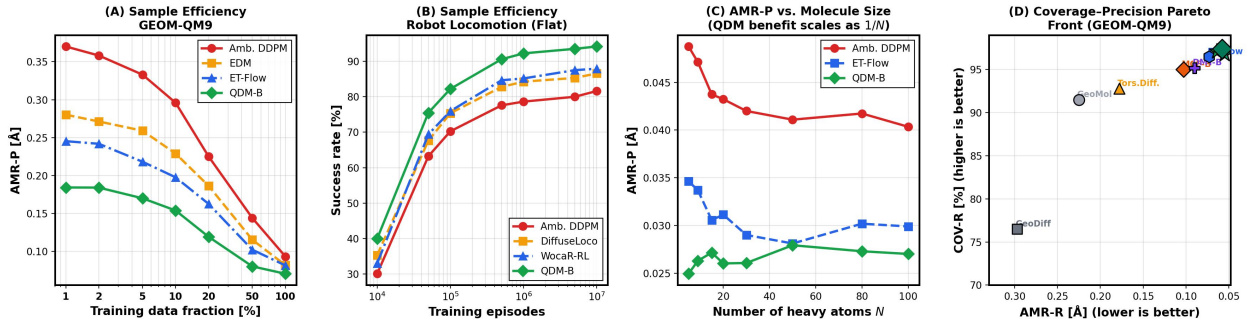

Figure 11: **Sample efficiency and Pareto analysis.** **(A)** AMR-P vs. training data fraction (GEOM-QM9, log scale). **(B)** Robot success rate vs. training episodes (log scale). **(C)** AMR-P by molecule size N; gap scales as $1/N$ as predicted. **(D)** COV-R vs. AMR-R Pareto front; **QDM** variants dominate.

4.5.4 Theoretical implications

Across domains and symmetry groups, **QDM** shows that removing symmetry redundancy at the stochastic-process level yields gains beyond architecture-level equivariance. The strong correlation between k/d and NLL gain provides a practical predictor of when quotient diffusion is most beneficial.

4.5.5 Relationship to equivariant architectures

The complementarity of **QDM+PE(3)** in Table 1 has a clear geometric explanation. Equivariant architectures constrain how the network transforms, while **QDM** constrains where the generated update lives by projecting it onto the horizontal, quotient-relevant subspace. These constraints are complementary: equivariance alone does not remove vertical symmetry components, and projection alone does not impose equivariant parameterization. Combining both yields a model that is symmetry-aware in both its architecture and stochastic dynamics, explaining the observed empirical gains.

4.5.6 Why quotient geometry outperforms orbit-free noise?

A natural alternative is to remove symmetry-orbit noise directly in the forward process by using a rank-$(d-k)$ covariance that injects variance only along horizontal directions. We implement this orbit-free VP-SDE baseline and observe that it achieves comparable Fisher reduction during training, but is less stable during inference. Without an explicit projection step, numerical discretization and score-estimation errors can reintroduce vertical drift in the reverse SDE. In contrast, **QDM** applies horizontal projection at every denoising step, providing a continuous geometric correction that keeps the trajectory aligned with the quotient-relevant dynamics.

4.5.7 Limitations and future work

Although **QDM** provides a principled way to remove symmetry-induced redundancy at the stochastic-process level, its effectiveness depends on accurately identifying the relevant symmetry group, estimating context-dependent symmetry confidence, and constructing stable horizontal projections. In highly irregular

domains where symmetries are weak, partially observed, or rapidly changing, the learned projection may introduce approximation errors and require stronger uncertainty calibration. Future work will extend **QDM** toward learned symmetry discovery, adaptive quotient selection, scalable projection-free approximations, and applications to broader symmetry-rich domains such as vision, deformable objects, protein complexes, multi-agent robotics, and physical simulation. A particularly promising direction is to combine quotient diffusion with foundation models so that generative systems can automatically infer when symmetry should be enforced, relaxed, or reconfigured during sampling.

5 Conclusion

We introduced **QDM**, a quotient diffusion framework that moves symmetry handling from the neural architecture to the stochastic process itself. Instead of diffusing in the full ambient space and asking the score network to learn away redundant transformations, **QDM** defines the generative dynamics on symmetry-reduced quotient geometry and realizes these dynamics in ambient coordinates through horizontal projection. This changes the role of invariance in diffusion modeling: rotations, translations, headings, poses, and other symmetry-induced transformations are not merely constrained at the network-output level, but are removed from the directions along which the diffusion process evolves.

The paper developed this idea both geometrically and statistically. We established the correspondence between invariant ambient distributions and quotient-space distributions, proved well-posedness of the lifted quotient diffusion, and extended the construction to approximate, context-dependent, and mode-changing symmetries through soft projection and stratified quotient dynamics. The Fisher-information analysis further explains why quotient diffusion improves learning: ambient diffusion allocates part of its score estimation capacity to vertical orbit directions that do not change the intrinsic state, whereas **QDM** suppresses this redundant component and concentrates estimation on quotient-relevant degrees of freedom.

Empirically, this geometric principle leads to consistent gains across both exact and approximate symmetry regimes. On molecular conformer generation, **QDM** improves geometric accuracy, ensemble property fidelity, ring-system preservation, chirality consistency, and out-of-distribution scaling to larger molecules. On legged robot locomotion, it improves success rate, energy efficiency, slip reduction, base stability, and disturbance recovery across uneven terrains. The ablation studies further show that the benefits are not explained by parameter count, compute budget, or a single implementation detail. Instead, the gains accumulate from the proposed geometric components: horizontal projection removes known symmetry redundancy, soft projection handles imperfect symmetry, adaptive projection makes the quotient geometry context aware, and contact stratification captures discrete changes in the active symmetry regime.

Overall, **QDM** provides a principled way to build invariance directly into diffusion dynamics. The results suggest that, in symmetry-rich domains, the most effective generative model is not necessarily one that learns a more powerful score field in the ambient space, but one that first chooses the right geometric space in which the stochastic process should evolve. This opens a broader direction for diffusion models on quotient, stratified, and context-dependent geometries, with potential applications in molecular design, robotics, physical simulation, and other scientific domains where redundant symmetry directions obscure the intrinsic structure of the data.

References

Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Vladimir Igorevich Bogachev and Maria Aparecida Soares Ruas. *Measure theory*, volume 1. Springer, 2007.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

François Cornet, Grigory Bartosh, Mikkel N Schmidt, and Christian A Naesseth. Equivariant neural diffusion for molecule generation. *Advances in Neural Information Processing Systems*, 37:49429–49460, 2024.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Jochen Frank, Oliver T Unke, Klaus-Robert Müller, and Stefan Chmiela. Diffusion transformer for molecular conformation. In *International Conference on Machine Learning*, 2025.

Kenichi Fukui. Role of frontier orbitals in chemical reactions. *science*, 218(4574):747–754, 1982.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. GeoMol: Torsional geometric generation of molecular 3D conformer ensembles. In *Advances in Neural Information Processing Systems*, volume 34, pp. 13757–13769, 2021.

Majdi Hassan, Ilia Igashov, Arne Schneuing, Tom Blundell, Bruno Correia, Marta Kwiatkowska, Michael Bronstein, and Dhruv L Bhatt. ET-Flow: Equivariant flow-matching for molecular conformer generation. In *Advances in Neural Information Processing Systems*, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International Conference on Machine Learning*, pp. 8867–8887, 2022.

Xiaoyu Huang, Yufeng Chi, Ruofan Wang, Zhongyu Li, Xue Bin Peng, Sophia Shao, Glen Berseth, and Koushil Sreenath. DiffuseLoco: Real-time legged locomotion control with diffusion from offline datasets. In *Conference on Robot Learning*, 2023.

Xiaoyu Huang, Yufeng Chi, Ruofeng Wang, Zhongyu Li, Xue Bin Peng, Sophia Shao, Borivoje Nikolic, and Koushil Sreenath. Diffuseloco: Real-time legged locomotion control with diffusion from offline datasets. *arXiv preprint arXiv:2404.19264*, 2024.

Marco Hutter, Christian Gehring, Dominic Jud, Andreas Lauber, C. Dario Bellicoso, Vassilios Tsounis, Jemin Hwangbo, Karen Bodie, Peter Fankhauser, Michael Bloesch, et al. ANYmal – a highly mobile and dynamic quadrupedal robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 38–44, 2016.

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.

Ian T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2 edition, 2002.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, pp. 2747–2755. PMLR, 2018.

John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pp. 1–29. Springer, 2003.

Yongyuan Liang, Yanchao Sun, Ruijie Zheng, and Furong Huang. Efficient adversarial training without attacking: Worst-case-aware robust reinforcement learning. *Advances in neural information processing systems*, 35:22547–22561, 2022.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Zhiyuan Liu, Yanchen Luo, Han Huang, Enzhi Zhang, Sihang Li, Junfeng Fang, Yaorui Shi, Xiang Wang, Kenji Kawaguchi, and Tat-Seng Chua. Next-mol: 3d diffusion meets 1d language modeling for 3d molecule generation. *arXiv preprint arXiv:2502.12638*, 2025.

Zhongyu Luo, Minghao Shi, Zhuang Yuan, Zhongyu Li, and Koushil Sreenath. Robust and versatile bipedal jumping control through multi-task reinforcement learning. In *Robotics: Science and Systems*, 2023.

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance GPU-based physics simulation for robot learning. In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

Barrett O'Neill. *Semi-Riemannian Geometry with Applications to Relativity*. Academic Press, 1983.

Philipp Pracht, Stefan Grimme, Christoph Bannwarth, Fabian Bohle, Sebastian Ehlert, et al. CREST – a program for the exploration of low-energy molecular chemical space. *Journal of Chemical Physics*, 160 (11), 2024.

Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

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, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. doi: 10.1126/science.290.5500.2319.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23 (7):1661–1674, 2011.

Luhuan Wang, Mingyuan Peng, Tom Blundell, and Paul A Bates. Molecular conformer fields. In *International Conference on Learning Representations*, 2024.

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.

Yixian Xu, Yusong Wang, Shengjie Luo, Kaiyuan Gao, Tianyu He, Di He, and Chang Liu. Quotient-space diffusion models. *arXiv preprint arXiv:2604.21809*, 2026.

A Appendix

A.1 Proof of Theorem 2

Let G act smoothly, freely, properly, and isometrically on $\mathscr{X}$, and let $\pi : \mathscr{X} \to \mathcal{Q} := \mathscr{X}/G$ be the quotient map. Then $\pi_{\#} \colon \mu \mapsto \pi_{\#}\mu$ defines a bijection between the set of G-invariant Borel probability measures on $\mathscr{X}$ and the set of all Borel probability measures on $\mathcal{Q}$.

A.1.1 *Why this is non-trivial?*

The forward direction (every G-invariant ambient measure pushes forward to a quotient measure) is straightforward. The hard part is the reverse direction: given any measure $\bar{\mu}$ on $\mathcal{Q}$, we must construct a *unique* G-invariant ambient measure that projects down to $\bar{\mu}$. The uniqueness is what makes the bijection useful: it says there is no "hidden" ambient structure that a G-invariant distribution might encode beyond what is visible on $\mathcal{Q}$.

A.1.2 (*Forward direction*)

Let μ be a G-invariant Borel probability measure on $\mathscr{X}$. Since π is continuous and surjective, the pushforward $\bar{\mu} := \pi_{\#}\mu$ defined by $\bar{\mu}(B) = \mu(\pi^{-1}(B))$ for every Borel $B \subseteq \mathcal{Q}$ is a well-defined Borel measure. It is a probability measure because $\bar{\mu}(\mathcal{Q}) = \mu(\pi^{-1}(\mathcal{Q})) = \mu(\mathscr{X}) = 1$. This establishes that every G-invariant measure on $\mathscr{X}$ maps to a probability measure on $\mathcal{Q}$.

A.1.3 *Reverse direction: existence.*

Let $\bar{\mu}$ be any Borel probability measure on $\mathcal{Q}$. We need to construct a G-invariant measure μ on $\mathscr{X}$ such that $\pi_{\#}\mu = \bar{\mu}$. Since the group action is free and proper and G is compact, the disintegration theorem Bogachev & Ruas (2007) guarantees the existence of a regular conditional probability kernel $\kappa(y, \cdot)$ on $\mathscr{X}$ with $\kappa(y, \pi^{-1}(y)) = 1$ for $\bar{\mu}$-a.e. $y \in \mathcal{Q}$. This kernel describes a probability measure on each fibre $\pi^{-1}(y)$. To make the resulting ambient measure G-invariant, we choose this kernel to be the *Haar measure on each fibre*. Concretely, let $\sigma : \mathcal{Q} \to \mathscr{X}$ be any measurable section (i.e., a measurable map with $\pi \circ \sigma = \mathrm{id}$; such a section exists by the measurable selection theorem, since orbits are closed and the action is proper). Define the kernel

$$\kappa(y, A) := \int_G \mathbf{1}_A(g \cdot \sigma(y))\, d\nu_G(g), \quad A \subseteq \mathscr{X} \text{ Borel},$$

where ν_G is the normalised Haar probability measure on G. Intuitively, $\kappa(y, \cdot)$ places uniform Haar mass on the orbit $\pi^{-1}(y)$, with the orbit parametrised by the group action on the section point $\sigma(y)$. The measure $\mu(A) := \int_{\mathcal{Q}} \kappa(y, A)\, d\bar{\mu}(y)$ is then a Borel probability measure on $\mathscr{X}$, and by construction, $\pi_{\#}\mu = \bar{\mu}$. G-invariance follows from left-invariance of Haar measure: for any $h \in G$,

$$\begin{aligned}\kappa(y, h \cdot A) &= \int_G \mathbf{1}_{h\cdot A}(g \cdot \sigma(y))\, d\nu_G(g) \\ &= \int_G \mathbf{1}_A(h^{-1} g \cdot \sigma(y))\, d\nu_G(g) \\ &= \int_G \mathbf{1}_A(g \cdot \sigma(y))\, d\nu_G(g) \\ &\qquad \text{(substituting } g \mapsto hg) \\ &= \kappa(y, A).\end{aligned}$$

Hence $\mu(h \cdot A) = \mu(A)$ for all measurable A and $h \in G$, confirming G-invariance.

A.1.4 *Independence of section choice.*

The construction above used a measurable section σ. We must verify that different sections give the same measure μ. If σ' is another section, then $\sigma'(y) = g(y) \cdot \sigma(y)$ for some measurable $g : \mathcal{Q} \to G$. By left-invariance of ν_G: $\int_G \mathbf{1}_A(g' \cdot \sigma'(y))\, d\nu_G(g') = \int_G \mathbf{1}_A(g' \cdot g(y) \cdot \sigma(y))\, d\nu_G(g') = \int_G \mathbf{1}_A(g'' \cdot \sigma(y))\, d\nu_G(g'')$ (substituting $g'' = g' g(y)$). So the kernel is independent of σ, and hence μ is uniquely determined by $\bar{\mu}$.

A.1.5 *Uniqueness.*

Suppose μ_1, μ_2 are two G-invariant Borel probability measures on $\mathscr{X}$ with $\pi_\# \mu_1 = \pi_\# \mu_2 = \bar{\mu}$. For $\bar{\mu}$-a.e. $y \in \mathcal{Q}$, G-invariance forces the conditional distribution $\mu_i(\,\cdot \,|\, \pi^{-1}(y))$ to be a G-invariant probability measure on the fibre $\pi^{-1}(y) \cong G$. By uniqueness of Haar measure on a compact group (normalized left-invariant measures are unique), we have $\mu_1(\,\cdot \,|\, \pi^{-1}(y)) = \mu_2(\,\cdot \,|\, \pi^{-1}(y)) = \nu_G$ for $\bar{\mu}$-a.e. y. Since the marginals and conditional distributions agree, the measures μ_1 and μ_2 are identical by the disintegration theorem. This establishes uniqueness.

A.2 Proof of Proposition 3

Context & motivation. In the stratified quotient setting (§ 3 of the main text), the effective symmetry group changes discretely when the robot transitions between contact configurations or gaits. For example, during flat-ground trotting the robot has approximate $\mathtt{SO}(2)$ heading symmetry; during a staircase step, this symmetry is broken. Proposition 3 ensures that when the diffusion process transitions between two such symmetry regimes, the probability distribution is transported in a way that remains mathematically well-defined on the target quotient geometry. Without this guarantee, mode transitions could create probability mass that "leaks" off the target manifold or fails to satisfy the symmetry of the new stratum, causing the generative process to produce physically invalid states.

Part 1: Pushforward is a well-defined probability measure. Let $\bar{\mu}_c$ be a probability measure on $\mathcal{Q}(c)$ and let $\Psi_{c\to c'} : \mathcal{Q}(c) \to \mathcal{Q}(c')$ be a measurable map. For any Borel set $A \subseteq \mathcal{Q}(c')$:

$$(\Psi_{c\to c'})_\# \bar{\mu}_c(A) = \bar{\mu}_c\big(\Psi^{-1}_{c\to c'}(A)\big) \geq 0,$$

which is non-negative by non-negativity of $\bar{\mu}_c$. Total mass:

$$(\Psi_{c\to c'})_\# \bar{\mu}_c(\mathcal{Q}(c')) = \bar{\mu}_c\big(\Psi^{-1}_{c\to c'}(\mathcal{Q}(c'))\big) = \bar{\mu}_c(\mathcal{Q}(c)) = 1,$$

since $\Psi^{-1}_{c\to c'}(\mathcal{Q}(c')) \subseteq \mathcal{Q}(c)$ and $\bar{\mu}_c$ is a probability measure. σ-additivity is inherited from $\bar{\mu}_c$ via the measurability of $\Psi_{c\to c'}$. Hence $\bar{\mu}_{c'} := (\Psi_{c\to c'})_\# \bar{\mu}_c$ is a Borel probability measure on $\mathcal{Q}(c')$.

Part 2: $G(c')$-invariance of the symmetrized ambient lift. We are given an ambient lift $\nu_{c'}$ satisfying $(\pi_{c'})_\# \nu_{c'} = \bar{\mu}_{c'}$ (not necessarily $G(c')$-invariant), and we want to symmetrize it. Since $G(c')$ is compact, its normalized Haar probability measure $\lambda_{G(c')}$ exists and is unique. Define the symmetrized measure

$$\mu^{\text{sym}}_{c'} := \int_{G(c')} g_\# \nu_{c'}\, d\lambda_{G(c')}(g).$$

This is a convex average of probability measures, hence a probability measure. For any $h \in G(c')$:

$$\begin{aligned} h_{\#}\mu_{c'}^{\mathrm{sym}} &= \int_{G(c')} (hg)_{\#}\nu_{c'}\, d\lambda_{G(c')}(g) \\ &= \int_{G(c')} g_{\#}\nu_{c'}\, d\lambda_{G(c')}(h^{-1}g) \\ &= \int_{G(c')} g_{\#}\nu_{c'}\, d\lambda_{G(c')}(g) \\ &= \mu_{c'}^{\mathrm{sym}}, \end{aligned}$$

where the second equality uses the substitution $g \mapsto h^{-1}g$ and the third uses left-invariance of Haar measure ($\lambda_{G(c')}(h^{-1}A) = \lambda_{G(c')}(A)$ for all Borel A). Hence $\mu_{c'}^{\mathrm{sym}}$ is $G(c')$-invariant.

Part 3: Quotient consistency of the symmetrized lift. Since $\pi_{c'}$ is constant on $G(c')$-orbits, i.e., $\pi_{c'} \circ g = \pi_{c'}$ for all $g \in G(c')$, we have:

$$\begin{aligned} (\pi_{c'})_{\#}\mu_{c'}^{\mathrm{sym}} &= \int_{G(c')} (\pi_{c'})_{\#}(g_{\#}\nu_{c'})\, d\lambda_{G(c')}(g) \\ &= \int_{G(c')} (\pi_{c'} \circ g)_{\#}\nu_{c'}\, d\lambda_{G(c')}(g) \\ &= \int_{G(c')} (\pi_{c'})_{\#}\nu_{c'}\, d\lambda_{G(c')}(g) \\ &= \bar{\mu}_{c'}, \end{aligned}$$

since $\lambda_{G(c')}$ is a probability measure so the integral over it of a λ-independent integrand is just the integrand itself. The last equality uses $(\pi_{c'})_{\#}\nu_{c'} = \bar{\mu}_{c'}$ by assumption. Thus the symmetrized ambient lift induces the same quotient distribution $\bar{\mu}_{c'}$, confirming probability consistency.

Practical interpretation. Parts 1–3 together say: (a) any measurable mode transition preserves total probability mass; (b) the result can always be made symmetry-consistent with the new regime's group $G(c')$ by Haar averaging; and (c) this symmetrization does not alter the quotient-level distribution, so it is free, we pay no statistical cost for restoring symmetry. This is why the **QDM** framework handles hybrid, mode-switching dynamical systems without requiring careful probability-mass bookkeeping at every transition. □

Proposition 3 guarantees that transitions between symmetry strata preserve the probabilistic structure of the model. A measurable transition map $\Psi_{c \mapsto c'}$ pushes any probability measure on the source quotient stratum $\mathcal{Q}(c)$ to a valid probability measure on the target stratum $\mathcal{Q}(c')$. When the target regime imposes a compact symmetry group $G(c')$, Haar averaging further produces a $G(c')$-invariant ambient lift without altering the induced quotient-level distribution. Hence, probability mass remains well defined and symmetry-consistent across mode transitions.

A.3 Proof of Theorem 4

What "well-posedness" means and why it must be proved. A diffusion process is *well-posed* if it admits a unique, non-explosive strong solution. Non-explosion means the process does not blow up to infinity in finite time. Strong solution means the solution is adapted to the same filtration as the driving Brownian motion, not just in distribution. These properties are not automatic: the horizontal projection Π^{soft} introduces a state-dependent, low-rank diffusion coefficient that standard textbook results do not directly cover. In particular, the diffusion matrix is *not* the identity, and the Stratonovich correction terms arising from the state-dependence of Π^{soft} must be accounted for. The proof below verifies that the standard SDE theory applies once the Lipschitz and linear-growth conditions on the projected coefficients are established.

Since $\bar{\mu}_c$ is a probability measure and $\Psi_{c\to c'}$ is measurable, the pushforward measure is well defined by

$$(\Psi_{c\to c'})_{\#}\bar{\mu}_c(A) = \bar{\mu}_c\big(\Psi^{-1}_{c\to c'}(A)\big)$$

for every measurable set $A \subseteq \mathcal{Q}(c')$. In particular,

$$(\Psi_{c\to c'})_{\#}\bar{\mu}_c(\mathcal{Q}(c')) = \bar{\mu}_c(\mathcal{Q}(c)) = 1.$$

Hence $\bar{\mu}_{c'}$ is a probability measure on $\mathcal{Q}(c')$.

For the second claim, compactness of $G(c')$ guarantees the existence of a normalized Haar probability measure $\lambda_{G(c')}$. Since $\mu^{\text{sym}}_{c'}$ is an average of probability measures, it is itself a probability measure. For any $h \in G(c')$,

$$\begin{aligned} h_{\#}\mu^{\text{sym}}_{c'} &= \int_{G(c')} (h\mathrm{g})_{\#}\nu_{c'}\, d\lambda_{G(c')}(\mathrm{g}) \\ &= \int_{G(c')} \mathrm{g}_{\#}\nu_{c'}\, d\lambda_{G(c')}(\mathrm{g}) = \mu^{\text{sym}}_{c'}, \end{aligned}$$

where the second equality follows from the left-invariance of the Haar measure. Thus $\mu^{\text{sym}}_{c'}$ is $G(c')$-invariant. Finally, since $\pi_{c'}$ is constant on $G(c')$-orbits, we have

$$\pi_{c'} \circ \mathrm{g} = \pi_{c'} \quad \forall \mathrm{g} \in G(c').$$

Therefore,

$$\begin{aligned} (\pi_{c'})_{\#}\mu^{\text{sym}}_{c'} &= \int_{G(c')} (\pi_{c'})_{\#}\mathrm{g}_{\#}\nu_{c'}\, d\lambda_{G(c')}(\mathrm{g}) \\ &= \int_{G(c')} (\pi_{c'})_{\#}\nu_{c'}\, d\lambda_{G(c')}(\mathrm{g}) = \bar{\mu}_{c'}. \end{aligned}$$

Hence the symmetrized ambient measure is invariant under the target symmetry group while preserving the same quotient-level probability distribution.

A.4 Proof of Theorem 5

Why a Fisher information decomposition matters? Fisher information quantifies how much statistical information a distribution carries about its parameters. In the score-matching context, higher Fisher information in the vertical (symmetry-orbit) directions means the score network must "use up" capacity estimating motion that is statistically vacuous. Theorem 5 makes this informal statement precise: it shows that the ambient Fisher information decomposes into an intrinsic quotient component (statistically useful) and a vertical component (statistically wasted), and that **QDM**'s horizontal projection removes the vertical component. The theorem also quantifies how much variance reduction remains available when symmetry is only approximate ($\varepsilon_0 > 0$).

Let

$$s_t(x) := \nabla_{\mathscr{X}} \log p_t(x)$$

denote the ambient score. Since the action of G is isometric, the quotient projection induces an orthogonal vertical–horizontal splitting

$$T_x\mathscr{X} = \mathcal{H}_x \oplus \mathcal{V}_x,$$

where $\mathcal{V}_x = \ker d\pi_x$ is tangent to the group orbit and $\mathcal{H}_x = \mathcal{V}_x^{\perp}$ is the quotient-relevant horizontal space. Let $\Pi^{\mathcal{H}}_x$ and $\Pi^{\mathcal{V}}_x$ denote the corresponding orthogonal projections. Then

$$s_t(x) = s^{\mathcal{H}}_t(x) + s^{\mathcal{V}}_t(x), \qquad s^{\mathcal{H}}_t := \Pi^{\mathcal{H}}_x s_t, \quad s^{\mathcal{V}}_t := \Pi^{\mathcal{V}}_x s_t.$$

By orthogonality,

$$\|s_t(x)\|^2 = \|s^{\mathcal{H}}_t(x)\|^2 + \|s^{\mathcal{V}}_t(x)\|^2.$$

Taking expectation with respect to $X_t \sim p_t$, the ambient Fisher information decomposes as

$$I_{\mathscr{X}} := \mathbb{E}\|s_t(X_t)\|^2 = \mathbb{E}\|s_t^{\mathcal{H}}(X_t)\|^2 + \mathbb{E}\|s_t^{\mathcal{V}}(X_t)\|^2.$$

In the exact-symmetry case, the horizontal score is the lift of the intrinsic quotient score, while the vertical component corresponds to orbit-level nuisance variability. Hence,

$$I_{\mathscr{X}} = I_{\mathcal{Q}} + I_{\mathcal{V}},$$

where

$$I_{\mathcal{Q}} := \mathbb{E}\|s_t^{\mathcal{H}}(X_t)\|^2, \qquad I_{\mathcal{V}} := \mathbb{E}\|s_t^{\mathcal{V}}(X_t)\|^2.$$

We now consider the context-dependent soft projection

$$\Pi_x^{\text{soft}}(\tau, c) = \lambda(\tau)\Pi_x^{\mathcal{H}} + \big(1 - \lambda(\tau)\big)I.$$

Since $I = \Pi_x^{\mathcal{H}} + \Pi_x^{\mathcal{V}}$, we equivalently have

$$\Pi_x^{\text{soft}}(\tau, c) = \Pi_x^{\mathcal{H}} + \big(1 - \lambda(\tau)\big)\Pi_x^{\mathcal{V}}.$$

Applying this operator to the score gives

$$s_t^{\text{soft}}(x, \tau, c) := \Pi_x^{\text{soft}}(\tau, c)s_t(x) = s_t^{\mathcal{H}}(x) + \big(1 - \lambda(\tau)\big)s_t^{\mathcal{V}}(x).$$

Again using orthogonality,

$$\|s_t^{\text{soft}}(x, \tau, c)\|^2 = \|s_t^{\mathcal{H}}(x)\|^2 + \big(1 - \lambda(\tau)\big)^2\|s_t^{\mathcal{V}}(x)\|^2.$$

Thus, the Fisher information retained after soft projection is

$$\begin{aligned} \bar{I}_{\mathcal{Q}}^{\text{soft}} := \mathbb{E}\|s_t^{\text{soft}}(X_t, \tau, c)\|^2 &= \mathbb{E}\|s_t^{\mathcal{H}}(X_t)\|^2 \\ &+ \mathbb{E}\Big[\big(1 - \lambda(\tau)\big)^2\|s_t^{\mathcal{V}}(X_t)\|^2\Big]. \end{aligned} \tag{14}$$

Therefore, the amount of Fisher information removed by soft projection is

$$I_{\mathcal{V}}^{\text{rem}} := I_{\mathscr{X}} - \bar{I}_{\mathcal{Q}}^{\text{soft}} = \mathbb{E}\Big[\Big(1 - \big(1 - \lambda(\tau)\big)^2\Big)\|s_t^{\mathcal{V}}(X_t)\|^2\Big].$$

Since

$$1 - \big(1 - \lambda(\tau)\big)^2 = \lambda(\tau)\big(2 - \lambda(\tau)\big),$$

we obtain

$$I_{\mathcal{V}}^{\text{rem}} = \mathbb{E}\big[\lambda(\tau)\big(2 - \lambda(\tau)\big)\|\Pi_x^{\mathcal{V}}\nabla_{\mathscr{X}} \log p_t(X_t)\|^2\big].$$

This expression has the correct limiting behavior: if $\lambda = 1$, all vertical Fisher information is removed; if $\lambda = 0$, no vertical Fisher information is removed and the method recovers ambient diffusion.

It remains to account for approximate symmetry and learned context-dependent projections. Let $\epsilon(\tau) \leq \epsilon_0$ quantify the deviation from exact G-invariance. Under the assumed smoothness of the context encoder, uniform positive definiteness of the regularized Gram matrix, and regularity of the score field, the learned soft projector differs from the ideal soft horizontal projector by a perturbation whose contribution to the Fisher energy is bounded by a second-order residual. Thus, for some constant $C_\lambda > 0$ depending on projection regularity and $\lambda_{\min} := \inf_\tau \lambda(\tau)$,

$$|\Delta(\epsilon_0, \lambda)| \leq C_\lambda \epsilon_0^2 I_{\mathscr{X}}.$$

Consequently,

$$I_{\mathscr{X}} = \bar{I}_{\mathcal{Q}}^{\text{soft}} + I_{\mathcal{V}}^{\text{rem}} + \Delta(\epsilon_0, \lambda),$$

which proves the controlled Fisher decomposition.

Finally, let $\widehat{s}_t$ be an unbiased ambient score estimator with covariance $\Sigma_{\mathscr{X}}$, and define its soft-projected version by

$$\widehat{s}_t^{\text{soft}} = \Pi_x^{\text{soft}}(\tau, c)\widehat{s}_t.$$

Then

$$\text{Cov}(\widehat{s}_t^{\text{soft}}) = \Pi_x^{\text{soft}}(\tau, c)\Sigma_{\mathscr{X}}\Pi_x^{\text{soft}}(\tau, c)^\top.$$

Taking traces and using the orthogonal horizontal–vertical splitting gives

$$\text{Tr}\big[\text{Cov}(\widehat{s}_t^{\text{soft}})\big] \leq \text{Tr}(\Sigma_{\mathscr{X}}) - \lambda_{\min}^2 I_{\mathcal{V}}^{\text{rem}} + C_\lambda \epsilon_0^2 \text{Tr}(\Sigma_{\mathscr{X}}).$$

The first negative term represents the variance removed by suppressing vertical score fluctuations, while the final term accounts for error due to imperfect symmetry and approximate projection. Therefore, strict variance reduction over the ambient estimator is obtained whenever

$$\lambda_{\min}^2 I_{\mathcal{V}}^{\text{rem}} > C_\lambda \epsilon_0^2 \text{Tr}(\Sigma_{\mathscr{X}}).$$

This completes the proof. □

A.4.1 Estimating $I_{\mathcal{V}}^{\text{rem}}$.

In practice, the removable vertical Fisher component is estimated by projecting the learned ambient score onto the vertical symmetry-orbit directions and averaging its squared norm over noisy samples. For a sample x_t at diffusion time t, let

$$V_{x_t}(\tau) = \big[\xi_{\mathscr{X}}^1(x_t), \ldots, \xi_{\mathscr{X}}^k(x_t)\big]$$

denote the vertical generator matrix and define the regularized vertical projector

$$\Pi_{x_t}^{\mathcal{V}}(\tau) = V_{x_t}(\tau)\big(V_{x_t}(\tau)^\top V_{x_t}(\tau) + \varepsilon I_k\big)^{-1} V_{x_t}(\tau)^\top.$$

Given the trained score network $\widehat{s}_\theta(x_t, t)$, we estimate

$$\widehat{I}_{\mathcal{V}}^{\text{rem}} = \frac{1}{N}\sum_{i=1}^{N} \lambda(\tau_i)(2 - \lambda(\tau_i)) \left\|\Pi_{x_{t,i}}^{\mathcal{V}}(\tau_i)\widehat{s}_\theta(x_{t,i}, t_i)\right\|_2^2.$$

Here, $\lambda(\tau)$ accounts for approximate symmetry: when symmetry is exact, $\lambda = 1$ and the full vertical Fisher component is removable; when symmetry is weak, the removable component is down-weighted. Thus, $\widehat{I}_{\mathcal{V}}^{\text{rem}}$ measures how much score energy lies in nuisance symmetry directions that do not affect the quotient state. A larger value indicates that ambient diffusion wastes more learning signal on redundant orbit directions, and therefore QDM is expected to yield stronger variance reduction.

A.5 Proof of Corollary 6

From Theorem 4, the covariance of the soft-projected score estimator

$$\widehat{s}_t^{\text{soft}} = \Pi_x^{\text{soft}}(\tau, c)\widehat{s}_t$$

satisfies

$$\text{Tr}\big[\text{Cov}\big(\widehat{s}_t^{\text{soft}}\big)\big] \leq \text{Tr}(\Sigma_{\mathscr{X}}) - \lambda_{\min}^2 I_{\mathcal{V}}^{\text{rem}} + C_\lambda \epsilon_0^2 \text{Tr}(\Sigma_{\mathscr{X}}),$$

where

$$I_{\mathcal{V}}^{\text{rem}} := \mathbb{E}\big[\lambda(\tau)(2 - \lambda(\tau))\|\Pi_x^{\mathcal{V}} \nabla_{\mathscr{X}} \log p_t(X_t)\|^2\big]$$

is the removable vertical Fisher component under the soft projection.

The ambient estimator has total score-estimation variance

$$\text{Tr}[\text{Cov}(\widehat{s}_t)] = \text{Tr}(\Sigma_{\mathscr{X}}).$$

Therefore, **QDM** achieves strict variance reduction whenever

$$\mathrm{Tr}\big[\mathrm{Cov}\big(\widehat{s}_t^{\mathrm{soft}}\big)\big] < \mathrm{Tr}[\mathrm{Cov}(\widehat{s}_t)]\,.$$

Using the bound above, this strict inequality is guaranteed if

$$\mathrm{Tr}(\Sigma_{\mathscr{X}}) - \lambda_{\min}^2 I_{\mathcal{V}}^{\mathrm{rem}} + C_\lambda \epsilon_0^2 \mathrm{Tr}(\Sigma_{\mathscr{X}}) < \mathrm{Tr}(\Sigma_{\mathscr{X}}).$$

Canceling $\mathrm{Tr}(\Sigma_{\mathscr{X}})$ from both sides gives

$$-\lambda_{\min}^2 I_{\mathcal{V}}^{\mathrm{rem}} + C_\lambda \epsilon_0^2 \mathrm{Tr}(\Sigma_{\mathscr{X}}) < 0,$$

which is equivalent to

$$\lambda_{\min}^2 I_{\mathcal{V}}^{\mathrm{rem}} > C_\lambda \epsilon_0^2 \mathrm{Tr}(\Sigma_{\mathscr{X}}).$$

Thus, the soft-projected estimator has strictly smaller total variance than the ambient estimator whenever the removable vertical Fisher component dominates the approximation error caused by imperfect symmetry.

In the exact-symmetry limit $\epsilon_0 = 0$ with $\lambda_{\min} > 0$, the right-hand side vanishes, so any nonzero removable vertical Fisher component yields strict variance reduction. Conversely, as the symmetry residual ϵ_0 increases, the approximation term grows and the guaranteed gain weakens. Therefore, approximate quotient projection remains beneficial precisely when the symmetry-related variance removed by **QDM** exceeds the error introduced by imperfect symmetry. □

Corollary 6 implies even when symmetry is not perfect, **QDM** can still reduce score-estimation variance as long as the removable symmetry-related variance is larger than the error caused by imperfect symmetry.

A.6 Proof of Proposition 8

Smoothness. The generator vectors $\xi^i_{\mathscr{X}}(x)$ are smooth in x because the group action Φ is smooth by assumption, and the infinitesimal generator is the derivative of Φ with respect to the group parameter at the identity. The context encoder $\phi_\psi : \mathcal{T} \to \mathbb{R}^{d\times k}$ is smooth as a finite-dimensional neural network with smooth activation functions. Their element-wise product $V_x(\boldsymbol{\tau}) = \phi_\psi(\boldsymbol{\tau}) \odot [\xi^1_{\mathscr{X}}(x)|\cdots|\xi^k_{\mathscr{X}}(x)]$ is therefore smooth in both x and $\boldsymbol{\tau}$.

The regularized Gram matrix $G_\varepsilon = V_x^\top V_x + \varepsilon I_k$ is smooth in x and $\boldsymbol{\tau}$ (smooth arithmetic operations on smooth functions). Since its eigenvalues are all bounded below by $\varepsilon > 0$, it is uniformly positive definite. By the matrix implicit function theorem (smooth inverse function theorem applied to the matrix equation $G_\varepsilon M = I_k$), the inverse G_ε^{-1} is smooth. Hence $\Pi_x(\boldsymbol{\tau}) = I_d - V_x G_\varepsilon^{-1} V_x^\top$ is smooth in both arguments, as a composition of smooth operations. The soft projection $\Pi^{\mathrm{soft}} = \lambda\Pi + (1-\lambda)I$ is smooth because $\lambda(\boldsymbol{\tau}) = \exp(-\gamma\hat{\varepsilon}(\boldsymbol{\tau}))$ is smooth in $\boldsymbol{\tau}$.

Constant rank $d-k$. For a free group action, the k generator vectors $\xi^1_{\mathscr{X}}(x), \ldots, \xi^k_{\mathscr{X}}(x)$ are linearly independent at every $x \in \mathscr{X}$: if they were linearly dependent at some x, the stabilizer G_x would be non-trivial, contradicting the freeness assumption. Hence V_x has rank exactly k for all x, so Π_x is the orthogonal projector onto a rank-$(d-k)$ subspace (the orthogonal complement of the k-dimensional column space of V_x). The rank is constant: it does not change with x because the column space of V_x (the vertical space) is a smooth rank-k vector sub-bundle of $T\mathscr{X}$, which follows from freeness and properness of the action.