On the implicit regularization of Langevin dynamics with projected noise

Govind Menon GOVIND_MENON@BROWN.EDU
Division of Applied Mathematics, Brown University and Institute for Advanced Study, Princeton

Austin J. Stromme AUSTIN.STROMME@ENSAE.FR
Department of Statistics, CREST, ENSAE, IP Paris

Adrien Vacher ADRIEN.VACHER@ENSAE.FR
Department of Statistics, CREST, ENSAE, IP Paris

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study Langevin dynamics with noise projected onto the directions orthogonal to an isometric group action. This mathematical model is introduced to shed new light on the effects of symmetry on stochastic gradient descent for over-parametrized models. Our main result identifies a novel form of implicit regularization: when the initial and target density are both invariant under the group action, Langevin dynamics with projected noise is equivalent in law to Langevin dynamics with isotropic diffusion but with an additional drift term proportional to the negative log volume of the group orbit. We prove this result by constructing a coupling of the two processes via a third process on the group itself, and identify the additional drift as the mean curvature of the orbits.

Keywords: implicit regularization; isometric group action; Langevin dynamics

1. Introduction

A central feature of modern machine learning is that models are often heavily over-parameterized yet still achieve excellent generalization performance. For example, when the model has more parameters than training data points, classical statistical theory suggests that in this case one must add regularization to the training method to avoid overfitting. However, this theory is widely known to conflict with practice, where such models can and do generalize well without explicit regularization Zhang et al. (2021). Even when the model has fewer parameters than data points, there is typically still over-parameterization coming from the model architecture itself, in the sense that there are parameter changes which do not affect the model output.

The by now widely accepted explanation for the fact that over-parameterized models can generalize well is that the optimization methods used to train the model *implicitly regularize* it by biasing it towards simpler solutions which generalize better Neyshabur et al. (2015). However, despite significant efforts, the precise nature and mechanism of this implicit regularization is not fully understood. And importantly, most theory for implicit regularization applies either only to gradient descent or equally to stochastic gradient descent (SGD), largely leaving open the question of the effect of stochasticity in particular on implicit regularization. This is a fundamental question due to the ubiquity of stochastic gradient descent in practice.

In this work, we introduce a continuous time model for SGD in the case where the over-parameterization arises from the model architecture itself. The primary motivation for our work is the simple observation that when the model is over-parameterized in this way, *the stochastic gradients will only point in directions which are orthogonal to the over-parameterization*.

Over-parameterization via group symmetries. Our main structural assumption is that the over-parameterization can be described by a group symmetry. For concreteness, suppose that $f(x) = \mathbb{E}_{z\sim\mathcal{P}}[L(x,z)]$ where $L(x,z)$ is the loss incurred on the data $z \sim \mathcal{P}$ with model parameter x. We model the over-parameterization by assuming that there is a group G acting on $\mathbb{R}^d$ such that $L(x,z) = L(g\cdot x, z)$ for all $g \in G$ and $z \in \operatorname{supp}(P)$. Define the orbit $\mathcal{O}_x := \{g \cdot x \colon g \in G\}$. Then the training gradients $\{\nabla_x L(x,z)\colon z \in \operatorname{supp}(P)\}$ will all be orthogonal to the tangent space $T_x\mathcal{O}_x$, and thus so will be any stochastic gradient. Many deep learning models indeed have group symmetries. For example, ReLU units are homogeneous, and attention layers are invariant under certain matrix multiplications. Our setting includes, in particular, some well-known simplified models of over-parameterization reviewed in Section 3.1.

Analyzing implicit bias for SGD in this setting is challenging because it requires a fine-grained understanding of the training dynamics for highly non-convex objectives. We thus propose a simplified continuous-time model of SGD which nonetheless permits us to gain insight into the effect of noise in directions orthogonal to the over-parameterization. In particular, let P_x denote the orthogonal projection on $(T_x\mathcal{O}_x)^\perp$, and $Q_x := I - P_x$ its complement. We consider

$$dX_t = -\nabla f(X_t)dt + \sqrt{2}(\alpha(X_t)P_{X_t} + \beta(X_t)Q_{X_t})dB_t, \tag{1}$$

where α, β control the strength of the projection onto the P and Q directions.

Observe that when $\alpha = \beta = 1$, the diffusion matrix is the identity, and so (1) reduces to standard Langevin dynamics, which is a common simplified model of stochastic gradient descent Raginsky et al. (2017); in this case the stationary distribution is proportional to e^{-f}. By contrast, when the diffusion matrix is not the identity, the stationary distribution can generally only be described indirectly as the solution of a PDE. The goal of this paper is to understand the effect on implicit regularization of noise in directions orthogonal to the over-parameterization by studying (1) in the case where $\alpha \neq \beta$.

Main contribution. Assuming that the group G acts by isometries, our main result, Theorem 3, establishes a novel equivalence between the dynamics (1) and different dynamics with an isotropic diffusion matrix but with an additional entropic drift term. In particular, we show that under certain regularity assumptions on f, α and β, and assuming that the dynamics are initialized at a G-invariant measure, the dynamics (1) are equivalent in marginal law to the dynamics

$$dY_t = -(\nabla f(Y_t) + (\alpha(Y_t)^2 - \beta(Y_t)^2)\nabla \log \operatorname{vol} \mathcal{O}_{Y_t})dt + \sqrt{2}\alpha(Y_t)dB_t,$$

where the orbit volume $\operatorname{vol} \mathcal{O}_y$ is computed with respect to the induced Riemannian geometry on the orbit $\mathcal{O}_y$. In other words, the equivalent dynamics have an isotropic diffusion matrix but an additional drift term proportional to the *negative* gradient of the log volume of $\mathcal{O}_x$. Our result reveals that having more noise in the P versus Q directions leads to a novel type of implicit regularization of the SDE (1) towards points with a small orbit as measured by their embedded volume. Furthermore, our proofs identify that this new phenomenon is essentially geometric: the drift term $\nabla \log \operatorname{vol} \mathcal{O}_x$ is exactly the negative mean curvature of the orbit $\mathcal{O}_x$ at x. In summary, our results shed new light on the relationship between the architecture of a model and implicit regularization, by using tools from geometry and stochastic calculus to identify a novel type of symmetry-specific regularization.

Organization of the paper. This paper is organized as follows. In the remainder of this section we discuss related work. In Section 2 we cover relevant background material from geometry, the theory of group actions, and stochastic calculus. In Section 3, we state our main result, Theorem 3, provide discussion, and consider several examples. In Section 4, we give an overview of our proof of Theorem 3. The Appendix then collects proofs that were omitted from the main text.

Related work. For models trained with a logistic-type loss, there is a large literature dedicated to understanding implicit bias by showing convergence in direction to max-margin solutions; first for logistic regression Soudry et al. (2018) and then for deep linear networks Ji and Telgarsky (2019), homogeneous networks Lyu and Li (2020), and most recently for so-called nearly-homogeneous networks Cai et al. (2025). The square loss setting is not as well understood, with the prevailing model being deep linear neural networks. An influential conjecture Gunasekar et al. (2017) states that certain two layer linear neural networks implicitly minimize nuclear norm; this conjecture was established under some conditions Li et al. (2018). The work Arora et al. (2019) suggested that, for general deep linear networks, implicit regularization is not captured by any matrix norm.

Comparatively less is known about the effects of stochasticity in particular on implicit bias. A widely known concept is that stochastic algorithms favor so-called flat minima Keskar et al. (2017), which tend to generalize better. Due to the difficulty of developing a detailed understanding of stochastic gradient descent dynamics, a common approach, and the one we take in this paper, is to invoke the central limit theorem and resort to an SDE approximation Cheng et al. (2020). An early work studied the implicit bias of such an SDE for linear regression Ali et al. (2020). In the case of diagonal linear neural networks, the implicit bias can be exactly characterized Pesme et al. (2021); Even et al. (2023). For an SDE model of two layer linear neural network training, formulas for the evolution of the singular values and determinant were derived in Varre et al. (2024); in particular, it was shown that the determinant decreases deterministically, in contrast to the gradient flow dynamics, suggesting implicit regularization.

As we discuss more in Section 2, the symmetries of G lead to a natural quotient space $\mathbb{R}^d/G$, and, away from certain degenerate orbits, the quotient map is a Riemannian submersion. Our work is therefore closely related to the study of Langevin dynamics on manifolds in the presence of a Riemannian submersion. Early work studied the image of Brownian motion (with identity diffusion) under a Riemannian submersion and identified the presence of a mean curvature correction in the image space Pauwels (1990); Carne (1990). The Riemannian submersion machinery has been extensively applied to analyze the geometry of certain quotients of Euclidean space in the literature on shape spaces, see Le and Kendall (1993) and subsequent references.

This submersion approach was later used to construct Dyson Brownian motion Huang et al. (2023) (see also Example 2), and Langevin dynamics on the Bures-Wasserstein space Yu et al. (2023), as well as to study the effective dynamics of deep linear networks Menon (2025). In the settings of both Huang et al. (2023); Menon (2025), it was found that for α, β constant, an SDE with anisotropic diffusion matrix $\alpha P + \beta Q$ projects to an SDE in the quotient space with an additional drift of the form $-\beta^2 \nabla \log \operatorname{vol} \mathcal{O}$ and diffusion scaled by α.

Our work is inspired by the works Huang et al. (2023); Menon (2025), but differs from them in the following important ways. First of all, our main observation Theorem 3 is novel. Secondly, rather than passing to the quotient space $\mathbb{R}^d/G$, we work on $\mathbb{R}^d$ and thus use entirely different proof techniques. One benefit of staying on $\mathbb{R}^d$ is that it allows us to avoid technical issues coming from the fact that the quotient space is, in general, a stratified singular space (see Section 2.2). Theorem 3 is also free of stopping times, while those previous works required stopping times to deal with the SDE hitting the boundary of the quotient space. Another important benefit of working on $\mathbb{R}^d$ is conceptual: our coupling based proof of Theorem 3 in Section 4 provides a new and complementary intuition to that coming from the submersion picture which may be of further use. Lastly, as in these previous works, we rely on Proposition 2, connecting the mean curvature to the log-volume, from Carne (1990).

The concurrent work Xu et al. (2026) takes a submersion perspective similar to Huang et al. (2023); Menon (2025) and obtains results closely related to our own. Their main result states that, in the presence of a Lie group H with proper, free isometric action on a Riemannian manifold M, Langevin dynamics with projected noise is equivalent, on the quotient space M/H, to the projection of Langevin dynamics with identity diffusion but an additional drift corresponding to the mean curvature. As we review in Section 2.2, there are no non-trivial subgroups of $O(d)$ with free action on $\mathbb{R}^d$, so their theory does not apply to our setting. Moreover, they neither conclude equivalence at the level of M, nor do they relate the mean curvature to the orbit volume, nor do they interpret the result as implicit regularization. Finally, the concurrent work Aladrah et al. (2026) studies implicit regularization of Langevin dynamics in the presence of possibly non-compact group symmetries, but for an isotropic diffusion matrix. They conjecture an implicit bias term after a certain symmetry breaking map is applied, and propose to design architectures based on the desired regularization.

Notation. We write the set of $d \times d$ real orthogonal matrices $O(d)$, and those with determinant one as $SO(d)$. When discussing an SDE of the form $dZ_t = -A(Z_t)dt + \sqrt{2}B(Z_t)dB_t$ with $Z_t, A(Z_t) \in \mathbb{R}^k$ and $B(Z_t) \in \mathbb{R}^{k \times k}$, we refer to $B(Z_t)$ as the *diffusion matrix*. For a random variable $X \in \mathbb{R}^d$, we write its law $\mathcal{L}(X)$.

2. Preliminaries

In this section, we collect some preliminary material on isometric group actions on $\mathbb{R}^d$, the orbit volume, and stochastic calculus. For general background in Riemannian geometry and the basics of Lie group theory, we refer to the books Lee (2013, 2018); Hall (2013). For background on stochastic calculus, we refer the reader to the books Hsu (2002); Oksendal (2013).

2.1. Isometric group action assumption and background on Lie groups

For a thorough introduction to Lie group theory, we refer to the textbook Hall (2013). A Lie group G is a smooth manifold with a group structure such that the group operations are smooth. The only non-trivial fact we shall use from Lie group theory is that a compact Lie group G has a Haar measure: a measure μ, invariant under right and left-multiplication by group elements, which is unique up to scaling (Lee, 2013, Proposition 16.10). We write the unique Haar measure with $\int_G \mu = 1$ as unif_G.

Throughout this work, we make the following assumption.

Assumption 1 (f is invariant under an isometric group action) *We assume that there exists a closed Lie group $G \subset O(d)$ acting by isometries on $\mathbb{R}^d$, such that $f(g \cdot x) = f(x)$ for all $g \in G$.*

For simplicity, we will identify G with its matrix representation, so one may concretely think of G as an embedded submanifold of matrices $G \subset \mathbb{R}^{d \times d}$, closed under matrix multiplication and inversion, such that $gg^\top = I$ for all $g \in G$. The projection matrix P is defined as the orthogonal projection onto $(T_x\mathcal{O}_x)^\perp$.

2.2. Stratification of $\mathbb{R}^d$ by orbit type

In this work we will use the theory of *orbit types* to deal with certain singular orbits, and to this end here recall some basic elements from the theory of group actions on Riemannian manifolds, following (Berndt et al., 2016, Chapter 2).

The *orbit* of a point $x \in \mathbb{R}^d$ is defined by

$$\mathcal{O}_x := \{g \cdot x \colon g \in G\}.$$

The stabilizer of a point $x \in \mathbb{R}^d$ is

$$G_x := \{g \in G \colon g \cdot x = x\}.$$

For compact groups acting on manifolds, the best case is when the group acts freely, meaning that the stabilizer is the identity, $G_x = \{e\}$, for all x (Lee, 2018, Theorem 21.10). The main benefit of free actions in this case is that the quotient space can be given a Riemannian structure such that the projection map is a Riemannian submersion.

Unfortunately, since the stabilizer of 0 is the whole group, there are no non-trivial subgroups of $O(d)$ with free action on $\mathbb{R}^d$. The lack of free actions means that the quotient space $\mathbb{R}^d/G$ will always have a boundary, and potentially a singular boundary, which rules out a simple Riemannian submersion picture (see also Example 3). In practical terms, this means that the matrices P, Q and the volume $\operatorname{vol} \mathcal{O}_x$ will not be smooth on all of $\mathbb{R}^d$, having instead singularities, which then must be dealt with to establish rigorous theory.

These singularities can be succinctly described by considering *orbit types*. For two orbits $\mathcal{O}_x$ and $\mathcal{O}_y$, we say they have the same type if and only if G_x and G_y are conjugate. This yields a partial order by defining $\mathcal{O}_x < \mathcal{O}_y$ if and only if G_x is conjugate to a subgroup of G_y. With this partial order, there is a unique minimal orbit type, the *principal* orbit type. These orbits have maximal dimension and form an open and dense subset of $\mathbb{R}^d$ (Berndt et al., 2016, Proposition 2.2.4). Orbit types which have maximal dimension but are not principal are termed *exceptional*, and the remaining orbit types are *singular*.

We write the set of principal orbits as $\mathbb{R}^d_{\text{pr}}$, and the set of singular orbits as $\mathbb{R}^d_{\text{sing}}$. We call the union of the principal and exceptional orbits the "regular" orbits, and write it as $\mathbb{R}^d_{\text{reg}}$. Notice that $\mathbb{R}^d_{\text{reg}}$ is exactly the set of orbits of maximal dimension.

The main fact we shall require about orbit types is that the projection operator P (and thus Q) is smooth on the regular orbits $\mathbb{R}^d_{\text{reg}}$.

Proposition 1 (Smoothness of P and Q away from singular orbits) *Suppose $x \in \mathbb{R}^d_{\text{reg}}$. Then there is a smooth orthonormal frame $V_1, \ldots, V_m$, defined on a neighborhood of x, such that for each $y \in U$, $V_1, \ldots, V_m$ spans $T_y\mathcal{O}_y$. In particular, P and Q are smooth in a neighborhood of x, and $\mathbb{R}^d_{\text{reg}}$ is an open set.*

The proof of this result is included in Appendix A.

2.3. Mean curvature and orbit volume

For a regular orbit $\mathcal{O}_x$, we write its volume as an embedded submanifold of $\mathbb{R}^d$ as $\operatorname{vol} \mathcal{O}_x$, and then define $\operatorname{vol} \mathcal{O}_x := 0$ if $\mathcal{O}_x$ is a singular orbit. The orbit volume $\operatorname{vol} \mathcal{O}_x$ is closely related to the geometry of the orbit as an embedded Riemannian manifold, and in particular to its mean curvature.

Recall that for an embedded Riemannian manifold $N \subset \mathbb{R}^d$, its second fundamental form is a tensor which maps $v, w \in T_x N$ to

$$h_x^N(v, w) := P_x^N(\nabla_V W),$$

where V, W are smooth extensions of v, w respectively, and P_x^N is the projection onto the normal space $(T_x N)^\perp$ at x. Using the second fundamental form, one can define the *mean curvature*, a fundamental extrinsic geometric object. Given an orthonormal basis $v_1, \ldots, v_m$ for $T_x N$, the mean curvature is defined to be

$$H^N(x) := \sum_{i=1}^m h_x^N(v_i, v_i).$$

When N is an orbit $\mathcal{O}_x$, we will generally suppress it and simply write the mean curvature and second fundamental form as $H(x)$ and h_x, respectively.

We then have the following important observation on the regularity of $\operatorname{vol} \mathcal{O}_x$ and its connection to mean curvature, which seems to have first appeared in Carne (1990), see also Pacini (2003).

Proposition 2 (Smoothness and gradient of the log-volume) *The orbit volume* $\operatorname{vol} \mathcal{O}_x$ *is smooth on the regular orbits* $\mathbb{R}^d_{\text{reg}}$*. Moreover, on the regular orbits* $\mathbb{R}^d_{\text{reg}}$*, we have the identity*

$$H(x) = -\nabla \log \operatorname{vol} \mathcal{O}_x, \tag{2}$$

where $H(x)$ *is the mean curvature vector of* $\mathcal{O}_x$ *at* x*.*

We give a complete proof of this result in Appendix A.

3. Main result

3.1. Main result and examples

Recall that we assume $f \colon \mathbb{R}^d \to \mathbb{R}$ is a smooth function and G is a Lie group such that f and G satisfy Assumption 1. We write the orthogonal projection onto $(T_x \mathcal{O}_x)^\perp$ as P_x and its complement as $Q_x := I - P_x$. We consider the SDE

$$dX_t = -\nabla f(X_t)dt + \sqrt{2}(\alpha(X_t)P_{X_t} + \beta(X_t)Q_{X_t})dB_t, \tag{3}$$

where $\alpha \colon \mathbb{R}^d \to \mathbb{R}_{\geqslant 0}$ and $\beta \colon \mathbb{R}^d \to \mathbb{R}_{\geqslant 0}$ are non-negative functions which control the strength of the noise in the horizontal versus vertical directions. Our main result applies to the solution of the SDE

$$dY_t = -(\nabla f(Y_t) + (\alpha(Y_t)^2 - \beta(Y_t)^2)\nabla \log \operatorname{vol} \mathcal{O}_{Y_t})dt + \sqrt{2}\alpha(Y_t)dB_t. \tag{4}$$

Theorem 3 (Main result) *Suppose Assumption 1 holds. Suppose* f, α, β *are smooth and* G*-invariant,* ∇f *and* α *are globally Lipschitz,* α *is positive everywhere, and the difference* $\alpha - \beta$ *is compactly supported within the set of regular orbits* $\mathbb{R}^d_{\text{reg}}$*. Suppose further that the SDEs* (3) *and* (4) *are initialized at a* G*-invariant distribution with finite second moment. Then both* (3) *and* (4) *have unique strong solutions* $(X_t)_t, (Y_t)_t$*, and* $\mathcal{L}(X_t) = \mathcal{L}(Y_t)$ *for all* $t \geqslant 0$*.*

We emphasize several aspects of the above result. First and most importantly, the equivalent SDE has an additional drift term proportional to the *negative* gradient of the log volume of $\mathcal{O}_x$. This volume is measured with respect to the geometry of $\mathcal{O}_x$ as an embedded manifold of $\mathbb{R}^d$, and is a novel type of implicit regularization for the SDE (1), biasing the SDE towards points with a small orbit. As noted above, the $\nabla \log \operatorname{vol} \mathcal{O}_x$ term is fundamentally geometric: it is exactly the (negative) mean curvature of the orbits $\mathcal{O}_x$. Second, although the SDE (3) has non-isotropic noise, the equivalent SDE (4) has isotropic diffusion. Third, we assume that the initial distribution is invariant under G. This is necessary for our results, but is mild, as it is satisfied by initializing at a normal distribution with isotropic covariance. Fourth, we assume that $\alpha - \beta$ is compactly supported within the set of regular orbits $\mathbb{R}^d_{\text{reg}}$. This assumption is necessary to apply standard strong well-posedness results for SDEs, as it allows us to avoid the non-smoothness of P, Q and $\nabla \log \operatorname{vol} \mathcal{O}_x$ at the singular orbits.

In the case where α is larger than β, Theorem 3 reveals that the SDE (3) is biased towards points with a small orbit. This identifies an intriguing new type of implicit regularization based on the volume $\operatorname{vol} \mathcal{O}_x$ rather than a more standard norm (see the examples below). The implicit regularization in our model (3) is thus tightly connected to the symmetries of the over-parameterization, and therefore the model architecture, in the sense that different architectures will lead to different symmetry groups with different $\nabla \log \operatorname{vol} \mathcal{O}_x$ terms. Theorem 3 therefore suggests the intriguing possibility that when one is selecting a particular model architecture one is also, implicitly, selecting a particular regularizer.

The intuition behind Theorem 3 is based on the observation that, because (3) is initialized at a G-invariant distribution, it will remain G-invariant at all times. Hence for any process $g_t \in G$ such that $g_t \cdot X_t$ remains G-invariant, the marginal law of $g_t \cdot X_t$ must be that of X_t. By choosing g_t appropriately, we can thus introduce additional noise in the Q directions without changing the marginal law. The $\nabla \log \operatorname{vol} \mathcal{O}_x$ term is the mean curvature of $\mathcal{O}_x$ at x (Proposition 2), and arises from the constraint that g_t remain in G. This proof strategy is developed in detail in Section 4.

Let us now consider several illustrative examples.

Example 1 (Radial symmetries) *The simplest setting is when $G = SO(d)$ acts on $\mathbb{R}^d$ via $U \cdot x = Ux$. The orbits are the spheres $\mathcal{O}_x = \mathbb{S}^{d-1}(\|x\|)$, $\mathbb{R}^d_{\text{reg}} = \mathbb{R}^d \setminus \{0\}$, and the volume is $\operatorname{vol} \mathcal{O}_x = c_d \|x\|^{d-1}$ for a dimension-dependent constant c_d. Functions f which are symmetric under this action are radial.*

Example 2 (Projection onto eigenvalues) *Consider the conjugation by $O(d)$ over symmetric matrices identified with $\mathbb{R}^{d(d+1)/2}$: for all $O \in O(d)$ and $M \in Sym(d)$, $O \cdot M = O^\top M O$. Two matrices are in the same orbit if and only if they have the same eigenvalues. The set of regular orbits $\mathbb{R}^d_{\text{reg}}$ is then exactly the matrices with distinct eigenvalues. In Appendix E, we show that for M with eigenvalues $(\lambda_1, \cdots, \lambda_d)$, $\operatorname{vol} \mathcal{O}_M = c_d \prod_{i<j} |\lambda_i - \lambda_j|$, for a dimension-dependent constant c_d. The class of functions which are symmetric under this action are the spectral functions.*

Example 3 (Bures–Wasserstein case) *Consider the right multiplication by $O(d)$ over $d \times d$ real matrices identified with $\mathbb{R}^{d^2}$: for all $O \in O(d)$ and $M \in \mathbb{R}^{d \times d}$, $O \cdot M = MO$. Two matrices M_1, M_2 are in the same orbit if and only if $M_1^\top M_1 = M_2^\top M_2$. It is straightforward to verify that the regular orbits are rank d matrices, the exceptional orbits are rank $d-1$ matrices, and singular orbits are matrices with rank strictly less than $d-1$. Furthermore, we show in Appendix E that, denoting $(\sigma_1, \ldots, \sigma_d)$ the singular values of M, the volume of the orbit at M is given by $\operatorname{vol} \mathcal{O}_M = c_d \prod_{i<j} \sqrt{\sigma_i^2 + \sigma_j^2}$ with c_d a dimension-dependent constant; related formulas appeared in Carne (1990); Ching-Peng (2023); Yu et al. (2023). In particular, the orbit volume is singular precisely at the singular orbits.*

This group action is closely related to the Bures–Wasserstein manifold on positive-definite matrices, as well as the study of over-parameterization in linear neural networks. Indeed, the quotient space $\mathbb{R}^{d^2}/O(d)$ is exactly the Bures–Wasserstein manifold, which is known to have a singular boundary Massart and Absil (2020). Functions f which are symmetric under this action are those of the form $f(X) = g(XX^\top)$. Such functions have been studied as a model of over-parameterization in neural networks since the influential work Gunasekar et al. (2017), see for example Li et al. (2018); Arora et al. (2019).

Finally, we remark here that the parameterization $XX^\top$ can be generalized to the deep linear network Arora et al. (2018), where one instead considers a product of N real matrices $X_N X_{N-1} \cdots X_1$. This model has $\mathrm{GL}_d(\mathbb{R})$ symmetries, but one may nonetheless apply our model to this case by considering only the subgroup of $O(d)$ symmetries, see also Menon and Yu (2025).

3.2. Identity diffusion and inverse log volume factors

To better illustrate Theorem 3, it is useful to consider the case where $\alpha \equiv 1$ and β is a function of $\log \operatorname{vol} \mathcal{O}_y$, since in this case the stationary distribution of the SDE is explicit.

Corollary 4 (Identity diffusion) *Suppose Asssumption 1 holds, f is smooth and ∇f is globally Lipschitz. Consider the SDE*

$$dX_t = -\nabla f(X_t)dt + \sqrt{2}(P_{X_t} + \beta(X_t) Q_{X_t})dB_t, \tag{5}$$

where $\beta(y) = \varphi(\log \operatorname{vol} \mathcal{O}_y)$ for a smooth function $\varphi \colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}$. Suppose that $\varphi(\mathbb{R}) \subset [c, C]$ for $0 < c < C$ and φ is 1 on a ray $(-\infty, \tau_0]$. Suppose further that the (5) is initialized at a G-invariant distribution with finite second moment. Then there is a unique strong solution to (5), and it has stationary distribution

$$\rho \propto e^{-\int_0^{\log \operatorname{vol} \mathcal{O}_y} (1-\varphi(s)^2)ds} e^{-f(y)}.$$

In particular, if $\varphi = \varepsilon > 0$ on a ray $[\tau_1, \infty)$ with $\tau_1 < 0$, then for any y such that $\log \operatorname{vol} \mathcal{O}_y \geqslant \tau_1$, $\rho(y) = \frac{1}{A}(\operatorname{vol} \mathcal{O}_y)^{\varepsilon^2 - 1} e^{-f(y)}$, where A is a normalizing constant. The regularizing effect of the log volume therefore manifests as an inverse factor in the stationary distribution.

This Corollary imposes slightly different regularity assumptions on β than Theorem 3, so we explain the minor modifications which must be made to the proof of Theorem 3 in Appendix D.

Let us finally remark that, ignoring regularity issues, Corollary 4 suggests that the fully projected SDE

$$dX_t = -\nabla f(X_t)dt + \sqrt{2} P_{X_t} dB_t, \tag{6}$$

has stationary distribution proportional to $\frac{1}{\operatorname{vol} \mathcal{O}_x} e^{-f}$. However, to the best of our knowledge, the well-posedness of the dynamics (6) falls outside known results in SDE theory due to the non-smoothness of the projection P at singular orbits, or equivalently the singularities in the drift of the equivalent dynamics. The only result we are aware of along these lines uses an analysis tailored to shape spaces in certain parameter regimes Le (1994). We therefore leave the rigorous study of (6) for general groups G as an interesting direction for future work.

4. Proof overview: coupling the two SDEs

In this section we overview our proof approach, which is based on introducing a third stochastic process g_t on the group G such that $g_t \cdot X_t$ is a solution to (4). Full details for this proof are provided in Appendix C.

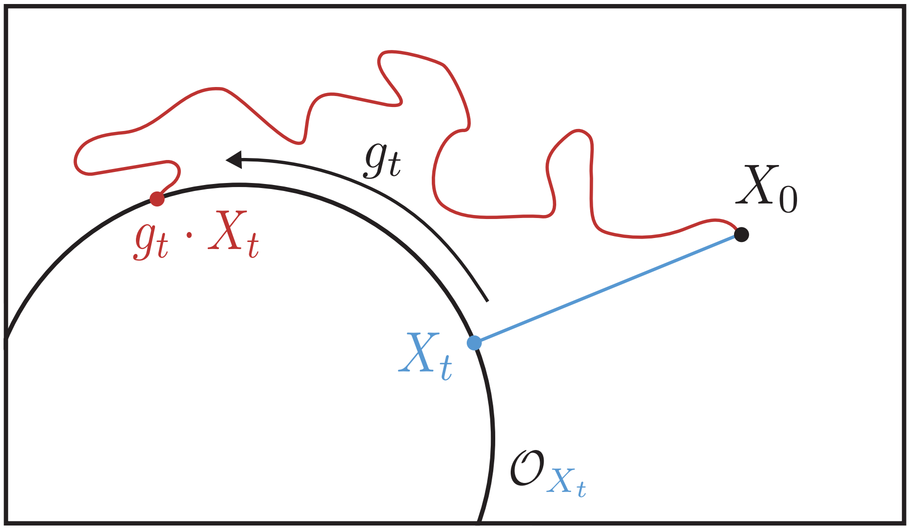

Figure 1: By introducing an appropriate process $g_t \in G$, we can create additional movement in the $T_x\mathcal{O}_x$ directions without changing the G-invariance property, and thus the marginal distributions.

4.1. G-invariance of the SDE and proof approach

The first key idea is that, so long as the SDEs are initialized at a G-invariant distribution, they must remain G-invariant for all times. This observation is formalized in the following Lemma, proved in Appendix B.

Lemma 5 (SDE remains G-invariant) *Suppose f, α, β are as in Theorem 3 and Assumption 1 holds. If (3) is initialized at a distribution which is G-invariant, then the law of X_t remains G-invariant at all times. Similarly, if (4) is initialized at a distribution which is G-invariant, then the law of Y_t remains G invariant at all times.*

The idea of the proof of Theorem 3 is to use the G-invariance of Lemma 5 to design a process g_t with the following properties:

P1. the process $(g_t)_t$ remains on the Lie group G, and

P2. the process $(g_t \cdot X_t)_t$ is a weak solution to the SDE (4).

Given g_t with the above properties, Lemma 5 then implies that $(g_t \cdot X_t)_t$ has the same marginal law as X_t, and thus yields Theorem 3. We essentially introduce rotation to the particle X_t via a process g_t, which will not affect the marginal law, but will still introduce noise in the Q directions so as to make $g_t \cdot X_t$ match with Y_t from (4), see Figure 1.

4.2. Constructing Brownian motion on an orbit

To better explain our construction, we consider a simpler but related question: given a *fixed* point x, how can we construct uniform Brownian motion on the group orbit $\mathcal{O}_x$ of x by a process of the form $g_t \cdot x$ for g_t a stochastic process in G? We thus solve a particular "stochastic geometric control theory" problem. Interestingly, it turns out that the answer to this question is *not* to take g_t as uniform Brownian motion on the group G itself (see Appendix E.4 for a simple example).

To approach this question, it is useful to recall the Itô form of Brownian motion on a manifold embedded in $\mathbb{R}^d$ Lewis (1986); Stroock (2000); Inauen and Menon (2023).

Theorem 6 (Brownian motion on an embedded manifold) *Suppose $M \subset \mathbb{R}^l$ is complete and connected smooth manifold without boundary embedded in $\mathbb{R}^l$. For all $p \in M$, let $H^M(p) \in N_pM$ be the mean curvature vector at p, and let Q_p^M be the projection from $T_p\mathbb{R}^d$ to T_pM. Then the SDE*

$$dZ_t = H^M(Z_t)dt + \sqrt{2}Q_{Z_t}^M dB_t, \qquad Z_0 = p_0 \tag{7}$$

has a unique strong solution and is a Brownian motion on M initialized at $p_0 \in M$.

The idea is then to try and find a process g_t in G such that $g_t \cdot x$ solves (7) for $M = \mathcal{O}_x$. Recall that we regard $g_t \in G$ as an element of $\mathbb{R}^{d\times d}$. Suppose g_t is following a diffusion with (to be specified) drift V and diffusion matrix J:

$$dg_t = V(g_t)dt + \sqrt{2}J(g_t)dB'_t \qquad g_0 = I, \tag{8}$$

where B'_t is an independent Brownian motion on the space $\mathbb{R}^{d\times d}$.

For ease of notation, let $F(g, x) := g \cdot x$, and let $L_{g,x} \colon T_g\mathbb{R}^{d\times d} \to T_{g\cdot x}\mathcal{O}_x$ be the linear map

$$L_{g,x}A := dF_{g,x}[Q_g^G A, 0], \tag{9}$$

where Q_g^G is the projection onto the tangent space of G at g. Assuming that $\mathrm{im}(J(g)) \subset T_gG$, Itô's lemma then implies that $Y_t := F(g_t, x) = g_t \cdot x$ follows

$$dY_t = L_{g_t,x}V(g_t)dt + \sqrt{2}L_{g_t,x}J(g_t)dB'_t. \tag{10}$$

Using the fact that $H(x) = \nabla \log \mathrm{vol}\, \mathcal{O}_x$ from Proposition 2, combined with Theorem 8, we know that Brownian motion on $\mathcal{O}_x$ can be constructed with the following SDE:

$$dZ_t = -\nabla \log \mathrm{vol}\, \mathcal{O}_x dt + \sqrt{2}Q_{Z_t}dB''_t, \tag{11}$$

where $H(Z_t)$ is the mean curvature of the orbit $\mathcal{O}_x$. We thus try and match terms between (10) and (11).

Beginning with the diffusion terms, this suggests we take J such that

$$L_{g_t,x}J(g_t)J(g_t)^\top L_{g_t,x}^\top = Q_{g_t\cdot x}.$$

This can be ensured by taking

$$J_0(g_t) := (L_{g_t,x}^\top L_{g_t,x})^{\dagger/2},$$

where we use $(\cdot)^\dagger$ to denote the pseudo-inverse.

Although the range of J_0 is always in T_gG, this doesn't yet guarantee that the process will remain in G due to the Itô correction. This is similar to the standard construction of Brownian motion on embedded manifolds encapsulated in Theorem 6: even if the drift and noise are tangential to the manifold, the Itô terms mean that the particle could still "fall off" the manifold, and therefore one must add certain terms involving the second fundamental form of G to stay on the manifold. Let h_g^G denote the second fundamental form of G (as an embedded manifold of $\mathbb{R}^{d\times d}$) at $g \in G$. Then put

$$V_0(g) := \mathrm{tr}(h_g^G(L_{g,x}^\dagger \cdot, L_{g,x}^\dagger \cdot)),$$

where we take the trace over an orthonormal basis of $T_{g\cdot x}\mathcal{O}_x$; note that $V_0(g) \in \mathbb{R}^{d\times d}$ is a matrix not a scalar. With this definition, it can be shown that (8), with $V = V_0$ and $J = J_0$ as above, remains on G.

The following observation, proved in Appendix C.1, shows that this drift term is closely related to the second fundamental form, and thus the mean curvature, of the orbit $\mathcal{O}_x$.

Lemma 7 (Relationship between second fundamental forms) *Let $g \in G$ and $x \in \mathbb{R}^d$. Let $h_{g\cdot x}^{\mathcal{O}_x}$ be the second fundamental form of $\mathcal{O}_x$ at $g \cdot x$ and h_g^G be the second fundamental form of G at g. Then for all $A \in T_gG$ we have*

$$h_{g\cdot x}^{\mathcal{O}_x}(L_{g,x}A, L_{g,x}A) = P_{g\cdot x}(h_g^G(A, A)x).$$

This result implies that the mean curvature $H(g\cdot x)$ of $\mathcal{O}_x$ at $g\cdot x$ satisfies $H(g\cdot x) = P_{g\cdot x}(V_0(g)x)$. Taking $V = V_0$ and $J = J_0$, and applying Proposition 2 we obtain

$$dY_t = d(g_t \cdot x) = (-\nabla \log \mathrm{vol}\, \mathcal{O}_{g_t\cdot x} - Q_{g_t\cdot x}(V_0(g_t)x)))dt + \sqrt{2}L_{g_t,x}J(g_t)dB_t'.$$

We therefore nearly have what we wanted: it only remains to get rid of the $Q_{g_t\cdot x}(V_0(g_t)x)$ term. But because this term is in $T_x\mathcal{O}_x$, it is in the image of $L_{g_t,x}$, and so we can remove it by introducing

$$V_1(g_t) := -L_{g_t,x}^\dagger Q_{g_t\cdot x}(V_0(g_t)x).$$

Since V_1 is always in $T_{g_t}G$, including it in the drift will not violate property 1 above. The following result sums up this discussion. A detailed proof is provided in Appendix C.

Theorem 8 (Brownian motion on an orbit) *Let J_0, V_0, V_1 be as above, and consider (8) with $V = V_0 + V_1$ and $J = J_0$. Then g_t is a well-defined stochastic process on $\mathbb{R}^{d\times d}$, which remains in G at all times, and $g_t \cdot x$ is a Brownian motion on $\mathcal{O}_x$ initialized at x.*

To the best of our knowledge, the above construction of Brownian motion on the orbit $\mathcal{O}_x$ is novel.

4.3. The full construction

The full construction uses the same ideas as the above with one significant difference. The above proof ansatz can only *add* noise in Q directions, while Theorem 3 states an equivalence between processes with *a priori* different amounts of noise in the Q directions. The full construction gets around this by proving equivalence of the SDEs (3) and (4) to a third process, with more noise on Q directions than either (3) and (4). Finally, once the full process g_t has been constructed, Theorem 3 follows from Lemma 5, which allows us to reduce to testing against G-invariant test functions. Full details are provided in Appendix C.

Alternative proof by PDE analysis. In Appendix D we also provide an alternative proof by a PDE based analysis. The main advantage of this proof is that it only requires well-posedness of the corresponding weak Fokker–Planck equations and is rather direct. Nonetheless, it does not provide a conceptual explanation for the equivalence of Theorem 3 as does the above approach. It is also interesting to note that the core geometric fact underlying the proof by PDE analysis, Lemma 12, is apparently completely distinct from that of the proof by SDE coupling, Lemma 7.

5. Discussion

Motivated by the study of SGD for over-parameterized models, in this paper we analyzed the SDE (3), which features a diffusion matrix with noise projected by different amounts in normal versus tangential directions to an isometric group action. Our main result, Theorem 3, identifies a novel type of implicit regularization for this SDE, where particles are pushed towards orbits with small volume. We proved this result by constructing a coupling of the SDE (3) with a second SDE on the group G, and critically using the fact that the mean curvature of the orbits is the negative gradient of the log-volume, Proposition 2. Overall, the results in this work provide evidence for a close connection between the symmetries of a model architecture and its implicit regularization, and we believe that further investigations into this connection is an exciting direction for future research.

Acknowledgments

GM was supported by the NSF grant DMS 2407055 and the Erik Ellentuck Fellow Fund at the Institute for Advanced Study, Princeton.

References

Nicola Aladrah, Emanuele Ballarin, Matteo Biagetti, Alessio Ansuini, Alberto d'Onofrio, and Fabio Anselmi. Understanding and inverse design of implicit bias in stochastic learning: a geometric perspective. *arXiv preprint arXiv:2601.06597*, 2026.

Alnur Ali, Edgar Dobriban, and Ryan Tibshirani. The implicit regularization of stochastic gradient flow for least squares. In *International Conference on Machine Learning*, 2020.

Sanjeev Arora, Nadav Cohen, and Elad Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. In *International Conference on Machine Learning*, 2018.

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. In *Advances in Neural Information Processing Systems*, 2019.

Jurgen Berndt, Sergio Console, and Carlos Enrique Olmos. *Submanifolds and holonomy*. CRC Press, 2016.

Vladimir I Bogachev, Nicolai V Krylov, Michael Röckner, and Stanislav V Shaposhnikov. *Fokker–Planck–Kolmogorov Equations*. American Mathematical Society, 2022.

Yuhang Cai, Kangjie Zhou, Jingfeng Wu, Song Mei, Michael Lindsey, and Peter L Bartlett. Implicit bias of gradient descent for non-homogeneous deep networks. In *International Conference on Machine Learning*, 2025.

TK Carne. The geometry of shape spaces. *Proceedings of the London Mathematical Society*, 3(2): 407–432, 1990.

Xiang Cheng, Dong Yin, Peter Bartlett, and Michael Jordan. Stochastic gradient and langevin processes. In *International Conference on Machine Learning*, 2020.

Huang Ching-Peng. A model of invariant control system using mean curvature drift from Brownian motion under submersions. *Quarterly of Applied Mathematics*, 81(1):175–202, 2023.

Denis Constales. A closed formula for the moore-penrose generalized inverse of a complex matrix of given rank. *Acta Mathematica Hungarica*, 80(1):83–88, 1998.

Mathieu Even, Scott Pesme, Suriya Gunasekar, and Nicolas Flammarion. (s) gd over diagonal linear networks: Implicit bias, large stepsizes and edge of stability. In *Advances in Neural Information Processing Systems*, 2023.

Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. In *Advances in Neural Information Processing Systems*, 2017.

Brian C Hall. Lie groups, lie algebras, and representations. In *Quantum Theory for Mathematicians*, pages 333–366. Springer, 2013.

Sigurdur Helgason. *Groups and geometric analysis: integral geometry, invariant differential operators, and spherical functions*, volume 83. American Mathematical Society, 1984.

Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Society, 2002.

Ching-Peng Huang, Dominik Inauen, and Govind Menon. Motion by mean curvature and dyson brownian motion. *Electronic Communications in Probability*, 28:1–10, 2023.

Dominik Inauen and Govind Menon. Stochastic Nash evolution. *arXiv preprint arXiv:2312.06541*, 2023.

Ziwei Ji and Matus Telgarsky. Gradient descent aligns the layers of deep linear networks. In *International Conference on Learning Representations*, 2019.

Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. In *International Conference on Learning Representations*, 2017.

Markus Kunze. *Stochastic Differential Equations*. University of Ulm, Germany, 2012. Lecture notes, Summer Term.

Huiling Le. Brownian motions on shape and size-and-shape spaces. *Journal of applied probability*, 31(1):101–113, 1994.

Huiling Le and David G Kendall. The riemannian structure of euclidean shape spaces: a novel environment for statistics. *The annals of Statistics*, 21(3):1225–1271, 1993.

John M Lee. *Introduction to Smooth Manifolds*, volume 218. Springer Science & Business Media, 2013.

John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.

JT Lewis. Brownian motion on a submanifold of euclidean space. *Bulletin of the London Mathematical Society*, 18(6):616–620, 1986.

Yuanzhi Li, Tengyu Ma, and Hongyang Zhang. Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In *Conference On Learning Theory*, 2018.

Kaifeng Lyu and Jian Li. Gradient descent maximizes the margin of homogeneous neural networks. In *International Conference on Learning Representations*, 2020.

Estelle Massart and P-A Absil. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM J. Matrix Anal. Appl.*, 41(1):171–198, 2020.

Govind Menon. The geometry of the deep linear network. In *XIV Symposium on Probability and Stochastic Processes*, volume 81 of *Progr. Probab.*, pages 1–47. Birkhäuser/Springer, 2025.

Govind Menon and Tianmin Yu. An entropy formula for the deep linear network. *arXiv preprint arXiv:2509.09088*, 2025.

Behnam Neyshabur, Ryota Tomioka, and Nathan Srebro. In search of the real inductive bias: On the role of implicit regularization in deep learning. In *International Conference on Learning Representations, Workshop Track Proceedings*, 2015.

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

Tommaso Pacini. Mean curvature flow, orbits, moment maps. *Transactions of the American Mathematical Society*, 355(8):3343–3357, 2003.

JE Pauwels. Riemannian submersions of brownian motions. *Stochastics: An International Journal of Probability and Stochastic Processes*, 29(4):425–436, 1990.

Scott Pesme, Loucas Pillaud-Vivien, and Nicolas Flammarion. Implicit bias of sgd for diagonal linear networks: a provable benefit of stochasticity. In *Advances in Neural Information Processing Systems*, 2021.

Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In *Conference on Learning Theory*, 2017.

Leon Simon. Introduction to geometric measure theory. *Tsinghua lectures*, 2(2):3–1, 2014.

Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *Journal of Machine Learning Research*, 19 (70):1–57, 2018.

Daniel W Stroock. *An introduction to the analysis of paths on a Riemannian manifold*. Number 74. American Mathematical Society, 2000.

Aditya Vardhan Varre, Margarita Sagitova, and Nicolas Flammarion. Sgd vs gd: Rank deficiency in linear networks. In *Advances in Neural Information Processing Systems*, 2024.

Yixian Xu, Yusong Wang, Shengjie Luo, Kaiyuan Gao, Tianyu He, Di He, and Chang Liu. Quotient-space diffusion model. In *International Conference on Learning Representations*, 2026.

Tianmin Yu, Shixin Zheng, Jianfeng Lu, Govind Menon, and Xiangxiong Zhang. Riemannian langevin monte carlo schemes for sampling psd matrices with fixed rank. *arXiv preprint arXiv:2309.04072*, 2023.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

Appendix A. Omitted proofs from Section 2

Proof [Proof of Proposition 1] Let $A_1, \dots, A_m \in \mathfrak{g}$ be such that $A_1 x, \dots, A_m x$ span $T_x\mathcal{O}_x$. For $y \in \mathbb{R}^d$, consider the vector fields $V_1(y) := A_1 y, \dots, V_m(y) := A_m y$. We first claim that by continuity, $(V_i(y))_{i=1}^m$ must remain linearly independent for y sufficiently close to x. Indeed, consider the $d \times m$ matrix B_y whose ith column is $V_i(y)$. Then B_y has rank m if and only if $\det(B_y^\top B_y) \neq 0$, and the latter is a continuous condition which holds at $y = x$, so must hold in a neighborhood around x.

Therefore, $(V_i(y))_{i=1}^m$ must remain linearly independent for y in some neighborhood of x. Notice that $V_i(y) \in T_y\mathcal{O}_y$ for each i, and by the fact that m is the maximal orbit dimension they must in fact span $T_y\mathcal{O}_y$. We may then take their Graham-Schmidt orthonormalization to obtain $(W_i(y))_{i=1}^m$ which are orthonormal and span $T_y\mathcal{O}_y$. Because the V_i are smooth functions of y, so are W_i, implying the first claim. But we can write

$$Q_y = \sum_{i=1}^m W_i(y) W_i(y)^\top, \qquad P_y = I - Q_y,$$

yielding the result. ∎

Proof [Proof of Proposition 2] We first establish smoothness, before turning to the identity $H_x = \nabla \log \operatorname{vol} \mathcal{O}_x$.

To show that $\operatorname{vol} \mathcal{O}_x$ is smooth, we first claim that there is a unique smooth structure and geometry on $\pi(\mathbb{R}^d_{\mathrm{pr}}) \subset \mathbb{R}^d/G$ such that the quotient map $\pi\colon \mathbb{R}^d \to \mathbb{R}^d/G$ is a smooth submersion on $\mathbb{R}^d_{\mathrm{pr}}$. Indeed, around any principal orbit there is a geodesic slice $\Sigma := \{x + v\colon v \in (T_x\mathcal{O}_x)^\perp,\ \|v\| < \varepsilon\}$ for some $\varepsilon > 0$, which has the property that for all $y \in \Sigma$, we have $G_x = G_y$ (Berndt et al., 2016, Section 2.1.6). The quotient G/G_x is a smooth manifold (Lee, 2013, Theorem 21.17), and there is then a diffeomorphism $G \cdot \Sigma \to G/G_x \times \Sigma$. Using this diffeomorphism, we can take coordinates $\varphi\colon U \to \varphi(U)$ around x with the property that $\varphi(U)$ is an open cube in $\mathbb{R}^m \times \mathbb{R}^n$ and an orbit intersects U either not at all or in a slice of the last n coordinates. With these coordinates, the rest of the proof goes in the standard way (Lee, 2013, Theorem 21.20); we sketch the details.

Let $V := \pi(\Sigma)$; because $G_x = G_y$ for all $y \in \Sigma$, $\pi|_\Sigma$ is a bijection onto its image. We then use the previous coordinates to define the smooth structure at V by first passing through $\pi|_\Sigma^{-1}$. In these coordinates, π becomes a projection onto its last n coordinates, so is certainly smooth. All that remains is to verify that the smooth structure is compatible, but this follows from the fact that the transition maps respect the orbits: points are in the same orbit if and only if they have the same last n coordinates, and so the transition maps are smooth when restricted to a slice of the first m coordinates. With this structure, $\pi|_{\mathbb{R}^d_{\text{pr}}}$ becomes a diffeomorphism onto its image. Using the maps $\psi := \pi|_\Sigma^{-1}$ we define the inner product of two tangent vectors $u, v \in T_p\mathbb{R}^d/G$ via $\langle u, v\rangle_p := \langle d\psi_p[u], d\psi_p[v]\rangle$; this definition is independent of the choice of ψ because G acts by isometries. This smooth structure and geometry is unique by the characteristic property of smooth Riemannian submersions.

By abuse of notation, we write $\operatorname{vol}\mathcal{O}_p\colon \mathbb{R}^d/G \to \mathbb{R}$ for $\operatorname{vol}\mathcal{O}_{\pi^{-1}(p)}$. With this geometry, we know that $\operatorname{vol}\mathcal{O}_p$ is a smooth function on the image $\pi(\mathbb{R}^d_{\text{pr}})$, as it is simply the volume of the fibers of a smooth Riemannian submersion whose fibers are compact, so is smooth. By composition with π, we find that $\operatorname{vol}\mathcal{O}_x$ is smooth at $x \in \mathbb{R}^d_{\text{pr}}$.

Let us now turn to the identity $\nabla \log \operatorname{vol}\mathcal{O}_x = -H(x)$. Let $\bar{h}(x)\colon \mathbb{R}^d \to \mathbb{R}$ be a smooth, G-invariant function compactly supported within $\mathbb{R}^d_{\text{pr}}$, and write $\bar{h} = h \circ \pi$. By the previous part of the proof, we know that the quotient map $\pi\colon \mathbb{R}^d \to \mathbb{R}^d/G$ is a smooth Riemannian submersion from $\mathbb{R}^d_{\text{pr}}$ to $M := \pi(\mathbb{R}^d_{\text{pr}})$, and since $\bar{h}$ is supported in this set, the smooth co-area formula (see, e.g. (Simon, 2014, Section 2.6)) yields

$$\int \bar{h}(x)dx = \int h(p)\operatorname{vol}\mathcal{O}_p d\operatorname{vol}_M(p).$$

The idea is to use this fact combined with integration by parts on M. Consider a smooth vector field $\bar{V}$ on $\mathbb{R}^d$, compactly supported within $\mathbb{R}^d_{\text{pr}}$. Assume for now that $\bar{V} \in (T_x\mathcal{O}_x)^\perp$ and $\bar{V}$ is G-equivariant, so that $\bar{V}_{g\cdot x} = g\bar{V}_x$. Then there exists a smooth vector field V on $M := \pi(\mathbb{R}^d_{\text{pr}})$ such that $V_{\pi(x)} = d\pi_x[\bar{V}_x]$ for all $x \in \mathbb{R}^d_{\text{pr}}$. We have

$$\operatorname{div}(\bar{V}) = \operatorname{div}(V) \circ \pi - \langle \bar{V}, H_x\rangle.$$

We include a proof of this fact for the reader's convenience as Lemma 9 below. Using this formula with the divergence theorem, we find that

$$\begin{aligned}\int \langle \bar{V}, \nabla \log \operatorname{vol}\mathcal{O}_x\rangle dx &= \int \langle V, \nabla \operatorname{vol}\mathcal{O}_p\rangle d\operatorname{vol}_M(p) = -\int \operatorname{div}(V)\operatorname{vol}\mathcal{O}_p d\operatorname{vol}_M(p) \\ &= -\int (\operatorname{div}(\bar{V}) + \langle \bar{V}, H_x\rangle)dx = -\int \langle \bar{V}, H_x\rangle dx.\end{aligned}$$

Given a general smooth vector field W, compactly supported in $\mathbb{R}^d_{\text{pr}}$, we can transform it into a G-equivariant vector field in $(T_x\mathcal{O}_x)^\perp$ by taking $\bar{V}(x) := \int g^\top W_{g\cdot x} d\operatorname{unif}_G(g)$, where $\operatorname{unif}_G(g)$ denotes the normalized Haar measure on G. Indeed, since $\nabla \log \operatorname{vol}\mathcal{O}_x \in (T_x\mathcal{O}_x)^\perp$ we can assume W is too, and notice that since g acts by isometries we can use Proposition 10 to find that

$$\int \langle W(x), \nabla \log \operatorname{vol}\mathcal{O}_x\rangle dx = \iint \langle W(g\cdot x), \nabla \log \operatorname{vol}\mathcal{O}_{g\cdot x}\rangle dx d\operatorname{unif}_G(g) = \int \langle \bar{V}(x), \nabla \log \operatorname{vol}\mathcal{O}_x\rangle dx.$$

Using the analogous formula for $\int \langle W, H_x \rangle dx$, we obtain

$$\int \langle W, \nabla \log \mathrm{vol}\, \mathcal{O}_x + H_x \rangle dx = 0,$$

for arbitrary smooth vector fields W supported in $\mathbb{R}^d_{\mathrm{pr}}$. Proposition 1 implies that the mean curvature is a smooth function on $\mathbb{R}^d_{\mathrm{reg}}$, and so the formula holds on $\mathbb{R}^d_{\mathrm{reg}}$ since $\mathbb{R}^d_{\mathrm{pr}}$ is dense in $\mathbb{R}^d$. ■

Lemma 9 (Divergence of horizontal vector field) *Suppose $\bar{V}$ is a smooth vector field on an open set $S \subset \mathbb{R}^d_{\mathrm{pr}}$, such that $\bar{V}_x \in (T_x\mathcal{O}_x)^\perp$ for all $x \in S$, and assume there is a smooth vector field V on $\pi(S)$ such that $V_{\pi(x)} = d\pi_x[\bar{V}]$. Then*

$$\mathrm{div}(\bar{V}) = \mathrm{div}(V) \circ \pi - \langle \bar{V}, H \rangle,$$

where H is the mean curvature vector field for the orbits $\mathcal{O}_x$.

Proof [Proof of Lemma 9] Let $(\bar{A}_i)_{i=1}^m$ be a smooth local orthonormal frame for $(T_x\mathcal{O}_x)^\perp$ and $(B_j)_{j=1}^n$ be a smooth local orthonormal frame for $T_x\mathcal{O}_x$. Using the fact, established in the proof of Proposition 2, that the projection $\pi \colon \mathbb{R}^d_{\mathrm{pr}} \to \pi(\mathbb{R}^d_{\mathrm{pr}})$ is a smooth Riemannian submersion, we find that

$$\begin{aligned}
\mathrm{div}(\bar{V}) &= \sum_{i=1}^m \langle \nabla_{\bar{A}_i} \bar{V}, \bar{A}_i \rangle + \sum_{j=1}^n \langle \nabla_{B_j} \bar{V}, B_j \rangle\,, \\
&= \sum_{i=1}^m \langle d\pi_x[\nabla_{\bar{A}_i} \bar{V}], d\pi_x[\bar{A}_i] \rangle_{\pi(x)} + \sum_{j=1}^n \langle \nabla_{B_j} \bar{V}, B_j \rangle\,.
\end{aligned}$$

Now, let $\tilde{\nabla}$ be the Levi-Civita connection on $\pi(\mathbb{R}^d_{\mathrm{pr}})$ and write $A_i := d\pi[\bar{V}_i]$. By (Lee, 2018, Problem 5-6(b)), we have $d\pi[\nabla_{\bar{A}_i} \bar{V}] = \tilde{\nabla}_{A_i} V$. Hence the first term is exactly $\sum_{i=1}^m \langle \tilde{\nabla}_{A_i} V, A_i \rangle_{\pi(x)} = \mathrm{div}(V) \circ \pi$. For the second term, we may calculate

$$\begin{aligned}
\sum_{j=1}^n \langle \nabla_{B_j} \bar{V}, B_j \rangle &= \sum_{j=1}^n B_j \cdot \langle \bar{V}, B_j \rangle - \sum_{j=1}^n \langle \bar{V}, \nabla_{B_j} B_j \rangle\,, \\
&= -\langle \bar{V}, P(\sum_{j=1}^n \nabla_{B_j} B_j)) \rangle\,, \\
&= -\langle \bar{V}, H_x \rangle\,,
\end{aligned}$$

where in the second line we used the fact that $\bar{V} \in (T_x\mathcal{O}_x)^\perp$ twice. ■

Appendix B. Dynamics remain G-invariant

In this section we first collect several useful calculations for G-invariant functions, and then prove Lemma 5 on the fact that the dynamics remain G-invariant.

B.1. Useful calculations

We first collect the following useful calculations, which describe how gradients and the projection matrices transform under the action of G.

Proposition 10 (Gradient of G-invariant functions) *Let h be invariant under the action of G. Then for all $g \in G$ and all $x \in \mathbb{R}^d$, it holds that $g \cdot \nabla h(x) = \nabla h(g \cdot x)$.*

Proof [Proof of Proposition 10] For all $g \in G$ and $x \in \mathbb{R}^d$, we have $h(g \cdot x) = h(x)$. Thus

$$\nabla h(x) = g^\top \nabla h(g \cdot x).$$

But since $g \in O(d)$ by Assumption 1, $g^\top = g^{-1}$. Hence the result. ■

Proposition 11 (Transformation of horizontal projection) *Let P_x be the horizontal projection onto the orbit $\mathcal{O}_x$ and $Q_x := I - P_x$. Then for all $x \in \mathbb{R}^d$ and $g \in G$:*

$$gP_x = P_{g \cdot x} g, \qquad gQ_x = Q_{g \cdot x} g\,.$$

Proof [Proof of Proposition 11] Observe that if $\mathfrak{g}$ is the Lie algebra of G, then $T_x \mathcal{O}_x = \{Ax \colon A \in \mathfrak{g}\}$. But notice that, for each $g \in G$, $\{gAg^{-1} \colon A \in \mathfrak{g}\} = \mathfrak{g}$, so that for any $A \in \mathfrak{g}$ and $v \in \mathbb{R}^d$

$$\langle g^{-1} P_{g \cdot x} g v, Ax \rangle = \langle P_{g \cdot x} g v, gAx \rangle = \langle gv, P_{g \cdot x} gAx \rangle = \langle gv, P_{g \cdot x} gAg^{-1}(g \cdot x) \rangle = 0,$$

where the first equality follows by the fact that $g \in O(d)$ by Assumption 1, the second equality is by symmetry of P since it is an orthogonal projection, and the fourth equality is because $gAg^{-1} \in \mathfrak{g}$, so that $gAx = gAg^{-1}(gx) \in T_{gx}\mathcal{O}_x$. On the other hand, it is clear that $L = g^{-1} P_{g \cdot x} g$ is such that $L^2 = L$; and finally that $\dim L = \dim \mathcal{O}_x$. This implies that $g^{-1} P_{g \cdot x} g = P_x$. ■

B.2. Proof of Lemma 5

Proof [Proof of Lemma 5] Under our assumptions, Equation (3) has a strong solution by standard SDE theory (Oksendal, 2013, Theorem 5.2.1) and the solution is weakly unique (Oksendal, 2013, Lemma 5.3.1). It then suffices to show that $g \cdot X_t$ is a weak solution of (3) with the same initial condition for all $g \in G$, since by weak uniqueness it must be the same in law as X_t itself, establishing G-invariance.

Indeed, by G-invariance and Itô's lemma, we may calculate that for any $g \in G$, we have

$$d(g \cdot X_t) = -g \cdot \nabla f(X_t) dt + g \cdot \sqrt{2}(\alpha(X_t) P_{X_t} + \beta(X_t) Q_{X_t}) dB_t.$$

By Proposition 10 and Proposition 11 this is

$$d(g \cdot X_t) = -\nabla f(g \cdot X_t) dt + \sqrt{2}(\alpha(X_t) P_{g \cdot X_t} + \beta(X_t) Q_{g \cdot X_t}) g dB_t.$$

Finally using G-invariance of α and β, we obtain

$$d(g \cdot X_t) = -\nabla f(g \cdot X_t) dt + \sqrt{2}(\alpha(g \cdot X_t) P_{g \cdot X_t} + \beta(g \cdot X_t) Q_{g \cdot X_t}) g dB_t.$$

But since $g \in O(d)$, it follows that gdB_t is a Brownian motion, and hence $g \cdot X_t$ is a weak solution of (3). Because the initial condition is G-invariant, it follows that X_t and $g \cdot X_t$ are weak solutions of the same SDE and thus must have the same marginal laws. This proves the first claim in the Lemma.

The proof of the second claim is analogous: by Proposition 2, $\nabla \log \operatorname{vol} \mathcal{O}_x$, is smooth away from the singular orbits. Given our assumption on $\alpha - \beta$ is follows that $(\alpha^2 - \beta^2)\nabla \log \operatorname{vol} \mathcal{O}_x$ is smooth and compactly supported, so is globally Lipschitz. Hence the drift in Equation (4) is globally Lipschitz, and the diffusion matrix is globally Lipschitz by assumption. Applying again (Oksendal, 2013, Theorem 5.2.1) and (Oksendal, 2013, Lemma 5.3.1) we find there exists a strong solution to (4) and it is weakly unique. The remainder of the argument is identical to the above and so is omitted. ■

Appendix C. Proof by SDE construction

C.1. Proof of second fundamental form identity

Proof [Proof of Lemma 7] We first compute the second fundamental form $h_g^G(A, A)$, for $A \in T_gG$. Let us begin by considering the case where $g = I$. Given any smooth vector field W on $\mathbb{R}^d$ with $W_I = A$ and such that $W_g \in T_gG$ for all $g \in G$, the second fundamental form is

$$h_I^G(A, A) = P_I^G(\nabla_W W),$$

where P_I^G is the projection form $T_I\mathbb{R}^d$ to the normal space N_IG and ∇ denotes the Euclidean connection. We may therefore take W to be $W_X := XA$ for all $X \in \mathbb{R}^{d\times d}$, which will be a smooth vector field on G and such that $W_I = A$. Then

$$(\nabla_W W)_I = \lim_{t\to 0} \frac{W_{I+tA} - W_I}{t} = A^2.$$

But then notice that, since $G \subset O(d)$, A must be anti-symmetric, so that in fact the above is a symmetric matrix. Since T_gG is a subset of anti-symmetric matrices, it follows that $\nabla_W W \in N_IG$ already. Hence

$$h_I^G(A, A) = A^2.$$

We may then deduce the general formula by the fact that $g\colon \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ which takes $X \mapsto gX$, is an isometry so that for $B \in T_gG$, we have

$$h_g^G(B, B) = g h_I^G(g^{-1}B, g^{-1}B) = Bg^{-1}B.$$

Next, we compute the second fundamental form of $\mathcal{O}_x$. Here, we know that each $v \in T_x\mathcal{O}_x$ is of the form Ax for $A \in T_eG$. We can then extend v to the vector field $V(x') := Ax'$, and again obtain

$$(\nabla_V V)_x = \lim_{t\to 0} \frac{V(x + tAx) - V(x)}{t} = A^2 x.$$

Hence if $v = Ax$ then

$$h_x^{\mathcal{O}_x}(v, v) = P_x(\nabla_V V)_x = P_x(A^2 x) = P_x(h_I^G(A, A)x) = P_x(g^{-1}h_g^G(gA, gA)x),$$

yielding the result. Applying Proposition 11, this implies

$$gh_x^{\mathcal{O}_x}(v,v) = P_{g\cdot x}(h_g^G(gA, gA)x).$$

Since $g\colon \mathbb{R}^d \to \mathbb{R}^d$ is an isometry the left-hand-side becomes

$$h_{g\cdot x}^{\mathcal{O}_x}(g\cdot v, g\cdot v) = P_{g\cdot x}(h_g^G(gA, gA)x).$$

Because v was an arbitrary element of $T_x\mathcal{O}_x$, the result follows. ■

C.2. Proof of Theorem 8

Proof [Proof of Theorem 8] Let $L_{g_t,x}$ be as defined in (9). Put

$$\begin{aligned} J_0(g) &:= (L_{g_t,x}^\top L_{g_t,x})^{\dagger/2} \\ V_0(g) &:= \operatorname{tr}(h_g^G(L_{g,x}^\dagger, L_{g,x}^\dagger)), \\ V_1(g) &:= -L_{g,x}^\dagger Q_{g\cdot x}(V_0(g)x), \end{aligned}$$

and consider the SDE

$$dg_t := (V_0(g_t) + V_1(g_t))dt + \sqrt{2}J_0(g)dB_t', \qquad g_0 = I. \tag{12}$$

We first address the question of existence and uniqueness of a strong solution to this SDE. To begin, we claim that J_0, V_0, V_1 are all smooth on G. Indeed, $L_{g,x}$ is a smooth function of g, and for all $g \in G$ it has rank $\dim \mathcal{O}_x$, so that $L_{g,x}^\dagger$ and $J_0(g)$ are smooth Constales (1998). Hence V_0 is also smooth on G. Similarly, $Q_{g\cdot x}$ is a smooth function of g, and thus V_1 is smooth as well. Hence J_0, V_0, V_1 are smooth on G, and thus globally Lipschitz on G by compactness. With a minor abuse of notation, we may then assume that J_0, V_0, V_1 have been extended to globally Lipschitz functions on $\mathbb{R}^{d\times d}$. Equation (12) then indeed has a unique strong solution (Oksendal, 2013, Theorem 5.2.1).

We next verify that g_t remains in G. Let $Z\colon \mathbb{R}^{d\times d} \to \mathbb{R}$ be a smooth function such that $Z(g) = 0$ if and only if $g \in G$ (e.g. a smoothing of $d^2(x, G)$). Then observe that $\nabla Z(g) \in (T_gG)^\perp$ for all $g \in G$, and thus for all $v \in T_gG$,

$$\nabla^2 Z_g[v,v] + \langle h_g^G(v,v), \nabla Z(g)\rangle = 0. \tag{13}$$

Using Itô's Lemma, we compute

$$dZ(g_t) = (\langle \nabla Z(g_t), \operatorname{tr}(h_{g_t}^G(L_{g_t,x}^\dagger\cdot, L_{g_t,x}^\dagger\cdot))\rangle + \operatorname{tr}(\nabla^2 Z(g_t)J_0(g_t)^2))dt,$$

where the V_1 and diffusion terms vanish because they are in $T_{g_t}G$ and thus in the kernel of dZ_{g_t}. Using the SVD decomposition, we may write

$$L_{g_t,x} = \sum_{i=1}^m \sigma_i u_i \langle A_i, \cdot\rangle,$$

where $(u_i)_{i=1}^m$ are orthonormal vectors spanning $T_{g_t\cdot x}\mathcal{O}_x$ and $(A_i)_{i=1}^m$ are orthonormal vectors in $T_{g_t}G$. Writing the above using this basis, we find

$$\langle \nabla Z(g_t), \operatorname{tr}(h_{g_t}^G(L_{g_t,x}^\dagger\cdot, L_{g_t,x}^\dagger\cdot))\rangle = \sum_{i=1}^m \frac{1}{\sigma_i^2}\langle \nabla Z(g_t), h_{g_t}^G(A_i, A_i)\rangle,$$

and

$$\operatorname{tr}(\nabla^2 Z(g_t) J_0(g_t)^2) = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \nabla^2 Z_{g_t}[A_i, A_i].$$

Applying (13), we thus find that $dZ(g_t) = 0$, and so combined with the initialization at $g_0 = I$ we conclude that g_t remains in G at all times.

We now verify that $(g_t \cdot x)_t$ is a Brownian motion on $\mathcal{O}_x$. We may calculate using Itô's formula and the definitions that

$$\begin{aligned} d(g_t \cdot x) &= L_{g_t,x}(V_0(g_t) + V_1(g_t))dt + \sqrt{2} L_{g_t,x} J_0(g_t) dB'_t \\ &= P_{g_t,x}(V_0(g_t)x)dt + \sqrt{2} L_{g_t,x} J_0(g_t) dB'_t. \end{aligned}$$

By Lemma 7, this is exactly

$$d(g_t \cdot x) = H(g_t \cdot x)dt + \sqrt{2} L_{g_t,x} J_0(g_t) dB'_t.$$

Notice that

$$h_{g_t,x} J_0(g_t) J_0(g_t)^\top L_{g_t,x}^\top = Q_{g_t \cdot x}.$$

Hence by Theorem 8 we have that $(g_t \cdot x)$ has the same drift and diffusion as a Brownian motion on $\mathcal{O}_x$. Under our assumptions, weak uniqueness holds for (11) (Bogachev et al., 2022, Theorem 9.4.3), and so the marginal law of $(g_t \cdot x)_t$ must therefore be identical to that of Brownian motion on $\mathcal{O}_x$. ■

C.3. Full details of proof by constructing a stochastic process on G

Proof [Proof of Theorem 3] *Proof approach.* Let $S := \operatorname{supp}(\alpha - \beta)$. Since S is compact and contained within the set of regular orbits $\mathbb{R}^d_{\text{reg}}$, which is open by Proposition 1, there exists a smooth bump function ψ which is 1 on S but supported on a compact set within $\mathbb{R}^d_{\text{reg}}$ (Lee, 2013, Proposition 2.25). Let $\varphi(x) := \mathbb{E}_{g \sim \text{unif}_G}[\psi(g \cdot x)]$, so that it is a G-invariant bump function for S.

Using this bump function φ, the proof of Theorem 3 by stochastic calculus consists in showing that the two stochastic processes (3) and (4) are equivalent in marginal law to a *third* process:

$$dZ_t = -(\nabla f + \varphi^2(\alpha^2 + \beta^2)\nabla \log \operatorname{vol} \mathcal{O})dt + \sqrt{2}(\alpha P + \sqrt{\beta^2 + \varphi^2(\alpha^2 + \beta^2)} Q) dB_t, \tag{14}$$

where we suppress Z_t in the right-hand side. We use this auxiliary process because the ansatz of multiplying $g_t \cdot X_t$ can only increase the amount of noise in the Q directions, at least when the Brownian motion of g_t is independent of that of X_t.

X_t, Y_t, and Z_t have unique strong solutions. We begin by checking that both (3) and (4) have a unique strong solution; to this end, let $S := \operatorname{supp}(\alpha - \beta)$. First consider X_t, namely (3). The drift ∇f is globally Lipschitz by assumption, and on S^c, $\alpha = \beta$, so the diffusion matrix is αI which is globally Lipschitz by assumption. Since $S \subset \mathbb{R}^d_{\text{reg}}$, Proposition 1 implies that the matrix $\alpha P + \beta Q$ is smooth; since S is compact it must be Lipschitz on S, and therefore it is globally Lipschitz. The equation (3) for X_t therefore has a unique strong solution by (Oksendal, 2013, Theorem 5.2.1).

For Y_t, namely (4), we again check global Lipschitz-ness of the drift and diffusion matrix. Here, the drift is globally Lipschitz because ∇f is by assumption, and because the additional term $(\alpha^2 - \beta^2)\nabla \log \operatorname{vol} \mathcal{O}$ is smooth on $\mathbb{R}^d_{\text{reg}}$ by Proposition 2 and it is supported in $S \subset \mathbb{R}^d_{\text{reg}}$. The diffusion matrix is αI which is globally Lipschitz because α is. The equation (4) therefore has a unique strong solution by (Oksendal, 2013, Theorem 5.2.1).

For Z_t in (14), we verify that the drift and diffusion are globally Lipschitz. Here, the drift is globally Lipschitz because ∇f is by assumption, and because the additional term $\varphi^2(\alpha^2 + \beta^2)\nabla \log \mathrm{vol}\, \mathcal{O}$ is smooth on and supported in $\mathbb{R}^d_{\mathrm{reg}}$ by Proposition 2. For the diffusion matrix, first observe that

$$\beta^2 + \varphi^2(\alpha^2 + \beta^2) = \alpha^2 + 2\varphi^2\beta^2,$$

and since we assume $\alpha > 0$ everywhere, the square root is smooth. Hence by Proposition 1 the diffusion matrix is smooth on $\mathbb{R}^d_{\mathrm{reg}}$ and thus smooth everywhere. It is globally Lipschitz because it is αI outside the support of φ and otherwise smooth on a compact set. The equation (14) for Z_t therefore has a unique strong solution by (Oksendal, 2013, Theorem 5.2.1). In particular, this solution must be G-invariant by the same argument as in the proof of Lemma 5.

Equivalence of X_t and Z_t. Let us begin by showing that (3) is equivalent in marginal law to (14). Recall the definition of $L_{g,x}$ from (9), and put

$$\begin{aligned}
V_0(g,x) &:= \varphi^2(g\cdot x)(\alpha^2(g\cdot x) + \beta^2(g\cdot x))\,\mathrm{tr}(h^G_g(L^\dagger_{g,x}\cdot, L^\dagger_{g,x}\cdot)),\\
V_1(g,x) &:= -L^\dagger_{g,x}(V_0(g,x)x),\\
J_0(g,x) &:= \varphi(g\cdot x)\sqrt{\alpha(g\cdot x)^2 + \beta(g\cdot x)^2}\cdot (L^\top_{g,x}L_{g,x})^{\dagger/2}.
\end{aligned}$$

We then consider the system of SDEs

$$\begin{aligned}
dX_t &= -\nabla f(X_t)dt + \sqrt{2}(\alpha(X_t)P_{X_t} + \beta(X_t)Q_{X_t})dB_t, \qquad X_0 \sim \mu\\
dg_t &= (V_0(g_t,X_t) + V_1(g_t,X_t))dt + \sqrt{2}J_0(g_t,X_t)dB'_t, \qquad g_0 = I.
\end{aligned} \tag{15}$$

We first verify that equation (15) has a unique strong solution. Observe that $L_{g,x}$ is a smooth function of g and x, and for all $g \in G$ and $x \in \mathbb{R}^d_{\mathrm{reg}}$, it has the same rank. It follows that $L^\dagger_{g,x}$ and $(L^\top_{g,x}L_{g,x})^{\dagger/2}$ are smooth for $x \in \mathbb{R}^d_{\mathrm{reg}}$ and $g \in G$ Constales (1998). Hence V_0 and V_1 are smooth on $G \times \mathbb{R}^d_{\mathrm{reg}}$. The matrix J_0 is also smooth on $G \times \mathbb{R}^d_{\mathrm{reg}}$ because we assume $\alpha > 0$ everywhere. But because φ is supported in $\mathbb{R}^d_{\mathrm{reg}}$, we find that V_0, V_1, J_0 are smooth on all of $G \times \mathbb{R}^d$. Because G and $\mathrm{supp}(\varphi)$ are compact, they must be globally Lipschitz in $G \times \mathbb{R}^d$, and thus can be extended to globally Lipschitz functions on all of $\mathbb{R}^d \times \mathbb{R}^{d\times d}$. By the same argument as before, the drift and diffusion for dX_t are globally Lipschitz as well. Hence (15) has a unique strong solution.

The same argument as in the proof of Theorem 8 shows that g_t then remains in G at all times. For property P2, we calculate using Itô's Lemma that if g_t follows (15) and X_t follows (3), then

$$\begin{aligned}
d(g_t\cdot X_t) &= (-g_t\nabla f(X_t) + L_{g_t,X_t}(V_0(g_t,X_t) + V_1(g_t,X_t)))dt\\
&\quad + g_t\sqrt{2}(\alpha(g_t\cdot X_t)P_{X_t} + \beta(g_t\cdot X_t)Q_{X_t})dB_t + L_{g_t,X_t}J_0(g_t,X_t)dB'_t.
\end{aligned}$$

Observe that

$$L_{g,x}(V_0(g,x) + V_1(g,x)) = P_{g\cdot x}(V_0(g,x)x).$$

Taking an orthonormal basis $v_1, \dots, v_m$ for $T_{g\cdot x}\mathcal{O}_x$ as well as orthonormal $A_1, \dots, A_m \in T_gG$ such that $L_{g,x}A_i = \sigma_i v_i$, we find by Lemma 7 that

$$\begin{aligned} P_{g\cdot x}(V_0(g,x)x) &= \varphi(x)^2(\alpha(x)^2+\beta(x)^2)\sum_{i=1}^m \frac{1}{\sigma_i^2} P_{g\cdot x}(h_g^G(A_i,A_i)x) \\ &= \varphi(x)^2(\alpha(x)^2+\beta(x)^2)\sum_{i=1}^m h_{g\cdot x}^{\mathcal{O}_x}(v_i,v_i) \\ &= \varphi(x)^2(\alpha(x)^2+\beta(x)^2)H(g\cdot x) = \varphi(g\cdot x)^2(\alpha(g\cdot x)^2+\beta(g\cdot x)^2)H(g\cdot x). \end{aligned}$$

By Proposition 2, $H(g\cdot x) = -\nabla \log \operatorname{vol} \mathcal{O}_{g\cdot x}$ for $x \in \mathbb{R}^d_{\text{reg}}$, so that

$$\begin{aligned} d(g_t\cdot X_t) = &-(\nabla f(g_t\cdot X_t) + \varphi(g_t\cdot X_t)^2(\alpha(g_t\cdot X_t)^2+\beta(g_t\cdot X_t)^2)\nabla\log\operatorname{vol}\mathcal{O}_{g_t\cdot X_t})dt \\ &+ g_t\sqrt{2}(\alpha(X_t)P_{X_t}+\beta(X_t)Q_{X_t})dB_t + L_{g_t,X_t}J_0(g_t,X_t)dB'_t. \end{aligned}$$

Now, observe that by our choice of J_0,

$$L_{g,x}J_0(g,x)J_0(g,x)^\top L_{g,x}^\top = \varphi(g\cdot x)^2(\alpha(g\cdot x)^2+\beta(g\cdot x)^2)Q_{g\cdot x}$$

The covariance is therefore

$$\begin{aligned} &g_t(\alpha(X_t)^2P_{X_t}+\beta(X_t)^2Q_{X_t})g_t^{-1} + \varphi(X_t)^2(\alpha(X_t)^2+\beta(X_t)^2)Q_{g_t\cdot X_t} \\ &= \alpha(X_t)^2P_{g_t\cdot X_t} + \beta(X_t)^2Q_{g_t\cdot X_t} + \varphi(X_t)^2(\alpha(X_t)^2+\beta(X_t)^2)Q_{g_t\cdot X_t}. \end{aligned}$$

where the equality follows by Proposition 11. Observe that for any $g \in G$, with $\tilde{X}_0 := g\cdot X_0$ By (Bogachev et al., 2022, Theorem 9.4.3), the marginal law of $(g_t\cdot X_t)_t$ must be identical to that of Z_t solving (14), so that, in particular, it is G-invariant.

To conclude, we argue that the law of X_t must be the same as that of $g_t\cdot X_t$, and thus Z_t, since both are G-invariant. Given a test function $h\colon \mathbb{R}^d \to \mathbb{R}$, let $\bar{h}(x) := \mathbb{E}_{g\sim\text{unif}_G}[h(g\cdot x)]$ where unif_G denotes the normalized Haar measure on G. Observe that $\bar{h}(g\cdot x) = \bar{h}(x)$ for any $g \in G$. Since $(g_t\cdot X_t)$ and X_t are both G-invariant, we have

$$\mathbb{E}[h(Z_t)] = \mathbb{E}[h(g_t\cdot X_t)] = \mathbb{E}[\bar{h}(g_t\cdot X_t)] = \mathbb{E}[\bar{h}(X_t)] = \mathbb{E}[h(X_t)].$$

Since h was an arbitrary test function, it follows that X_t and Z_t have the same marginal distributions at all times.

Equivalence of Y_t and Z_t. This proof is completely analogous to that of X_t and Z_t, with the only difference being the scaling of J_0, V_0 for g_t. We let

$$\begin{aligned} V_0(g,x) &:= 2\varphi(g\cdot x)^2\beta(g\cdot x)^2\operatorname{tr}(h_I^G(L_{g,x}^\dagger\cdot, L_{g,x}^\dagger\cdot)), \\ V_1(g,x) &:= -L_{g,x}^\dagger(V_0(g,x)x), \\ J_0(g,x) &:= \sqrt{2}\varphi(g\cdot x)\beta(g\cdot x)(L_{g,x}^\top L_{g,x})^{\dagger/2}. \end{aligned}$$

And then consider the system of SDEs

$$\begin{aligned} dY_t &= -(\nabla f(Y_t) + (\alpha(Y_t)^2-\beta(Y_t)^2)\nabla\log\operatorname{vol}\mathcal{O}_{Y_t})dt + \sqrt{2}\alpha(X_t)PdB_t, \qquad Y_0\sim\mu \\ dg_t &= (V_0(g_t,Y_t)+V_1(g_t,Y_t))dt + \sqrt{2}J_0(g_t,Y_t)dB'_t, \qquad g_0 = I. \end{aligned} \tag{16}$$

As before, we calculate that the drift of $g_t \cdot Y_t$ is

$$-\nabla f(g_t \cdot Y_t) - (\alpha(g_t \cdot Y_t)^2 - \beta(g_t \cdot Y_t)^2 + 2\varphi(g_t \cdot Y_t)^2 \beta(g_t \cdot Y_t)^2) \nabla \log \operatorname{vol} \mathcal{O}_{g_t \cdot Y_t}.$$

But since $\alpha^2 - \beta^2 = \varphi^2(\alpha^2 - \beta^2)$ we have

$$\alpha^2 - \beta^2 + 2\varphi^2\beta^2 = \varphi^2(\alpha^2 - \beta^2) + 2\varphi^2\beta^2 = \varphi^2(\alpha^2 + \beta^2),$$

hence the drift is the same as that of Z_t. As before, we calculate that the covariance matrix of $g_t \cdot Y_t$ is

$$\alpha(g_t \cdot Y_t)^2 P_{g_t \cdot X_t} + (\alpha(g_t \cdot Y_t)^2 + 2\varphi^2(g_t \cdot Y_t)\beta(g_t \cdot Y_t)^2) Q_{g_t \cdot Y_t}.$$

This is the same as the covariance matrix of Z_t since

$$\alpha^2 + 2\varphi^2\beta^2 = \beta^2 + \alpha^2 - \beta^2 + 2\varphi^2\beta^2 = \beta^2 + \varphi^2(\alpha^2 - \beta^2) + 2\varphi^2\beta^2 = \beta^2 + \varphi^2(\alpha + \beta^2).$$

The rest of the proof is identical to the previous one: Z_t and $g_t \cdot Y_t$ must have the same marginal laws, and by G-invariance we can conclude that Y_t has the same marginal law as $g_t \cdot Y_t$ and thus Z_t.

Conclusion. We have shown that X_t and Y_t have the same marginal law as the auxiliary process Z_t, and thus X_t and Y_t have the same marginal law, yielding Theorem 3. ■

Appendix D. Proof by PDE analysis

Proof [Proof of Theorem 3 by PDE analysis] We verified in the proof by SDE coupling that both (3) and (4) have unique strong solutions, so we do not repeat that argument here.

Let the law of X_t be ρ_t. By Itô's formula, ρ_t is a solution of the *weak Fokker–Planck equation*: for all positive times $T > 0$ and all smooth compactly supported $\varphi\colon \mathbb{R}^d \times (0, T) \to \mathbb{R}$,

$$\int_{\mathbb{R}^d \times (0,T)} (\partial_t \varphi + \operatorname{tr}(R\nabla^2\varphi) - \langle \nabla f, \nabla \varphi \rangle)\rho_t = 0, \tag{17}$$

where for ease of notation, we let $R := \alpha^2 P + \beta^2 Q$. The idea is then to show that this ρ_t also satisfies the weak Fokker–Planck equation associated to the SDE (4).

Suppressing the time t, observe that

$$\nabla\varphi(g \cdot x) = g\nabla(\varphi \circ g)(x), \qquad \nabla^2\varphi(g \cdot x) = g^\top \nabla^2(\varphi \circ g)(x) g.$$

So by Lemma 5 we have

$$\int_{\mathbb{R}^d} \operatorname{tr}(R\nabla^2\varphi) - \langle \nabla f, \nabla\varphi \rangle)\rho_t$$
$$= \int_{\mathbb{R}^d} \int (\operatorname{tr}(R_{g \cdot x} \nabla^2 \varphi(g \cdot x)) - \langle \nabla f(g \cdot x), \nabla \varphi(g \cdot x) \rangle) \rho_t(x) d \operatorname{unif}_G(g) dx.$$

Applying Prop. 10 and Prop. 11 this is exactly

$$\int_{\mathbb{R}^d} \operatorname{tr}(R\nabla^2\varphi) - \langle \nabla f, \nabla\varphi \rangle)\rho_t = \int_{\mathbb{R}^d} (\operatorname{tr}(R\nabla^2\bar{\varphi}) - \langle \nabla f, \nabla\bar{\varphi} \rangle)\rho_t,$$

where $\bar{\varphi}(x) := \int \varphi(g \cdot x) d \operatorname{unif}_G(g)$. We now come to the main geometric observation driving the current proof, we remark that a similar observation appears in (Helgason, 1984, Theorem 3.7).

Lemma 12 (Hessian of G-invariant functions) *If $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is G-invariant and $v \in T_x\mathcal{O}_x$ then*

$$\nabla^2\varphi[v,v] = -\langle \nabla\varphi, h_x(v,v)\rangle.$$

Using this result, we obtain

$$\int_{\mathbb{R}^d\times(0,T)} (\partial_t\bar{\varphi} + \alpha^2\Delta\bar{\varphi} - \langle \nabla\bar{\varphi}, (\beta^2-\alpha^2)H + \nabla f\rangle)\rho_t = 0.$$

Since $H(x) = -\nabla \log \operatorname{vol}\mathcal{O}_x$ by Prop. 2 and the volume $\operatorname{vol}\mathcal{O}_x$ is G-invariant, we can again apply Prop. 10 to go from $\bar{\varphi}$ to φ and obtain

$$\int_{\mathbb{R}^d\times(0,T)} (\partial_t\varphi + \alpha^2\Delta\varphi - \langle \nabla\varphi, (\alpha^2-\beta^2)\nabla \log \operatorname{vol}\mathcal{O} + \nabla f\rangle)\rho_t = 0.$$

This is the weak Fokker–Planck equation associated to equation (4). But by Itô's Lemma, this equation is also satisfied by the law of Y_t of (4). The result (Bogachev et al., 2022, Theorem 9.4.3) ensures that there can only be one such solution, yielding the result. ■

Proof [Proof of Lemma 12] By taking appropriate local coordinates, one can construct a smooth vector field V on $\mathbb{R}^d$ such that $V_x = v$ and $V_y \in T_y\mathcal{O}_y$ for all y in a neighborhood of x. Then

$$\nabla^2\varphi[v,v] = \langle \nabla_V\nabla\varphi, V\rangle = \nabla_V\langle\nabla\varphi, V\rangle - \langle\nabla\varphi, \nabla_V V\rangle = -\langle\nabla\varphi, \nabla_V V\rangle = -\langle\nabla\varphi, h_x(v,v)\rangle.$$

■

D.1. Proof of Corollary 4

Proof [Proof of Corollary 4] Under the assumptions of Corollary 4, the SDE (5) has smooth drift and diffusion matrices, and thus locally Lipschitz drift and diffusion matrices. Because we assume that φ is bounded, the diffusion matrix is also bounded in Frobenius norm. We may therefore apply (Kunze, 2012, Proposition 4.2.4) to conclude that Equation (5) has a unique strong solution. This implies weak uniqueness, and thus the proof of Lemma 5 goes through as before to show that X_t remains G-invariant at all times. On the other hand, our assumptions guarantee that the SDE converge to a unique stationary distribution, which must be G-invariant. It is straightforward to check using the calculations in the PDE proof that this stationary distribution is the claimed one, proving the result. ■

Appendix E. Proofs for examples

E.1. Auxiliary lemmas

We collect several elementary linear algebra results that will prove useful for volume computations.

Lemma 13 *Denote E_{ij} the canonical basis of $\mathbb{R}^{d\times d}$ and D a diagonal matrix with entries λ_k. It holds that*

$$E_{ij}D = \lambda_i E_{ij}\,,$$
$$DE_{ij} = \lambda_j E_{ij}\,.$$

Proof We have

$$[E_{ij}D]_{kl} = \sum_{p=1}^{d}(E_{ij})_{kp}D_{pl} = \sum_{p=1}^{d}(E_{ij})_{kp}\lambda_k\delta_{p=l} = (E_{ij})_{kl}\lambda_k = \lambda_k\delta_{(i,j)=(k,l)}\,.$$

Hence $E_{ij}D = \lambda_i E_{ij}$ and by taking the transpose, $DE_{ij} = \lambda_j E_{ij}$. ■

Lemma 14 *Let $O \in O(d)$. We have $T_O O(d) = \{OA \mid A^\top = -A\} = \{AO \mid A^\top = -A\}$.*

Proof Let $O(t)$ be a curve on $O(d)$ such that $O(0) = O$. We have that $O(t)^\top O(t) = I_d$, hence differentiating w.r.t. t yields

$$O^\top \dot{O}(0) + \dot{O}(0)^\top O = 0\,.$$

A necessary and sufficient condition on $\dot{O}(0)$ is thus $\dot{O}(0) = OA$ with A a skew matrix. ■

Lemma 15 *Suppose (M, g) is a smooth Riemannian manifold with or without boundary and $\tilde{g}$ is another metric tensor such that $(M, \tilde{g})$ is also a smooth Riemannian manifold. Suppose that there exists a global frame $V_1, \ldots, V_n$ which is orthonormal for g and such that $\tilde{g}(V_i, V_j) = c_i\delta_{ij}$ for all $i, j \in [n]$ for $c_i > 0$ constant on M. Then*

$$\mathrm{vol}(M, \tilde{g}) = \Big(\prod_{i=1}^{n}\sqrt{c_i}\Big)\,\mathrm{vol}(M, g).$$

Proof Let us begin by assuming that the manifold (M, g) is oriented. Let $\omega^1, \ldots, \omega^n$ denote the covectors corresponding to $V_1, \ldots, V_n$. If necessary, swap labels such that $\omega^1 \wedge \cdots \wedge \omega^n$ is positively oriented. Define $W_i := \frac{1}{\sqrt{c_i}}V_i$. Then note that

$$\tilde{V} := \Big(\prod_{i=1}^{n}\sqrt{c_i}\Big)\omega^1 \wedge \cdots \omega^n$$

evaluates to 1 at $(W_1, \ldots, W_n)$. Because $(W_1, \ldots, W_n)$ is an orthonormal frame, V must be the Riemannian volume form on $(M, \tilde{g})$ (Lee, 2013, Proposition 15.29). But the Riemannian volume form on (M, g) is $V := \omega^1 \wedge \cdots \wedge \omega^n$, so that

$$\mathrm{vol}(M, \tilde{g}) = \int \tilde{V} = \prod_{i=1}^{n}\sqrt{c_i}\int V = \Big(\prod_{i=1}^{n}\sqrt{c_i}\Big)\,\mathrm{vol}(M, g).$$

To extend to non-orientable manifolds, we may take a partitition of unity and apply the above formula on each partition. ■

E.2. Conjugation over symmetric matrices

For the action $(O \in O(d), M \in \text{Sym}(d)) \mapsto O \cdot M = O^\top M O$, the orbits are given by the set of eigenvalues: two matrices are in the same orbit if and only if they have the same eigenvalues. In particular, the principal orbits are the symmetric matrices with d distinct eigenvalues. Let us compute the volume of these principal orbits. We first show that an orbit $\mathcal{O}_D$ with D a diagonal matrix with d distinct eigenvalues, can be parametrized by a subset of $O(d)$ via Φ defined as

$$\begin{cases} \Phi : & \mathcal{F} \to O_D \,, \\ & O \mapsto O^\top D O \,, \end{cases}$$

with $\mathcal{F} = \{O \in O(d) \mid \text{the first non-zero element of each column is positive}\}$. Let us prove that Φ is indeed invertible.

Injectivity. Assume that there exists $(O_1, O_2) \in \mathcal{F}^2$ such that $O_1^\top D O_1 = O_2^\top D O_2$. In particular it holds that $O_2 O_1^\top D = D O_2 O_1^\top$. Denoting $O = O_2 O_1^\top$ and (e_i) the canonical basis, we decompose Oe_i as $\sum_{k=1}^d \alpha_k e_k$. Applying the previous equality yields $\sum_{k=1}^d \alpha_k e_k(\lambda_k - \lambda_i) = 0$. Hence, since all eigenvalues are distinct, we recover that $\alpha_k = 0$ for $k \neq i$ and $Oe_i = \alpha_i e_i$. Since $\|Oe_i\| = 1$, we get that $\alpha_i = \pm 1$ so that O is a diagonal matrix with only ± 1 entries that we shall denote Δ. In particular we get $O_1 = \Delta O_2$. Since both O_1, O_2 belong to $\mathcal{F}$, the sign constraint implies $\Delta = I_d$.

Surjectivity. For $M \in \mathcal{O}_D$, there exits $O \in O(d)$ such that $M = O^\top D O$. Now, for all Δ a diagonal matrix with ± 1 entries, it holds that $\Delta^\top D \Delta = D$ so that $M = (\Delta O)^\top D \Delta O$ for all Δ. In particular, Δ can be chosen such that $\Delta O \in \mathcal{F}$.

Hence we have $\operatorname{vol} \mathcal{O}_D = \operatorname{vol} \mathcal{F}$ with the metric on $\mathcal{F}$ given by the pullback metric $\tilde{g}_O$ that reads for all $(HO, VO) \in T_O\mathcal{F}$ with (H, V) skew matrices,

$$\begin{aligned} \tilde{g}_O(HO, VO) &= \operatorname{tr}(d\Phi_O(OH), (d\Phi_O(OV))^\top) \,, \\ &= \operatorname{tr}((O^\top H^\top D O + O^\top D H O)(O^\top V^\top D O + O^\top D V O)) \,, \\ &= \operatorname{tr}(O^\top H^\top D V^\top D O + O^\top H^\top D^2 V O + O^\top D H V^\top D O + O^\top D H D V O) \,, \\ &= \operatorname{tr}(HDVD - HDDV - DHVD + DHDV) \,, \\ &= \operatorname{tr}([H, D][V, D]) \,, \end{aligned}$$

where we denoted $[\cdot, \cdot]$ the Lie bracket $[H, D] = HD - DV$. Let us compute this metric in the basis $\tilde{G}_{ij} = (E_{ij} - E_{ji})O$ of $T_O\mathcal{F}$ (see Lemma 14). First we compute the Lie bracket

$$\begin{aligned} [E_{ij} - E_{ji}, D] &= (E_{ij} - E_{ji})D - D(E_{ij} - E_{ji}) \,, \\ &= (\lambda_i E_{ij} - \lambda_j E_{ji}) - (\lambda_j E_{ij} - \lambda_i E_{ji}) \,, \\ &= (E_{ij} + E_{ji})(\lambda_i - \lambda_j) \,. \end{aligned}$$

It follows that

$$\begin{aligned} \tilde{g}_O(\tilde{G}_{ij}, \tilde{G}_{kl}) &= (\lambda_i - \lambda_j)(\lambda_k - \lambda_l) \operatorname{tr}((E_{ij} + E_{ji})(E_{kl} + E_{lk})) \,, \\ &= (\lambda_i - \lambda_j)(\lambda_k - \lambda_l) \operatorname{tr}(\delta_{j=k} E_{il} + \delta_{i=} E_{jk} + \delta_{i=k} E_{jl} + \delta_{j=l} E_{ik}) \,, \\ &= 2(\lambda_i - \lambda_j)^2 \delta_{(i,j)=(k,l)} \,. \end{aligned}$$

Observe also that the vector fields $\frac{1}{\sqrt{2}}\tilde{G}_{ij}$ are orthonormal with respect to the standard metric on $\mathcal{F}$. Hence we can apply Lemma 15 to conclude that

$$\operatorname{vol}\mathcal{O}_D \propto \prod_{i<j} |\lambda_i - \lambda_j| \operatorname{vol}(\mathcal{F}) .$$

E.3. Bures-Wasserstein

Consider the action over the space of squared matrices $X \in \mathbb{R}^{d\times d} \mapsto XO$. Let us compute the volume of the orbit at a given invertible matrix X: $\mathcal{O}_X = \{XO \mid O \in O(d)\}$. In that case,

$$\begin{cases} \Phi : & O(d) \to \mathcal{O}_X \,, \\ & O \mapsto XO \,, \end{cases}$$

is a diffeomorphism. In particular, $\operatorname{vol}(\mathcal{O}_X) = \operatorname{vol}(O(d))$ where $O(d)$ is equipped with the pullback metric. Let us compute the pullback metric for well-chosen coordinates. Denote $X^\top X = P^\top \Sigma P$ with $P \in O(d)$ and Σ diagonal with entries (σ_i^2) and consider the following basis of $T_O O(d)$: $\tilde{G}_{ij} = P^\top(E_{ij} - E_{ji})PO$ with $i < j$. In this basis, applying Lemma 13 implies that the pullback metric reads

$$\begin{aligned}
\tilde{g}_{ij,kl} &= \operatorname{tr}(X\tilde{G}_{ij}(X\tilde{G}_{kl})^\top) \,, \\
&= -\operatorname{tr}(P^\top \Sigma P P^\top (E_{ij} - E_{ji}) P O O^\top P^\top (E_{kl} - E_{lk}) P P^\top \Sigma P) \,, \\
&= -\operatorname{tr}((E_{ij} - E_{ji})\Sigma^2 (E_{kl} - E_{lk})) \,, \\
&= -\operatorname{tr}((\sigma_i^2 E_{ij} - \sigma_j^2 E_{ji})(E_{kl} - E_{lk})) \,, \\
&= -\operatorname{tr}(\sigma_i^2 E_{ij}E_{kl} - \sigma_i^2 E_{ij}E_{lk} - \sigma_j^2 E_{ji}E_{kl} + \sigma_j^2 E_{ji}E_{lk}) \,, \\
&= -\operatorname{tr}(\sigma_i^2 \delta_{j=k} E_{il} - \sigma_i^2 \delta_{j=l} E_{ik} - \sigma_j^2 \delta_{i=k} E_{jl} + \sigma_j^2 \delta_{i=l} E_{jk}) \,, \\
&= -\sigma_i^2 \delta_{(j,i)=(k,l)} + \sigma_i^2 \delta_{(i,j)=(k,l)} + \sigma_j^2 \delta_{(i,j)=(k,l)} - \sigma_j^2 \delta_{(i,j)=(l,k)} \,, \\
&= (\sigma_i^2 + \sigma_j^2)\delta_{(i,j)=(k,l)} \,.
\end{aligned}$$

Observe that $\tilde{G}_{ij}$ are also a smooth orthonormal frame for $O(d)$ with respect to the standard metric. Hence we may apply Lemma 15 to determine that

$$\operatorname{vol}(\mathcal{O}_X) \propto \prod_{i<j} \sqrt{\sigma_i^2 + \sigma_j^2} \operatorname{vol}(O(d)) \,.$$

E.4. Brownian motion on the group does not necessarily yield Brownian on the orbit

Consider the case $G = SO(2)$, acting by left multiplication on $\mathbb{R}^2$. Brownian motion on G is explicitly given by

$$g_t = \begin{pmatrix} \cos(\theta_t) & -\sin(\theta_t) \\ \sin(\theta_t) & \cos(\theta_t) \end{pmatrix} \,,$$

with θ_t a 1D Brownian motion (Hsu, 2002, Example 3.3.1). Now for x fixed, consider $y_t = g_t \cdot x$. The process y_t verifies

$$dy_t = -y_t dt + \sqrt{2} \begin{pmatrix} -\sin(\theta_t) & -\cos(\theta_t) \\ \cos(\theta_t) & -\sin(\theta_t) \end{pmatrix} x d\theta_t \,.$$

Hence, while the drift term $-y_t$ does correspond to mean curvature, the diffusion term, whose norm grows with $\|x\|$, cannot be expressed as $Q_{y_t} dB_t$ with B_t a 2D Brownian motion since its norm is unitary and independent of x.