Diffusion Models for Constrained Domains

Anonymous authors
Paper under double-blind review

Abstract

Denoising diffusion models are a novel class of generative algorithms that achieve state-of-the-art performance across a range of domains, including image generation and text-to-image tasks. Building on this success, diffusion models have recently been extended to the Riemannian manifold setting, broadening their applicability to a range of problems from the natural and engineering sciences. However, these Riemannian diffusion models are built on the assumption that their forward and backward processes are well-defined for all times, preventing them from being applied to an important set of tasks that consider manifolds defined via a set of inequality constraints. In this work, we introduce a principled framework to bridge this gap. We present two distinct noising processes based on (i) the *logarithmic barrier* metric and (ii) the *reflected* Brownian motion induced by the constraints. As existing diffusion model techniques cannot be applied in this setting, we proceed to derive new tools to define such models in our framework. We then empirically demonstrate the scalability and flexibility of our methods on a number of synthetic and real-world tasks, including applications from robotics and protein design.

1 Introduction

Diffusion models (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Song et al., 2021; Ho et al., 2020) have recently been introduced as a powerful new paradigm for generative modelling. They work as follows: noise is progressively added to data following a Stochastic Differential Equation (SDE)—the forward *noising* process—until it is approximately Gaussian. The generative model is given by an approximation of the associated *time-reversed process* called the backward *denoising* process. This is also an SDE whose drift depends on the gradient of the logarithmic densities of the forward process, referred as the Stein score. This score is approximated by leveraging techniques from deep learning and score matching (Hyvärinen, 2005; Vincent, 2011). Building on the success of diffusion models in domains such as images and text, this framework has recently been extended to a wide range of Riemannian manifolds (De Bortoli et al., 2022; Huang et al., 2022), broadening their applicability to various important modelling domains from the natural and engineering sciences—including Lie groups such as the group of rotations SO(3), the group of rigid body motions SE(3), and many others (see e.g. Trippe et al., 2022; Corso et al., 2022; Watson et al., 2022; Leach et al., 2022; Urain et al., 2022; Yim et al., 2023).

However, a key assumption of the Riemannian diffusion models introduced in De Bortoli et al. (2022) and Huang et al. (2022) is that the stochastic processes they consider are defined *for all times*. While this holds for a large class of stochastic processes, it is not the case for most manifolds defined via a set of inequality constraints. For instance, in the case of the hypercube $(-1,1)^d$ equipped with the Euclidean metric, the Riemannian Brownian motion coincides with the Euclidean d-dimensional Brownian motion $(\mathbf{B}_t)_{t\in[0,T]}$ as long as $\mathbf{B}_t \in (-1,1)$. With probability one, $(\mathbf{B}_t)_{t\geq 0}$ escapes from $(-1,1)$, meaning that the Riemannian Brownian motion is not defined for all times and the frameworks introduced in De Bortoli et al. (2022) and Huang et al. (2022) do not apply. Such constrained manifolds comprise a wide variety of settings—including polytopes and convex sets of Euclidean spaces—and are studied across a large number of disciplines, ranging from computational statistics (Morris, 2002), over robotics (Han & Rudolph, 2006) and quantum physics (Lukens et al., 2020), to computational biology (Thiele et al., 2013). Deriving principled diffusion models that are able to operate directly on these manifolds is thus of significant practical importance, as they enable

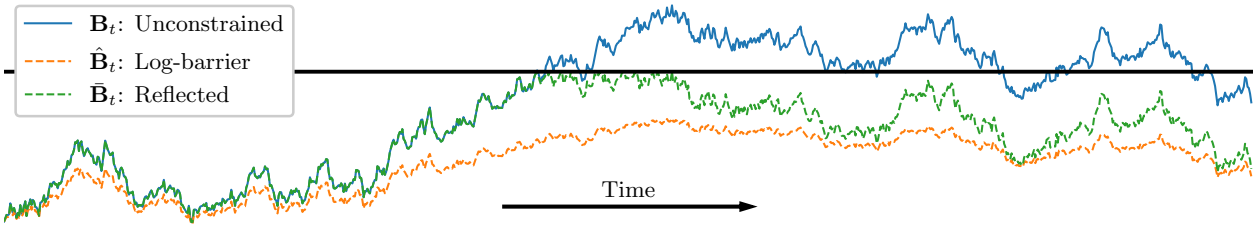


Figure 1: The behaviour of different types of noising processes considered in this work defined on the unit interval. $\mathbf{B}_t$: Euclidean (unconstrained) Brownian motion. $\hat{\mathbf{B}}_t$: Log barrier forward noising process. $\bar{\mathbf{B}}_t$: Reflected Brownian motion. All sampled with the same initial point and driving noise. Black line indicates the boundary.

generative modelling in data-scarce and safety-critical settings in which constraints on the modelled domain may reduce the number of degrees of freedom or prevent unwanted behaviour.

As sampling problems on such manifolds are important (Kook et al., 2022; Heirendt et al., 2019), a flurry of Markov chain based methods have been developed to sample from unnormalised densities. Successful algorithms include the reflected Brownian motion (Williams, 1987; Petit, 1997; Shkolnikov & Karatzas, 2013), log-barrier methods (Kannan & Narayanan, 2009; Lee & Vempala, 2017; Noble et al., 2022; Kook et al., 2022; Gatmiry & Vempala, 2022; Lee & Vempala, 2018) and hit-and-run approaches in the case of polytopes (Smith, 1984; Lovász & Vempala, 2006). In this work, we study the generative modelling counterparts of these algorithms through the lens of diffusion models. Among existing methods for statistical sampling on constrained manifolds, the geodesic Brownian motion (Lee & Vempala, 2017) and the reflected Brownian motion (Williams, 1987) are continuous stochastic processes, and thus well suited for extending the continuous Riemannian diffusion framework developed by De Bortoli et al. (2022) and Huang et al. (2022). In particular, we introduce two principled diffusion models for generative modelling on constrained domains based on (i) the geodesic Brownian motion, leveraging tools from the log-barrier methods, and (ii) the reflected Brownian motion. In both cases, we show how one can extend the ideas of time-reversal and score matching to these settings. We empirically demonstrate the scalability and flexibility of these methods on a range of tasks defined on high-dimensional convex polytopes and the space of symmetric positive definite matrices, including the constrained conformational modelling of proteins and robotic arms.

2 Background

Riemannian manifolds. A Riemannian manifold is a tuple $(\mathcal{M}, \mathfrak{g})$ with $\mathcal{M}$ a smooth manifold and $\mathfrak{g}$ a metric which defines an inner product on tangent spaces. The metric $\mathfrak{g}$ induces key quantities on the manifold, such as an exponential map $\exp_x : \mathrm{T}_x\mathcal{M} \to \mathcal{M}$, defining the notion of following straight lines on manifolds, a gradient operator ∇[1] and a divergence operator div[2]. It also induces the Laplace-Beltrami operator Δ and consequently a Brownian motion with density (w.r.t. the volume form[3]) whose density is given by the heat equation $\partial_t p_t = \Delta p_t$. We refer the reader to Appendix A for a brief introduction to differential geometry, to Lee (2013) for a thorough treatment and to Hsu (2002) for details on stochastic analysis on manifolds.

Constrained manifolds. In this work, we are concerned with *constrained* manifolds. More precisely, given a Riemannian manifold $(\mathcal{N}, \mathfrak{h})$, we consider a family of real functions $\{f_i : \mathcal{N} \to \mathbb{R}\}_{i \in \mathcal{I}}$ indexed by $\mathcal{I}$. We then define

$$\mathcal{M} = \{x \in \mathcal{N} \,:\, f_i(x) < 0,\ i \in \mathcal{I}\}. \tag{1}$$

In this scenario, $\{f_i\}_{i \in \mathcal{I}}$ is interpreted as a set of constraints on $\mathcal{N}$. For example, choosing $\mathcal{N} = \mathbb{R}^d$ and affine constraints $f_i(x) = \langle a_i, x \rangle - b_i$, $x \in \mathbb{R}^d$, we get that $\mathcal{M}$ is an open polytope as illustrated in Figure 2.

[1]The (Riemannian) gradient ∇ is defined s.t. for any smooth $f \in \mathrm{C}^\infty(\mathcal{M})$, $x \in \mathcal{M}, v \in \mathrm{T}_x\mathcal{M}$, $\mathfrak{g}(x)(f(x), v) = \mathrm{d}f(x)(v)$.
[2]The Riemannian divergence acts vector fields and can be defined using the volume form of $\mathcal{M}$.
[3]We assume that $\mathcal{M}$ is orientable and therefore that a volume form exists.

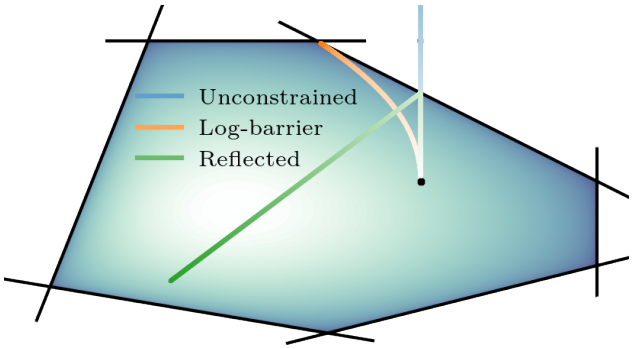


Figure 2: A convex polytope defined by six constraints $\{f_i\}_{i\in\mathcal{I}}$, along with the log barrier potential, and 'straight trajectories' under the log-barrier metric and under the Euclidean metric with and without reflection at the boundary.

This setting naturally appears in many areas of engineering, biology, and physics (Boyd et al., 2004; Han & Rudolph, 2006; Lukens et al., 2020).

While the two methods we introduce in Section 3 can be applied to the general framework (1), in our applications, we focus on two specific settings: (a) polytopes—$\mathcal{N} = \mathbb{R}^d$, (b) symmetric positive definite matrices (SPD) under trace conditions—$\mathcal{N} = \mathcal{S}^d_{++}$.

Continuous diffusion models. We briefly recall the framework for constructing continuous diffusion processes introduced by Song et al. (2021) in the context of generative modelling over $\mathbb{R}^d$. At minimum diffusion models need four things: (i) A forward noising process converging to an invariant distribution. (ii) A time reversal for the reverse process. (iii) A discretization of the continuous-time process for the forward/reverse process. (iv) A score matching loss — in this paper we will focus on the implicit score matching loss. Song et al. (2021) consider a forward *noising process* $(\mathbf{X}_t)_{t\in[0,T]}$ which progressively noises a data distribution p_0 into a Gaussian $\mathrm{N}(0, \mathrm{Id})$. More precisely $(\mathbf{X}_t)_{t\in[0,T]}$ is an Ornstein–Uhlenbeck (OU) process which is given by the following stochastic differential equation (SDE)

$$\mathrm{d}\mathbf{X}_t = -\tfrac{1}{2}\mathbf{X}_t\mathrm{d}t + \mathrm{d}\mathbf{B}_t, \qquad \mathbf{X}_0 \sim p_0.$$

Under mild conditions on p_0, the time-reversed process $(\overleftarrow{\mathbf{X}}_t)_{t\in[0,T]} = (\mathbf{X}_{T-t})_{t\in[0,T]}$ also satisfies an SDE (Cattiaux et al., 2021; Haussmann & Pardoux, 1986) given by

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \{\tfrac{1}{2}\overleftarrow{\mathbf{X}}_t + \nabla \log p_{T-t}(\overleftarrow{\mathbf{X}}_t)\}\mathrm{d}t + \mathrm{d}\mathbf{B}_t, \qquad \overleftarrow{\mathbf{X}}_0 \sim p_T, \tag{2}$$

where p_t denotes the density of $\mathbf{X}_t$. This construction allows direct sampling of the forward process and leverages the Euler-Maruyama discretisation to facilitate sampling of the reverse process. Finally, the quantity $\nabla \log p_t$ is referred as the Stein score and is unavailable in practice. It can be approximated with a score network $s_\theta(t, \cdot)$ trained by minimising a denoising score matching (dsm) loss

$$\mathcal{L}(\theta) = \mathbb{E}[\lambda_t \left\| \nabla \log p_{t|0}(\mathbf{X}_t|\mathbf{X}_0) s_\theta(t, \mathbf{X}_t) \right\|^2],$$

or an equivalent score matching (ism) loss

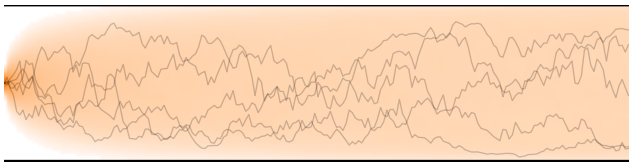
$$\mathcal{L}(\theta) = \mathbb{E}[\lambda_t\{\tfrac{1}{2}\|s_\theta(t, \mathbf{X}_t)\|^2 + \mathrm{div}(s_\theta)(t, \mathbf{X}_t)\}] + C, \tag{3}$$

where $C \geq 0$ and $\lambda_t > 0$ is a weighting function, and the expectation is taken over $t \sim \mathcal{U}([0, T])$ and $(\mathbf{X}_0, \mathbf{X}_t)$. For an arbitrarily flexible score network, the global minimiser $\theta^\star = \mathrm{argmin}_\theta \mathcal{L}(\theta)$ satisfies $s_{\theta^\star}(t, \cdot) = \nabla \log p_t$.

3 Inequality-constrained diffusion models

We are now ready to introduce our methodology to deal with manifolds defined via inequality *constraints* (1). In Section 3.1, we propose a Riemannian diffusion model endowed with a metric induced by a log-barrier potential. In Section 3.2, we introduce a *reflected* diffusion model. While both models extend classical diffusion models to inequality-constrained settings, they exhibit very different behaviours. We discuss their practical differences in Section 5.1.

3.1 Log-barrier diffusion models



$$\mathrm{d}\mathbf{X}_t = \tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1})(\mathbf{X}_t)\mathrm{d}t + \mathfrak{g}^{-1/2}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t$$

Figure 3: Convergence of the Barrier Langevin dynamics on the unit interval to the uniform distribution.

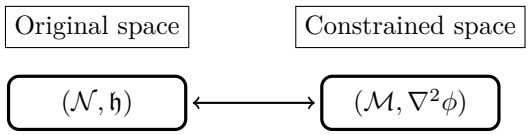


Figure 4: Illustrative diagram of the barrier method and the change of metric.

Barrier Langevin Dynamics. Barrier methods work by constructing a smooth potential $\phi : \mathcal{M} \to \mathbb{R}$ such that it blows up on the boundary of a desired set, see Nesterov et al. (2018). Such potentials are at the basis of interior methods in optimisation (Boyd et al., 2004). Among these functions, the *logarithmic barrier* is the most popular among practitioners (Lee & Vempala, 2017). For a convex polytope $\mathcal{M}$ defined by the constraints $\mathrm{A}x < b$, with $\mathrm{A} \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$, the logarithmic barrier $\phi : \mathcal{M} \to \mathbb{R}_+$ is given for any $x \in \mathcal{M}$ by

$$\phi(x) = -\textstyle\sum_{i=1}^m \log(\langle \mathrm{A}_i, x\rangle - b_i). \tag{4}$$

Assuming that $\|\mathrm{A}_i\| = 1$, we have that for any $x \in \mathcal{M}$, $\phi(x) = -\sum_{i=1}^m \log(d(x, \partial\mathcal{M}_i))$, where $\partial\mathcal{M}_i = \{x \in \mathbb{R}^d : \langle \mathrm{A}_i, x\rangle = b\}$. More generally we can define for any $x \in \mathcal{M}$

$$\phi(x) = -\textstyle\sum_{i=1}^m \log(d(x, \partial\mathcal{M}_i))$$

where $d(x, \partial\mathcal{M}_i)$ computes the minimum geodesic distance from x to the boundary $\partial\mathcal{M}$. In general this is a highly non-trivial optimization problem, contrary to the polytope case which admits a simple closed form.

While developed and most commonly used in optimisation, barrier methods can also be used for sampling (Lee & Vempala, 2017). The core idea of barrier methods is to 'warp the geometry' of the constrained space, stretching it as the process approaches the boundary so it never hits it, hence bypassing the need to explicitly deal with the boundary. Assuming ϕ to be strictly convex and smooth, its Hessian $\nabla^2\phi$ is positive definite and thus defines a valid Riemannian metric on $\mathcal{M}$. The formal approach to 'warping the geometry' of the convex space with the boundary is to endow $\mathcal{M}$ with the Hessian as a Riemannian metric $\mathfrak{g} = \nabla^2\phi$, making it into a Hessian Manifold, see Shima & Yagi (1997). In the special case where the barrier is given by (4), we get for any $x \in \mathcal{M}$

$$\mathfrak{g}(x) = \mathrm{A}^\top S^{-2}(x)\mathrm{A} \text{ with } S(x) = \mathrm{diag}(b_i - \langle \mathrm{A}_i, x\rangle)_i.$$

Equipped with this Riemannian metric, we consider the following Langevin dynamics as a forward process

$$\mathrm{d}\mathbf{X}_t = \tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1})(\mathbf{X}_t)\mathrm{d}t + \mathfrak{g}(\mathbf{X}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t, \tag{5}$$

with $\mathrm{div}(F)(x) \triangleq (\mathrm{div}(F_1)(x), \ldots, \mathrm{div}(F_d)(x))^\top$, for any smooth $F : \mathbb{R}^d \to \mathbb{R}^d$. Under mild assumptions on $\mathcal{M}$, $\mathrm{div}(\mathfrak{g}^{-1})$ and $\mathfrak{g}^{-1}$ we get that $(\mathbf{X}_t)_{t\geq 0}$ is well-defined and for any $t \geq 0$, $\mathbf{X}_t \in \mathcal{M}$. In particular, for any $t \geq 0$, $\mathbf{X}_t$ does not reach the boundary. This stochastic process was first proposed by Lee & Vempala (2017) in the context of efficient sampling from the uniform distribution over a polytope. Under similar conditions, $\mathbf{X}_t$ admits a density p_t w.r.t. the Lebesgue measure and we have that $\partial_t p_t = \frac{1}{2}\mathrm{Tr}(\mathfrak{g}^{-1}\nabla^2 p_t)$. In addition, $(\mathbf{X}_t)_{t\geq 0}$ is irreducible. Hence, assuming that $\mathcal{M}$ is compact, the uniform distribution on $\mathcal{M}$ is the unique invariant measure of the process $(\mathbf{X}_t)_{t\geq 0}$ and $(\mathbf{X}_t)_{t\geq 0}$ converges to the uniform distribution in some sense. We refer the reader to Appendix C for a proof of these results.

Time-reversal. Assuming that $\mathfrak{g}^{-1}$ and its derivative are bounded on $\mathcal{M}$, the time-reversal of (5) is given by Cattiaux et al. (2021), in particular we have

$$\begin{aligned}\mathrm{d}\overleftarrow{\mathbf{X}}_t &= [-\tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1}) + \mathrm{div}(\mathfrak{g}^{-1}) + \mathfrak{g}^{-1}\nabla \log p_{T-t}](\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathfrak{g}(\overleftarrow{\mathbf{X}}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t,\\ &= \left[\tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1}) + \mathfrak{g}^{-1}\nabla \log p_{T-t}\right](\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathfrak{g}(\overleftarrow{\mathbf{X}}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t.\end{aligned} \tag{6}$$

$\overleftarrow{\mathbf{X}}_0$ is initialised with the uniform distribution on $\mathcal{M}$ (which is close to the one of $\mathbf{X}_T$ for large T).

The estimation of the score term $\nabla \log p_t$ is done by minimising the ism loss function (3). We refer to Section 3.3 for details on the training and parameterisation.

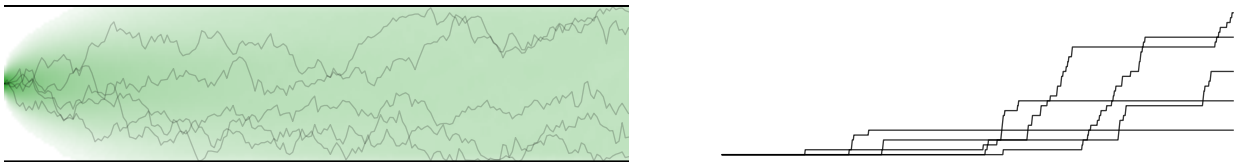
Sampling. Sampling from the forward (5) and backward (6) processes, once the score is learnt, requires a discretisation scheme. We use Geodesic Random Walks (GRW) (Jørgensen, 1975) for this purpose, see Algorithm 1. This discretisation is a generalisation of the Euler-Maruyama discretisation of SDE in Euclidean spaces, where the $+$ operator is replaced by the exponential mapping on the manifold, computing the geodesics.

Algorithm 1 *Geodesic Random Walk.* Discretisation of the SDE $\mathrm{d}\mathbf{X}_t = d(t, \mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t$.

Require: T (simulation time), N (number of steps), X_0^γ (initial point), d (drift function)
$\gamma = T/N$
for $k \in \{0, \dots, N-1\}$ **do**
  $Z_{k+1} \sim \mathrm{N}(0, \mathrm{Id})$
  $W_{k+1} = \gamma d(k\gamma, X_k) + \sqrt{\gamma} Z_{k+1}$
  $X_{k+1}^\gamma = \exp_{X_k}[W_{k+1}] \approx \mathrm{proj}_{\mathcal{M}}(X_k + W_{k+1})$
return $\{X_k\}_{k=0}^N$

However, contrary to De Bortoli et al. (2022), we will not have access explicitly to the exponential mapping of the Hessian manifold. Instead, we rely on an approximation, using a *retraction* (see Ablin & Peyré (2022) for a definition and alternative schemes).

3.2 Reflected diffusion models



$\mathrm{d}\bar{\mathbf{B}}_t = \mathrm{d}\mathbf{B}_t - \mathrm{d}\mathbf{k}_t$ $\qquad$ $|\mathbf{k}|_t$

Figure 5: *Left:* Convergence of the reflected Brownian motion on the unit interval to the uniform distribution. *Right:* Value of $|\mathbf{k}|_t$ for the trajectory samples on the left through time.

Another approach to deal with the geometry of $\mathcal{M}$ is to use the standard metric $\mathfrak{h}$ and forward dynamics of $\mathcal{N}$ and constraining it to $\mathcal{M}$ by *reflecting* the process whenever it would encounter a boundary. We will first assume that $\mathcal{M}$ is compact and convex. To simplify the presentation we focus on the Euclidean case $\mathcal{N} = \mathbb{R}^d$ with smooth boundary ∂M[4].

The key difference between this approach and the barrier approach is that in the reflected case we leave the geometry unchanged, so all we need to do is show that the dynamics induced by reflecting the forward process whenever it hits the boundary leads to an invariant distribution and admit a time-reversal.

It is worth noting that while barrier approaches have received considerable theoretical attention in the sampling literature (Lee & Vempala, 2017; Noble et al., 2022), reflected methods have remained comparatively undeveloped from the methodological and practical point of view. In the next section, we recall the basics of reflected stochastic processes.

Skorokhod problem. The reflected Brownian motion is defined as the solution to the Skorokhod problem. Roughly speaking a solution to the Skorokhod problem consists into two coupled processes $(\bar{\mathbf{B}}_t, \mathbf{k}_t)_{t \geq 0}$ such

[4]We refer to Appendix G for a definition of smooth boundary.

that $(\bar{\mathbf{B}}_t)_{t\geq 0}$ acts *locally* as a Euclidean Brownian motion $(\mathbf{B}_t)_{t\geq 0}$ and $\mathbf{k}_t$ compensates for the excursion of $(\mathbf{B}_t)_{t\geq 0}$ so that $(\bar{\mathbf{B}}_t)_{t\geq 0}$ remains in $\mathcal{M}$. We say that $(\bar{\mathbf{B}}_t, \mathbf{k}_t)_{t\geq 0}$ is a solution to the *Skorokhod problem* (Skorokhod, 1961) if $(\mathbf{k}_t)_{t\geq 0}$ and $(\bar{\mathbf{B}}_t)_{t\geq 0}$ are two processes satisfying mild conditions, see Appendix D for a rigorous introduction, such that for any $t \geq 0$,

$$\bar{\mathbf{B}}_t = \bar{\mathbf{B}}_0 + \mathbf{B}_t - \mathbf{k}_t \in \mathcal{M}, \tag{7}$$

and $|\mathbf{k}|_t = \int_0^t \mathbf{1}_{\bar{\mathbf{B}}_s \in \partial\mathcal{M}} \mathrm{d}|\mathbf{k}|_s$, $\mathbf{k}_t = \int_0^t \mathbf{n}(\bar{\mathbf{B}}_s)\mathrm{d}|\mathbf{k}|_s$, where $(|\mathbf{k}|_t)_{t\geq 0}$ is the total variation of $(\mathbf{k}_t)_{t\geq 0}$ and we recall that $\mathbf{n}$ is the outward normal to $\mathcal{M}$[5]. When $(\bar{\mathbf{B}}_t)_{t\geq 0}$ hits the boundary, the condition $\mathbf{k}_t = \int_0^t \mathbf{n}(\bar{\mathbf{B}}_s)\mathrm{d}|\mathbf{k}|_s$, tells us that $-\mathbf{k}_t$ "compensates" for $\bar{\mathbf{B}}_t$ by pushing the process back into $\mathcal{M}$ along the inward normal $-\mathbf{n}$, while the condition $|\mathbf{k}|_t = \int_0^t \mathbf{1}_{\bar{\mathbf{B}}_s \in \partial\mathcal{M}} \mathrm{d}|\mathbf{k}|_s$ can be interpreted as $\mathbf{k}_t$ being constant when $(\bar{\mathbf{B}}_t)_{t\geq 0}$ does not hit the boundary. As a result $(\bar{\mathbf{B}}_t)_{t\geq 0}$ can be understood as the continuous-time counterpart to the reflected Gaussian random walk. The process $(\mathbf{k}_t)_{t\geq 0}$ can be related to the notion of *local time* (Revuz & Yor, 2013) and quantifies the amount of time $(\bar{\mathbf{B}}_t)_{t\geq 0}$ spends at the boundary $\partial\mathcal{M}$. Lions & Sznitman (1984, Theorem 2.1) ensure the existence and uniqueness of a solution to the Skorokhod problem. One key observation is that the event $\{\bar{\mathbf{B}}_t \in \partial\mathcal{M}\}$ has probability zero (Harrison & Williams, 1987, Section 7, Lemma 7). As in the *unconstrained* setting, one can describe the dynamics of the density of $\bar{\mathbf{B}}_t$.

Proposition 3.1 (Burdzy et al. (2004))**.** *For any $t > 0$, the distribution of $\bar{\mathbf{B}}_t$ admits a density w.r.t. the Lebesgue measure denoted p_t. In addition, we have for any $x \in \mathrm{int}(\mathcal{M})$ and $x_0 \in \partial\mathcal{M}$*

$$\partial_t p_t(x) = \tfrac{1}{2}\Delta p_t(x), \quad \partial_{\mathbf{n}} p_t(x_0) = 0, \tag{8}$$

where we recall that $\mathbf{n}$ is the outward normal to $\mathcal{M}$.

Note that contrary to the unconstrained setting, the heat equation has *Neumann* boundary conditions. Similarly to the compact Riemannian setting (Saloff-Coste, 1994) it can be shown that the reflected Brownian motion converges to the uniform distribution on $\mathcal{M}$ exponentially fast (Loper, 2020), see Section 3.2. Hence, $(\bar{\mathbf{B}}_t)_{t\geq 0}$ is a candidate for a forward noising process in the context of diffusion models.

Time-reversal. In order to extend the diffusion model approach to the reflected setting, we need to derive a *time-reversal* for $(\bar{\mathbf{B}}_t)_{t\in[0,T]}$. Namely, we need to characterise the evolution of $(\overleftarrow{\mathbf{X}}_t)_{t\in[0,T]} = (\bar{\mathbf{B}}_{T-t})_{t\in[0,T]}$. It can be shown that the time-reversal of $(\bar{\mathbf{B}}_t)_{t\in[0,T]}$ is also the solution to a Skorokhod problem.

Theorem 3.2. *There exist $(\overleftarrow{\mathbf{k}}_t)_{t\geq 0}$ a bounded variation process and a Brownian motion $(\mathbf{B}_t)_{t\geq 0}$ such that*

$$\overleftarrow{\mathbf{X}}_t = \overleftarrow{\mathbf{X}}_0 + \mathbf{B}_t + \int_0^t \nabla \log p_{T-s}(\overleftarrow{\mathbf{X}}_s)\mathrm{d}s - \overleftarrow{\mathbf{k}}_t.$$

In addition, for any $t \in [0, T]$ we have

$$\overleftarrow{|\mathbf{k}|}_t = \int_0^t \mathbf{1}_{\overleftarrow{\mathbf{X}}_s \in \partial\mathcal{M}} \mathrm{d}\overleftarrow{|\mathbf{k}|}_s, \quad \overleftarrow{\mathbf{k}}_t = \int_0^t \mathbf{n}(\overleftarrow{\mathbf{X}}_s)\mathrm{d}\overleftarrow{|\mathbf{k}|}_s.$$

The proof, see Appendix G, follows Petit (1997) which provides a time-reversal in the case where $\mathcal{M}$ is the positive orthant. It is based on an extension of Haussmann & Pardoux (1986) to the reflected setting, with a careful handling of the boundary conditions. In particular, contrary to Petit (1997), we do not rely on an explicit expression of p_t but instead use the intrinsic properties of $(\mathbf{k}_t)_{t\geq 0}$. Informally, Theorem 3.2 means that the process $(\overleftarrow{\mathbf{X}}_t)_{t\in[0,T]}$ satisfies

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \nabla \log p_{T-t}(\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t - \mathrm{d}\overleftarrow{\mathbf{k}}_t, \tag{9}$$

which echoes the usual time-reversal formula (2). In practice, in order to sample from $(\overleftarrow{\mathbf{X}}_t)_{t\in[0,T]}$, one needs to consider the *reflected* version of the *unconstrained* dynamics $\mathrm{d}\overleftarrow{\mathbf{X}}_t = \nabla \log p_{T-t}(\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t$.

[5] We extend the normal $\mathbf{n}$ to the whole space by letting $\mathbf{n}(x) = 0$ if $x \notin \partial\mathcal{M}$.

Algorithm 2 *Reflected step algorithm.* The algorithm operates by repeatedly taking geodesic steps until one of the constraints is violated, or the step is fully taken. Upon hitting the boundary we parallel transport the tangent vector to the boundary and then reflect it against the boundary. We then start a new geodesic from this point in the new direction. The $\arg\operatorname{intersect}_t$ function computes the distance one must travel along a geodesic in direction $\boldsymbol{s}$ til constraint f_i is intersected. For a discussion of paralleltransport, $\exp_{\mathfrak{g}}$ and reflect please see Appendix A.

Input: $x \in \mathcal{M}$, $\boldsymbol{v} \in \mathrm{T}_x\mathcal{M}$, $\{f_i\}_{i\in\mathcal{I}}$
  $\ell \leftarrow \|\boldsymbol{v}\|_{\mathfrak{g}}$
  $\boldsymbol{s} \leftarrow \boldsymbol{v}/\|\boldsymbol{v}\|_{\mathfrak{g}}$
  while $\ell \geq 0$ **do**
    $d_i = \arg\operatorname{intersect}_t\big[\exp_{\mathfrak{g}}(x, t\boldsymbol{s}), f_i\big]$
    $i \leftarrow \arg\min_i \; d_i \; s.t. \; d_i > 0$
    $\alpha \leftarrow \min(d_i, \ell)$
    $x' \leftarrow \exp_{\mathfrak{g}}(x, \alpha\boldsymbol{s})$
    $\boldsymbol{s} \leftarrow \text{paralleltransport}_{\mathfrak{g}}(x, \boldsymbol{s}, x')$
    $\boldsymbol{s} \leftarrow \text{reflect}(\boldsymbol{s}, f_i)$
    $\ell \leftarrow \ell - \alpha$
    $x \leftarrow x'$
  return x

Algorithm 3 *Reflected Random Walk.* Discretisation of the SDE $\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t - \mathrm{d}\mathbf{k}_t$.

  Require: T (simulation time), N (number of steps), X_0^γ (initial point), $\{f_i\}_{i\in\mathcal{I}}$ (boundary functions)
  $\gamma = T/N$
  for $k \in \{0, \dots, N-1\}$ **do**
    $Z_{k+1} \sim \mathrm{N}(0, \mathrm{Id})$
    $X_{k+1}^\gamma = \text{ReflectedStep}[X_k^\gamma, \sqrt{\gamma}Z_{k+1}, \{f_i\}_{i\in\mathcal{I}}]$
  return $\{X_k^\gamma\}_{k=0}^N$

Sampling. In practice, we approximately sample the reflected dynamics by considering the Markov chain given by Algorithm 3. We refer to Pacchiarotti et al. (1998); Bossy et al. (2004) for weak convergence results on this numerical scheme in the Euclidean setting.

Likelihood evaluation. In the case of a reflected Brownian motion, it is possible to compute an equivalent ODE in order to perform likelihood evaluation. The associated ODE was first derived in Lou & Ermon (2023). The form of the ODE and the proof that it remains in $\mathcal{M}$ are postponed to Appendix F.

3.3 Loss and score network parameterisation

In order to train log-barrier and reflected diffusion models we prove that we can use a *tractable* score matching loss in constrained manifolds. We will prove that the implicit score-matching loss leads to the recovery of the correct score when we have a boundary, so long as we enforce that the score is zero on the boundary (see Appendix E.1). This proof holds for both the log-barrier and the reflected process.

Proposition 3.3. *Let $s \in \mathrm{C}^\infty([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ such that for any $x \in \partial\mathcal{M}$ and $t \geq 0$, $s_t(x) = 0$. Then, there exists $C > 0$ such that*

$$\mathbb{E}[\|\nabla \log p_t - s_t\|^2] = \mathbb{E}[\|s_t\|^2 + 2\,\operatorname{div}(s_t)] + C,$$

where $\mathbb{E}$ is taken over $\mathbf{X}_t \sim p_t$ and $t \sim \mathcal{U}([0,T])$.

This result immediately implies we can optimise the score network using the ism loss function so long as we enforce a Neumann boundary condition. The estimation of the score term $\nabla \log p_t$ is also done by minimising the ism loss function (3).

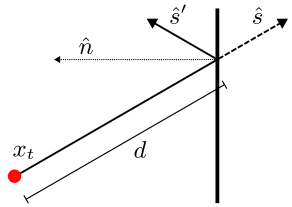


Figure 6: Reflection against a linear boundary. For a step s with magnitude $|s|$ and direction $\hat{s}$ the distance to the boundary described by the normal $\hat{n}$ and offset b is $d = \frac{\langle \hat{s}, x_t \rangle - b}{\langle \hat{s}, \hat{n} \rangle}$. The reflected direction is given by $\hat{s}' = \hat{s} - 2\langle \hat{s}, \hat{n} \rangle \hat{n}$.

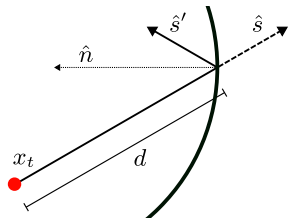


Figure 7: Reflection against a spherical boundary. For a sphere of radius r, the distance to the boundary from x_t in direction s is given by $d = \frac{1}{2}(\langle \hat{s}, x_t \rangle^2 + 4(r^2 - \|x_t\|^2))^{1/2} - \frac{1}{2}\langle \hat{s}, x_t \rangle$. The normal at the intersection can be computed as the unit vector in the direction $-2(d\hat{s} + x_t)$, and then $\hat{s}'$ as above.

The score term $\nabla \log p_{T-t}$ appearing in both time-reversal processes (6) and (9) is intractable. It is thus approximated with a *score network* $s_\theta(t, \cdot) \approx \nabla \log p_t$. We use a multi-layer perceptron architecture with sin activation functions, see Appendix J for more details on the experimental setup. Due to boundary condition (8), the normal component of the score is zero at the boundary in the reflected case. A similar result holds in the log-barrier setting. This is additionally required for our proof of the ism loss.

Following Liu et al. (2022) we can accommodate this in our score parameterisation by additionally scaling the score output of the neural network by a monotone function $h(d(x, \partial\mathcal{M}))$ where d is the distance from x to the boundary, where $h(0) = 0$. In particular we use a clipped ReLU function: $s_\theta(t, x) = \min(1, \mathrm{ReLU}(d(x, \partial\mathcal{M}) - \delta)) \cdot \mathrm{NN}_\theta(t, x)$ with $\delta > 0$ a threshold so the model is forced to be zero "close" to the boundary as well as exactly on the boundary, see Appendix J for an illustration. The inclusion of this scaling function is necessary to produce reasonable results as we show in Appendix E.2 The weighting function in (3) is set to $\lambda_t = (t + 1)$.

The forward processes (5) and (7) for the barrier and reflected methods cannot be sampled in closed form, so at training time samples from the conditional marginals $p(\mathbf{X}_t | \mathbf{X}_0)$ are obtained by discretising these processes. As to take the most of this computational overhead, we use several samples from the discretised forward trajectory $(\mathbf{X}_{t_1}, \cdots, \mathbf{X}_{t_k} | \mathbf{X}_0)$ instead of only using the last sample.

4 Related Work

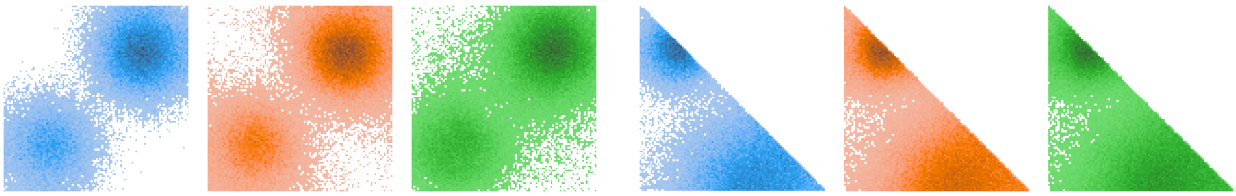
Sampling on constrained manifolds. Sampling from a distribution on a space defined by a set of constraints is an important ingredient in several computational tasks, such as computing the volume of a polytope (Lee & Vempala, 2017). Incorporating such constraints inside MCMC algorithms while preserving fast convergence properties is an active field of research (Kook et al., 2022; Lee & Vempala, 2017; Noble et al., 2022). In this work, we are interested in sampling from the uniform distribution defined on the constrained set in order to define a proper *forward process* for our diffusion model. Log-barrier methods such as the Dikin walk or Riemannian Hamiltonian Monte Carlo Kannan & Narayanan (2009); Lee & Vempala (2017); Noble et al. (2022) change the geometry of the underlying space and define stochastic processes which never violate the constraints, see Kannan & Narayanan (2009); Noble et al. (2022) for instance. If we keep the Euclidean metric, then *unconstrained* stochastic processes might not be well-defined for all times. To counter this effect, it has been proposed to *reflect* the Brownian motion Williams (1987); Petit (1997); Shkolnikov & Karatzas (2013). Finally, we also highlight hit-and-run approaches (Smith, 1984; Lovász & Vempala, 2006) which generalise Gibbs' algorithm and enjoy fast convergence properties provided that one knows how to sample from the one-dimensional marginals.

Diffusion models on manifolds. De Bortoli et al. (2022) extended the work of Song et al. (2021) to Riemannian manifolds by defining forward and backward stochastic processes in this setting. Concurrently, a similar framework was introduced by Huang et al. (2022) extending the maximum likelihood approach of

Table 1: MMD metrics between synthetic data of varying dimension and samples from trained constrained diffusion models. Bold indicates best results.

Space	Dimension	Log-barrier	Reflected
$[-1, 1]^d$	2	$0.066_{\pm 0.006}$	$0.055_{\pm 0.015}$
	3	$0.209_{\pm 0.077}$	$0.080_{\pm 0.004}$
	10	$0.330_{\pm 0.004}$	$0.313_{\pm 0.048}$

Space	Dimension	Log-barrier	Reflected
Δ^d	2	$0.050_{\pm 0.012}$	$0.043_{\pm 0.002}$
	3	$0.238_{\pm 0.009}$	$0.181_{\pm 0.007}$
	10	$0.275_{\pm 0.015}$	$0.290_{\pm 0.009}$



(a) 2D square data. (b) 2D Dirichlet data.

Figure 8: Histograms of samples from the data distribution and from trained constrained diffusion models.

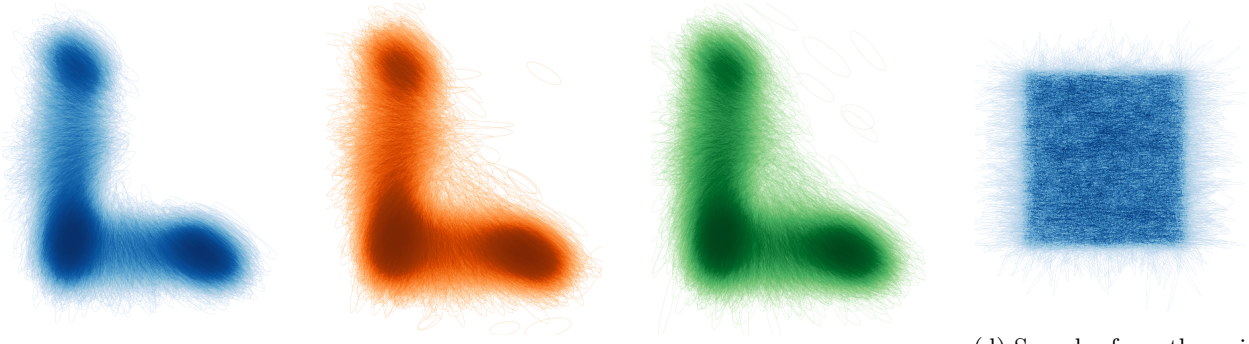
Huang et al. (2021). Existing applications of denoising diffusion models on Riemannian manifolds have been focused on well-known manifolds for which one can find metrics so that the framework of De Bortoli et al. (2022) applies. In particular, on compact Lie groups, geodesics and Brownian motions can be defined in a canonical manner. Their specific structure can be leveraged to define efficient diffusion models Yim et al. (2023). Leach et al. (2022) define diffusion models on SO(3) for rotational alignment, while Jing et al. (2022) consider the product of tori for molecular conformer generation. Corso et al. (2022) use diffusion models on $\mathbb{R}^3 \times \mathrm{SO}(3) \times \mathrm{SO}(2)$ for protein docking applications. RFDiffusion (Watson et al., 2022) and FrameDiff also incorporate SE(3) diffusion models. Finally, Urain et al. (2022) introduce a methodology for SE(3) diffusion models with applications to robotics.

5 Experimental results

To illustrate the scalability and flexibility of the constrained diffusion models introduced in Section 3, we first construct a series of increasingly difficult synthetic tasks on convex polytopes, including the hypercube and the simplex, in Section 5.1. We then demonstrate their applicability to more practical settings by considering two real-world problems from robotics and protein design. In particular, we show that our models are able to learn distributions over the space of $d \times d$ symmetric positive definite (SPD) matrices S^d_{++} under trace constraints in Section 5.2—a setting that is essential to describing and controlling the motions and exerted forces of robotic platforms (Yoshikawa, 1985). In Section 5.3, we use the parametrisation introduced in Han & Rudolph (2006) to map the problem of modelling the conformational ensembles of proteins under positional constraints on their endpoints to the product manifold of a convex polytope and a torus. We refer the reader to Appendix J for more details on the experimental setup. Additionally, the codebase will be made public upon publication.

5.1 Method characterisation on convex polytopes

First, we aim to assess the scalability of our methods on constrained manifolds of increasing dimensionality. For this, we focus on polytopes and construct synthetic datasets that represent bimodal distributions—in particular, we investigate two specific instances of polytopes: hypercubes and simplices. In Appendix J.1 we also present results on the Birkhoff polytope. We measure the performance of each model via the Maximum Mean Discrepancy (MMD) (Gretton et al., 2012), which is a kernel-based metric between distributions.



(a) Samples from the data distribution.

(b) Samples from our log-barrier diffusion model.

(c) Samples from our reflected diffusion model.

(d) Samples from the uniform distribution.

Figure 9: Samples of joint distributions over SPD matrices (illustrated as ellipsoids) and their location from (a) the data distribution, (b) our log-barrier diffusion model, (c) our reflected diffusion model and (d) the uniform distribution. For visual clarity, a small fraction of samples with a large determinant was omitted from the plot, see Appendix J.2 for additional results and full correlation plots.

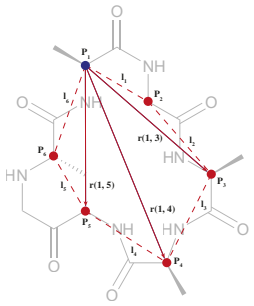
We present a qualitative comparison of the logarithmic barrier and reflected Brownian motion models in Figure 8, and observe that both methods recover the data distribution on the two-dimensional hypercube and simplex, although the reflected method seems to produce a better fit. In Table 1, we report the MMD between the data distribution and the learnt diffusion models. We similarly observe that the reflected method consistently gives better results than the log-barrier one. One key drawback of the log-barrier method is the slow convergence of the forward process, as already illustrated in Figure 3. It should also be highlighted that in order to implement the log-barrier method, one needs to invert a $d \times d$ matrix, which requires $O(d^3)$ operations. This prevents the use of the log-barrier method for very high-dimensional applications. Additionally, although we can derive the likelihood evaluation for both the log-barrier and reflected models (see Appendix F), in practice the numerical issues in the log-barrier approach prevent actual evaluation.

5.2 Modelling robotic arms under force and velocity constraints

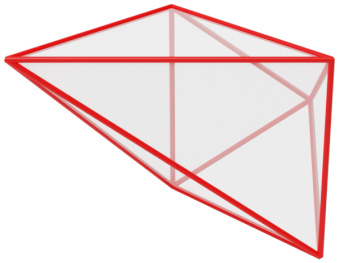
Accurately determining and controlling the movement and exerted forces of robotic platforms is a fundamental problem in many real-world robotics applications. A kinetostatic descriptor that is commonly used to quantify the ability of a robotic arm to move and apply forces along certain dimensions is the so-called manipulability ellipsoid $E \in \mathbb{R}^d$ (Yoshikawa, 1985) which is naturally described as a symmetric positive definite (SPD) matrix $\mathrm{M} \in \mathbb{R}^{d \times d}$ (Jaquier et al., 2021). The manifold of such $d \times d$ SPD matrices, denoted as S^d_{++}, is defined as the set of matrices $\{x^\top \mathrm{M} x \geq 0,\ x \in \mathbb{R}^d : \mathrm{M} \in \mathbb{R}^{d \times d}\}$.

In many practical settings, it may be desirable to constrain the volume of E to retain flexibility or limit the magnitude of an exerted force, which can be expressed as an upper bound on the trace of M, i.e. as the inequality constraint $\sum_{i=1}^d \mathrm{M}_{ii} < C$ with $C > 0$. Constraining the rest of the entries of the matrix to ensure it is SPD is non-trivial. Alternatively, we can parameterise the SPD matrices via their Cholesky decomposition. Each SPD matrix has a unique decomposition of the form $\mathrm{M} = \mathrm{L}\mathrm{L}^\top$, where L is a lower triangular matrix with strictly positive diagonal (Golub & Van Loan, 2013, p.143). Constraining the entries of this matrix simply requires ensuring the diagonal is positive. The trace of the SPD matrix is given by $\mathrm{Tr}(\mathrm{M}) = \mathrm{Tr}(\mathrm{L}\mathrm{L}^\top) = \sum_{ij} \mathrm{L}^2_{ij}$, and results in the constraint on the entries of L to live in a ball of radius C.

We additionally model the two-dimensional position of the arm. In summary, the space over which we parameterise the diffusion models is defined as $\{\mathrm{L} \in \mathbb{R}^{d(d+1)/2} : \mathrm{L}_{i,i} > 0, \sum_{i,j} \mathrm{L}^2_{i,j} < C\} \times \mathbb{R}^2$. Under the Euclidean metric, we can apply both our log-barrier and reflected approaches. The positive diagonal (linear) constraint is handled similarly to the polytope setting. The reflection on the ball boundary is defined and illustrated on Figures 6 and 7. Using this framework, we model the datasets presented in Jaquier et al. (2021) (see Appendix I for full experimental details). The joint distribution over the SPD matrices (represented as ellipsoids) and their positions is presented in Figure 9. We qualitatively observe that the reflected method is



(a) An illustrative diagram of the parameterisation used for the conformational modelling of the C_α trace of a cyclic peptide, introduced in Han & Rudolph (2006).



(b) The convex polytope constraining the diagonals of the triangles for the given bond lengths in the illustrated molecule. The total design space is the product of this polytope with the 4D flat torus.

Figure 10: Illustrations of the parametrisation used to model distribution of polypeptide backbone conformations under anchor point distance constraints as the product of a convex polytope $\mathbb{P}$ and torus $\mathbb{T}$: $\mathbb{P}^3 \times \mathbb{T}^4$.

able to fit the joint data distribution. Quantitatively, the reflected method achieves an MMD of 0.161 as opposed to 0.247 for the log-barrier one.

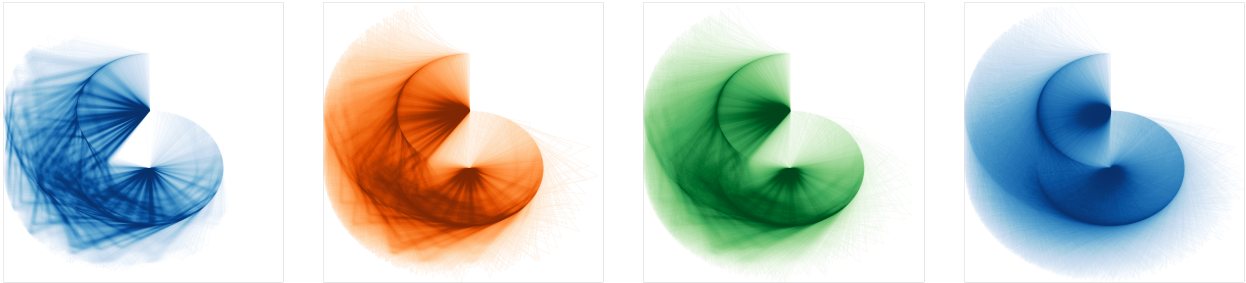
5.3 Modelling protein loops with anchored endpoints

Modelling the conformational ensembles of proteins is an important task in the field of molecular biology, particularly in the context of bioengineering and drug discovery. In many data-scarce practical settings such as antibody or enzyme design, it is often unnecessary or even undesirable to model the structure of an entire protein, as researchers are primarily interested in specific functional sites with distinct biochemical properties. However, generating conformational ensembles for such substructural elements necessitates positional constraints on their endpoints to ensure that they can be accommodated by the remaining scaffold.

This problem can be reformulated as modelling a spatial chain with spherical joints and fixed end points. Following the framework outlined in Han & Rudolph (2006), we parametrise the conformations of such chains with d fixed-length links and arbitrary end-point constraints as the product of a convex polytope $\mathbb{P}$ and torus $\mathbb{T}$: $\mathbb{P}^{(d-2)} \times \mathbb{T}^{(d-1)}$. The essential idea in this parameterisation is to fix one end point as an "anchor" and model the chain as the series of triangles formed by the anchor and each pair of adjacent joints in the chain. A point in the polytope corresponds to the lengths of the diagonals of these triangles, and a point in the torus to the angles between each subsequent triangle. See Figure 10 for an illustration and Appendix H.1 for a full description.

Using this framework, we model the conformational landscape of the cyclic peptide C-AAGAGG, consisting of a circular polypeptide chain with coinciding endpoints. We generate 10^6 backbone conformations using tools from molecular dynamics (Eastman et al., 2017; Hornak et al., 2006) and divide them into training and evaluation datasets, (see Appendix H.2 for full experimental details). Drawing on the definition above, the space on which we effectively parameterise our constrained diffusion models for a circular polypeptide chain of length $d = 6$ is given by the product manifold $\mathbb{P}^3 \times \mathbb{T}^4$.

To learn a distribution over this space, we leverage the methodology introduced in Section 3 for the polytope component $\mathbb{P}^3$ and in De Bortoli et al. (2022) for the torus component $\mathbb{T}^4$. A qualitative comparison of samples from the data distribution, our log-barrier and reflected diffusion models, and the uniform distribution is presented in Figures 11a to 11d. For enhanced visual clarity, we project the modelled spatial chain onto the 2D plane by removing the (unconstrained) torus component of the product manifold and only plotting the planar chains encoded by the (constrained) polytope component (a correlation plot of the full product manifold is presented in Figure 26). It is immediately obvious that the data distribution is highly multimodal, encompassing a large number of locally optimal conformational clusters. Nevertheless, both our reflected and log-barrier diffusion models are able to robustly approximate this challenging energetic landscape, producing



(a) Samples from the data distribution.

(b) Samples from our log-barrier diffusion model.

(c) Samples from our reflected diffusion model.

(d) Samples from the uniform distribution.

Figure 11: Planar projection of the modelled C_α chains from (a) the training dataset, (b) our log-barrier diffusion model, (c) our reflected diffusion model and (d) the uniform distribution. Additional results and full correlation plots are postponed to Appendix J.3.

samples that reflect key conformational states. Interestingly when we quantiatively evaluate the sample the log-barrier and the reflected models both learn about the same with respect to an RBF MMD: $0.0319_{\pm 0.001}$ for the log-barrier and $0.321_{\pm 0.021}$ for the reflected. The improved relative performance of the log-barrier method seems likely due to the fact that the data for this experiment is largely in the interior of the constrained set, not close to the boundary.

6 Discussion

Learning complex distributions supported on constrained spaces is a crucial task in many natural and engineering sciences, including computational statistics (Morris, 2002), robotics (Han & Rudolph, 2006), quantum physics (Lukens et al., 2020) and computational biology (Thiele et al., 2013). In this work we extend continuous diffusion models to this setting, proposing two complementary approaches—one based on log-barrier methods and the other on the reflected Brownian motion. For both approaches, we derive the time-reversal formula, propose discretisation schemes and extend the score-matching toolbox. We demonstrate the utility of our methods on a range of synthetic and real-world tasks, including the constrained conformational modelling of proteins and robotic arms, and find that reflected methods, while enjoying fewer theoretical guarantees than their log-barrier counterparts, often yield preferable results.

Concurrently to our work, Lou & Ermon (2023) introduced a reflected diffusion model for image applications. They introduce reflected stochastic processes in order to give a methodological and theoretical justification to the *thresholding* procedure commonly used in large image diffusion models such as (Saharia et al., 2022).

We conclude by highlighting important future directions. First, the computational cost of performing the reflection when discretising the reflected Brownian motion is high. Finding numerically efficient approximations of the reflected process is therefore necessary to extend this methodology to very high dimensional settings. Second, the retraction used in place of the exponential map for the barrier method leads to a high number of discretisation steps to ensure a good approximation. Designing a faster forward process for the log-barrier method is key to targeting more complex distributions.

References

Pierre Ablin and Gabriel Peyré. Fast and accurate optimization on the orthogonal manifold without retraction. In *International Conference on Artificial Intelligence and Statistics*, pp. 5636–5657. PMLR, 2022.

RN Bhattacharya. Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *The Annals of Probability*, pp. 541–553, 1978.

Mireille Bossy, Emmanuel Gobet, and Denis Talay. A symmetrized euler scheme for an efficient approximation of reflected diffusions. *Journal of applied probability*, 41(3):877–889, 2004.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Krzysztof Burdzy, Zhen-Qing Chen, and John Sylvester. The heat equation and reflected brownian motion in time-dependent domains. *The Annals of Probability*, 32(1B):775–804, 2004.

Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking, October 2022. URL http://arxiv.org/abs/2210.01776.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modeling, 2022.

Patrick G Dougherty, Ashweta Sahni, and Dehua Pei. Understanding cell penetration of cyclic peptides. *Chemical Reviews*, 119(17):10241–10287, 2019.

Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7): e1005659, 2017.

Khashayar Gatmiry and Santosh S Vempala. Convergence of the riemannian langevin algorithm. *arXiv preprint arXiv:2204.10818*, 2022.

Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(null):723–773, March 2012. ISSN 1532-4435.

Li Han and Lee Rudolph. Inverse kinematics for a serial chain with joints under distance constraints. In *Robotics: Science and systems*, 2006.

J Michael Harrison and Ruth J Williams. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics: An International Journal of Probability and Stochastic Processes*, 22(2): 77–115, 1987.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pp. 1188–1205, 1986.

Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N Mendoza, Anne Richelle, Almut Heinken, Hulda S Haraldsdóttir, Jacek Wachowiak, Sarah M Keating, Vanja Vlasov, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nature protocols*, 14(3):639–702, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 2020.

Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

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

Chin-Wei Huang, Jae Hyun Lim, and Aaron C Courville. A variational perspective on diffusion-based generative models and score matching. *Advances in Neural Information Processing Systems*, 34:22863–22876, 2021.

Chin-Wei Huang, Milad Aghajohari, Avishek Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian Diffusion Models, August 2022.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*. Elsevier, 2014.

John Ingraham, Max Baranov, Zak Costello, Vincent Frappier, Ahmed Ismail, Shan Tie, Wujie Wang, Vincent Xue, Fritz Obermeyer, Andrew Beam, et al. Illuminating protein space with a programmable generative model. *bioRxiv*, pp. 2022–12, 2022.

Noémie Jaquier, Leonel Rozo, Darwin G Caldwell, and Sylvain Calinon. Geometry-aware manipulability learning, tracking, and transfer. *The International Journal of Robotics Research*, 40(2-3):624–650, 2021.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional Diffusion for Molecular Conformer Generation, June 2022. URL http://arxiv.org/abs/2206.01729.

Erik Jørgensen. The central limit problem for geodesic random walks. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1-2):1–64, 1975.

Weining Kang and Kavita Ramanan. On the submartingale problem for reflected diffusions in domains with piecewise smooth boundaries. 2017.

Ravi Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, pp. 561–570, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585062. doi: 10.1145/1536414.1536491. URL https://doi.org/10.1145/1536414.1536491.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL http://arxiv.org/abs/1412.6980.

Yunbum Kook, Yin Tat Lee, Ruoqi Shen, and Santosh S Vempala. Sampling with riemannian hamiltonian monte carlo in a constrained space. *arXiv preprint arXiv:2202.01908*, 2022.

Thomas J Lane. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, pp. 1–4, 2023.

Adam Leach, Sebastian M Schmon, Matteo T Degiacomi, and Chris G Willcocks. Denoising Diffusion Probabilistic Models on SO(3) for Rotational Alignment. *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, pp. 8, 2022.

John M Lee. Smooth manifolds. In *Introduction to Smooth Manifolds*, pp. 1–31. Springer, 2013.

Yin Tat Lee and Santosh S. Vempala. Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 927–940, Montreal Canada, June 2017. ACM. ISBN 978-1-4503-4528-6. doi: 10.1145/3055399.3055416.

Yin Tat Lee and Santosh S Vempala. Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1115–1121, 2018.

Pierre-Louis Lions and Alain-Sol Sznitman. Stochastic differential equations with reflecting boundary conditions. *Communications on pure and applied Mathematics*, 37(4):511–537, 1984.

Song Liu, Takafumi Kanamori, and Daniel J Williams. Estimating density models with truncation boundaries using score matching. *Journal of Machine Learning Research*, 23(186):1–38, 2022.

Jackson Loper. Uniform ergodicity for brownian motion in a bounded convex set. *Journal of Theoretical Probability*, 33(1):22–35, 2020.

Aaron Lou and Stefano Ermon. Reflected diffusion models. *arXiv preprint arXiv:2304.04740*, 2023.

László Lovász and Santosh Vempala. Hit-and-run from a corner. *SIAM Journal on Computing*, 35(4): 985–1005, 2006. doi: 10.1137/S009753970544727X.

Joseph M Lukens, Kody JH Law, Ajay Jasra, and Pavel Lougovski. A practical and efficient approach for bayesian quantum state estimation. *New Journal of Physics*, 22(6):063038, 2020.

Sean P Meyn and Richard L Tweedie. Stability of markovian processes ii: Continuous-time processes and sampled chains. *Advances in Applied Probability*, 25(3):487–517, 1993.

Nina Miolane, Nicolas Guigui, Alice Le Brigant, Johan Mathe, Benjamin Hou, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, Hatem Hajri, Yann Cabanes, Thomas Gerald, Paul Chauchat, Christian Shewmake, Daniel Brooks, Bernhard Kainz, Claire Donnat, Susan Holmes, and Xavier Pennec. Geomstats: A python package for riemannian geometry in machine learning. *Journal of Machine Learning Research*, 21(223):1–9, 2020. URL http://jmlr.org/papers/v21/19-027.html.

Ben J. Morris. Improved bounds for sampling contingency tables. *Random Structures & Algorithms*, 21(2): 135–146, 2002. doi: https://doi.org/10.1002/rsa.10049. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.10049.

Yurii Nesterov et al. *Lectures on convex optimization*, volume 137. Springer, 2018.

Maxence Noble, Valentin De Bortoli, and Alain Durmus. Barrier Hamiltonian Monte Carlo, October 2022.

David Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006.

Barbara Pacchiarotti, Cristina Costantini, and Flavio Sartoretto. Numerical approximation for functionals of reflecting diffusion processes. *SIAM Journal on Applied Mathematics*, 58(1):73–102, 1998.

Frédérique Petit. Time reversal and reflected diffusions. *Stochastic Processes and their Applications*, 69(1): 25–53, July 1997. ISSN 03044149. doi: 10.1016/S0304-4149(97)00035-5.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

Piotr Rotkiewicz and Jeffrey Skolnick. Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of computational chemistry*, 29(9):1460–1465, 2008.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35: 36479–36494, 2022.

Laurent Saloff-Coste. Precise estimates on the rate at which certain diffusions tend to equilibrium. *Mathematische Zeitschrift*, 217(1):641–677, 1994.

Hirohiko Shima and Katsumi Yagi. Geometry of hessian manifolds. *Differential geometry and its applications*, 7(3):277–290, 1997.

Mykhaylo Shkolnikov and Ioannis Karatzas. Time-reversal of reflected Brownian motions in the orthant, July 2013. URL http://arxiv.org/abs/1307.4422.

Anatoliy V Skorokhod. Stochastic equations for diffusion processes in a bounded region. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.

Robert L. Smith. Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions. *Operations Research*, 32(6):1296–1308, December 1984. doi: 10.1287/opre.32.6.1296. URL https://ideas.repec.org/a/inm/oropre/v32y1984i6p1296-1308.html.

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 and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, 2019.

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

Ines Thiele, Neil Swainston, Ronan M T Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, Stefan G Thorleifsson, Rasmus Agren, Christian Bölling, Sergio Bordel, Arvind K Chavali, Paul Dobson, Warwick B Dunn, Lukas Endler, David Hala, Michael Hucka, Duncan Hull, Daniel Jameson, Neema Jamshidi, Jon J Jonsson, Nick Juty, Sarah Keating, Intawat Nookaew, Nicolas Le Novère, Naglis Malys, Alexander Mazein, Jason A Papin, Nathan D Price, Evgeni Selkov, Martin I Sigurdsson, Evangelos Simeonidis, Nikolaus Sonnenschein, Kieran Smallbone, Anatoly Sorokin, Johannes H G M van Beek, Dieter Weichart, Igor Goryanin, Jens Nielsen, Hans V Westerhoff, Douglas B Kell, Pedro Mendes, and Bernhard Ø Palsson. A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, March 2013. ISSN 1087-0156. doi: 10.1038/nbt.2488.

Brian L. Trippe, Jason Yim, Doug Tischer, Tamara Broderick, David Baker, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem, June 2022.

Julen Urain, Niklas Funk, Georgia Chalvatzaki, and Jan Peters. Se (3)-diffusionfields: Learning cost functions for joint grasp and motion optimization through diffusion. *arXiv preprint arXiv:2209.03855*, 2022.

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

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models, December 2022.

R. J. Williams. Reflected Brownian motion with skew symmetric data in a polyhedral domain. *Probability Theory and Related Fields*, 75(4):459–485, August 1987. ISSN 0178-8051, 1432-2064. doi: 10.1007/BF00320328.

Kevin E Wu, Kevin K Yang, Rianne van den Berg, James Y Zou, Alex X Lu, and Ava P Amini. Protein structure generation via folding diffusion. *arXiv preprint arXiv:2209.15611*, 2022.

Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023.

Tsuneo Yoshikawa. Manipulability of robotic mechanisms. *The international journal of Robotics Research*, 4 (2):3–9, 1985.

Supplementary to: Diffusion Models for Constrained Domains

Introduction

In this supplementary, we first recall some key concepts from Riemannian geometry in Appendix A. In Appendix B we remind the expression of the Brownian motion in local coordinates. Details about the geodesic Brownian motion are given in Appendix C. Background on the Skorokhod problem is given in Appendix D. In Appendix E we derive the implicit score matching loss. We give details about the likelihood evaluation in Appendix F. Then in Appendix G we prove the time-reversal formula of reflected Brownian motion. In Appendix H we give some background on the conformational modelling of proteins backbone for the experiment in Section 5.3. In Appendix I we detail the geometrical constraint arising from the configurational robotics arm modelling experiment from Section 5.2. Additional results, training and miscalleneous experimental details are reported in Appendix J.

A Manifold concepts

For readers unfamiliar with Riemannian geometry here we give a brief overview of some of the key concepts. This is not a technical introduction, but a conceptual one for the understanding of terms. For a technical introduction, we refer readers to Lee (2013). A Riemannian manifold is a tuple $(\mathcal{M}, \mathfrak{g})$ with $\mathcal{M}$ a smooth manifold and $\mathfrak{g}$ a metric which defines an inner product on tangent spaces.

A *smooth manifold* is a topological space which locally can be described by Euclidean space. It is characterised by a family of *charts* $\{U \subset \mathcal{M}, \phi : U \to \mathbb{R}^d\}$, homeomorphic mappings between subsets of the manifold and Euclidean space. The collection of charts must cover the whole manifold. For the manifolds to be smooth the charts must be smooth, and the transition between charts where their domains overlap must also be smooth.

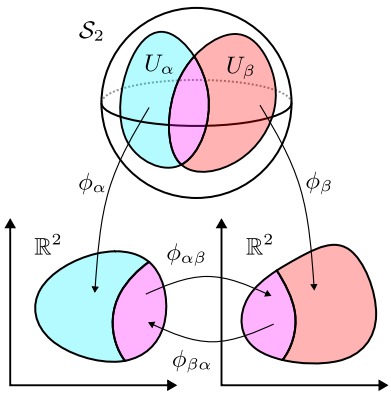


Figure 12: Example charts of the 2D manifold $\mathcal{S}^2$.

The *metric* on a Riemannian manifold gives the manifold a notion of distance and curvature. The same underlying smooth manifold with different metrics can look wildly different. The metric is defined as a smooth choice of positive definite inner product on each of the tangent spaces of the manifold. That is we have at every point a symmetric bilinear map

$$\mathfrak{g}(p) : \mathrm{T}_p\mathcal{M} \times \mathrm{T}_p\mathcal{M} \to \mathbb{R}$$

The tangent space of a point on a manifold is the generalisation of the notion of tangent planes and can be thought of as the space of derivatives of scalar functions on the manifold at that point. The collection of all tangent spaces is written as $\mathrm{T}\mathcal{M} = \bigcup_{p\in\mathcal{M}} \mathrm{T}_p\mathcal{M}$ and is called the tangent bundle. It is also manifold. Vector fields on manifolds are defined by making a choice of tangent vector at every point on the manifold in a smooth fashion. The space of vector fields is written at $\Gamma(\mathrm{T}\mathcal{M})$ and is more technically the space of *sections* of the tangent bundle.

One thing that the metric itself does not immediately define is how different tangent spaces at points on the manifold relate to one another. For this, we need additionally the concept of a *connection*. A connection is a map that takes two vector fields and produces a derivative of the first with respect to the second, that is a function $\nabla : \Gamma(\mathrm{T}\mathcal{M}) \times \Gamma(\mathrm{T}\mathcal{M}) \to \Gamma(\mathrm{T}\mathcal{M})$ and it typically written as $\nabla(X, Y) = \nabla_X Y$. Such a connection must for $X, Y, Z \in \Gamma(\mathrm{T}\mathcal{M})$ and smooth functions on the manifold $a, b : \mathcal{M} \to \mathbb{R}$ obey the following conditions:

(a) $\nabla_{aX+bY} Z = a\nabla_X Z + b\nabla_Y Z$,

(b) $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$,

(c) $\nabla_X(aY) = \partial_X aY + a\nabla_X Y$,

where $\partial_X aY$ is the regular *directional derivative* of aY in the direction X. These conditions ensure the connection is a well-defined derivative.

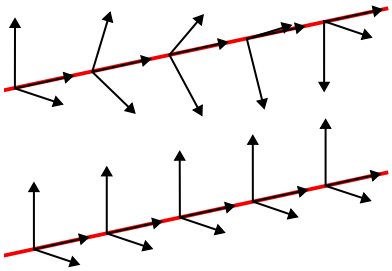


Figure 13: *Top:* Parallel transport of vectors along the red path with non-zero torsion. *Bottom:* Parallel transport of vectors along the red path with zero torsion. Both under the Euclidean metric.

On a given manifold, there are infinitely many connections. Fortunately, there is a natural choice, called the *Levi-Cevita* connection if we impose two additional conditions:

(a) $X \cdot (\mathfrak{g}(Y,Z)) = \mathfrak{g}(\nabla_X Y, Z) + \mathfrak{g}(Y, \nabla_X Z)$,

(b) $[X,Y] = \nabla_X Y - \nabla_Y X$,

where $[\cdot,\cdot]$ is the Lie bracket. The first condition ensures that the metric is preserved by the connection. That is to say, the *parallel transport* (to be defined shortly) using the connection leaves inner products unchanged on the manifold. The second ensures the connection is *torsion-free*. The change in tangent space along a geodesic (again to be defined shortly) can be described in two parts, the *curvature*, how the tangent space rotates perpendicular to the direction of travel, and the torsion, how the tangent space rotates around the axis of the direction of travel. The curvature of the connection is uniquely fixed by the other 4 conditions (the well-defined derivative and preservation of the metric). The torsion however is not fixed. By requiring it to be zero we ensure a unique connection. The requirement of zero torsion also has implications for ensuring integrability on the manifold.

With the metric and the Levi-Cevita connection in hand, we can define a number of key concepts.

We say that a vector field X is *parallel* to a curve $\gamma : (0,1) \to \mathcal{M}$ if

$$\nabla_{\gamma'} X = 0$$

where $\gamma' : (0,1) \to \mathrm{T}_{\gamma(t)}\mathcal{M}$ is the derivative of the path. For two points on the manifold $p, q \in \mathcal{M}$ and a curve between them, γ, $\gamma(0) = p$, $\gamma(1) = q$, for an initial vector $X_0 \in \mathrm{T}_p\mathcal{M}$ there is a unique vector field X that is parallel to γ such that $X(p) = X_0$. This induces a map between the tangent spaces at p and q

$$\tau_\gamma : \mathrm{T}_p\mathcal{M} \to \mathrm{T}_q\mathcal{M}$$

This map is referred as the *parallel transport* of tangent vectors between p and q, and this satisfies the condition that for $\boldsymbol{v}, \boldsymbol{u} \in \mathrm{T}_p\mathcal{M}$

$$\mathfrak{g}(p)(\boldsymbol{v}, \boldsymbol{u}) = \mathfrak{g}(q)(\tau_\gamma(\boldsymbol{v}), \tau_\gamma(\boldsymbol{u})).$$

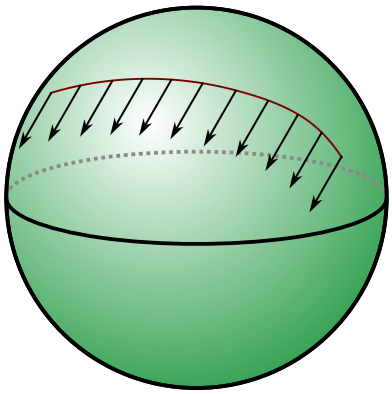


Figure 14: Parallel transport of vectors along a path on the sphere.

A *geodesic* on a manifold is the unique path on the manifold $\gamma : (0,1) \to \mathcal{M}$ such that $\nabla_{\gamma'}\gamma' = 0$. It is also the shortest path between two points on a manifold in the sense that

$$L(\gamma) = \int_0^1 \sqrt{\mathfrak{g}(\gamma(t))(\gamma'(t), \gamma(t))},$$

is minimal out of any path between the start and end of the geodesic. Geodesics give the notion of 'straight lines' on manifolds. We define the *exponential map* on a manifold as the mapping between an element of the tangent space at point p, $\boldsymbol{v} \in T_p\mathcal{M}$ and the endpoint of the unique geodesic γ with $\gamma(0) = p$ and $\gamma'(0) = \boldsymbol{v}$. In the tangent space of a manifold, we require the notion of a reflection in order to reflect geodesics off of boundary constraints. If at a point in the manifold with have $\boldsymbol{v} \in T_p\mathcal{M}$ and a unit vector normal to the constraint $\boldsymbol{n} \in T_p\mathcal{M}$, then the reflection of this vector is given by $\boldsymbol{v}' = \boldsymbol{v} - 2\mathfrak{g}(\boldsymbol{v}, \boldsymbol{n})\boldsymbol{n}$.

B Brownian motion in local coordinates

We consider a smooth function $f \in \mathrm{C}^\infty(\mathcal{M})$. The Laplace-Beltrami operator $\Delta_\mathcal{M}$ is given by $\Delta_\mathcal{M}(f) = \mathrm{div}(\mathrm{grad}(f)))$. In local coordinates we have

$$\mathrm{div}(X) = (\det(\mathfrak{g})^{-1/2}) \sum_{i=1}^d \partial_i(\det(\mathfrak{g})^{1/2} X_i),$$

as well as

$$\mathrm{grad}(f) = \mathfrak{g}^{-1}\nabla f.$$

Therefore, the Laplace-Beltrami operator is given by

$$\Delta_\mathcal{M}(f) = \sum_{i,j=1}^d \mathfrak{g}_{i,j}^{-1}\partial_{i,j} f + (\det(\mathfrak{g})^{-1/2}) \sum_{i,j=1}^d \partial_i(\det(\mathfrak{g})^{1/2}\mathfrak{g}_{i,j}^{-1})\partial_j f.$$

Therefore, in local coordinates the infinitesimal generator associated with the Laplace-Beltrami operator is given by

$$\mathcal{A}(f) = \sum_{i,j=1}^d \mathfrak{g}_{i,j}^{-1}\partial_{i,j} f + \langle b^i, \nabla f\rangle,$$

with

$$b^i = (\det(\mathfrak{g})^{-1/2}) \sum_{j=1}^d \partial_j(\det(\mathfrak{g})^{1/2}\mathfrak{g}_{i,j}^{-1}). \tag{10}$$

Therefore, the dual operator associated with $\mathcal{A}$ is given by

$$\mathcal{A}^\star(f) = \sum_{i,j=1}^d \partial_{i,j}(\mathfrak{g}_{i,j}^{-1} f) - \sum_{i=1}^d \partial_i(b^i f).$$

Note that by letting $f = \det(\mathfrak{g})^{1/2}$ we get that $\mathcal{A}^\star(f) = 0$ and therefore we recover that $p \propto \det(\mathfrak{g})$ is the invariant distribution of the Brownian motion.

Langevin dynamics on $\mathcal{M}$. We know that the Brownian motion targets $\det(\mathfrak{g})^{1/2}$. Therefore in order to correct and sample from the uniform distribution we consider the Langevin dynamics

$$\mathrm{d}\mathbf{X}_t = -\mathrm{grad}\log(\det(\mathfrak{g})^{1/2})(\mathbf{X}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{B}_t^\mathcal{M}.$$

Note that in the previous equation grad and $\mathbf{B}_t^\mathcal{M}$ are defined w.r.t. the metric of the manifold. In local coordinates we have

$$\mathrm{d}\mathbf{X}_t = \{b - \mathrm{grad}\log(\det(\mathfrak{g})^{1/2})\}(\mathbf{X}_t)\mathrm{d}t + \sqrt{2}\mathfrak{g}(\mathbf{X}_t)^{-1/2}\mathrm{d}\mathbf{B}_t. \tag{11}$$

where $b = \{b^i\}_{i=1}^d$ is given in equation 10. In addition, we have

$$\mathrm{grad}\log(\det(\mathfrak{g})^{1/2}) = \det(\mathfrak{g})^{-1/2}\mathfrak{g}^{-1}\nabla\det(\mathfrak{g}). \tag{12}$$

Using equation 10 we have

$$b^i = (\det(\mathfrak{g})^{-1/2}) \sum_{j=1}^d \partial_j(\det(\mathfrak{g})^{1/2}\mathfrak{g}_{i,j}^{-1}) = \sum_{j=1}^d \partial_j \mathfrak{g}_{i,j}^{-1} + \mathrm{grad}\log(\det(\mathfrak{g})^{1/2})_i$$

This can also be rewritten as

$$\mathrm{div}_\mathcal{M}(\mathfrak{g}^{-1}) = \mu + \mathrm{grad}\log(\det(\mathfrak{g})^{1/2}),$$

with

$$\mu_i = \sum_{j=1}^d \partial_j \mathfrak{g}_{i,j}^{-1}.$$

Combining this result and equation 12 we get that equation 11 can be rewritten as

$$\mathrm{d}\mathbf{X}_t = \mu(\mathbf{X}_t) + \sqrt{2}\mathrm{d}\mathbf{B}_t.$$

Note that (up to a factor 2) this is the same SDE as the one considered in Lee & Vempala (2017).

C Geodesic Brownian Motion

In this section, we provide some details on the geodesics Brownian motion introduced in Section 3.1. In the rest of the section, we make the following assumption.

***A*1.** $\overline{\mathcal{M}} \subset \mathbb{R}^d$ is compact and $\mathfrak{g}^{-1} : \mathcal{M} \to \mathcal{S}_d^{++}$ can be $\mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^{d\times d})$.

First, we start by showing that the process $(\mathbf{X}_t)_{t\geq 0}$ defined in equation 5 exists and that we have for any $t \geq 0$, $\mathbf{X}_t \in \mathcal{M}$. We recall that $\mathfrak{g} = \nabla^2\phi$ and $\lim_{x\to\partial\mathcal{M}} \Phi(x) = +\infty$.

Proposition C.1. *Assume* **A**1*. For any $x_0 \in \mathcal{M}$, there exists a unique strong solution to equation 5 denoted $(\mathbf{X}_t)_{t\geq 0}$. In addition, we have that for any $t \geq 0$, $\mathbf{X}_t \in \mathcal{M}$ almost surely. More precisely, we have $\mathbb{E}[\phi(\mathbf{X}_t)] \leq \phi(x_0) + t$.*

Proof. A unique strong solution $(\mathbf{X}_t)_{t\geq 0}$ of equation 5 with starting point $x_0 \in \mathcal{M}$ exists since the coefficients are smooth, see (Ikeda & Watanabe, 2014, Theorem 3.1, p.165). For any $A \geq 0$, we define $\tau_A = \inf\{t \geq 0 : \Phi(\mathbf{X}_t) \geq A\}$. Note that for any $t \in [0, \tau_A]$, $\Phi(\mathbf{X}_t) \in \mathcal{M}$. Using Itô formula, we have for any $t \geq 0$

$$\mathbb{E}[\Phi(\mathbf{X}_{t\wedge\tau_A})] = \Phi(x_0) + \mathbb{E}[\textstyle\int_0^{t\wedge\tau_A} \mathrm{Tr}(\mathfrak{g}^{-1}(\mathbf{X}_s)\nabla^2\Phi(\mathbf{X}_s))\mathrm{d}s] = \Phi(x_0) + \mathbb{E}[t \wedge \tau_A].$$

Using Fatou's lemma, and letting $A \to +\infty$, we conclude the proof. □

In the next result, we show that the uniform distribution is the *unique* invariant probability distribution for $(\mathbf{X}_t)_{t\geq 0}$ and that $(\mathbf{X}_t)_{t\geq 0}$ converges to this invariant distribution. We refer to (Meyn & Tweedie, 1993, Section 2, p.490) for a definition of irreducibility. We recall that the total variation of a finite (not necessarily non-negative) measure μ over $\mathbb{R}^d$ is given by $\|\mu\|_{\mathrm{TV}} = \sup\{\mu(\mathsf{A}) : \mathsf{A} \in \mathcal{B}(()\mathbb{R}^d)\}$.

Proposition C.2. *Assume* **A**1*. $(\mathbf{X}_t)_{t\geq 0}$ is π-irreducible, the uniform distribution over $\mathcal{M}$ is the only invariant probability distribution and $\lim_{t\to+\infty} \|\mathrm{P}_t - \pi\|_{\mathrm{TV}} = 0$, where P_t is the distribution of $\mathbf{X}_t$ for any $t \geq 0$ and π is the uniform distribution over $\mathcal{M}$.*

Proof. Since $x \mapsto \mathrm{div}(\mathfrak{g}^{-1})(x)$ and $x \mapsto \mathfrak{g}^{-1}$ are smooth and $\mathfrak{g}^{-1}(x)$ is positive definite for any $x \in \mathcal{M}$, we have that $(\mathbf{X}_t)_{t\geq 0}$ is π-irreducible, extending (Bhattacharya, 1978, Lemma 1.4) to $\mathcal{M}$ and using (Meyn & Tweedie, 1993, Proposition 2.1). In addition, $(\mathbf{X}_t)_{t\geq 0}$ is T-Feller using (Meyn & Tweedie, 1993, Proposition 3.3). Combining these results and the fact that $\mathcal{M}$ is bounde, we get that $(\mathbf{X}_t)_{t\geq 0}$ is positive Harris recurrent (Meyn & Tweedie, 1993, Theorem 3.2). The uniform distribution π is an invariant distribution for equation 5. Since $(\mathbf{X}_t)_{t\geq 0}$ is π-irreducible, we get that this invariant measure is unique. Hence, we conclude using (Meyn & Tweedie, 1993, Theorem 6.1). □

Note that the convergence result in total variation could be improved. In particular, quantitative geometric results could be derived. We finish this section, by applying results from the Malliavin calculus to show that for any $t > 0$, $\mathbf{X}_t$ admits a density w.r.t. the Lebesgue measure.

Proposition C.3. *Assume* **A**1*. Then, for any $t \geq 0$, $\mathbf{X}_t$ admits a smooth density p_t w.r.t. the Lebesgue measure.*

Proof. This is a direct consequence of (Nualart, 2006, Theorem 2.3.3). □

D Reflected Brownian Motion and Skorokhod problems

In this section, we provide the basic definitions and results to derive the time-reversal of the reflected Brownian motion in Appendix G. We follow closely the presentation of Lions & Sznitman (1984) and Burdzy et al. (2004). We first define the *Skorokhod problem* for deterministic problems. We consider $\mathcal{M}$ to be a smooth open bounded domain. We recall that the normal vector n is defined on $\partial\mathcal{M}$ and we set $n(x) = 0$ for any $x \notin \partial\mathcal{M}$.

Before giving the definition of the *Skorokhod problem*, we recall what is the space of functions of *bounded variations*.

Definition D.1. Let $a, b \in (-\infty, +\infty)$ and $f :\in \mathrm{C}([a,b], \mathbb{R})$. We define the *total variation* of f as

$$\mathrm{V}_{a,b}(f) = \sup\{\textstyle\sum_{i=0}^{n-1} \|f(x_{i+1}) - f(x_i)\| \ : \ (x_i)_{i=0}^{n-1}, \ a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b, \ n \in \mathbb{N}\}.$$

f has bounded variations over $[a,b]$ if $\mathrm{V}_{a,b}(f) < +\infty$. Let $f \in \mathrm{C}([0,+\infty), \mathbb{R})$. f has bounded variations over $[0,+\infty)$ if for any $b > 0$, f has bounded variations over $[0,b]$.

The notion of bounded variation is a relaxation of the differentiability requirement. In particular, if $f \in \mathrm{C}^1([a,b], \mathbb{R})$, we have $\mathrm{V}_{a,b}(f) = \int_a^b \|f'(t)\| \mathrm{d}t$. In the definition of the *Skorokhod problem*, we will see that this relaxation is necessary, even in the deterministic setting.

For any function of bounded variation $f \in \mathrm{C}([a,b], \mathbb{R})$ on $[a,b]$, we define $|f| : \ [a,b] \to [0,+\infty)$ given for any $t \in [a,b]$ by $|f|_t = \mathrm{V}_{a,t}(f)$. Note that $|f|$ is non-decreasing and right-continuous. Therefore, we can define the measure $\mu_{|f|}$ on $[a,b]$, given for any $s, t \in [a,b]$ with $t \geq s$ by $\mu_{|f|}([s,t]) = |f|(t) - |f|(s)$. In particular, for any $\varphi : [a,b] \to \mathbb{R}_+$, we define

$$\textstyle\int_a^b \varphi(t) \mathrm{d}|f|_t = \int_a^b \varphi(t) \mathrm{d}\mu_{|f|}(t).$$

In addition, f can be decomposed in a difference of two non-decreasing processes right continuous processes g_1, g_2, where for any $t \in [a,b]$, $f(t) = g_1(t) - g_2(t)$, $g_1(t) = |f|_t$ and $g_2(t) = |f|_t - f(t)$. Hence, for every φ bounded on $[a,b]$, we can define

$$\textstyle\int_a^b \varphi(t) \mathrm{d}f(t) = \int_a^b \varphi(t) \mathrm{d}g_1(t) - \int_a^b \varphi(t) \mathrm{d}g_2(t).$$

Note that these definitions can be extended to the setting where $f : \ [a,b] \to \mathbb{R}^d$.

We begin with the following result, see Lions & Sznitman (1984).

Theorem D.2. *Let $(x_t)_{t \geq 0} \in \mathrm{C}([0,+\infty), \mathbb{R})$. Then, there exists a unique couple $(\bar{x}_t, k_t)_{t \geq 0}$ such that*

(a) $(k_t)_{t \geq 0}$ has bounded variation over $[0,+\infty)$.

(b) $(\bar{x}_t)_{t \geq 0} \in \mathrm{C}([0,+\infty), \overline{\mathcal{M}})$.

(c) For any $t \geq 0$, $x_t + k_t = \bar{x}_t$.

(d) For any $t \geq 0$, $|k|_t = \int_0^t \mathbf{1}_{\bar{x}_s \in \partial\mathcal{M}}(\bar{x}_s) \mathrm{d}|k|_s$ and $k_t = \int_0^t n(\bar{x}_s) \mathrm{d}|k|_s$.

Let us discuss Theorem D.2. First, Theorem D.2-(c) states the original (unconstrained) process $(x_t)_{t \geq 0}$ can be decomposed into a constrained version $(\bar{x}_t)_{t \geq 0}$ and a bounded variation process $(k_t)_{t \geq 0}$. The process $(|k|_t)_{t \geq 0}$ counts the number of times the constrained process $(\bar{x}_t)_{t \geq 0}$ hits the boundary. More formally, we have $|k|_t = \int_0^t \mathbf{1}_{x \in \partial\mathcal{M}}(\bar{x}_s) \mathrm{d}|k|_s$. When, we hit the boundary, we reflect the process. This condition is expressed in $k_t = \int_0^t n(\bar{x}_s) \mathrm{d}|k|_s$.

We now consider the extension to stochastic processes. We are given $(\mathbf{X}_t)_{t \geq 0}$ such that

$$\mathrm{d}\mathbf{X}_t = b(\mathbf{X}_t)\mathrm{d}t + \sigma(t)\mathrm{d}\mathbf{B}_t,$$

where $(\mathbf{B}_t)_{t \geq 0}$ is a d-dimensional Brownian motion. We also assume that b and σ are Lipschitz which implies the existence and strong uniqueness of $(\mathbf{X}_t)_{t \geq 0}$. We have the following result Lions & Sznitman (1984).

Theorem D.3. *There exists a unique process $(\bar{\mathbf{X}}_t, \mathbf{k}_t)_{t \geq 0}$ such that*

(a) $(\mathbf{k}_t)_{t \geq 0}$ has bounded variation over $[0,+\infty)$ almost surely.

(b) $(\bar{\mathbf{X}}_t)_{t \geq 0} \in \mathrm{C}([0,+\infty), \overline{\mathcal{M}})$.

(c) For any $t \geq 0$, $\bar{\mathbf{X}}_t = \bar{\mathbf{X}}_0 + \int_0^t b(\bar{\mathbf{X}}_s)\mathrm{d}s + \int_0^t \sigma(\bar{\mathbf{X}}_s)\mathrm{d}\mathbf{B}_s - \mathbf{k}_t$.

(d) For any $t \geq 0$, $|\mathbf{k}|_t = \int_0^t \mathbf{1}_{\bar{x}_s \in \partial\mathcal{M}}(\bar{x}_s) \mathrm{d}|\mathbf{k}|_s$ and $\mathbf{k}_t = \int_0^t n(\bar{x}_s) \mathrm{d}|\mathbf{k}|_s$.

The process $(\mathbf{X}_t)_{t\geq 0}$ is almost surely continuous, so we could apply the previous theorem almost surely for all the realizations of the process,. However, this does not tell us if the obtained solutions $(\bar{\mathbf{X}}_t, \mathbf{k}_t)_{t\geq 0}$ form themselves a process. The main difference with Theorem D.2 is in Theorem D.3-(c) which differs from Theorem D.3-(c). Note that in the case where $b = 0$ and $\sigma = \mathrm{Id}$ we recover Theorem D.3-(c). This is not true in the general case. However, it can be seen that for any realization of the process $(\bar{\mathbf{X}}_t)_{t\geq 0}$, we have that $(\bar{\mathbf{X}}_t, \mathbf{k}_t)_{t\geq 0}$ is solution of the *deterministic* Skorokhod problem by letting $x_t = \bar{\mathbf{X}}_0 + \int_0^t b(\bar{\mathbf{X}}_s)\mathrm{d}s + \int_0^t \sigma(\bar{\mathbf{X}}_s)\mathrm{d}\mathbf{B}_s$. The backward and forward Kolmogorov equations can be found in Burdzy et al. (2004). Note that the presence of the process $(\mathbf{k}_t)_{t\geq 0}$ incurs notable complications compared to unconstrained processes. In particular, there is no martingale problem associated with weak solutions of reflected SDEs but only sub-martingale problems, see Kang & Ramanan (2017) for instance.

E Implicit Score Matching Loss

E.1 Proof of ISM

Using the divergence theorem, we have

$$
\begin{aligned}
&(1/2)\int_{\mathcal{M}} \|\mathbf{s}_\theta(x) - \nabla \log p_t(x)\|^2 p_t(x)\mathrm{d}\mu(x) \\
&= (1/2)\int_{\mathcal{M}} \|\mathbf{s}_\theta(x)\|^2 p_t(x)\mathrm{d}\mu(x) - \int_{\mathcal{M}} \langle \mathbf{s}_\theta(x), \nabla \log p_t(x)\rangle p_t(x)\mathrm{d}\mu(x) + (1/2)\int_{\mathcal{M}} \|\nabla \log p_t(x)\|^2 p_t(x)\mathrm{d}\mu(x) \\
&= (1/2)\int_{\mathcal{M}} \|\mathbf{s}_\theta(x)\|^2 p_t(x)\mathrm{d}\mu(x) - \int_{\partial\mathcal{M}} \langle \mathbf{s}_\theta(x), \mathbf{n}\rangle p_t(x)\mathrm{d}\nu(x) \\
&\quad + \int_{\mathcal{M}} \mathrm{div}(\mathbf{s}_\theta)(x) p_t(x)\mathrm{d}\mu(x) + (1/2)\int_{\mathcal{M}} \|\nabla \log p_t(x)\|^2 p_t(x)\mathrm{d}\mu(x).
\end{aligned}
$$

Using that $\mathbf{s}_\theta(x) = 0$ for all $x \in \partial\mathcal{M}$, we get that

$$
\begin{aligned}
&(1/2)\int_{\mathcal{M}} \|\mathbf{s}_\theta(x) - \nabla \log p_t(x)\|^2 p_t(x)\mathrm{d}\mu(x) \\
&\qquad = (1/2)\int_{\mathcal{M}} \|\mathbf{s}_\theta(x)\|^2 p_t(x)\mathrm{d}\mu(x) + \int_{\mathcal{M}} \mathrm{div}(\mathbf{s}_\theta)(x) p_t(x)\mathrm{d}\mu(x) + (1/2)\int_{\mathcal{M}} \|\nabla \log p_t(x)\|^2 p_t(x)\mathrm{d}\mu(x),
\end{aligned}
$$

which concludes the proof.

E.2 Importance of scaling function

As discussed in Section 3.3, we include a monotone scaling function h which is zero close to the boundary to ensure the relevant conditions are met for the score matching loss and the boundary conditions. This may seem like a technical detail, but is of significant practical importance. Upon removal of the scaling function, we observe that the learned score functions behave strangely around the boundary in the reverse process, leading to samples that do not match the forward process. The problems are apparent when comparing the top three plots of Fig. 15 and Fig. 16. Interestingly, we found that these issues early on in the sampling are smoothed out by the end of the reverse process, but still lead to a failure to recover the target density.

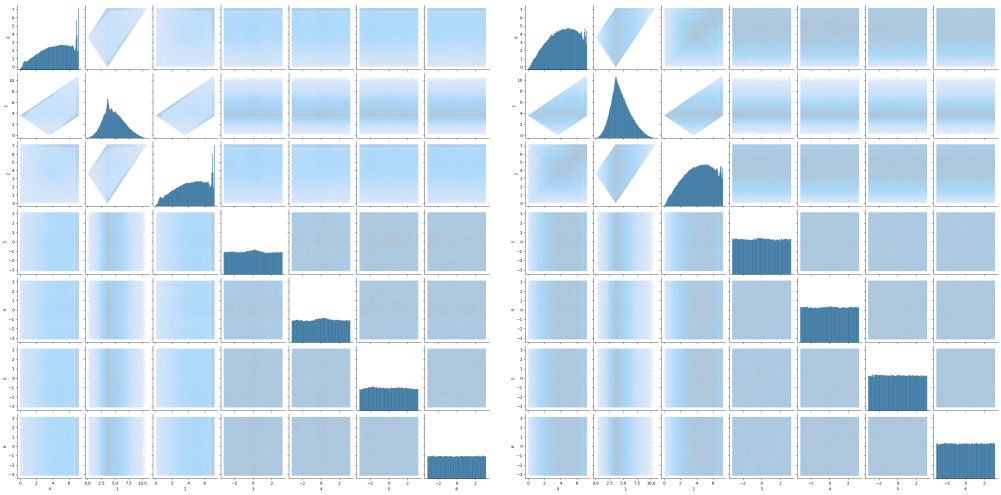


Figure 15: Reverse process samples for the cyclic peptide dataset from Section 5.3 at $t = 1.0, 0.9$ (left and right respectively) trained without the scaling function.

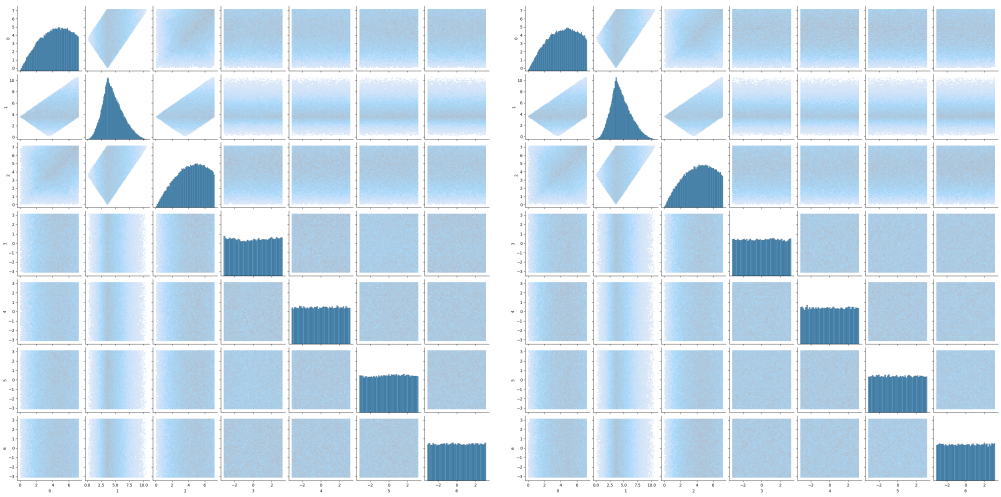


Figure 16: Reverse process samples for the cyclic peptide dataset from Section 5.3 at $t = 1.0, 0.9$ (left and right respectively) trained with the scaling function.

F Likelihood evaluation

One key advantage of constructing a continuous noising process is that, similarly to Song et al. (2021), we can evaluate the model's likelihood via the following *probability flow* Ordinary Differential Equation (ODE). In particular, for the Langevin dynamics (5) which we recall

$$\mathrm{d}\mathbf{X}_t = \tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1})(\mathbf{X}_t)\mathrm{d}t + \mathfrak{g}(\mathbf{X}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t,$$

the following ODE has the same marginal density

$$\mathrm{d}\mathbf{Y}_t = \left[\tfrac{1}{2}\nabla \cdot \mathfrak{g}^{-1}(\mathbf{Y}_t) - \tfrac{1}{2}\mathfrak{g}^{-1}(\mathbf{Y}_t)\nabla \log p_t(\mathbf{Y}_t)\right] \mathrm{d}t.$$

We conclude this section with a derivation of the equivalent ODE. We highlight that the ODE representation for reflected diffusion models was first derived in Lou & Ermon (2023). We recall that if $\mathcal{M} \subset \mathbb{R}^d$ is a bounded

open set with smooth boundary then (Burdzy et al., 2004, Theorem 2.2) ensures that the reflected Brownian motion admits a density w.r.t. the Lebesgue measure. We denote p_t this smooth density.

Proposition F.1. *Assume that $\mathcal{M} \subset \mathbb{R}^d$ is a bounded open set with smooth boundary. Assume that $(t,x) \mapsto \nabla \log p_t(x)$ is smooth on $[0,+\infty) \times \partial\mathcal{M}$. Let $(\bar{\mathbf{B}}_t)_{t\geq 0}$ be the reflected Brownian motion with $\bar{\mathbf{B}}_0 \sim p_0$ smooth and supported in $\mathcal{M}$. Let $(\mathbf{X}_t)_{t\geq 0}$ be given for any $t \geq 0$ by $\mathrm{d}\mathbf{X}_t = \frac{1}{2}\nabla \log p_t(\mathbf{X}_t)\mathrm{d}t$ and $\mathbf{X}_0 \sim p_0$, where p_t denotes the density of $\bar{\mathbf{B}}_t$ w.r.t. the Lebesgue measure for any $t > 0$. Then for any $t \in [0,T]$, $\bar{\mathbf{B}}_t$ and $\mathbf{Y}_t$ have the same distribution.*

Proof. Since the distributions of $(\bar{\mathbf{B}}_t)_{t\in[0,T]}$ and $(\mathbf{Y}_t)_{t\in[0,T]}$ satisfy the same Fokker-Planck equation whenever these processes are well-defined. Therefore, we first show that the process $(\mathbf{Y}_t)_{t\in[0,T]}$ is well-defined and stay in $\mathcal{M}$ at all times. Using (Burdzy et al., 2004, Theorem 2.2), we have that $\partial p_t(x) = \frac{1}{2}\mathrm{div}(\nabla \log p_t)(x)$, for any $t > 0$ and $x \in \mathcal{M}$. Next, we define $\mathrm{d}\mathbf{X}_t = \frac{1}{2}\nabla \log p_t(\mathbf{X}_t)\mathrm{d}t$. Note that $(\mathbf{X}_t)_{t\geq 0}$ is defined up to an explosion time T_∞ after which we fix $\mathbf{X}_t = \infty$. Denote T_0 the first time such that $\mathbf{X}_t \in \partial\mathcal{M}$. Note that since p_0 is supported on $\mathcal{M}$ we have $T_0 > 0$. We denote $(\mathbf{Y}_t)_{t\in[0,T_0]} = (\mathbf{X}_{T_0-t})_{t\in[0,T_0]}$. We have that for any $t \in [0,T_0]$, $\mathrm{d}\mathbf{Y}_t = -\frac{1}{2}\nabla \log p_{T_0-t}(\mathbf{Y}_t)\mathrm{d}t$. Since $(t,x) \mapsto \nabla \log p_t(x)$ is smooth on $[0,+\infty) \times \partial\mathcal{M}$, we get that for any $t \in [0,T_0]$, $\mathbf{Y}_t \in \partial\mathcal{M}$. In particular, we have that $\mathbf{Y}_{T_0/2} = \mathbf{X}_{T_0/2} \in \partial\mathcal{M}$ which is absurd. Therefore $T_0 = +\infty$ (which also implies that $T_\infty = +\infty$). Hence, $(\mathbf{X}_t)_{t\geq 0}$ is a flow on $\mathcal{M}$ and therefore for any $t \geq 0$, the density q_t of $\mathbf{X}_t$ is smooth and satisfies $\partial_t q_t(x) = -\frac{1}{2}\mathrm{div}(q_t \nabla \log p_t)(x)$. We conclude using the uniqueness of the solutions to the transport equation for smooth initialisation and coefficients on $\mathbb{R}^d$. □

G Time-reversal for reflected Brownian motion

We start with the following definition.

Definition G.1. Let $\mathcal{M} \subset \mathbb{R}^d$ be an open set. $\mathcal{M}$ has a smooth boundary if for any $x \in \partial\mathcal{M}$, there exists $\mathsf{U} \subset \mathbb{R}^d$ open and $f \in \mathrm{C}^\infty(\mathsf{U},\mathbb{R})$ such that $x \in \mathsf{U}$ and (a) $\mathrm{cl}(\mathcal{M}) \cap \mathsf{U} = \{x \in \mathsf{U} : f(x) \leq 0\}$, (b) $\nabla f(x) \neq 0$ for any $x \in \mathsf{U}$ where $\mathrm{cl}(\mathcal{M})$ is the closure of $\mathcal{M}$.

We will make the use of the following lemma which is a straightforward extension of Burdzy et al. (2004, Theorem 2.6).

Lemma G.2. *Let u such that $s \mapsto u(s,x) \in \mathrm{C}^1((0,T),\mathbb{R})$, for any $s \in (0,T)$, $x \mapsto u(s,x) \in \mathrm{C}^2(\mathcal{M},\mathbb{R})$ and $u \in \mathrm{C}^1(\mathrm{cl}(\mathcal{M}),\mathbb{R})$. Then for any $T \geq 0$, $s,t \in [0,T]$, we have*

$$\mathbb{E}[\textstyle\int_s^t u(w,\bar{\mathbf{B}}_w)\mathrm{d}\mathbf{k}_w] = \frac{1}{2}\int_s^t \int_{\partial\mathcal{M}} u(x)p_w(x)\mathrm{d}\sigma(x)\mathrm{d}w.$$

Note that we recover Burdzy et al. (2004, Theorem 2.6) if we set $u = 1$. We also emphasize that the result of Burdzy et al. (2004, Theorem 2.6) is stronger than Theorem G.2 as it holds not only in expectation but also in L^2 and almost surely.

We are now ready to prove Theorem 3.2. We follow the approach of Petit (1997) which itself is based on an extension of Haussmann & Pardoux (1986). We refer to Cattiaux et al. (2021) for recent entropic approaches of time-reversal. Recall that $(\bar{\mathbf{B}}_t, \mathbf{k}_t)_{t\geq 0}$ is a solution to the *Skorokhod problem* (Skorokhod, 1961) if $(\mathbf{k}_t)_{t\geq 0}$ a bounded variation process and $(\bar{\mathbf{B}}_t)_{t\geq 0}$ a continuous adapted process such that for any $t \geq 0$, $\mathbf{B}_t = \bar{\mathbf{B}}_t + \mathbf{k}_t \in \mathcal{M}$, $(\bar{\mathbf{B}}_t)_{t\geq 0}$ and

$$|\mathbf{k}|_t = \textstyle\int_0^t \mathbf{1}_{\bar{\mathbf{B}}_s \in \partial\mathcal{M}}\mathrm{d}|\mathbf{k}|_s, \quad \mathbf{k}_t = \int_0^t \mathbf{n}(\bar{\mathbf{B}}_s)\mathrm{d}|\mathbf{k}|_s, \tag{13}$$

In what follows, we define $(\mathbf{Y}_t)_{t\in[0,T]}$ such that for any $t \in [0,T]$, $\mathbf{Y}_t = \bar{\mathbf{B}}_{T-t}$. Let us consider the process $(\tilde{\mathbf{B}}_t)_{t\in[0,T]}$ defined for any $t \in [0,T]$ by

$$\tilde{\mathbf{B}}_t = -\bar{\mathbf{B}}_T + \bar{\mathbf{B}}_{T-t} + \mathbf{k}_T - \mathbf{k}_{T-t} - \textstyle\int_{T-t}^T \nabla \log p_s(\bar{\mathbf{B}}_s)\mathrm{d}s.$$

First, note that $t \mapsto \tilde{\mathbf{B}}_t$ is continuous. Denote by $\mathcal{F}$, the filtration associated with $(\bar{\mathbf{B}}_{T-t})_{t\in[0,T]}$. We have that $(\tilde{\mathbf{B}}_t)_{t\in[0,T]}$ is adapted to $(\bar{\mathbf{B}}_{T-t})_{t\in[0,T]}$. Even more so, we have that $(\tilde{\mathbf{B}}_t)_{t\in[0,T]}$ satisfies the strong Markov

property since $(\bar{\mathbf{B}}_t)_{t\in[0,T]}$ also satisfies the strong Markov property. Let $g \in \mathrm{C}_c^\infty(\mathrm{cl}(\mathcal{M}))$ and consider for any $0 \leq s \leq t \leq T$, $\mathbb{E}[(\tilde{\mathbf{B}}_t - \tilde{\mathbf{B}}_s)g(\bar{\mathbf{B}}_{T-t})]$. For any $0 \leq s \leq t \leq T$ we have

$$\mathbb{E}[(\tilde{\mathbf{B}}_t - \tilde{\mathbf{B}}_s)g(\bar{\mathbf{B}}_{T-t})] = \mathbb{E}[(-\bar{\mathbf{B}}_{T-s} + \bar{\mathbf{B}}_{T-t} + \mathbf{k}_{T-s} - \mathbf{k}_{T-t} - \int_{T-t}^{T-s} \nabla \log p_u(\bar{\mathbf{B}}_u)\mathrm{d}u)g(\bar{\mathbf{B}}_{T-t})]. \tag{14}$$

In what follows, we prove that for any $0 \leq s \leq t \leq T$ we have $\mathbb{E}[(\tilde{\mathbf{B}}_t - \tilde{\mathbf{B}}_s)g(\bar{\mathbf{B}}_{T-t})] = 0$. Therefore, we only need to prove that for any $0 \leq s \leq t \leq T$ we have

$$\mathbb{E}[(-\bar{\mathbf{B}}_t + \bar{\mathbf{B}}_s + \mathbf{k}_t - \mathbf{k}_s - \int_s^t \nabla \log p_u(\bar{\mathbf{B}}_u)\mathrm{d}u)g(\bar{\mathbf{B}}_t)] = 0. \tag{15}$$

Let $t \in (0,T]$. We introduce $u: [0,t] \times \mathcal{M}$ such that for any $s \in [0,t]$ and $x \in \mathcal{M}$, $u(s,x) = \mathbb{E}[g(\bar{\mathbf{B}}_t)|\bar{\mathbf{B}}_s = x]$. Using Burdzy et al. (2004, Theorem 2.8) we get that for any $x \in \mathcal{M}$, $s \mapsto u(s,x) \in \mathrm{C}^1((0,t),\mathbb{R})$ and for any $s \in (0,t)$, $x \mapsto u(s,x) \in \mathrm{C}^2(\mathcal{M},\mathbb{R})$ and $x \mapsto u(s,x) \in \mathrm{C}^1(\mathrm{cl}(\mathcal{M}),\mathbb{R})$. In addition, we have that for any $s \in (0,t)$ and for any $x \in \mathcal{M}$ and $x_0 \in \partial\mathcal{M}$

$$\partial_s u(s,x) + \tfrac{1}{2}\Delta u(s,x) = 0, \qquad \langle \nabla u(s,x_0), \mathbf{n}(x_0)\rangle = 0. \tag{16}$$

This equation is called the backward Kolmogorov equation. Using equation 16, $\bar{\mathbf{B}}_t = \mathbf{B}_t - \mathbf{k}_t$ for any $t \geq 0$ and the Itô formula for semimartingale (Revuz & Yor, 2013, Chapter IV, Theorem 3.3) we have that for any $s \in (0,t)$

$$\begin{aligned}
\mathbb{E}[u(t,\bar{\mathbf{B}}_t)\bar{\mathbf{B}}_t] &= \mathbb{E}[u(s,\bar{\mathbf{B}}_s)\bar{\mathbf{B}}_s] + \mathbb{E}[\tfrac{1}{2}\int_s^t \bar{\mathbf{B}}_w \Delta u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] + \mathbb{E}[\int_s^t \nabla u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] \\
&\quad - \mathbb{E}[\int_s^t \bar{\mathbf{B}}_w \langle \nabla u(w,\bar{\mathbf{B}}_w), \mathbf{n}(\bar{\mathbf{B}}_w)\rangle \mathrm{d}|\mathbf{k}|_w] \\
&\quad - \mathbb{E}[\int_s^t u(w,\bar{\mathbf{B}}_w)\mathbf{n}(\bar{\mathbf{B}}_w)\mathrm{d}|\mathbf{k}|_w] \\
&\quad + \mathbb{E}[\int_s^t \bar{\mathbf{B}}_w \partial_w u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] \\
&= \mathbb{E}[u(s,\bar{\mathbf{B}}_s)\bar{\mathbf{B}}_s] + \mathbb{E}[\int_s^t \nabla u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] - \mathbb{E}[\int_s^t u(w,\bar{\mathbf{B}}_w)\mathbf{n}(\bar{\mathbf{B}}_w)\mathrm{d}|\mathbf{k}|_w]
\end{aligned} \tag{17}$$

In addition, using the Fubini theorem and the definition of $\mathbf{k}_t$ we have that for any $s \in (0,t)$

$$\mathbb{E}[\int_s^t u(w,\bar{\mathbf{B}}_w)\mathbf{n}(\bar{\mathbf{B}}_w)\mathrm{d}|\mathbf{k}|_w] = \mathbb{E}[\int_s^t \mathbb{E}[g(\bar{\mathbf{B}}_t)|\bar{\mathbf{B}}_w]\mathbf{n}(\bar{\mathbf{B}}_w)\mathrm{d}|\mathbf{k}|_w] = \mathbb{E}[g(\bar{\mathbf{B}}_t)(\mathbf{k}_t - \mathbf{k}_s)]. \tag{18}$$

Finally, using the divergence theorem and Burdzy et al. (2004, Theorem 2.6) we have that for any $s \in (0,t)$

$$\begin{aligned}
\mathbb{E}[\int_s^t \nabla u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] &= \int_s^t \int_{\mathcal{M}} \nabla u(w,x) p_w(x)\mathrm{d}x\mathrm{d}w \\
&= -\int_s^t \int_{\mathcal{M}} u(w,x)\nabla \log(p_w(x)) p_w(x)\mathrm{d}x\mathrm{d}w + \int_s^t \int_{\partial\mathcal{M}} u(w,x)p_w(x)\mathrm{d}x\mathrm{d}\sigma(w),
\end{aligned}$$

where σ is the surface area measure on $\partial\mathcal{M}$, see Burdzy et al. (2004). Using Theorem G.2 and the Fubini theorem we get that

$$\begin{aligned}
\mathbb{E}[\int_s^t \nabla u(w,\bar{\mathbf{B}}_w)\mathrm{d}w] &= -\int_s^t \int_{\mathcal{M}} u(w,x)\nabla \log(p_w(x))p_w(x)\mathrm{d}x\mathrm{d}w + \mathbb{E}[\int_s^t u(w,\bar{\mathbf{B}}_w)\mathrm{d}\mathbf{k}_w] \\
&= -\mathbb{E}[\int_s^t g(\bar{\mathbf{B}}_t)\nabla \log(p_w(\bar{\mathbf{B}}_w))\mathrm{d}w] + 2\mathbb{E}[g(\bar{\mathbf{B}}_t)(\mathbf{k}_t - \mathbf{k}_s)]
\end{aligned} \tag{19}$$

Combining equation 17, equation 18 and equation 19 we get that

$$\mathbb{E}[u(t,\bar{\mathbf{B}}_t)] = \mathbb{E}[u(s,\bar{\mathbf{B}}_s)] - \mathbb{E}[g(\bar{\mathbf{B}}_t)\int_s^t \nabla \log p_w(\bar{\mathbf{B}}_w)\mathrm{d}w] + \mathbb{E}[g(\bar{\mathbf{B}}_t)(\mathbf{k}_t - \mathbf{k}_s)].$$

Therefore, we get equation 15 and equation 14. Hence, $(\tilde{\mathbf{B}}_t)_{t\in[0,T]}$ is a continuous martingale. In addition, we have that for any $t \in [0,T]$, $\mathbb{E}[\tilde{\mathbf{B}}_t\tilde{\mathbf{B}}_t^\top] = t\,\mathrm{Id}$ and therefore, $(\tilde{\mathbf{B}}_t)_{t\in[0,T]}$ is a Brownian motion using the Lévy characterisation of Brownian motion (Revuz & Yor, 2013, Chapter IV, Theorem 3.6). Denote $(\mathbf{j}_t)_{t\in[0,T]} = (\mathbf{k}_T - \mathbf{k}_{T-t})_{t\in[0,T]}$. Using equation 14, we have that for any $t \in [0,T]$

$$\bar{\mathbf{B}}_{T-t} = \bar{\mathbf{Y}}_0 + \tilde{\mathbf{B}}_t + \int_0^t \nabla \log p_{T-s}(\mathbf{Y}_s)\mathrm{d}s - \mathbf{j}_t.$$

Using equation 13, we have for any $t \in [0,T]$

$$|\mathbf{j}|_t = \int_0^t \mathbf{1}_{\mathbf{Y}_s \in \partial\mathcal{M}}\mathrm{d}|\mathbf{j}|_s, \quad \mathbf{j}_t = \int_0^t \mathbf{n}(\bar{\mathbf{Y}}_s)\mathrm{d}|\mathbf{j}|_s,$$

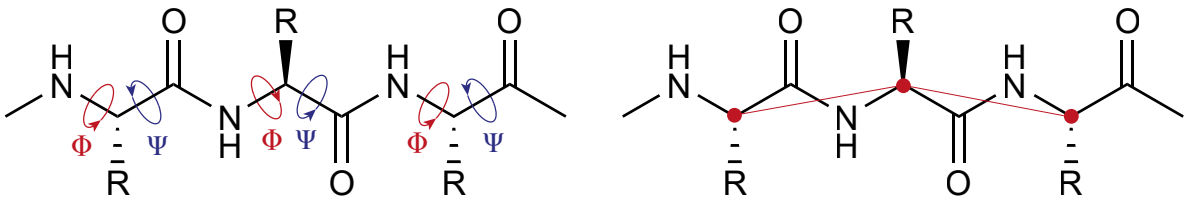
which concludes the proof.

H Conformational modelling of polypeptide backbones under end point constraints

Polypeptides and proteins constitute an important class of biogenic macromolecules that underpin most aspects of organic life. Accurately modelling their conformational ensembles, i.e. the set of three-dimensional structures they assume under physiological conditions, is essential to both understanding the biological function of existing and designing the enzymatic or therapeutic properties of novel proteins (Lane, 2023). Motivated by the success of diffusion models in computer vision and natural language processing, there has been considerable interest in applying them to learn and sample from distributions over the conformational space of protein structures (Watson et al., 2022; Trippe et al., 2022; Wu et al., 2022).

H.1 Problem parameterisation

Proteins are biopolymers in which a sequence of N amino acids is joined together through $N-1$ peptide bonds, resulting in a so-called polypeptide backbone with protruding amino acid residues. As the deviation of chemical bond lengths and angles from their theoretical optimum is generally negligible, the problem of modelling the three-dimensional structure of this polypeptide chain is often reframed in the space of the internal torsion angles Φ and Ψ (see Figure 17a for an illustration), which can be modelled on a $(2N-2)$-dimensional torus $\mathbb{T}^{2N-2}$.



(a) A commonly-used approximate parameterisation of backbone geometry only considers the C_α torsion angles Φ and Ψ.

(b) As peptide bond orientations can be inferred relatively reliably, researchers often only model the C_α traces.

Figure 17: Standard approaches to modelling the conformations of polypeptide backbones.

In many data-scarce practical settings such as antibody or enzyme design, it is often unnecessary or even undesirable to model the structure of an entire protein, as researchers are primarily interested in specific functional sites with distinct biochemical properties. However, generating conformational ensembles for such substructural elements necessitates positional constraints on their endpoints to ensure that they can be accommodated by the remaining scaffold. While it is conceivable that a diffusion model could derive such constraints from limited experimental data, we argue that it is much more efficient and precise to encode them explicitly.

For this purpose, we adopt the distance constraint formulation from Han & Rudolph (2006) and interpret the backbone as a spatial chain with N spherical joints and fixed-length links (see Figure 18a for an illustration). After selecting a suitable anchor point, the geometry of the polypeptide chain is fully specified by (a) the set of link lengths $\ell = \{\ell_j\}_{j=1}^{N}$, (b) the set of vectors $\mathbf{r} = \{r(1,j)\}_{j=2}^{N}$ between the anchor point and each atom in the chain, and (c) the set of dihedral angles

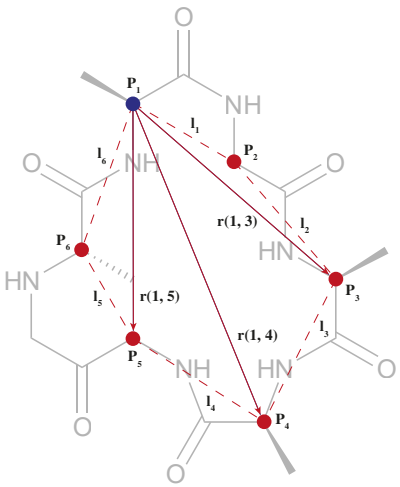
$$\mathbf{T} = \left\{ \arccos\left(\frac{|\langle r(1,j) \times r(1,j+1), r(1,j+1) \times r(1,j+2) \rangle|}{|r(1,j) \times r(1,j+1)||r(1,j+1) r(1,j+2)|} \right) \right\}_{j=2}^{N-2} \in \mathbb{T}^{N-3}$$

between any three consecutive vectors. After specifying the fixed bond lengths ℓ, including an arbitrary anchor point distance $d_{\text{anchor}} = \ell_N = r(1,N)$, the set of valid vectors $\mathbf{r}$ is given by the convex polytope $\mathbb{P} \subseteq \mathbb{R}^3$ defined by the following linear constraints (see Figure 18b for an illustration):

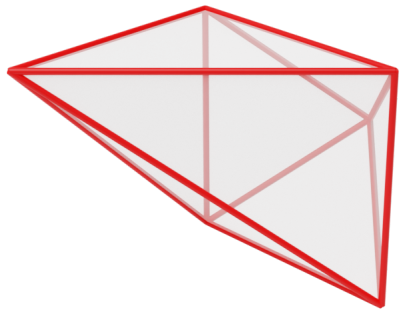
$$\begin{aligned} r(1,3) &\leq \ell_1 + \ell_2, \\ -r(1,3) &\leq -|\ell_1 - \ell_2|, \end{aligned}$$

$$\left.\begin{array}{rcl} r(1,j) - r(1,j+1) & \leq & \ell_j, \\ -r(1,j) + r(1,j+1) & \leq & \ell_j, \\ -r(j) - r(j+1) & \leq & -\ell_j, \end{array}\right\} 3 \leq j \leq N-2,$$

$$\begin{array}{rcl} r(1,N-1) & \leq & \ell_{N-1} + d_{\text{anchor}}, \\ -r(1,N-1) & \leq & -\left|\ell_{N-1} - d_{\text{anchor}}\right|. \end{array}$$



(a) An illustrative diagram of the proposed parameterisation for modelling the C_α trace geometry of the cyclic peptide C-AAGAGG.



(b) The convex polytope constraining the diagonals of the triangles for the given bond lengths in the illustrated molecule. The total design space is the product of this polytope with the 4D flat torus.

Figure 18: Parameterising the conformational space of polypeptide backbones under anchor point distance constraints.

This means that the set of all valid polypeptide backbone conformations is defined by the product manifold $\mathbb{P} \times \mathbb{T}^{N-3}$, enabling us to train diffusion models that exclusively generate conformations with a fixed anchor point distance d_{anchor}.

H.2 Data generation and model training

As a proof-of-concept for the practicality of our methods, we chose to model the conformational distribution of the cyclic peptide C-AAGAGG. Cyclic peptides are an increasingly important drug modality with therapeutic uses ranging from antimicrobials to oncology, exhibiting circular polypeptide backbones (i.e. $d_{\text{anchor}} = 0$) that confer a range of desirable pharmacodynamic and pharmacokinetic properties (Dougherty et al., 2019). To reduce the dimensionality of the problem, we only consider the C_α traces (with fixed C_α-C_α link distances of 3.6 Å) instead of the full polypeptide backbone (see Figure 17b), although we note that our framework is fully applicable to both settings.

To derive a suitable dataset, the product manifold $\mathbb{P} \times \mathbb{T}^3$ describing the conformations of cyclic C_α traces of length $N = 6$ was constructed (see Figures 18a and 18b for an illustration) and used to generate 10^7 uniform samples satisfying the anchor point distance constraint $d_{\text{anchor}} = 0$. Subsequently, an estimate of the free energy E_i of each sample i was obtained by (1) reconstructing the full-atom peptide from each C_α trace using the PULCHRA algorithm (Rotkiewicz & Skolnick, 2008), (2) relaxing all non-C_α backbone and side-chain atoms (keeping the C_α positions fixed), and (3) quantifying the potential energy of each of the resulting conformations using the OpenMM suite of molecular dynamics tools (Eastman et al., 2017), and the AMBER force field (Hornak et al., 2006). These free energy estimates were then used to approximate the Boltzmann distribution over conformational states

$$p_B(i) \propto \exp\left(-\frac{E_i}{k_B T}\right),$$

where temperature was set to $T = 273.15\,\mathrm{K}$ and $k_B = 1.380\,649 \times 10^{-23}\,\mathrm{J\,K^{-1}}$ is the Boltzmann constant. We then apply a very minor amount of smoothing to the resulting distribution by running forward Brownian motion on both the polytope and the torus for 10 steps, using a small step size of 5×10^{-3} and the respective metrics. Finally, a subsample of 10^6 C_α traces was drawn from this distribution and used for training and evaluating our models.

I Configurational modelling of robotic arms under manipulability constraints

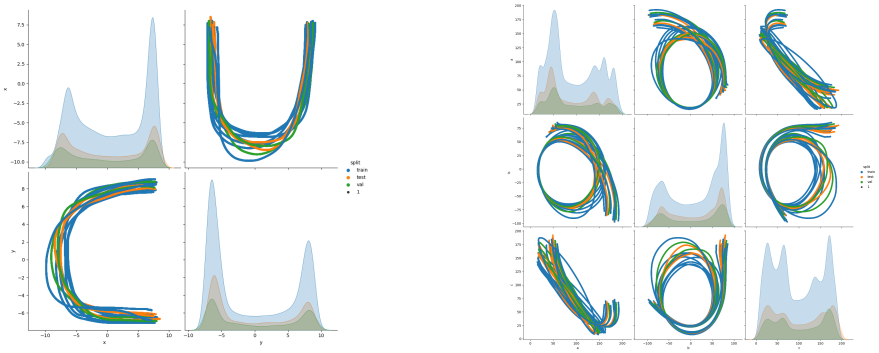
Accurately determining and specifying the movement of a robotic arm and the forces it exerts is a fundamental problem in many real-world robotics applications. A widely-used set of descriptors for modelling the flexibility of a given joint configuration are so-called manipulability ellipsoids (Yoshikawa, 1985), which are kinetostatic performance measures that quantify the ability to move or exert forces along different directions. Jaquier et al. (2021) present a geometric framework to learn trajectories of manipulability ellipsoids by making use of the fact any ellipsoid $M \in \mathbb{R}^N$ is defined by the set of points $\{\mathbf{x} | \mathbf{x}^T \mathbf{A} \mathbf{x} = 1\}$ where $\mathbf{A}$ lies on the manifold of $N \times N$ symmetric positive definite matrices S^N_{++}.

In many practical settings, it is desirable to constrain the minimal or maximal volume of a manipulability ellipsoid to retain motional flexibility or limit the magnitude of the exerted force. This necessitates lower or upper limits on the determinant of $\mathbf{A}$, translating into constraints on S^N_{++}. To model this, we make use of one of the datasets introduced by Jaquier et al. (2021), containing demonstrations of a robotic arm drawing different letters in the plane, providing the respective positional trajectories ($\mathbb{R}^2$) and velocity manipulability ellipsoids (S^2_{++}).

We use the processing routines provided by Jaquier et al. (2021) to interpolate the trajectories into 10^4 distinct points, for each of which we derive the position $\mathbf{x} \in \mathbb{R}^2$ and the PSD matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in S^2_{++}.$$

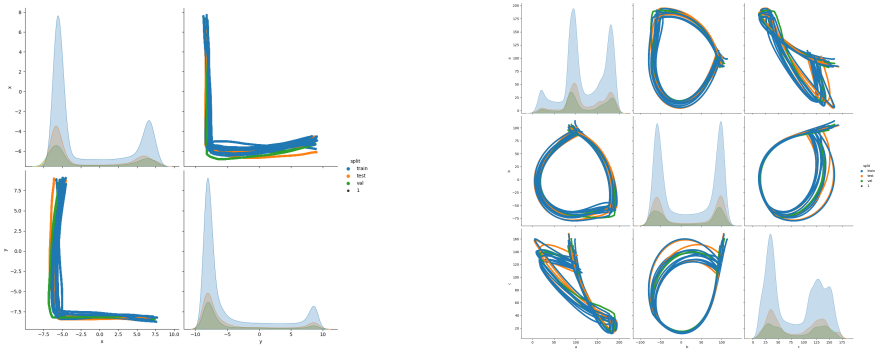
parametrizing the velocity manipulability ellipsoid $M \in \mathbb{R}^2$. The resulting data is split into training, validation, and test sets by trajectory and visualized in Figures 19 and 20.



(a) (b)

Figure 19: Positional trajectories $\mathbf{x} \in \mathbb{R}^2$ (a) and the parameters a, b, c of the the SPD matrix $\mathbf{A} \in S^2_{++}$ (b) for the letter C.

We add a small amount of Gaussian noise to these trajectories, which is shown as the target distribution in J.2.



(a) (b)

Figure 20: Positional trajectories $\mathbf{x} \in \mathbb{R}^2$ (a) and the parameters a, b, c of the the SPD matrix $\mathbf{A} \in S^2_{++}$ (b) for the letter L.

J Experimental details

In what follows we describe the experimental settings used to generate results introduced in Section 5. The models and experiments have been implemented in Jax (Bradbury et al., 2018), using a modified version of the Riemannian geometry library Geomstats (Miolane et al., 2020).

Architecture. The architecture of the score network $\boldsymbol{s}_\theta$ is given by a multilayer perceptron with 6 hidden layers with 512 units each. We use sinusoidal activation functions.

Training. All models are trained by the stochastic optimizer Adam (Kingma & Ba, 2014) with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch-size of 256 data-points. The learning rate is annealed with a linear ramp from 0 to 1000 steps, reaching the maximum value of $2e-4$, and from then with a cosine schedule down to 0 after $100k$ iterations in total.

Diffusion. Following Song et al. (2021), the diffusion models diffusion coefficient is parametrized as $g(t) = \sqrt{\beta(t)}$ with $\beta: \ t \mapsto \beta_{\text{min}} + (\beta_{\text{max}} - \beta_{\text{min}}) \cdot t$, where we found $\beta_{\text{min}} = 0.001$ and $\beta_{\text{max}} = 6$ to work best.

Metrics. We measure the performance of trained models via the Maximum Mean Discrepancy (MMD) (Gretton et al., 2012), which is a kernel based metric between two distributions P and Q. The MMD can be empirically approximated with the following U-statistics $\text{MMD}^2(P,Q) = \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(y_i, y_j) - 2\frac{1}{m^2} \sum_i \sum_j k(x_i, y_j)$ with $x_i \sim P$ and $y_i \sim Q$, where k is a kernel.

Low temperature sampling. As in Ingraham et al. (2022), we use low temperature sampling in order to trade-off fidelity for diversity. It combines to ingredients: 1. an approximation of the corrector sampling Song et al. (2021), 2. a low temperature approximation of the target p. Namely, instead of targeting p_0 we target $p_0^{\lambda_0}$ (unnormalised density). We refer to Ingraham et al. (2022, Appendix B) for details about the derivation. Recall that we consider the following SDE in the backward process

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \beta_{T-t}\{\tfrac{1}{2}\overleftarrow{\mathbf{X}}_t + s_\theta(T-t, \overleftarrow{\mathbf{X}}_t)\}\mathrm{d}t + \beta_t^{1/2}\mathrm{d}\mathbf{B}_t.$$

The updated scheme is given by

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \beta_{T-t}\{\tfrac{1}{2}\overleftarrow{\mathbf{X}}_t + (\lambda_t + \tfrac{\lambda_0 \Psi}{2}) s_\theta(T-t, \overleftarrow{\mathbf{X}}_t)\}\mathrm{d}t + (1+\Psi)^{1/2}\beta_t^{1/2}\mathrm{d}\mathbf{B}_t,$$

where $\lambda_t = \lambda_0/(\alpha_t + (1-\alpha_t)\lambda_0)$ and $\alpha_t = \exp[-2\int_0^t \beta_s \mathrm{d}s]$. In practice λ_0 controls the concentration of the measure while Ψ controls the strength of the corrector. We use $\Psi = 0$ as we do not consider a corrector term and perform a sweep over the parameters $\lambda_0 \in \{1, 2, 3, 5, 10\}$.

J.1 Synthetic data on polytopes

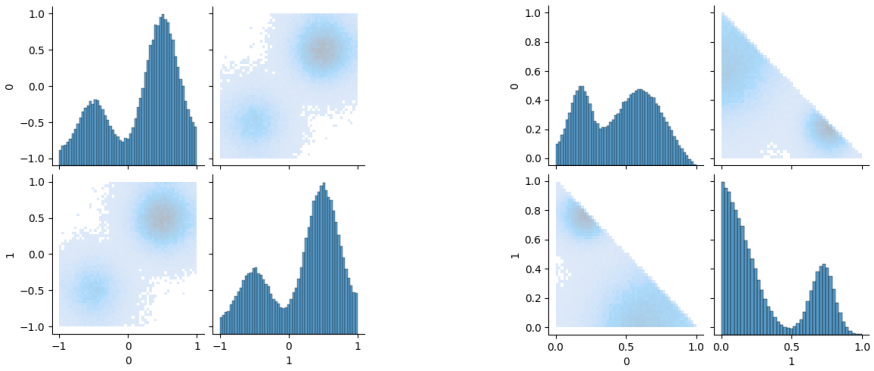
Hypercube $[-1,1]^n$. The hypercube is a specific instance of a convex polytope where the affine constraints are given by the following coefficients:

$$A = \begin{pmatrix} 1 & \dots & 0 \\ -1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ 0 & \dots & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}.$$

Where A is a $2n \times n$ matrix and b an n dimensional vector.

We construct the training and test datasets by sampling for both 100000 points from a mixture of 'wrapped normal' distributions illustrated in Figure 21a and which density is given by

$$p_0(x) = 0.7\,\text{ReflectedStep}[(0.5,0.5),\cdot,\{f_i\}_{i\in\mathcal{I}}]\#\mathcal{N}(0,0.25)+0.3\,\text{ReflectedStep}[(-0.5,-0.5),\cdot,\{f_i\}_{i\in\mathcal{I}}]\#\mathcal{N}(0,0.25).$$



(a) On the hypercube. (b) On the simplex.

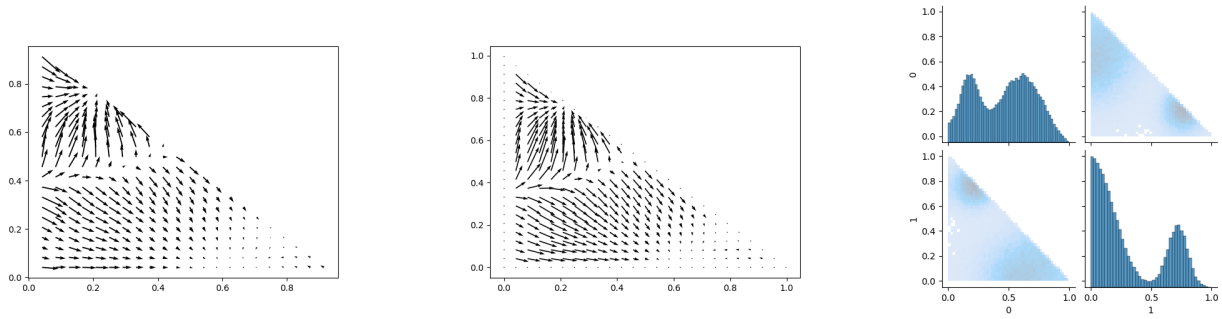
Figure 21: Pairwise and marginals samples from the synthetic data distribution.

Simplex Δ^n. Similarly, to parameterise the simplex as a convex polytope we set the matrix and constraints to be given by

$$A = \begin{pmatrix} -1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Where A will be a $n-1 \times n$ matrix. Essentially we perform diffusion over the first $n-1$ components of the simplex, allowing the last component to be determined by the one minus the sum of the first $n-1$.

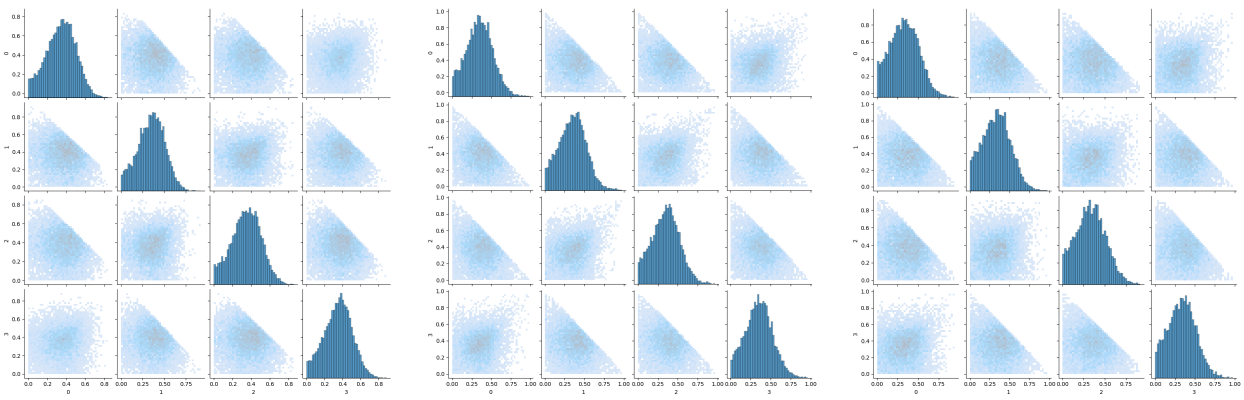
Similarly than for the hypercube, we construct the training and test datasets from generated data points which are illustrated in Figure 21b. The score network at different times is illustrated in Figure 22.



(a) $t = 0.01$. (b) $t = 0.5$. (c) Generated distribution.

Figure 22: Evolution of the score on the simplex and generated distribution.

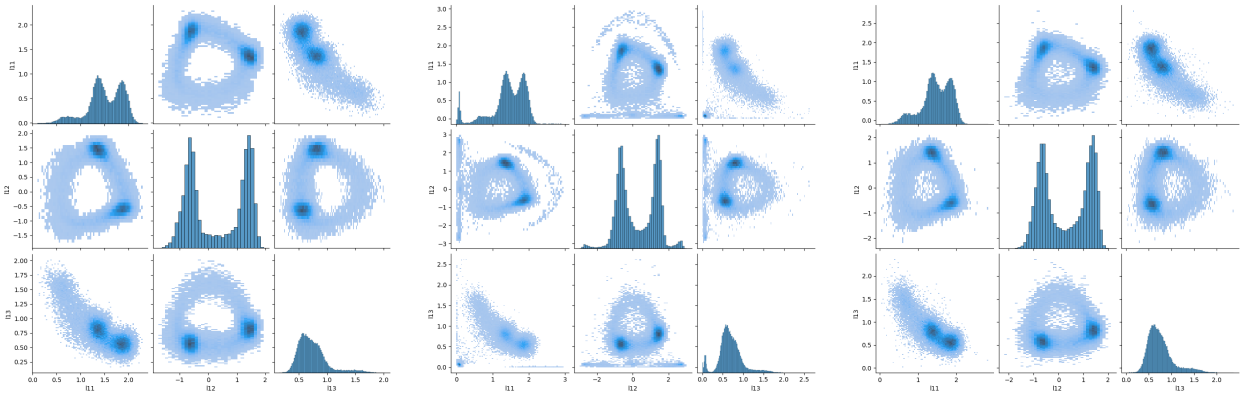
The Birkhoff polytope. The Birkhoff polytope is the space of doubly stochastic matrices, i.e. $B_n = \{\mathrm{P} \in [0,1]^{n\times n} : \sum_i^n P_{i,j} = 1, \sum_j^n P_{i,j} = 1\}$. It is a convex polytope in $\mathbb{R}^{n^2}$ and has dimension $d = (n-1)^2$.



(a) Data. (b) Log-barrier. (c) Reflected.

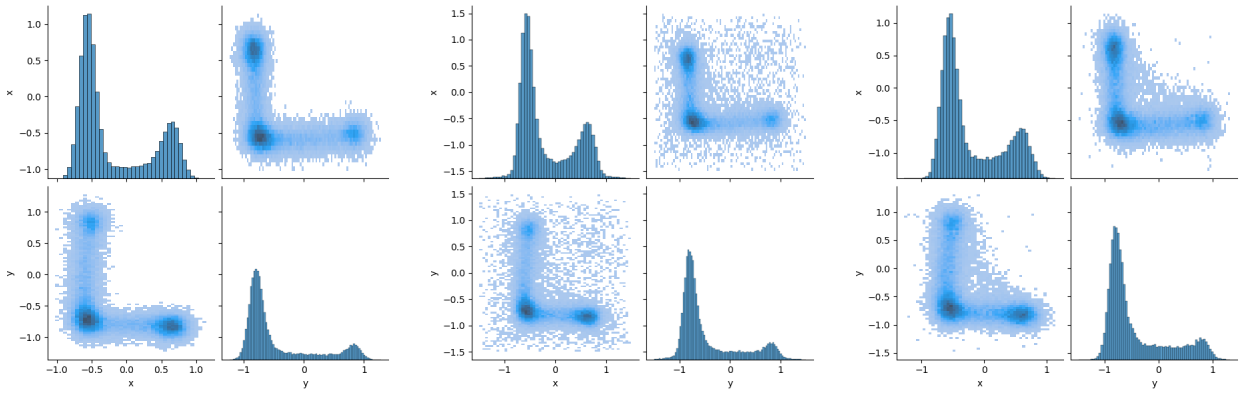
Figure 23: Pairwise and marginals samples on the Birkhoff polytope from synthetic data distribution and from trained constrained diffusion models.

J.2 Constrained SPD matrices for robotic arms modelling



(a) Data. (b) Log-barrier. (c) Reflected.

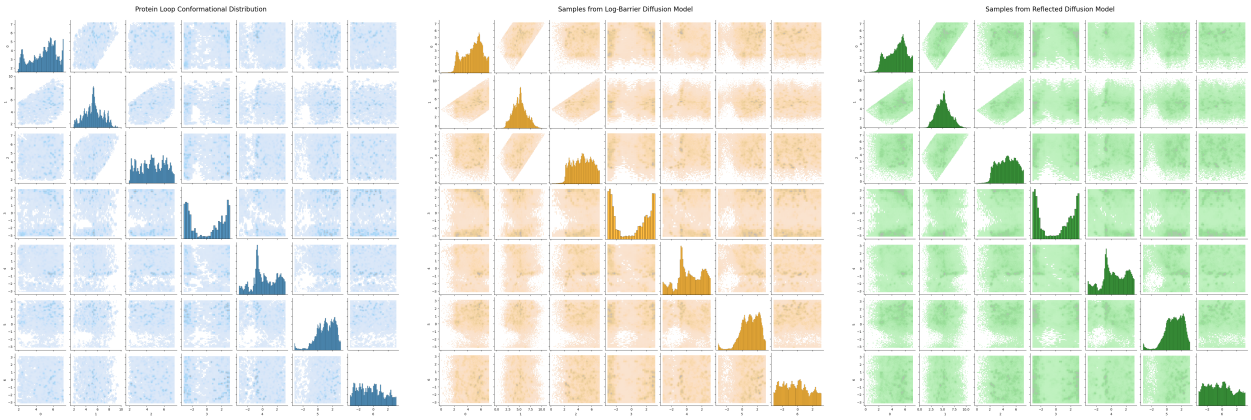
Figure 24: Pairwise and marginals distributions over the coefficients L_{11}, L_{21}, L_{22} of the lower triangle matrix parameterising SPD matrices $M = LL^\top$ (which represent the manipulability ellipsoids of the robotic arms).



(a) Data. (b) Log-barrier. (c) Reflected.

Figure 25: Pairwise and marginals distributions over the (x, y) locations of the robotic arms.

J.3 Conformational modelling of polypeptide backbones under anchor point constraints



(a) Data. (b) Log-barrier. (c) Reflected.

Figure 26: Pairwise and marginals distributions over the dimensions of the polytope and torus used to model the conformational ensembles of cyclic peptides generated by the reflected diffusion model.

J.4 Euclidean Models on the Hypercube and the Simplex

We train a set of Euclidean models on the hypercube and the simplex datasets. We use the same hyper-parameters as the main experiments, which were derived from best practices on Euclidean models in any case.

Space	Dimension	MMD	Portion outside $\mathcal{M}$
$[-1,1]^d$	2.00	$0.062_{\pm 0.011}$	0.988
	3.00	$0.076_{\pm 0.004}$	0.985
	10.00	$0.081_{\pm 0.005}$	0.964
Δ^d	2.00	$0.055_{\pm 0.013}$	0.964
	3.00	$0.068_{\pm 0.014}$	0.963
	10.00	$0.060_{\pm 0.003}$	0.926

Table 2: MMD and Proportion of samples outside the manifold for Euclidean models trained on the synthetic experiments.

In these experiments we can see that in lower dimensions the constrained models do better; in higher dimensions the Euclidean model performs better. There are a number of confounding factors here. First of all the data generating process for the synthetic experiments is actually a mixture of Normal distributions constrained within the boundary, as displayed in the plots. In higher dimensions our simulated distributions place much less mass near the boundary and more closely resemble a mixture of Normals. This is a very natural distribution for the Euclidean diffusion models to learn, and relatively more challenging for the log-barrier/reflected models which initialise at the uniform within the constraints. An alternative factor is that the dynamics of the Euclidean model remain simple in high dimensions, while the constrained processes become more complex in higher dimensions. Finally we note that, as Lou & Ermon (2023) comments, it may be that the ism loss is causing problems for the constrained methods in higher dimensions. Another factor may be relevant: in this paper we did not significantly explore the design space for constrained diffusion models. We use all the standard tricks derived for Euclidean models, including the std trick, the residual trick, gradient clipping, etc. It seems likely we could significantly improve the relative performance of the constrained models with more time working on the training dynamics in higher dimensions.