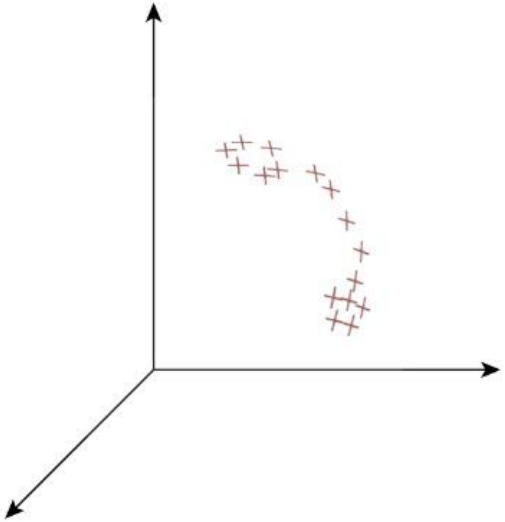
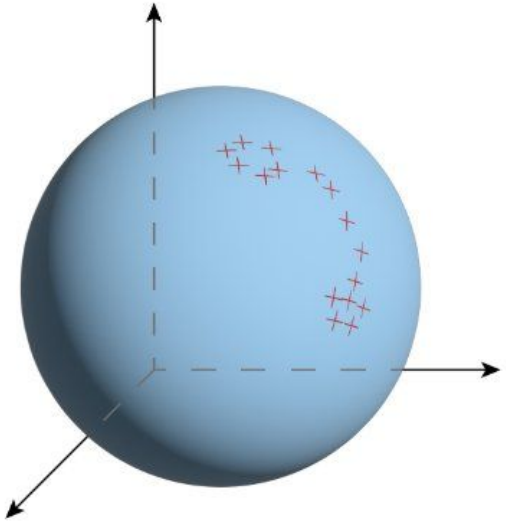
(Riemannian) Diffusion Models with Boundary

Joint work with Leo Klarner, Valentin de Bortoli, Emile Mathieu, and Michael Hutchinson

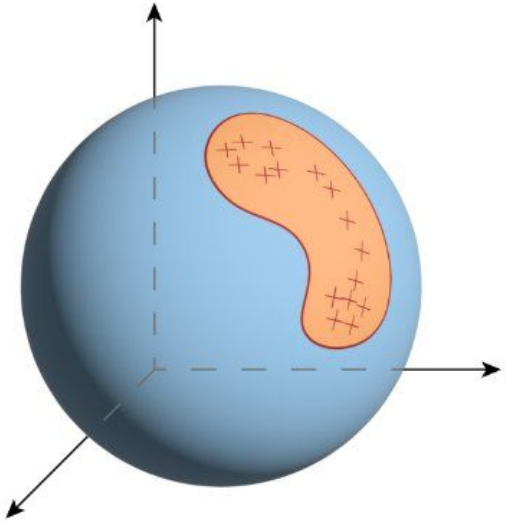
What is constrained diffusion?



(a) $\mathbb{R}^3$



(b) $\mathcal{S}^2 \subset \mathbb{R}^3$



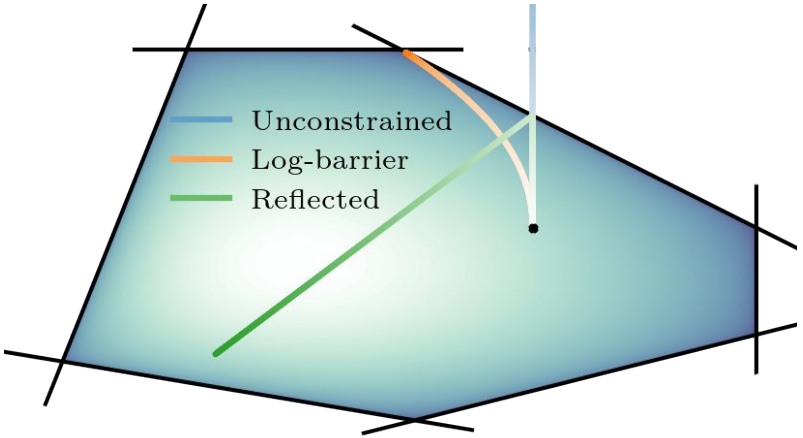
(c) $\mathcal{M} \subset \mathcal{S}^2 \subset \mathbb{R}^3$

Constrained Diffusion Models: Methods

How to define a forward noising process that stays in the polytope?

Two basic approaches:

1. **Warping methods**: change the geometry by warping space so that we never hit the boundary.
2. **Reflected methods**: leave the geometry as is and whenever we would hit the boundary we reflect

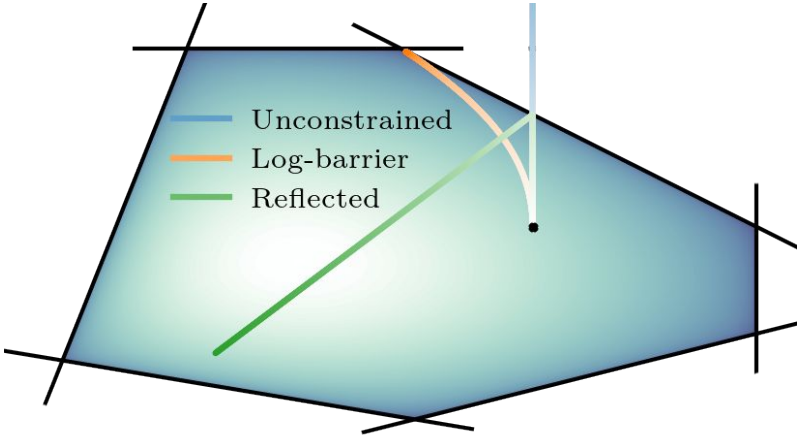


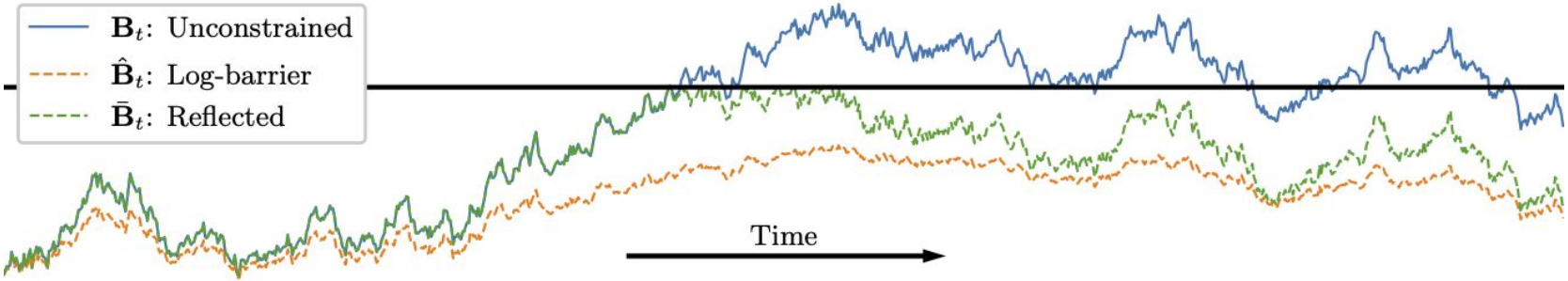
A note: the theory is better developed for barrier approaches than reflected methods. Most of the novel theory we develop is in the reflected methods.

How to define a forward noising process that stays in the polytope?

Two basic approaches:

1. **Warping methods**: change the geometry by warping space so that we never hit the boundary.
2. **Reflected methods**: leave the geometry as is and whenever we would hit the boundary we reflect





The recipe for a continuous diffusion model

1. A forward noising process converging to an invariant distribution
    - Both noising processes will converge to the uniform distribution over the compact set represented by the constraints

2. A time reversal giving a corresponding reverse process
    - This follows from Lee & Vempala (2017), Theorem 23 for the barrier approach.
    - We prove this by techniques from Petit (1997) and Haussman & Pardoux (1986).

3. A score matching loss
    - We prove the validity of the implicit score matching loss in the constrained domain under a Neumann boundary condition.

4. A discretization which converges to the continuous time forward/reverse processes in the limit of small step sizes

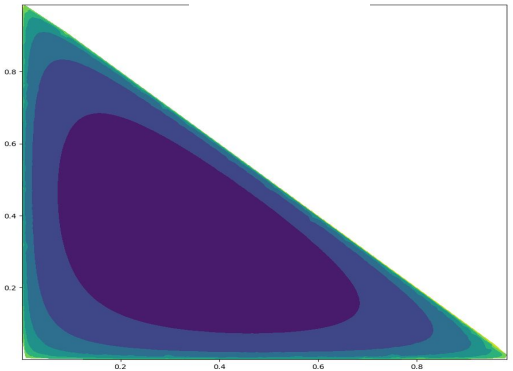
Discretization: Log-Barrier Approach

For the log-barrier approach our discretization is a Geodesic Random Walk on the Hessian manifold induced by a logarithmic barrier function. But since we do not have explicit access to the exponential we have to rely on a retraction, which is linearizes the geodesic.

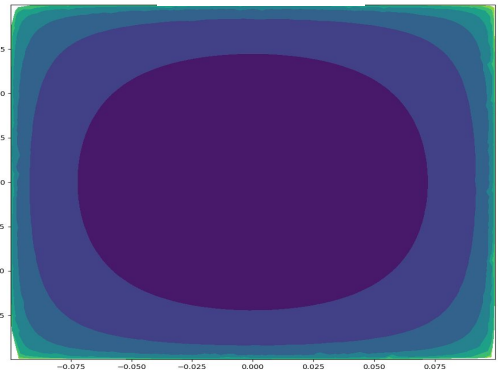
Algorithm 1 *Geodesic Random Walk.*

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

Simplex



Hypercube



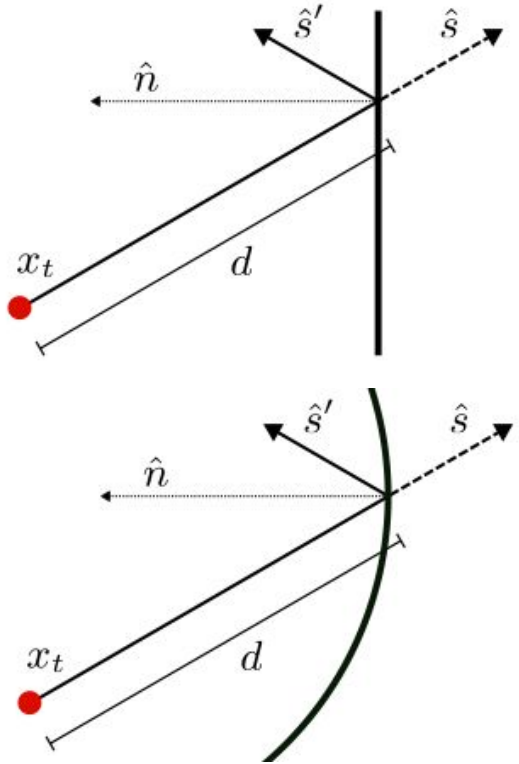
Discretization: Reflected Approach

Algorithm 2 *Reflected step algorithm.* The algorithm operates by repeatedly taking geodesic steps until one of the constraint 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 $\arg\text{intersect}_t$ 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 B

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

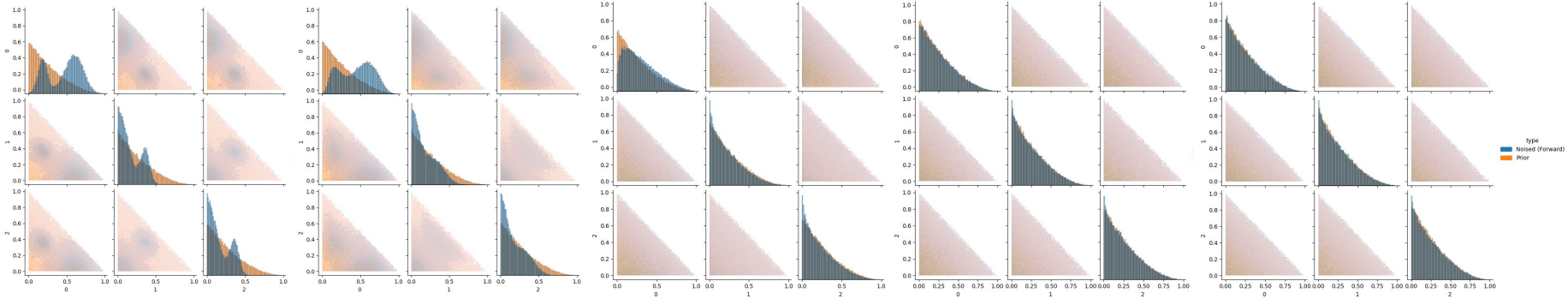
Algorithm 3 *Reflected Random Walk.* Discretisation of the SDE $\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t - \mathrm{d}\mathbf{k}_t$.

  Require: $T, N, X_0^\gamma, \{f_i\}_{i\in\mathcal{I}}$
  $\gamma = T/N$
  for $k \in \{0, \ldots, N-1\}$ **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 $\{X_k^\gamma\}_{k=0}^N$

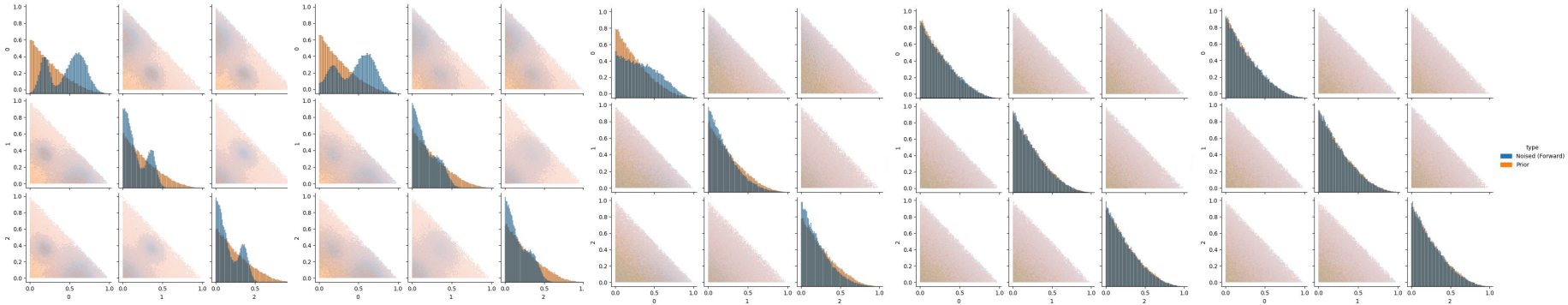


Invariant Distribution

Log Barrier Forward Process



Reflected Forward Process



t=0 → t=1

Constrained Diffusion Models: Applications

Applications for constrained diffusion models

1. **Mixtures of Gaussians in hypercubes and simplices**:
   We investigate performance on a set of synthetic distributions on hypercubes and simplices in various dimensions.

2. **Robotic arms under manipulability constraints**:
   When planning the trajectory of a robotic arm we often want to ensure the arm can manipulate certain areas as it moves. We model this as a 2D position and an ellipse representing the target manipulation area.

3. **Modeling protein chains with fixed end points**:
   Cyclic peptides and antibody loop structures represent very important clinical targets for generative protein modeling. Both of these are essentially chains with fixed endpoints, and we can model this as a product of a polytope and a torus.

Applications for constrained diffusion models

1. **Mixtures of Gaussians in hypercubes and simplices**:
   We investigate performance on a set of synthetic distributions on hypercubes and simplices in various dimensions.

2. **Robotic arms under manipulability constraints**:
   When planning the trajectory of a robotic arm we often want to ensure the arm can manipulate certain areas as it moves. We model this as a 2D position and an ellipse representing the target manipulation area.

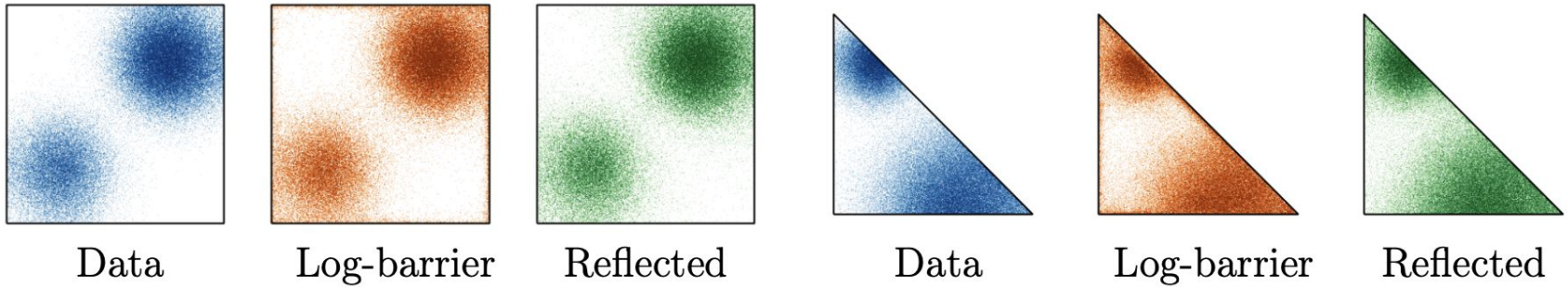
3. **Modeling protein chains with fixed end points**:
   Cyclic peptides and antibody loop structures represent very important clinical targets for generative protein modeling. Both of these are essentially chains with fixed endpoints, and we can model this as a product of a polytope and a torus.

Applications for constrained diffusion models

1. **Mixtures of Gaussians in hypercubes and simplices**:
   We investigate performance on a set of synthetic distributions on hypercubes and simplices in various dimensions.

2. **Robotic arms under manipulability constraints**:
   When planning the trajectory of a robotic arm we often want to ensure the arm can manipulate certain areas as it moves. We model this as a 2D position and an ellipse representing the target manipulation area.

3. **Modeling protein chains with fixed end points**:
   Cyclic peptides and antibody loop structures represent very important clinical targets for generative protein modeling. Both of these are essentially chains with fixed endpoints, and we can model this as a product of a polytope and a torus.

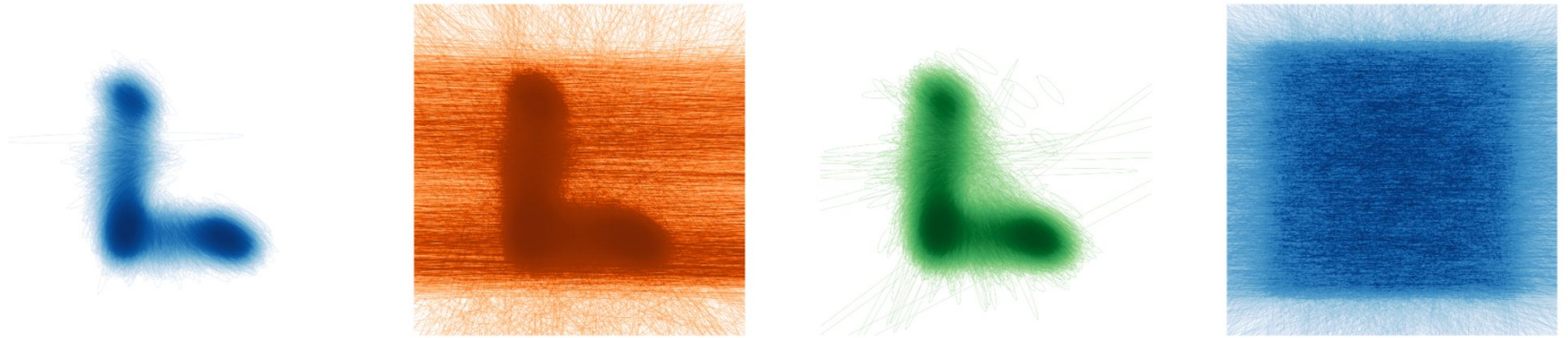


(a) 2D square data. (b) 2D Dirichlet data.

Figure 4.1: Histograms of samples from the data distribution and from trained constrained diffusion models.

Table 4.1: MMD metrics between samples from synthetic distributions and trained constrained and unconstrained (Euclidean) diffusion models. Means and confidence intervals are computed over 5 different runs.

Space	d	Log-barrier		Reflected		Euclidean	
		MMD	% in $\mathcal{M}$	MMD	% in $\mathcal{M}$	MMD	% in $\mathcal{M}$
$[-1,1]^d$	2	.066±.006	100.0	.055±.015	100.0	.062±.011	98.8
	3	.209±.077	100.0	.080±.004	100.0	.076±.004	98.5
	10	.330±.004	100.0	.313±.048	100.0	.081±.005	96.4
Δ^d	2	.050±.012	100.0	.043±.002	100.0	.055±.013	96.4
	3	.238±.009	100.0	.181±.007	100.0	.068±.014	96.3
	10	.275±.015	100.0	.290±.009	100.0	.060±.003	92.6



(a) Samples from the data distribution.

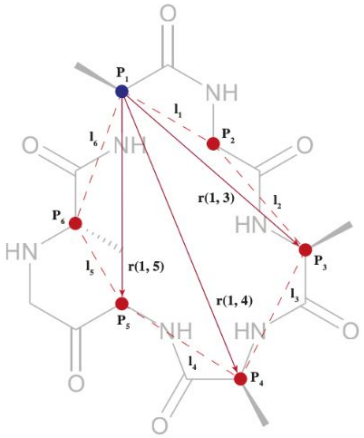
(b) Samples from our log-barrier diffusion.

(c) Samples from our reflected diffusion.

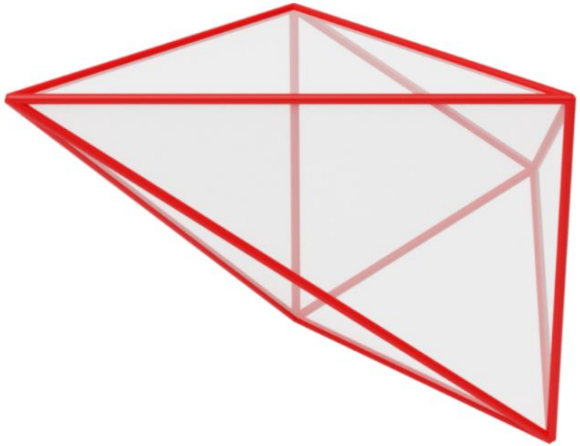
(d) Samples from the uniform distribution.

Figure 4.2: Samples in $S^2_{++} \times \mathbb{R}^2$ from (a) the data distribution, (b) our log-barrier diffusion model, (c) our reflected diffusion model and (d) the uniform distribution. Each sample is visualised as the manipulability ellipsoid encoded by the SPD matrix $M \in S^2_{++}$ placed at the corresponding location in $\mathbb{R}^2$.

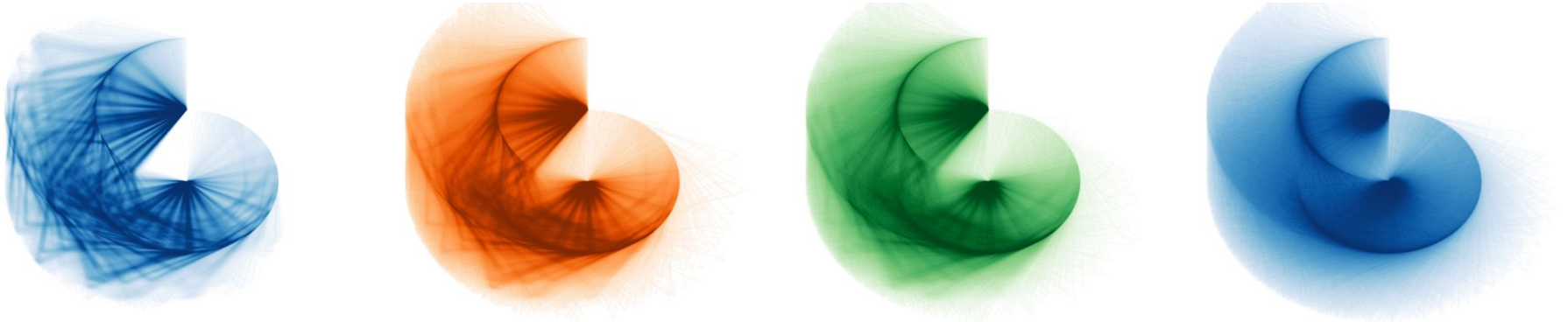
Generating Conformations of Cyclic Peptides



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



(a) Samples from the data distribution.

(b) Samples from our log-barrier diffusion.

(c) Samples from our reflected diffusion.

(d) Samples from the uniform distribution.

Figure 4.4: 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 C.3.3.

My contributions to this work

- I proposed and initially conceived of this work, including both methods we developed here, and wrote all of the code for the implementations.
- Valentin de Bortoli was mainly responsible of the technical proofs, though we discussed it at length and I reviewed all proofs.
- Leo Klarner was responsible for setting up the data for the experiments.
- Emile and Michael provided guidance throughout, and wrote the codebase which facilitated extending our methods to Riemannian manifolds.

Problems and Follow-up Work

There are three major issues with the approaches here:

1. Both approaches are quite slow even in low dimensions when compared to unconstrained diffusion models, and they scale terribly as dimension grows.
2. As seen in the loops case there is considerable work to be done in improving learning dynamics using these methods.
3. Unconstrained diffusion models lead to better results, contrary to expectations about incorporating prior information.

Problems and Follow-up Work

Two follow up papers address these issues in different ways:

Warping methods: Liu, G.H., Chen, T., Theodorou, E. and Tao, M., 2024. Mirror diffusion models for constrained and watermarked generation. NeurIPS, 36.

Reflected methods: Fishman, N., Klarner, L., Mathieu, E., Hutchinson, M. and De Bortoli, V., 2024. Metropolis sampling for constrained diffusion models. NeurIPS, 36.

1. The computational issues are completely resolved by these new approaches.
2. The training dynamics are significantly improved for both these methods.
3. The warping methods exploit the performance of Euclidean models to do well, so assuming there is not significant mass near the boundary these methods should do very well. Reflected methods lag behind, but generalize to manifold geometries in more straightforward ways.

Appendix

Log-Barrier

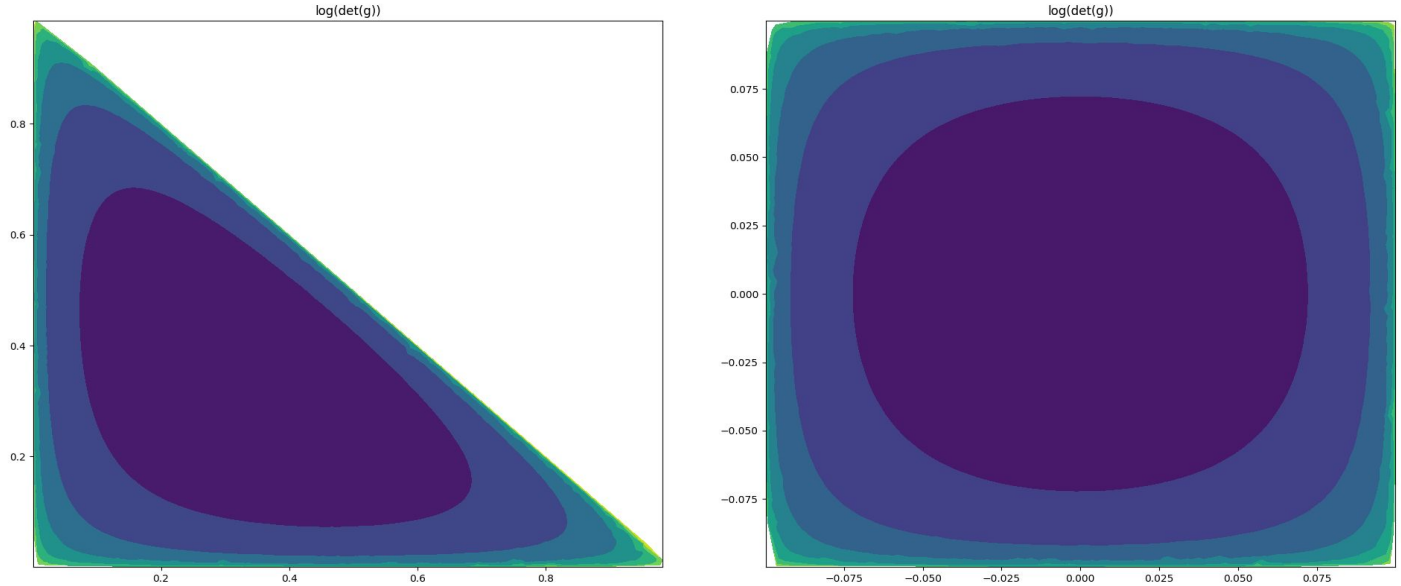
Barrier Langevin Dynamics. Barrier methods work by constructing a smooth potential $\phi : \mathcal{M} \to \mathbb{R}$ such that it blows up on the boundary of a desired set, see Nesterov et al. (2018). Such potentials are at the basis of interior methods in optimisation Boyd et al. (2004). Among these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

$$\phi(x) = -\sum_{i=1}^{m} \log(\langle \mathrm{A}_i, x \rangle - b_i). \tag{2.1}$$

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 $\partial\mathcal{M}_i = \{x \in \mathbb{R}^d : \langle \mathrm{A}_i, x \rangle = b\}$.

Assuming ϕ to be strictly convex and smooth its Hessian $\nabla^2\phi$ is positive definite thus it defines a valid Riemannian metric. The formal approach to 'warping the geometry' of the convex space with the boundary is to endow that space with the Hessian as a Riemannian metric $\mathfrak{g} = \nabla^2\phi$, making it into a Hessian Manifold. Using the barrier in equation 4, we have that

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



We can now define the forward process as the following Langevin dynamics as proposed by Lee & Vempala (2017)

$$\mathrm{d}\mathbf{X}_t = \tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1})(\mathbf{X}_t)\mathrm{d}t + \mathfrak{g}(\mathbf{X}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t, \tag{2.2}$$

with $\mathrm{div}(F)(x) \triangleq (\mathrm{div}(F_1)(x), \dots, \mathrm{div}(F_d)(x))^\top$, for any smooth $F : \mathbb{R}^d \to \mathbb{R}^d$. Denoting p_t the density of $\mathbf{X}_t$, we have that $\partial_t p_t = \mathrm{Tr}(\mathfrak{g}^{-1}\nabla^2 p_t)$. Hence, assuming that $\mathcal{M}$ is compact, the uniform distribution on $\mathcal{M}$, is the invariant measure and we have that the distribution of $(\mathbf{X}_t)_{t\geq 0}$ converges exponentially fast towards the uniform, see Lee & Vempala (2017, Theorem 23).

Time-reversal. Assuming that $\mathfrak{g}^{-1}$ and its derivative are bounded on $\mathcal{M}$, the time-reversal of (2.2) is given by Cattiaux et al. [2021], in particular we have

$$\begin{aligned} \mathrm{d}\overleftarrow{\mathbf{X}}_t &= [-\tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1}) + \mathrm{div}(\mathfrak{g}^{-1}) + \mathfrak{g}^{-1}\nabla \log p_{T-t}](\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathfrak{g}(\overleftarrow{\mathbf{X}}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t, \\ &= \left[\tfrac{1}{2}\mathrm{div}(\mathfrak{g}^{-1}) + \mathfrak{g}^{-1}\nabla \log p_{T-t}\right](\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathfrak{g}(\overleftarrow{\mathbf{X}}_t)^{-\frac{1}{2}}\mathrm{d}\mathbf{B}_t. \end{aligned} \tag{2.3}$$

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

Reflected

Skorokhod problem. The reflected Brownian motion is defined as the solution to the Skorokhod problem. We say that $(\bar{\mathbf{B}}_t, \mathbf{k}_t)_{t\geq 0}$ is a solution to the *Skorokhod problem* (Skorokhod, 1961) if $(\mathbf{k}_t)_{t\geq 0}$ is a bounded variation process and $(\bar{\mathbf{B}}_t)_{t\geq 0}$ a continuous adapted process such that for any $t \geq 0$,

$$\bar{\mathbf{B}}_t = \bar{\mathbf{B}}_0 + \mathbf{B}_t - \mathbf{k}_t \in \mathcal{M}, \tag{2.4}$$

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 $(\mathbf{k}_t)_{t\geq 0}$. 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. 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 $-\mathbf{n}$. 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 quantify 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 2.2.1(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 $x_0 \in \partial\mathcal{M}$*

$$\partial_t p_t(x) = \tfrac{1}{2}\Delta p_t(x), \quad \partial_{\mathbf{n}} p_t(x_0) = 0, \tag{2.5}$$

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

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 Figure 5. 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 2.2.2. *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 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

$$\mathrm{d}\overleftarrow{\mathbf{X}}_t = \nabla \log p_{T-t}(\overleftarrow{\mathbf{X}}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t - \mathrm{d}\overleftarrow{\mathbf{k}}_t, \tag{2.6}$$

which echoes the usual time-reversal formula equation 2.

Implicit Score Matching

Proposition 2.3.1. *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*

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

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 optimize the score network using the ism loss function so long as we enforce a Neumann boundary condition.