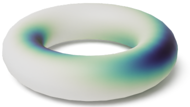
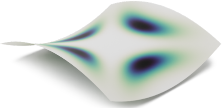
Score-based Generative Models on Riemannian manifolds

Émile Mathieu, University of Cambridge

December 15, 2022






Valentin De Bortoli


Émile Mathieu


Michael Hutchinson


James Thornton

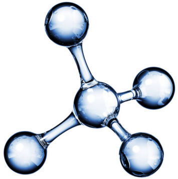

Yee Whye Teh


Arnaud Doucet

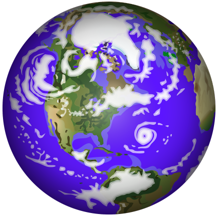
The rise of diffusion models

Figure 1: Images from stable diffusion.

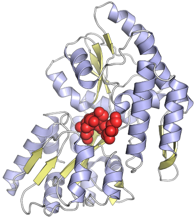
The need for geometric prior



Molecular conformation

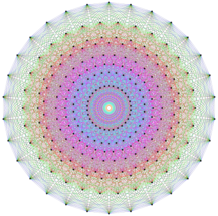


Weather Modelling

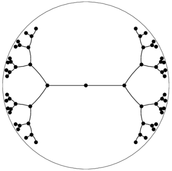


Protein Docking

Robotics



Lie groups



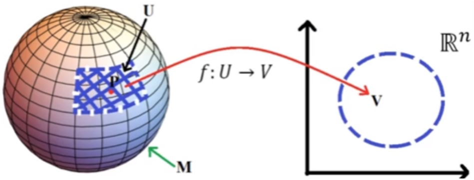
Tree embedding

What are Riemannian manifolds? A smooth manifold $\mathcal{M}$

A **Riemannian manifold** is a tuple $(\mathcal{M}, \mathfrak{g})$.

What are Riemannian manifolds? A smooth manifold $\mathcal{M}$

A **Riemannian manifold** is a tuple $(\mathcal{M}, \mathfrak{g})$.



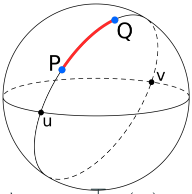
- A smooth **manifold** $\mathcal{M}$ is locally 'similar' (homeomorphic) to $\mathbb{R}^d$.
  - Exists homeomorphic *coordinate charts* (U, ϕ) s.t. $\phi : U \subset \mathcal{M} \rightarrow V \subset \mathbb{R}^d$.
  - Charts are suitably compatible (i.e. compositions are differentiable).

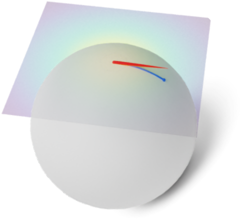
What are Riemannian manifolds? A metric g

A (Riemannian) **metric** $g(x)$ defines a (positive-definite) inner product on $\mathrm{T}\mathcal{M}_x$.

What are Riemannian manifolds? A metric g

A (Riemannian) **metric** $g(x)$ defines a (positive-definite) inner product on $\mathrm{T}\mathcal{M}_x$.



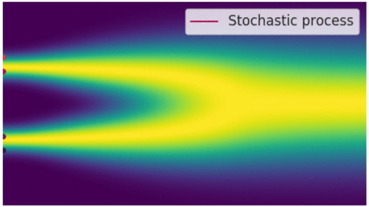


$\langle u, v\rangle_x = u^\top g(x)v$, and $g(x)$ varies smoothly with x.

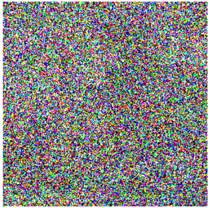
- Gradient $\nabla^{\mathcal{M}} f(x) = g(x)^{-1}\nabla f(x)$ and divergence $\mathrm{div}^{\mathcal{M}}$.
- Geodesic distance: $d^{\mathcal{M}}(x, y) = \inf\{L(\gamma) : \mathrm{C}^1 \text{ curve } \gamma \ s.t.\ \gamma(0) = x \text{ and } \gamma(1) = y\}$.
- Geodesic: argmin $\gamma(t)$ of geodesic distance.
- Exponential map: $\exp_x(tv) = \gamma(t)$ with $\gamma(0) = x$ and $\gamma'(0) = v$.

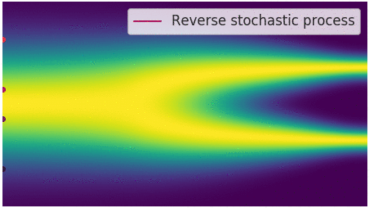
Ingredients of diffusion models

- **Forward process**



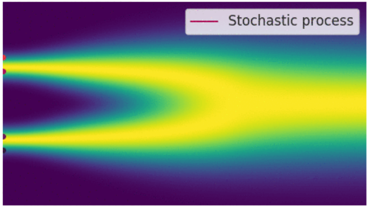
- **Backward process**



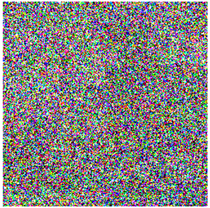


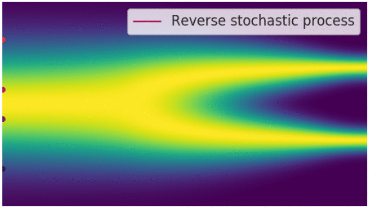
Ingredients of diffusion models

- **Forward process**



- **Backward process**





- The main ingredients:
  - Forward noising process
  - Time-reversal formula
  - Score-matching toolbox
  - Discretisation scheme

Video extracted from Yang Song's blog.

Forward processes on manifolds

Stochastic differential equation (SDE):

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t, \mathbf{X}_t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}.}$$

Forward processes on manifolds

Stochastic differential equation (SDE):

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t, \mathbf{X}_t)\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}.}$$

Langevin dynamics:

$$\boxed{\mathrm{d}\mathbf{X}_t = -\nabla_{\mathbf{X}_t} U(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}}$$

admits **invariant** density: $\mathrm{d}p_{\mathsf{ref}}/\mathrm{dVol}_{\mathcal{M}}(x) \propto \mathrm{e}^{-U(x)}$ (Durmus, 2016, Section 2.4).

Forward processes on manifolds

Stochastic differential equation (SDE):

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t, \mathbf{X}_t)\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}.}$$

Langevin dynamics:

$$\boxed{\mathrm{d}\mathbf{X}_t = -\nabla_{\mathbf{X}_t} U(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}}$$

admits **invariant** density: $\mathrm{d}p_{\text{ref}}/\mathrm{dVol}_{\mathcal{M}}(x) \propto \mathrm{e}^{-U(x)}$ (Durmus, 2016, Section 2.4).

What is this $\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}$ thing?

Intrinsic view

- Laplace-Beltrami Operator
  $\Delta_{\mathcal{M}} \triangleq \mathrm{div}_{\mathcal{M}} \nabla_{\mathcal{M}}$
- BM has density given by *heat equation*
  $\frac{\partial p_t}{\partial t} = \Delta_{\mathcal{M}} p_t$

Forward processes on manifolds

Stochastic differential equation (SDE):

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t, \mathbf{X}_t)\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}.}$$

Langevin dynamics:

$$\boxed{\mathrm{d}\mathbf{X}_t = -\nabla_{\mathbf{X}_t} U(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}}$$

admits **invariant** density: $\mathrm{d}p_{\text{ref}}/\mathrm{dVol}_{\mathcal{M}}(x) \propto \mathrm{e}^{-U(x)}$ (Durmus, 2016, Section 2.4).

What is this $\mathrm{d}\mathbf{B}_t{}^{\mathcal{M}}$ thing?

Intrinsic view

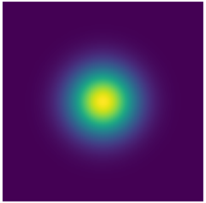
- Laplace-Beltrami Operator $\Delta_{\mathcal{M}} \triangleq \mathrm{div}_{\mathcal{M}} \nabla_{\mathcal{M}}$
- BM has density given by *heat equation* $\frac{\partial p_t}{\partial t} = \Delta_{\mathcal{M}} p_t$

Extrinsic view

- Embed the manifold in $\mathbb{R}^{p>d}$
- Projecting the $\mathbb{R}^p$ BM onto the manifold

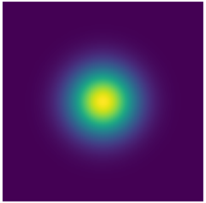
Invariant densities

- If $\mathcal{M} = \mathbb{R}^d$ is **Euclidean**
  - $U(x) = \|x\|^2/2$ (Gaussian distribution).
  - $\Rightarrow$ $b(t,x) = -x \Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = -\mathbf{X}_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t}$ (Ornstein-Uhlenbeck).



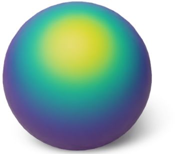
Invariant densities

- If $\mathcal{M} = \mathbb{R}^d$ is **Euclidean**
  - $U(x) = \|x\|^2/2$ (Gaussian distribution).
  - $\Rightarrow$ $b(t,x) = -x \Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = -\mathbf{X}_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t}$ (Ornstein-Uhlenbeck).



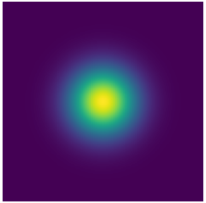
- If is $\mathcal{M}$ **non Euclidean**
  - $U(x) = d_{\mathcal{M}}(x,\mu)^2/(2\gamma^2)$ (*Riemannian* normal).
  - $\Rightarrow$ $b(t,x) = -\exp_x^{-1}(\mu)/\gamma^2$

    $\Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = -\exp_{\mathbf{X}_t}^{-1}(\mu)/\gamma^2 \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}}$.



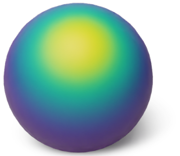
Invariant densities

- If $\mathcal{M} = \mathbb{R}^d$ is **Euclidean**
  - $U(x) = \|x\|^2/2$ (Gaussian distribution).
  - $\Rightarrow b(t,x) = -x \Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = -\mathbf{X}_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t}$ (Ornstein-Uhlenbeck).



- If is $\mathcal{M}$ **non Euclidean**
  - $U(x) = d_{\mathcal{M}}(x,\mu)^2/(2\gamma^2)$ (*Riemannian* normal).
  - $\Rightarrow b(t,x) = -\exp_x^{-1}(\mu)/\gamma^2$

    $\Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = -\exp_{\mathbf{X}_t}^{-1}(\mu)/\gamma^2 \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}}$.



- If $\mathcal{M}$ is **compact**:
  - $U(x) =$ constant (Uniform distribution)
  - $\Rightarrow b(t,x) = \mathbf{0} \Leftrightarrow \boxed{\mathrm{d}\mathbf{X}_t = \mathrm{d}\mathbf{B}_t^{\mathcal{M}}}$ (Brownian motion).

Time reversal process

Theorem 1: Time-reversal diffusion

Let $(\mathbf{X}_t)_{t\in[0,T]}$ associated with the SDE $\mathrm{d}\mathbf{X}_t = b(t,\mathbf{X}_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}$ and $(\mathbf{Y}_t)_{t\in[0,T]} = (\mathbf{X}_{T-t})_{t\in[0,T]}$ the time-reversal. Under mild assumptions on p_0 and on p_t the density of $\mathbb{P}_t = \mathcal{L}(\mathbf{X}_t)$, then $(\mathbf{Y}_t)_{t\in[0,T]}$ is associated with

$$\mathrm{d}\mathbf{Y}_t = \left\{ -b(T-t,\mathbf{Y}_t) + \sigma(T-t)^2\, \nabla \log p_{T-t}(\mathbf{Y}_t) \right\}\mathrm{d}t + \sigma(T-t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}. \tag{1}$$

This is an extension of Cattiaux et al. (2021, Theorem 4.9).

Time reversal process

Theorem 2: Time-reversal diffusion

Let $(\mathbf{X}_t)_{t\in[0,T]}$ associated with the SDE $\mathrm{d}\mathbf{X}_t = b(t,\mathbf{X}_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}$ and $(\mathbf{Y}_t)_{t\in[0,T]} = (\mathbf{X}_{T-t})_{t\in[0,T]}$ the time-reversal. Under mild assumptions on p_0 and on p_t the density of $\mathbb{P}_t = \mathcal{L}(\mathbf{X}_t)$, then $(\mathbf{Y}_t)_{t\in[0,T]}$ is associated with

$$\mathrm{d}\mathbf{Y}_t = \left\{-b(T-t,\mathbf{Y}_t) + \sigma(T-t)^2\,\nabla\log p_{T-t}(\mathbf{Y}_t)\right\}\mathrm{d}t + \sigma(T-t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}. \tag{1}$$

This is an extension of Cattiaux et al. (2021, Theorem 4.9).

Time reversal of **Langevin dynamics**:

$$\mathrm{d}\mathbf{Y}_t = \left\{\nabla_{\mathbf{X}_t}U(\mathbf{X}_t) + 2\,\nabla\log p_{T-t}(\mathbf{Y}_t)\right\}\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}. \tag{2}$$

Discretising SDEs: Geodesic Random Walk

- Given a **continuous time dynamics**

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}\ .} \tag{3}$$

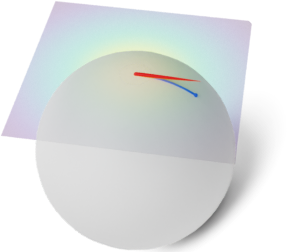
Discretising SDEs: Geodesic Random Walk

- Given a **continuous time dynamics**

$$\boxed{\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}\ .} \tag{3}$$

- **Geodesic Random Walk** approximation

$$\boxed{X_{k+1} = \exp_{X_k}[\gamma\, b(k\gamma, X_k) + \sqrt{\gamma}\,\sigma(k\gamma)\, Z_{k+1}]\ .} \tag{4}$$



Score approximation: Denoising score matching (DSM)

Denoising score matching (DSM) on manifolds is *very* similar to the Euclidean case

$$\ell_{t|s}(\mathbf{s}_t) = \int_{\mathcal{M}^2} \left\| \nabla_x \log p_{t|s}(x_t|x_s) - \mathbf{s}_t(x_t) \right\|^2 \mathrm{d}\mathbb{P}_{s,t}(x_s, x_t). \tag{5}$$

However need to evaluate $\nabla_x \log p_{t|s}(x_t|x_s)$.

Score approximation: Denoising score matching (DSM)

Denoising score matching (DSM) on manifolds is *very* similar to the Euclidean case

$$\ell_{t|s}(\mathbf{s}_t) = \int_{\mathcal{M}^2} \left\| \nabla_x \log p_{t|s}(x_t|x_s) - \mathbf{s}_t(x_t) \right\|^2 \mathrm{d}\mathbb{P}_{s,t}(x_s, x_t). \tag{5}$$

However need to evaluate $\nabla_x \log p_{t|s}(x_t|x_s)$.

- **Sturm–Liouville** decomposition (Chavel, 1984) (assuming compactness)

$$p_{t|0}(x_t|x_0) = \sum_{j \in \mathbb{N}} \mathrm{e}^{-\lambda_j t} \phi_j(x_0) \phi_j(x_t). \tag{6}$$

Truncation approximation to DSM

$$\nabla_{x_t} \log p_{t|0}(x_t|x_0) \approx S_{J,t}(x_0, x_t) \triangleq \nabla_{x_t} \log \sum_{j=0}^{J} \mathrm{e}^{-\lambda_j t} \phi_j(x_0) \phi_j(x_t). \tag{7}$$

Score approximation: Denoising score matching (DSM)

Denoising score matching (DSM) on manifolds is *very* similar to the Euclidean case

$$\ell_{t|s}(\mathbf{s}_t) = \int_{\mathcal{M}^2} \left\| \nabla_x \log p_{t|s}(x_t|x_s) - \mathbf{s}_t(x_t) \right\|^2 \mathrm{d}\mathbb{P}_{s,t}(x_s, x_t). \tag{5}$$

However need to evaluate $\nabla_x \log p_{t|s}(x_t|x_s)$.

- **Sturm–Liouville** decomposition (Chavel, 1984) (assuming compactness)

$$p_{t|0}(x_t|x_0) = \sum_{j \in \mathbb{N}} \mathrm{e}^{-\lambda_j t} \phi_j(x_0) \phi_j(x_t). \tag{6}$$

Truncation approximation to DSM

$$\nabla_{x_t} \log p_{t|0}(x_t|x_0) \approx S_{J,t}(x_0, x_t) \triangleq \nabla_{x_t} \log \sum_{j=0}^{J} \mathrm{e}^{-\lambda_j t} \phi_j(x_0) \phi_j(x_t). \tag{7}$$

- **Varadhan approximation** to DSM, 1st order taylor expansion

$$\lim_{t \to 0} t \nabla_{x_t} \log p_{t|0}(x_t|x_0) = \exp_{x_t}^{-1}(x_0). \tag{8}$$

Score approximation: Implicit Score Matching (ISM)

We can also avoid the need for $\nabla_{x_t} \log p_{t|0}(x_t|x_0)$ by using **implicit score matching**.

Proposition 1

Let $t, s \in (0, T]$ with $t > s$. Then, for any $\mathbf{s}_t \in \mathrm{C}^\infty(\mathcal{M})$, $\ell_{t|s}(\mathbf{s}_t) = 2\ell_t^{\mathrm{im}}(\mathbf{s}_t) + \int_{\mathcal{M}^2} \left\| \nabla_{x_t} \log p_{t|s}(x_t|x_s) \right\|^2 \mathrm{d}\mathbb{P}_{s,t}(x_s, x_t)$, where

$$\ell_t^{\mathrm{im}}(\mathbf{s}_t) = \int_{\mathcal{M}} \left\{ \frac{1}{2} \left\| \mathbf{s}_t(x_t) \right\|^2 + \mathrm{div}_{\mathcal{M}}(\mathbf{s}_t)(x_t) \right\} \mathrm{d}\mathbb{P}_t(x_t). \tag{9}$$

We need sufficient conditions on $p_{t|s}(x_t|x_s)\boldsymbol{s}(x_t)$, but these are easy to satisfy.

Score approximation: Summary

Loss	Approx	Loss function	Requirements $p_{t\|0}$	Requirements $\exp^{-1}_{\mathbf{X}_t}$	Complexity
$\ell_{t\|0}$ (DSM)	None	$\frac{1}{2}\mathbb{E}\left[\Vert \mathbf{s}(\mathbf{X}_t) - \nabla \log p_{t\|0}(\mathbf{X}_t\|\mathbf{X}_0)\Vert^2\right]$	✓	✗	$\mathcal{O}(1)$
	Truncation	$\frac{1}{2}\mathbb{E}\left[\Vert \mathbf{s}(\mathbf{X}_t) - S_{J,t}(\mathbf{X}_0, \mathbf{X}_t)\Vert^2\right]$	eigen system	✗	$\mathcal{O}(1)$
	Varhadan	$\frac{1}{2}\mathbb{E}\left[\Vert \mathbf{s}(\mathbf{X}_t) - \exp^{-1}_{\mathbf{X}_t}(\mathbf{X}_0)/t\Vert^2\right]$	✗	✓	$\mathcal{O}(1)$
$\ell_{t\|s}$ (DSM)	Varhadan	$\frac{1}{2}\mathbb{E}\left[\Vert \mathbf{s}(\mathbf{X}_t) - \exp^{-1}_{\mathbf{X}_t}(\mathbf{X}_s)/(t-s)\Vert^2\right]$	✗	✓	$\mathcal{O}(1)$
ℓ_t^{im} (ISM)	Deterministic	$\mathbb{E}\left[\frac{1}{2}\Vert \mathbf{s}(\mathbf{X}_t)\Vert^2 + \operatorname{div}_{\mathcal{M}}(\mathbf{s})(\mathbf{X}_t)\right]$	✗	✗	$\mathcal{O}(d)$
	Stochastic	$\mathbb{E}\left[\frac{1}{2}\Vert \mathbf{s}(\mathbf{X}_t)\Vert^2 + \varepsilon^\top \partial\, \mathbf{s}(\mathbf{X}_t)\varepsilon\right]$	✗	✗	$\mathcal{O}(1)$

Table 1: Computational complexity of score matching losses w.r.t. score network passes.

Parametrisation of score network

Approximate **Stein score** $(\nabla \log p_t)_{t\in[0,T]} \approx \mathbf{s}_\theta(t,\cdot)$ and $\mathbf{s}_\theta : [0,T] \to \mathcal{X}(\mathcal{M})$.

Parametrisation of score network

Approximate **Stein score** $(\nabla \log p_t)_{t\in[0,T]} \approx \mathbf{s}_\theta(t,\cdot)$ and $\mathbf{s}_\theta : [0,T] \to \mathcal{X}(\mathcal{M})$.

- **Generators** of vector fields:
  - $\mathbf{s}_\theta(t,x) \triangleq \sum_{i=1}^n \mathbf{s}_\theta^i(t,x) E_i(x)$.
  - *Definition*: Smooth vector fields $\{E_i(x)\}_{i=1}^n$ s.t. span $\left(\{E_i(x)\}_{i=1}^n\right) = \mathrm{T}_x\mathcal{M}$.
  - $\mathcal{M}$ is parallelisable $\Leftrightarrow$ there exists generators $\{E_i\}_{i=1}^n$ with $n = d$.

Parametrisation of score network

Approximate **Stein score** $(\nabla \log p_t)_{t\in[0,T]} \approx \mathbf{s}_\theta(t,\cdot)$ and $\mathbf{s}_\theta : [0,T] \to \mathcal{X}(\mathcal{M})$.

- **Generators** of vector fields:
  - $\mathbf{s}_\theta(t,x) \triangleq \sum_{i=1}^n \mathbf{s}_\theta^i(t,x) E_i(x)$.
  - *Definition*: Smooth vector fields $\{E_i(x)\}_{i=1}^n$ s.t. $\mathrm{span}\left(\{E_i(x)\}_{i=1}^n\right) = \mathrm{T}_x\mathcal{M}$.
  - $\mathcal{M}$ is parallelisable $\Leftrightarrow$ there exists generators $\{E_i\}_{i=1}^n$ with $n = d$.
- Examples for several class of manifolds:
  - If $\mathcal{M} = \mathbb{R}^d$, can choose $E_i(x) = e_i$ for $i = 1, \dots, d$.
  - If $\mathcal{M} = G$ is a Lie group, can choose $E_i(g) = g \cdot e_i$ with $\{e_i\}_i$ basis of Lie algebra.
  - If $\mathcal{M} \subset \mathbb{R}^n$ is submersion, can choose $E_i(x) = P_i(x)$ for $i = 1, \dots, n$ with P_i the ith column of the tangent projection matrix operator.

Theoretical guaranties on time-reversal

Theorem 3

Under mild assumption over p_0 and assuming that there exists $\mathtt{M} \geq 0$ such that for any $t \in [0, T]$ and $x \in \mathcal{M}$, $\left\| \mathbf{s}_{\theta^\star}(t, x) - \nabla \log p_t(x) \right\| \leq \mathtt{M}$, with $\mathbf{s}_{\theta^\star} \in \mathrm{C}([0, T], \mathcal{X}(\mathcal{M}))$. Then if $T > 1/2$, there exists $C \geq 0$ independent on T s.t.

$$\mathbb{W}_1(\mathcal{L}(Y_N), p_0) = C(\mathrm{e}^{-\lambda_1 T} + \sqrt{T/2}\mathtt{M} + \mathrm{e}^{T} \gamma^{1/2}), \tag{10}$$

where $\mathbb{W}_1$ is the Wasserstein distance of order one.

Experimental results

Prior work

- **Continuous normalising flows** (CNFs) (Mathieu and Nickel, 2020; Falorsi, 2021)

Train drift b_θ by maximising likelihood, solving the following **augmented** ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{X}_t \\ \log p(\mathbf{X}_t)\end{bmatrix} = \begin{bmatrix} b_\theta(t,\cdot) \\ -\mathrm{div}\left(b_\theta(t,\cdot)\right)\end{bmatrix}(\mathbf{X}_t)\,. \tag{11}$$

Prior work

- **Continuous normalising flows** (CNFs) (Mathieu and Nickel, 2020; Falorsi, 2021)

Train drift b_θ by maximising likelihood, solving the following **augmented** ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{X}_t \\ \log p(\mathbf{X}_t)\end{bmatrix} = \begin{bmatrix} b_\theta(t,\cdot) \\ -\mathrm{div}\left(b_\theta(t,\cdot)\right)\end{bmatrix}(\mathbf{X}_t)\,. \tag{11}$$

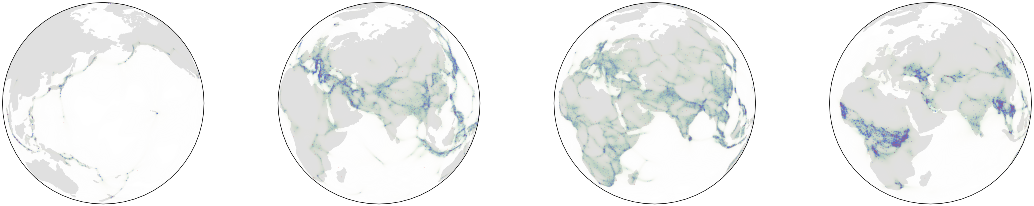
- **Moser flows** (Rozen et al., 2021)
  - Build on CNFs but uses a likelihood estimator which bypass solving an ODE.
  - Require a regularisation term involving an integral over the manifold...
  - Hence scale poorly with the manifold dimension.

Prior work: Summary

Method	Training	Likelihood evaluation	Sampling
RCNF	ODE $\mathcal{O}(dN)$	Augmented ODE $\mathcal{O}(dN)$	ODE $\mathcal{O}(N)$
Moser flow	div $\mathcal{O}(dk)$ or $\mathcal{O}(k)$	Augmented ODE $\mathcal{O}(dN)$	ODE $\mathcal{O}(N)$
RSGM	Score matching $\mathcal{O}(d)$ or $\mathcal{O}(1)$	Augmented ODE $\mathcal{O}(dN)$	SDE $\mathcal{O}(N^*)$

Table 2: Summary of computational complexity (w.r.t. neural network forward and backward passes) for different methods. d is the manifold dimension, k the number of Monte Carlo batches in Moser flow's regularizer, N is the number of steps in the (adaptive) ODE solver, whereas N^* is the number of steps in the SDE Euler-Maruyama solver.

Earth science data



(a) Volcano **(b)** Earthquake **(c)** Flood **(d)** Fire

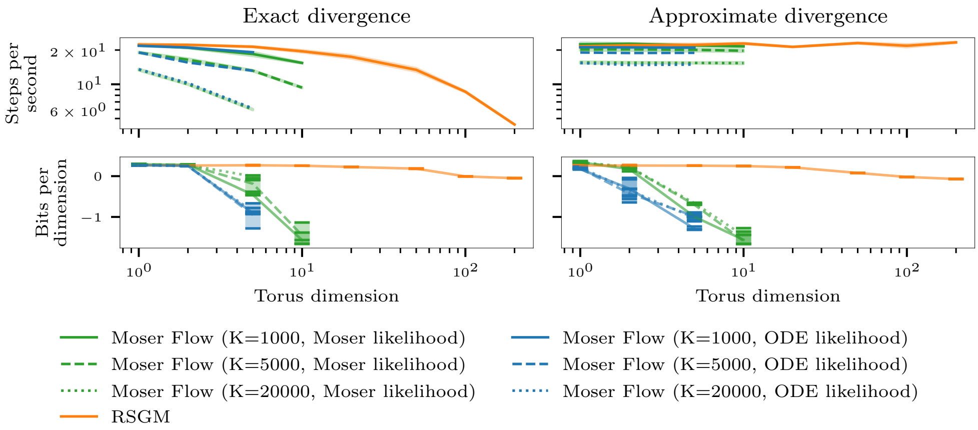
Method	Volcano	Earthquake	Flood	Fire
Mixture of Kent	$-0.80_{\pm 0.47}$	$0.33_{\pm 0.05}$	$0.73_{\pm 0.07}$	$-1.18_{\pm 0.06}$
Riemannian CNF	$\mathbf{-6.05}_{\pm \mathbf{0.61}}$	$0.14_{\pm 0.23}$	$1.11_{\pm 0.19}$	$\mathbf{-0.80}_{\pm \mathbf{0.54}}$
Moser Flow	$-4.21_{\pm 0.17}$	$\mathbf{-0.16}_{\pm \mathbf{0.06}}$	$\mathbf{0.57}_{\pm \mathbf{0.10}}$	$\mathbf{-1.28}_{\pm \mathbf{0.05}}$
Stereographic Score-Based	$-3.80_{\pm 0.27}$	$\mathbf{-0.19}_{\pm \mathbf{0.05}}$	$\mathbf{0.59}_{\pm \mathbf{0.07}}$	$\mathbf{-1.28}_{\pm \mathbf{0.12}}$
Riemannian Score-Based	$-4.92_{\pm 0.25}$	$\mathbf{-0.19}_{\pm \mathbf{0.07}}$	$\mathbf{0.45}_{\pm \mathbf{0.17}}$	$\mathbf{-1.33}_{\pm \mathbf{0.06}}$
Dataset size	827	6120	4875	12809

High dimensional torus

- We consider a wrapped Gaussian target distribution on $\mathbb{T}^d = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$.

High dimensional torus

- We consider a wrapped Gaussian target distribution on $\mathbb{T}^d = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$.



Synthetic data on Lie groups

- We consider a mixture of wrapped Gaussian target distribution on $\mathrm{SO}_3(\mathbb{R}) = \{\mathrm{Q} \in \mathrm{M}_3(\mathbb{R}) \,:\, \mathrm{Q}\mathrm{Q}^\top = I_3,\ \det(\mathrm{Q}) = 1\}$.

Synthetic data on Lie groups

- We consider a mixture of wrapped Gaussian target distribution on $\mathrm{SO}_3(\mathbb{R}) = \{\mathrm{Q} \in \mathrm{M}_3(\mathbb{R}) \ : \ \mathrm{Q}\mathrm{Q}^\top = I_3, \ \det(\mathrm{Q}) = 1\}$.

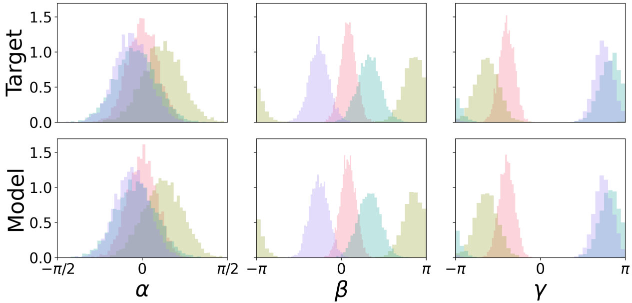


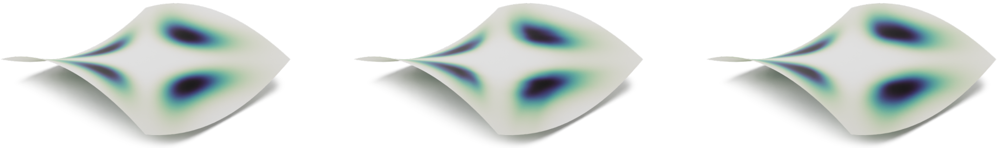
Figure 6: Histograms of $\mathrm{SO}_3(\mathbb{R})$ samples from a target mixture distribution.

Synthetic data on Lie groups (Cont'd)

Method	$M = 16$		$M = 32$		$M = 64$	
	$\log p$	NFE	$\log p$	NFE	$\log p$	NFE
Moser Flow	$0.85_{\pm 0.03}$	$2.3_{\pm 0.5}$	$0.17_{\pm 0.03}$	$2.3_{\pm 0.9}$	$\mathbf{-0.49_{\pm 0.02}}$	$7.3_{\pm 1.4}$
Exp-wrapped SGM	$\mathbf{0.87_{\pm 0.04}}$	$0.5_{\pm 0.1}$	$0.16_{\pm 0.03}$	$0.5_{\pm 0.0}$	$-0.58_{\pm 0.04}$	$0.5_{\pm 0.0}$
RSGM	$\mathbf{0.89_{\pm 0.03}}$	$\mathbf{0.1_{\pm 0.0}}$	$\mathbf{0.20_{\pm 0.03}}$	$\mathbf{0.1_{\pm 0.0}}$	$\mathbf{-0.49_{\pm 0.02}}$	$\mathbf{0.1_{\pm 0.0}}$

Table 4: Log-likelihood and neural function evaluations (NFE) in 10^3.

Hyperbolic experiments



(a) Target distribution. **(b)** Exp-wrapped SGM. **(c)** RSGM.

Figure 7: Samples from different probability distributions on $\mathbb{H}^2$ coloured w.r.t their density.

Conclusion

- Extension of **diffusion models** to the **Riemannian manifold** setting.

Ingredient \ Space	Euclidean	'Generic' Manifold	Compact
Forward process $\mathrm{d}\mathbf{X}_t =$	$-\mathbf{X}_t \,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}$	$-\nabla_{\mathbf{X}_t} U(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{B}_t^{\mathcal{M}}$	$\mathrm{d}\mathbf{B}_t^{\mathcal{M}}$
Base distribution	Gaussian	Wrapped Gaussian	Uniform
Time reversal	Cattiaux, 2021	Theorem 1	
Sampling forward	Direct	Geodesic Random Walk	
Sampling backward	Euler–Maruyama	Geodesic Random Walk	

Table 5: Differences between SGM on Euclidean spaces and RSGM on manifolds.

- Scalable, flexible, broadly applicable (sphere, $\mathrm{SO}(3)$, hyperbolic space).

References

P. Cattiaux, G. Conforti, I. Gentil, and C. Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021. Cited on pages 17, 18.

I. Chavel. *Eigenvalues in Riemannian Geometry*. Academic press, 1984. Cited on pages 21–23.

A. Durmus. *High Dimensional Markov Chain Monte Carlo Methods: Theory, Methods and Application*. PhD thesis, Paris-Sud XI, 2016. Cited on pages 10–13.

L. Falorsi. Continuous Normalizing Flows on Manifolds. Mar. 14, 2021. URL: `http://arxiv.org/abs/2104.14959`. Cited on pages 31, 32.

E. Mathieu and M. Nickel. Riemannian Continuous Normalizing Flows. In *Advances in Neural Information Processing Systems 33*. Curran Associates, Inc., 2020. Cited on pages 31, 32.

N. Rozen, A. Grover, M. Nickel, and Y. Lipman. Moser Flow: Divergence-based Generative Modeling on Manifolds. *Advances in Neural Information Processing Systems*, 2021. Cited on pages 31, 32.