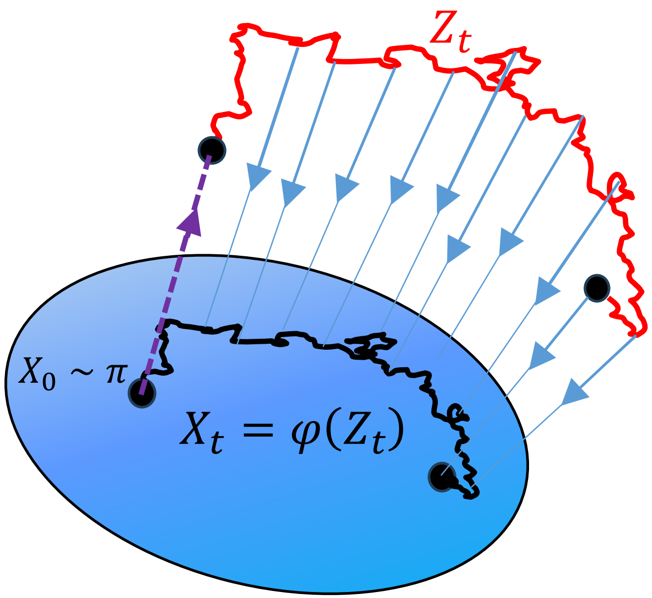
Efficient Diffusion Models for Symmetric Manifolds



Oren Mangoubi
WPI

Neil He
Yale

Nisheeth Vishnoi
Yale

ICML 2025

WPI

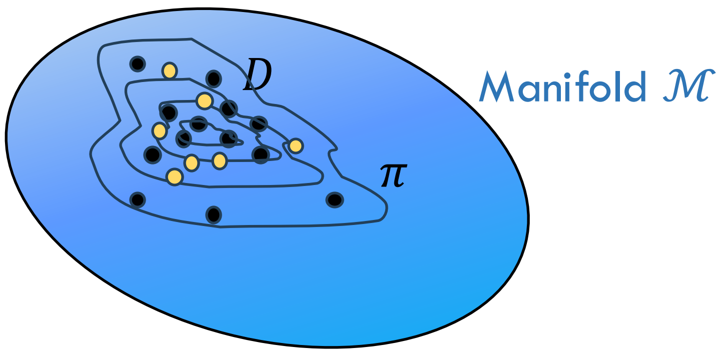
Yale University

Generative modeling on manifolds

Given:

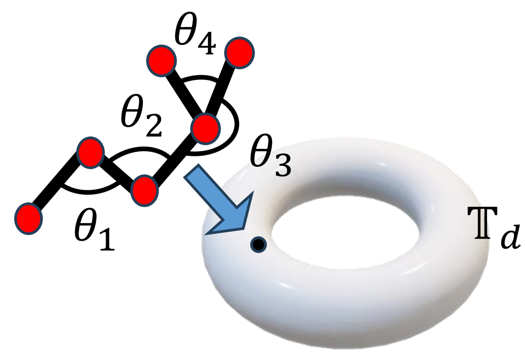
- a d-dimensional manifold $\mathcal{M}$
- dataset $D \subseteq \mathcal{M}$ sampled from a probability distribution π on $\mathcal{M}$.

Goal: Train a model to generate new samples ● (approximately) from π

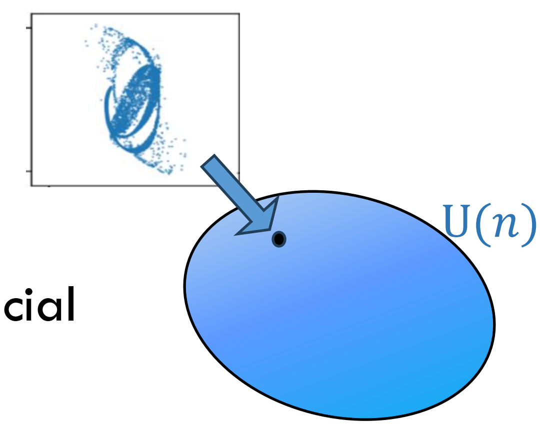


Many applications:

- Drug discovery: molecules with bond angles constrained to d-torus $\mathbb{T}_d$



- Quantum physics: time-evolution operators of quantum systems represented by $n \times n$ complex unitary matrices on unitary group $\mathrm{U}(n)$ (manifold dimension $d = \frac{n(n-1)}{2} \approx n^2$)



- Robotics: Robotic configurations constrained to Torus or special orthogonal group $\mathrm{SO}(n)$ of $n \times n$ orthogonal matrices

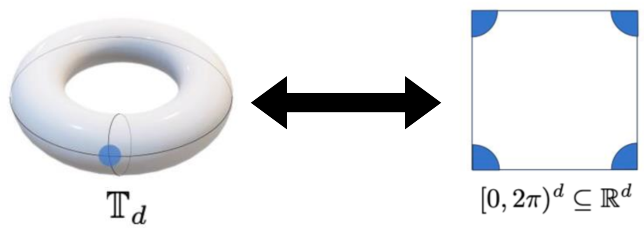
Previous work

Diffusion generative models in Euclidean space $\mathbb{R}^d$ (e.g., [Ho, Jain, Abbeel '20], [Rombach, Blattmann, Lorenz, Esser, Ommer '22])

- Very successful generating data in Euclidean space (e.g., images, videos)
- Fast training runtime: $O(d)$ arithmetic operations + $O(1)$ model *gradient evaluations per-iteration*

For data constrained to non-Euclidean manifold, can generate in Euclidean space & project onto manifold

Degrades sample quality due to distortions introduced by projection mapping



Many works provide diffusion models constrained to non-Euclidean Riemannian manifolds, e.g.

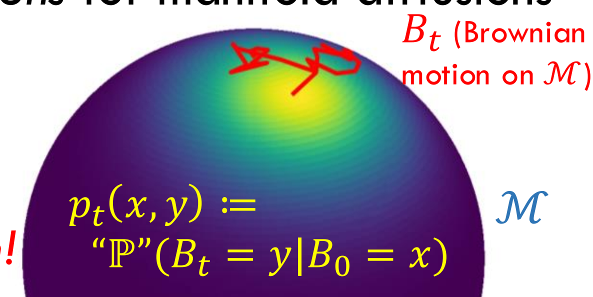
- Riemannian Score-Based Generative Models (RSGM) [De Bortoli, Mathieu, Hutchinson, Thornton, Teh, Doucet '22],
- Riemannian Diffusion (RDM) [Huang, Aghajohari, Bose, Panangaden, Courville '22],
- Scaling Riemannian Diffusion (SCRD) [Lou, Xu, Farris, Ermon '23],
- Trivialized Momentum Diffusion (TDM) [Zhu, Chen, Kong, Theodorou, Tao '25]

Generate samples *directly on manifold*, no distortion from post-processing projections

Large gap between per-iteration training runtime of Euclidean & manifold diffusions:
$\exp(d)$ arithmetic operations <u>or</u> $O(d)$ *model gradient evaluations* for manifold diffusions

Training objective relies on manifold's heat kernel p_t.

In Euclidean case, heat kernel is Gaussian. On manifolds with curvature, *in general, heat kernel has no closed-form expression!*



Our contributions

We consider the setting where the d-dimensional manifold $\mathcal{M}$ is a *symmetric-space*, such as a torus, sphere, special orthogonal group $\mathrm{SO}(n)$, or Unitary group $\mathrm{U}(n)$ $(\mathrm{d} \approx n^2)$

$\mathcal{M}$ has a *discontinuous* projection φ: $\mathbb{R}^m \to \mathcal{M}$, $m = O(d)$, *satisfying a symmetry property*

Algorithm: We give a new diffusion model on these manifolds with training objective computable in near-linear ($\leq O(d^{1.17})$) arith. ops. + $O(1)$ gradient evaluations

Improves on previous manifold diffusion algorithms, which require $\exp(d)$ arith. ops. or $\mathrm{O}(d)$ *model gradient evaluations* on symmetric manifolds such as $\mathrm{SO}(n), \mathrm{U}(n)$

Algorithm	Grad. Eval.	Arith. Ops.
RSGM, SCRD	1	2^d
RDM, RSGM, TDM	d	$\mathrm{poly}(d)$
This paper	1	$d^{1.17}$

Symmetry property: Each $z \in \mathbb{R}^m$ parametrizes as $z \equiv z(U, \Lambda)$, $U = \varphi(z)$
Average-case Lipschitzness: φ is L-Lipschitz on subset of $\mathbb{R}^m$ containing a Brownian motion w.h.p. *e.g., for* $\mathrm{U}(n)$*, we show* $L \leq O(d^2)$

Theoretical guarantees for sampling: Given an ε-minimizer of our training algorithm's objective function for a target distribution π on $\mathcal{M}$. Our sampling algorithm generates samples within $\tilde{O}(\varepsilon \times \mathrm{poly}(d))$ total variation of π, in $\mathrm{poly}(d)$ arithmetic operations.

Improves on sampling guarantees of, e.g., RSGM, which may not be polynomial-in-d

Empirical results:

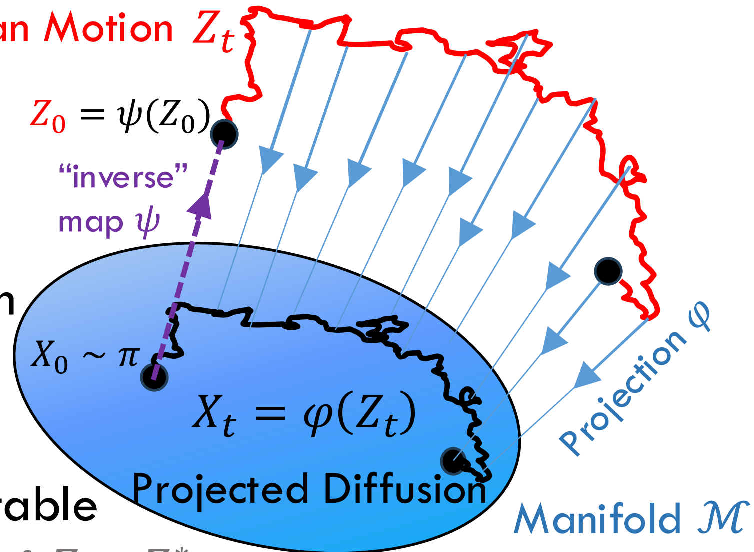
Runtime: Our model trains faster per-iteration than previous manifold diffusions on $\mathrm{SO}(n), \mathrm{U}(n)$, staying within factor of 3 of Euclidean diffusions even in high dimensions

Sample quality: Improves visual quality, C2ST scores of generated samples when trained on synthetic datasets on torus, $\mathrm{SO}(n)$, $\mathrm{U}(n)$. Improvements increase with dimension.

Algorithm derivation

Forward diffusion:

Key idea: Introduce diffusion with *spatially-varying* covariance, to efficiently handle manifold's curvature



- Covariance term allows diffusion to be a projection $X_t = \varphi(Z_t)$ of Euclidean Brownian motion Z_t onto manifold $\mathcal{M}$ *even when $\mathcal{M}$ has non-zero curvature*
- Choose projection map φ s.t. it is efficiently computable
  e.g., for $\mathrm{SO}(n)$, $\mathrm{U}(n)$, $\varphi(Z) = U$ where $U\Lambda U^*$ SVD of $Z + Z^*$

Forward diffusion adds noise to data until it is (nearly) distributed as the projection $\varphi(N(0, I_d))$ onto $\mathcal{M}$ of a standard Gaussian

---

Reverse diffusion: $Y_t = X_{T-t}$ (time-reversal of forward diffusion on manifold)

Starting from a sample $Y_0 \sim \varphi(N(0, I))$, removes noise to generate (approximate) samples from π

Training:

Expression for drift f and covariance g in SDE for Y_t can be derived as a projection of SDE of time-reversal H_t of Euclidean Brownian motion, via Ito's Lemma:

$$\mathrm{d}Y_t = f(Y_t, t)\mathrm{d}t + g(Y_t, t)\mathrm{d}B_t = \mathbb{E}\left[\left(\nabla\varphi(H_t)^\top + \mathrm{d}H_t^\top \nabla^2\varphi(H_t)\right)\mathrm{d}H_t \,|\, \varphi(H_t) = Y_t\right]$$

SDE expression for Y_t directly leads to an efficient training objective for f, g

- No need to compute manifold's heat Kernel (only Euclidean Gaussian heat Kernel)
- Projection φ can be computed in $O(d^{1.17})$ arithmetic operations, e.g. via SVD

Proof highlights

Previous diffusion models use Girsanov transforms to bound accuracy of generated samples

Girsanov transforms do not apply to our diffusion as it has spatially-varying covariance!

Instead, use optimal transport: construct coupling between "ideal" diffusion $\mathrm{d}Y_t = f^*(Y_t, t)\mathrm{d}t + \mathrm{g}^*(Y_t, t)\mathrm{dB}_t$ and a diffusion $\hat{Y}_t$ with f^*, g^* replaced with our model $\hat{f}, \hat{g}$ trained within error ε.

Applying curvature comparison theorem [Rauch, '51] to the coupled diffusions, we bound their Wasserstein distance $W_2(\hat{Y}_t, Y_t)$:

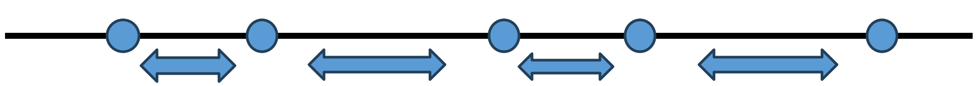
$$W_2(\hat{Y}_t, Y_t) \leq \left(W_2(\hat{Y}_0, Y_0) + \varepsilon\right)e^{ct}, \qquad \text{if } f^*, g^* \text{ c-Lipschitz on all of } \mathcal{M}$$

Projection φ is not in general Lipschitz! (e.g., on $\mathrm{SO}(n), \mathrm{U(n)}$, $\varphi(Z)$ has singularities at points where eigenvalue gaps of $Z + Z^*$ vanish)

Average-case L-Lipschitzness (stated here for $\mathrm{U}(n)$): There exists subset $\Omega_t \subseteq \mathbb{R}^m$ s.t.

- *Lispchitzness*: first two derivatives of φ have operator norms $\leq L$ on all of Ω_t
- *Symmetry*: indicator $\mathbf{1}_{\Omega_t}(Z)$ depends only on the eigenvalues of $Z + Z^*$
- *High probability set*: Euclidean Brownian motion Z_t remains in Ω_t w.h,p. for all $\mathrm{t} \in [0, T]$.

Use eigengap bounds from random matrix theory that say eigenvalues of matrix BM repel w.h.p. (e.g. [Anderson, Guionnet, Zeitouni '10], [M-V '24]) to show φ is average-case L-Lipschitz, $L = O(d^2)$



Average-case Lipschitzness of φ implies f^*, g^* are $\mathrm{poly(d)}$-Lipschitz ***on all of $\mathcal{M}$***:

$$f^*(U, t) \propto \int [\nabla\varphi(U\Lambda U^*)^\top \nabla \log \mathrm{p_{T-t}}(U\Lambda U^*) + \cdots]\, \mathbf{1}_{\Omega_t}(U\Lambda U^*)\mathrm{d}\Lambda$$

Symmetry implies integrand depends only on Λ, not U, "smoothing over" singularities of φ

Empirical results

Training runtime

When using same network architecture and GPU hardware for all models,

- Our training time is significantly faster than previous manifold diffusions on $\mathrm{U}(n)$, $\mathrm{SO}(n)$

Per-iteration training time on $\mathrm{U}(n)$ (seconds)

Method	$n = 5$ $(d = 10)$	$n = 50$ $(d = 1225)$
Euclidean	$0.19 \pm .01$	$0.21 \pm .01$
RSGM	$1.22 \pm .08$	$11.55 \pm .31$
TDM	$1.07 \pm .06$	$9.43 \pm .23$
This paper	$\mathbf{0.36 \pm .00}$	$\mathbf{0.60 \pm .01}$

- Runtime on torus, sphere, $\mathrm{U}(n)$, and $\mathrm{SO}(n)$ remains within small constant factor of Euclidean diffusion's, nearly closing gap in training runtime with Euclidean diffusion.

Sampling accuracy

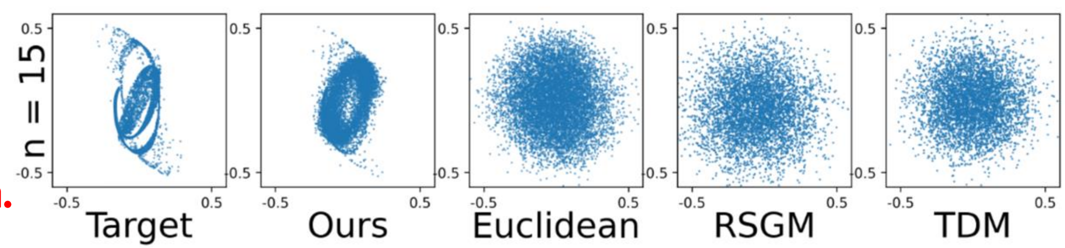
When trained on quantum operator datasets on $\mathrm{U}(n)$ and wrapped Gaussian datasets on the torus and $\mathrm{SO}(n)$,

- Except in very low dimensions, our model improves on the quality of generated samples over previous Euclidean and manifold diffusions

- Magnitude of runtime and quality improvements increases with dimension.

C2ST sample quality scores, and generated samples, on $\mathrm{U}(n)$

Method	$n = 5$ $(d = 10)$	$n = 12$ $(d = 66)$	$n = 15$ $(d = 105)$
Euclidean	$\mathbf{.75 \pm .04}$	$.97 \pm .04$	$1.00 \pm .01$
RSGM	$.92 \pm .04$	$1.00 \pm .02$	$1.00 \pm .01$
TDM	$.91 \pm .02$	$1.00 \pm .01$	$1.00 \pm .01$
This paper	$0.80 \pm .04$	$\mathbf{.88 \pm .05}$	$\mathbf{.90 \pm .04}$



Conclusion

Introduced new diffusion on symmetric manifolds, with spatially-varying covariance

Key idea: covariance allows diffusion to be projection from Euclidean space onto manifolds with non-zero curvature, bypassing computation of manifold's heat kernel

Training runtime: trains in near-linear ($O(d^{1.17})$) arithmetic operations + $O(1)$ model gradient evaluations per iteration

- improves on runtime of previous manifold diffusions by $\exp(d)$ arith. ops. or $O(d)$ *gradient evaluations* on symmetric manifolds such as $\mathrm{SO}(n), \mathrm{U}(n)$
- nearly closes gap with per-iteration runtime of Euclidean diffusions

Sampling guarantees: Manifold symmetries ensure reverse diffusion satisfies an "average-case" Lipschitz condition, ensuring accuracy & efficiency of sampling algorithm

Improves on guarantees of previous manifold diffusions which may not be polynomial-in-d

Empirical results: Outperforms prior methods in training runtime and sample quality, on datasets on torus, $\mathrm{SO}(n), \mathrm{U}(n)$. Magnitude of improvement increases with dimension.

Open problem: Can one extend our framework to more general manifolds?

Open problem: How to further tighten our theoretical *sampling accuracy* guarantees to generalize Euclidean diffusion guarantees?

Thanks!