PERMUTATION-SYMMETRIZED DIFFUSION FOR UNCONDITIONAL MOLECULAR GENERATION

Gyeonghoon Ko, Juho Lee
Korea Advanced Institute of Science and Technology
{kog, juholee}@kaist.ac.kr

ABSTRACT

Permutation invariance is fundamental in molecular point-cloud generation, yet most diffusion models enforce it indirectly via permutation-equivariant networks on an ordered space. We propose to model diffusion directly on the quotient manifold $\tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$, where all atom permutations are identified. We show that the heat kernel on $\tilde{\mathcal{X}}$ admits an explicit expression as a sum of Euclidean heat kernels over permutations, which clarifies how diffusion on the quotient differs from ordered-particle diffusion. Training requires a permutation-symmetrized score involving an intractable sum over S_N; we derive an expectation form over a posterior on permutations and approximate it using MCMC in permutation space. We evaluate on unconditional 3D molecule generation on QM9 under the EQGAT-Diff protocol, using SemlaFlow-style backbone and treating all variables continuously. The results demonstrate that quotient-based permutation symmetrization is practical and yields competitive generation quality with improved efficiency.

1 INTRODUCTION

Diffusion models provide a flexible framework for generative modeling and have recently achieved strong performance for 3D molecular generation when combined with equivariant architectures (Hoogeboom et al., 2022; Dunn & Koes, 2024; Vignac et al., 2023; Le et al., 2023). A key symmetry of molecules is permutation invariance: relabeling atoms does not change the underlying structure. Most diffusion approaches address this by representing molecules in an ordered space $\mathcal{X} = \mathbb{R}^{d\times N}$ and enforcing permutation *equivariance* of both the transition kernels and the score network (Zaheer et al., 2017). This guarantees that permuting the input permutes the output in the same way, but the diffusion trajectory still evolves in an ordered representation that implicitly tracks particle identities.

We instead formulate diffusion directly on the quotient space $\tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$, where all permutations are identified. We show that the heat kernel on $\tilde{\mathcal{X}}$ admits an explicit expression as a sum of Euclidean heat kernels over permutations. This characterization highlights a qualitatively different behavior: during noising and denoising, particle identities may be effectively exchanged. The resulting permutation-symmetrized score can be written as an expectation under a posterior distribution over permutations, which we approximate via MCMC. Related symmetrized diffusion ideas were introduced in Anonymous (2026), but their approach additionally requires training a physics-informed neural network (PINN). We evaluate on unconditional QM9 generation (Ramakrishnan et al., 2014) under the EQGAT-Diff framework (Le et al., 2023), using a lightweight SemlaFlow-style backbone (Irwin et al., 2024) and treating all variables continuously.

2 BACKGROUND

2.1 DIFFUSION MODELS ON RIEMANNIAN MANIFOLD

Consider an ambient space $\mathcal{X}$ in which the data lie. A *noising process* or *forward process* is a stochastic process $\{\boldsymbol{x}^{(t)}\}_{t=0}^{T}$ starting from $\boldsymbol{x}^{(0)} \sim p_0 = p_{\text{data}}$, usually defined by a stochastic differential equation (SDE) adding noises, with transition kernel $p_{t|s}(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(s)})$ for $0 \le s < t \le T$ and marginal distribution $p_t(\boldsymbol{x}^{(t)}) = \int p_{t|0}(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(0)})p_{\text{data}}(\boldsymbol{x}^{(0)})d\boldsymbol{x}^{(0)}$. A *denoising process* or *reverse*

process is a stochastic process designed to reverse the noising process, starting from the noise distribution $q_T(\boldsymbol{x}^{(T)}) \approx p_T$ at time T and aiming to recover the original data distribution $p_{\text{data}}(\boldsymbol{x}^{(0)})$. The reverse process is also often characterized by a (reverse-time) SDE, with a learnable transition kernel $q_{\theta,s|t}(\boldsymbol{x}^{(s)}|\boldsymbol{x}^{(t)})$ for $s < t$, parametrized by a model parameter θ.

For $\mathcal{X} = \mathbb{R}^d$, we use the Ornstein–Uhlenbeck (OU) forward SDE (Song et al., 2020)

$$d\boldsymbol{x}^{(t)} = -\tfrac{1}{2}\boldsymbol{x}^{(t)}dt + d\boldsymbol{w}^{(t)}, \tag{1}$$

with closed-form Gaussian transitions $p_{t|s}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}_{t|s}(\boldsymbol{x}), \boldsymbol{\Sigma}_{t|s})$ and marginals p_t. The associated reverse-time SDE is

$$d\boldsymbol{x}^{(t)} = \Big[-\tfrac{1}{2}\boldsymbol{x}^{(t)} - \nabla_{\boldsymbol{x}^{(t)}} \log p_t(\boldsymbol{x}^{(t)}) \Big] dt + d\boldsymbol{w}^{(t)}. \tag{2}$$

We learn the score with a network $\boldsymbol{s}_\theta(\boldsymbol{x}, t) \approx \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})$ via denoising score matching,

$$\theta^* = \arg\min_\theta \mathbb{E}_{t,\boldsymbol{x}^{(0)}} \left\| \boldsymbol{s}_\theta(\boldsymbol{x}^{(t)}, t) - \nabla_{\boldsymbol{x}^{(t)}} \log p_t(\boldsymbol{x}^{(t)} \mid \boldsymbol{x}^{(0)}) \right\|^2, \tag{3}$$

using the closed-form conditional score of the OU kernel. In practice we may use time-inhomogeneous variants and loss reweighting.

The noising and denoising processes can also be defined well in a non-Euclidean manifold, if the manifold $\mathcal{X}$ is granted with a Riemannian metric. The Wiener process $\boldsymbol{w}^{(t)}$ is replaced by its Riemannian version $\boldsymbol{w}_{\mathcal{X}}^{(t)}$, derived from the Laplace-Beltrami operator $\Delta^{\mathcal{X}}$ of $\mathcal{X}$ as an infinitesimal generator. Similar to the Euclidean case, one can define a forward SDE and its reverse

$$d\boldsymbol{x}^{(t)} = b(\boldsymbol{x}^{(t)})dt + d\boldsymbol{w}_{\mathcal{X}}^{(t)}, \quad d\boldsymbol{x}^{(t)} = -[b(\boldsymbol{x}^{(t)}) + \nabla_{\boldsymbol{x}^{(t)}} \log p_t(\boldsymbol{x}^{(t)})]dt + d\boldsymbol{w}_{\mathcal{X}}^{(t)} \tag{4}$$

where $b(\boldsymbol{x})$ is a tangent vector field on $\mathcal{X}$. However, unlike the Euclidean case, the closed-form equations for transition kernels and marginal densities are unknown in general, which becomes one of the major challenges of Riemannian diffusion models.

For later use, we also mention that having closed-form expressions for forward transition kernels is possible since the heat kernel $K(t, \boldsymbol{x}, \boldsymbol{y})$, i.e. the function satisfying $\frac{\partial}{\partial t} K = \Delta_{\boldsymbol{y}} K$ with $\lim_{t\to 0} K = \delta(\boldsymbol{x} - \boldsymbol{y})$, is given by the Gaussian

$$K(t, \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{(4\pi t)^{d/2}} \exp(-\frac{|\boldsymbol{x} - \boldsymbol{y}|^2}{4t}). \tag{5}$$

For the Riemannian manifold $\mathcal{X}$, the heat kernel $K^{\mathcal{X}}$ is defined similarly with the Laplacian Δ replaced by the Laplace-Beltrami operator $\Delta^{\mathcal{X}}$. The exact form of the heat kernel is unknown in general, and understanding them is a crucial component in modern mathematics.

3 METHOD

3.1 RETHINKING PERMUTATION SYMMETRY IN DIFFUSION MODELS

In molecule generation tasks, we usually consider a diffusion process on the space of point clouds $\mathcal{X} = \mathbb{R}^{d\times N} = \{\boldsymbol{x} = (x_1, \cdots, x_N) \mid x_i \in \mathbb{R}^N \text{ for } i = 1, \cdots, N\}$, with the space $\mathbb{R}^d$ consisting of coordinates and atom features, e.g. atom types. The molecules $\boldsymbol{x} = (x_1, \cdots, x_N)$ have a permutation symmetry defined by the permutation group S_N which acts by permuting the points as $\sigma(\boldsymbol{x}) = (x_{\sigma^{-1}(1)}, \cdots, x_{\sigma^{-1}(N)})$. A common recipe for incorporating the permutation symmetry is designing a permutation-equivariant transition kernel:

$$p_{t|s}(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(s)}) = p_{t|s}(\sigma(\boldsymbol{x}^{(t)})|\sigma(\boldsymbol{x}^{(s)})), \quad q_{\theta,s|t}(\boldsymbol{x}^{(s)}|\boldsymbol{x}^{(t)}) = q_{\theta,s|t}(\sigma(\boldsymbol{x}^{(s)})|\sigma(\boldsymbol{x}^{(t)})) \tag{6}$$

for all $s < t$ and $\sigma \in S_N$. In this formulation, the equivariance is achieved by using a permutation equivariant diffusion kernel (forward) and a permutation equivariant architecture (reverse). This formulation implies that if the input is altered by a permutation, then the output is altered accordingly by the same permutation, i.e. **the process behaves identically** under permutation.

On the other hand, we can think of a stronger notion of utilizing the symmetry. Instead of building an equivariant diffusion on $\mathcal{X} = \mathbb{R}^{d\times N}$, we model the diffusion process on the quotient manifold

$\tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$. Let $\pi : \mathcal{X} = \mathbb{R}^{d\times N} \to \tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$ denote the quotient map, and let's write $\tilde{\boldsymbol{x}} = \pi(\boldsymbol{x})$ for brevity. In this formulation, we have simply $\sigma(\tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{x}}$, i.e. **the elements are identical** under permutation. The manifold $\tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$ inherits the Euclidean metric, and hence has a well-defined heat kernel. The exact form of the heat kernel is given by Theorem 3.1:

Theorem 3.1. *The heat kernel $K^{\tilde{\mathcal{X}}}(t, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ associated to the quotient manifold $\tilde{\mathcal{X}} = \mathbb{R}^{d\times N}/S_N$ is given by, for $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \tilde{\mathcal{X}}$,*

$$K^{\tilde{\mathcal{X}}}(t, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \frac{1}{(4\pi t)^{dN/2}} \sum_{\sigma\in S_N} \exp(-\frac{|\boldsymbol{x} - \sigma(\boldsymbol{y})|^2}{4t}) = \frac{1}{(4\pi t)^{dN/2}} \sum_{\sigma\in S_N} \prod_{i=1}^{N} \exp(-\frac{(x_i - y_{\sigma^{-1}(i)})^2}{4t}),$$

i.e. the heat kernel is simply the sum of the Euclidean heat kernels over all permutations.

The proof of Theorem 3.1 is postponed to Section A.1, but it just comes from the fact that permutations are isometries of Euclidean spaces.

Now, with the heat kernel identified, it looks much clearer how the diffusion processes in the quotient manifold behave. The classic diffusion process in the Euclidean space inherently assumes that first particle x_1 diffuses to the first one y_1, second one x_2 diffuses to y_2, and so on. However, our diffusion opens the possibilities that first particle x_1 can be diffused to any y_i for $i \in \{1, \cdots, N\}$, and so on, over all the configurations $\sigma \in S_N$. Intuitively, in this diffusion process, the identities of particles are all shared, so their identities can be swapped with each other while noising or denoising. This behavior resembles the behaviors of *bosons* in particle physics (Feynman & Mechanics, 1979).

3.2 SDEs on the quotient manifold

Let's go back to the OU process as defined in Equation (1) and Equation (2). We'll consider the pushforward of the OU process by the quotient map $\pi : \mathcal{X} \mapsto \tilde{\mathcal{X}}$. Although working with pushforward process in arbitrary Riemannian manifolds would involve extremely complicated computations in general, the computations are significantly simplified in our case, because the OU process is invariant under permutations and permutations are isometries of the Euclidean space.

The forward and reverse equations of the pushforward of the OU process are given by

$$d\tilde{\boldsymbol{x}}^{(t)} = -\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)}dt + d\tilde{\boldsymbol{w}}^{(t)}, \quad d\tilde{\boldsymbol{x}}^{(t)} = \left[-\frac{1}{2}\tilde{\boldsymbol{x}}^{(t)} - \nabla_{\tilde{\boldsymbol{x}}^{(t)}} \log \tilde{p}_t(\tilde{\boldsymbol{x}}^{(t)})\right] dt + d\tilde{\boldsymbol{w}}^{(t)}. \tag{7}$$

where we use $\tilde{p}_t$ and $\tilde{p}_{s|t}$ for marginals and transition kernels. Rigorously speaking, the term $\tilde{\boldsymbol{x}}^{(t)}$ multiplied to dt is a vector field $\tilde{\boldsymbol{x}}^{(t)} = \tilde{V}(\tilde{\boldsymbol{x}}^{(t)})$ on $\tilde{\mathcal{X}}$ defined by quotient-mapping the vector field on Euclidean space $V(\boldsymbol{x}) = \boldsymbol{x}$, i.e. $\tilde{V} = \pi \circ V \circ \pi^{-1}$, well-defined since the vector field V is invariant under permutation. Similarly, the gradient $\nabla_{\tilde{\boldsymbol{x}}^{(t)}}$ is in fact the gradient operator on the Riemannian manifold $\tilde{\mathcal{X}}$, but it locally behaves identical to the Euclidean gradient since the permuatations are isometries.

The transition kernels $\tilde{p}_{t|s}$ and marginals $\tilde{p}_t$ can simply yielded by summing the original transition kernels $p_{t|s}$ and marginals p_t over all permutations:

$$\tilde{p}_{t|s}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) = \sum_{\sigma\in S_N} p_{t|s}(\sigma(\boldsymbol{y})|\boldsymbol{x}), \quad \tilde{p}_t(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) = \sum_{\sigma\in S_N} p_t(\sigma(\boldsymbol{y})|\boldsymbol{x}). \tag{8}$$

3.3 Training permutation symmetrized diffusion

The transition kernel p_t derived in Theorem 3.1 contains summmation over the permutation group S_N, hence naive computation would require at least $O(N!)$ computations. Instead, we look at the score function $\nabla \log \tilde{p}_t$, which will be the target function of diffusion loss. Letting $I(\sigma) = -\frac{|\boldsymbol{x}-\sigma(\boldsymbol{y})|^2}{4t}$, we have

$$\nabla_{\tilde{\boldsymbol{y}}} \log p_t(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) = \nabla_{\boldsymbol{y}} \log \sum_{\sigma\in S_N} \exp(I(\sigma)) = \sum_{\sigma\in S_N} \frac{\exp(I(\sigma))}{\sum_{\sigma'\in S_N} \exp(I(\sigma'))} \nabla_{\boldsymbol{y}} I(\sigma).$$

Table 1: Comparison of QM9 unconditional generation metrics (%).

Method	Atom-Stab	Mol-Stab	Valid	Uniq	Nov
EQGAT-Diff	99.87 ± 0.02	98.24 ± 0.11	98.58 ± 0.13	$\mathbf{100 \pm 0.00}$	61.62 ± 1.55
Ours	$\mathbf{99.89 \pm 0.02}$	$\mathbf{98.48 \pm 0.27}$	$\mathbf{98.74 \pm 0.25}$	$\mathbf{100 \pm 0.00}$	$\mathbf{67.72 \pm 1.55}$

Thus, if we consider a distribution $\mathcal{S}$ on the permutation group S_N whose probability mass function $q(\sigma)$ is proportional to $\exp I(\sigma)$, we have

$$\nabla_{\boldsymbol{y}} \log p_t(\tilde{\boldsymbol{x}}|\tilde{\boldsymbol{y}}) = \mathbb{E}_{\mathcal{S}}[\nabla_{\boldsymbol{y}} I(\sigma)]. \tag{9}$$

Note that $\nabla_{\boldsymbol{y}} I(\sigma) = \frac{\boldsymbol{x}-\sigma(\boldsymbol{y})}{2t}$ would be the score function for the original Euclidean diffusion model, associated with the original data $\boldsymbol{x}$ and its noised version $\sigma(\boldsymbol{y})$. We can interpret the probability $q(\sigma) = q(\sigma|\boldsymbol{x}, \boldsymbol{y})$ as the posterior of the permutation σ given the original and noised data, and the target function is the expected value of the original score function over the posterior $q(\sigma)$. In fact, from the perspective of variational inference, learning towards this target function can be seen as minimizing the lower bound for the evidence lower bound (ELBO) of this diffusion model. This argument is explained in detail in Section A.2.

Sampling permutations from $\mathcal{S}$ can be done using Markov chain Monte Carlo (MCMC). Define the cost matrix $C = (C_{ij})$ with $C_{ij} = -\frac{(x_i - y_j)^2}{4t}$. We use an MCMC starting from $\sigma_0 = \mathrm{id} \in S_N$ and the proposals yielded by swapping entries of i, j, for $i \in \{1, \cdots, N\}$ sampled uniformly at random and j sampled from distribution proportional to $\exp(C_{ij})$. Once the permutations $\sigma_1, \cdots, \sigma_K \sim \mathcal{S}$ are sampled, penalizing the model s_θ towards $\mathbb{E}_{\mathcal{S}}[\nabla_{\boldsymbol{y}} I(\sigma)] \approx \sum_{k=1}^{K} \nabla_{\boldsymbol{y}} I(\sigma_k)$ gives an unbiased estimate of the gradient:

$$\nabla_\theta \left\| s_\theta - \mathbb{E}_{\mathcal{S}}[\nabla_{\boldsymbol{y}} I(\sigma)] \right\|^2 = \mathbb{E}_{\sigma_1, \cdots, \sigma_K \sim \mathcal{S}} \left[\nabla_\theta \left\| s_\theta - \sum_{k=1}^{K} \nabla_{\boldsymbol{y}} I(\sigma) \right\|^2 \right] \tag{10}$$

since derivatives of quadratic functions are linear.

4 Experiment

We evaluate on unconditional 3D molecule generation on QM9 (Ramakrishnan et al., 2014). Our experimental pipeline follows EQGAT-Diff (Le et al., 2023) (data split, training schedule, sampling procedure, and evaluation), but we replace its backbone with the lighter SemlaFlow architecture (Irwin et al., 2024) that does not carry edge features for efficiency. Throughout, we treat all variables as continuous (i.e., no discrete diffusion over atom types). When computing the permutations using MCMC, edge features are ignored since they will require heavier computation.

Each molecule is represented as a point cloud with 3D coordinates and continuous node features. We train on the QM9 training split and generate samples from the learned reverse-time dynamics following EQGAT-Diff. We compare primarily against EQGAT-Diff under its default QM9 setup. All methods are evaluated using the same split and metric implementation as in EQGAT-Diff and its prior works (Vignac et al., 2023). We report the standard QM9 generation metrics used in EQGAT-Diff, including atom stability, molecule stability, validity, uniqueness, and novelty. The experiments are performed five times, and their averages and standard deviations are reported in Table 1. Our model wins all the stability/validity metrics by a small margin, and achieves much higher novelty.

5 Conclusion

We presented a quotient-manifold formulation of permutation symmetry for diffusion models on point clouds, focusing on $\tilde{\mathcal{X}} = \mathbb{R}^{d \times N}/S_N$. The resulting heat kernel admits a simple permutation-sum expression, and the corresponding symmetrized score can be estimated efficiently by MCMC over permutations. On QM9, following EQGAT-Diff framework (Le et al., 2023) with a lighter SemlaFlow architecture (Irwin et al., 2024), we show the approach is practical in a fully continuous setting. Future work includes improving permutation-sampler efficiency and extending the same quotient perspective to other symmetry quotients in geometric generative modeling.

REFERENCES

Anonymous. Riemannian diffusion models on general manifolds via physics-informed neural networks. Unpublished manuscript (under anonymous review)., 2026.

Ian Dunn and David Ryan Koes. Mixed continuous and categorical flow matching for 3d de novo molecule generation. *ArXiv*, pp. arXiv–2404, 2024.

RP Feynman and AR Hibbs Quantum Mechanics. Path integrals. *Lecture Notes Phys*, 106:64, 1979.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Ross Irwin, Alessandro Tibo, Jon Paul Janet, and Simon Olsson. Semlaflow–efficient 3d molecular generation with latent attention and equivariant flow matching. *arXiv preprint arXiv:2406.07266*, 2024.

Tuan Le, Julian Cremer, Frank Noe, Djork-Arné Clevert, and Kristof Schütt. Navigating the design space of equivariant diffusion-based generative models for de novo 3d molecule generation. *arXiv preprint arXiv:2309.17296*, 2023.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Clement Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 560–576. Springer, 2023.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

A MATHEMATICAL DETAILS

A.1 PROOF OF THEOREM 3.1

Proof. Let $K^{\mathcal{X}}(t, \boldsymbol{x}, \boldsymbol{y})$ be the Euclidean heat kernel on $\mathcal{X} \simeq \mathbb{R}^{dN}$:

$$K^{\mathcal{X}}(t, \boldsymbol{x}, \boldsymbol{y}) = \frac{1}{(4\pi t)^{\frac{dN}{2}}} \exp\Big(- \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{4t} \Big), \qquad t > 0.$$

Define, for representatives $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ of $\tilde{\boldsymbol{x}} = \pi(\boldsymbol{x})$ and $\tilde{\boldsymbol{y}} = \pi(\boldsymbol{y})$,

$$\widehat{K}(t, \boldsymbol{x}, \boldsymbol{y}) \ := \sum_{\sigma \in S_N} K^{\mathcal{X}}(t, \boldsymbol{x}, \sigma(\boldsymbol{y})).$$

Step 1: $\widehat{K}$ descends to the quotient. Each permutation $\tau \in S_N$ is an isometry of $\mathcal{X}$, hence $\|\tau(\boldsymbol{x}) - \tau(\boldsymbol{y})\| = \|\boldsymbol{x} - \boldsymbol{y}\|$ and therefore $K^{\mathcal{X}}(t, \tau(\boldsymbol{x}), \tau(\boldsymbol{y})) = K^{\mathcal{X}}(t, \boldsymbol{x}, \boldsymbol{y})$. Using this and the change of index $\sigma' = \tau^{-1}\sigma$, we get

$$\widehat{K}(t, \tau(\boldsymbol{x}), \boldsymbol{y}) = \sum_{\sigma \in S_N} K^{\mathcal{X}}(t, \tau(\boldsymbol{x}), \sigma(\boldsymbol{y})) = \sum_{\sigma' \in S_N} K^{\mathcal{X}}(t, \boldsymbol{x}, \sigma'(\boldsymbol{y})) = \widehat{K}(t, \boldsymbol{x}, \boldsymbol{y}),$$

and similarly $\widehat{K}(t, \boldsymbol{x}, \tau(\boldsymbol{y})) = \widehat{K}(t, \boldsymbol{x}, \boldsymbol{y})$. Hence $\widehat{K}(t, \boldsymbol{x}, \boldsymbol{y})$ depends only on the orbits $\tilde{\boldsymbol{x}} = \pi(\boldsymbol{x})$ and $\tilde{\boldsymbol{y}} = \pi(\boldsymbol{y})$, so we may define

$$K^{\tilde{\mathcal{X}}}(t, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) := \widehat{K}(t, \boldsymbol{x}, \boldsymbol{y}).$$

Step 2: heat equation. For each fixed σ, the function $\boldsymbol{x} \mapsto K^{\mathcal{X}}(t, \boldsymbol{x}, \sigma(\boldsymbol{y}))$ solves $(\partial_t - \Delta_{\mathcal{X}})u = 0$ for $t > 0$, hence so does the finite sum $\widehat{K}(t, \boldsymbol{x}, \boldsymbol{y})$. Since $\widehat{K}$ is S_N-invariant in $\boldsymbol{x}$, it corresponds to a function on $\tilde{\mathcal{X}}$, and thus $K^{\tilde{\mathcal{X}}}(\cdot, \cdot, \tilde{\boldsymbol{y}})$ solves the heat equation on the quotient.

Step 3: initial condition. Let $\tilde{f}$ be a smooth test function on $\tilde{\mathcal{X}}$ and let $f := \tilde{f} \circ \pi$ be its S_N-invariant lift to $\mathcal{X}$. With respect to the Riemannian volume on $\tilde{\mathcal{X}}$, the heat semigroup defined by $K^{\tilde{\mathcal{X}}}$ satisfies

$$\begin{aligned}\int_{\tilde{\mathcal{X}}} K^{\tilde{\mathcal{X}}}(t, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})\, \tilde{f}(\tilde{\boldsymbol{y}})\, d\tilde{\boldsymbol{y}} &= \frac{1}{|S_N|} \int_{\mathcal{X}} \Big(\sum_{\sigma \in S_N} K^{\mathcal{X}}(t, \boldsymbol{x}, \sigma(\boldsymbol{y})) \Big) f(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int_{\mathcal{X}} K^{\mathcal{X}}(t, \boldsymbol{x}, \boldsymbol{z})\, f(\boldsymbol{z})\, d\boldsymbol{z},\end{aligned}$$

where the last equality uses the change of variables $\boldsymbol{z} = \sigma(\boldsymbol{y})$ and the invariance $f \circ \sigma = f$ (each term in the sum is identical, cancelling the factor $|S_N|$). As $t \downarrow 0$, the right-hand side converges to $f(\boldsymbol{x}) = \tilde{f}(\tilde{\boldsymbol{x}})$ by the standard delta initial condition for the Euclidean heat kernel. Thus $K^{\tilde{\mathcal{X}}}$ has the correct initial condition on the quotient.

Conclusion. We have shown that the expression in Theorem 3.1 defines a kernel on $\tilde{\mathcal{X}}$ which solves the heat equation for $t > 0$ and converges to the delta distribution at $t = 0$. By uniqueness of the heat kernel, it must equal the heat kernel on $\tilde{\mathcal{X}}$. □

A.2 BAYESIAN / VARIATIONAL INTERPRETATION OF PERMUTATION SYMMETRIZATION

Posterior over permutations. From Theorem 3.1, the symmetrized transition density on the quotient satisfies

$$\tilde{p}_t(\tilde{\boldsymbol{y}} \mid \tilde{\boldsymbol{x}}) \ \propto \sum_{\sigma \in S_N} \exp\big(I(\sigma)\big), \qquad I(\sigma) := -\frac{\|\boldsymbol{x} - \sigma(\boldsymbol{y})\|^2}{4t}.$$

Introduce a latent permutation σ with a uniform prior $p(\sigma) = 1/|S_N|$ and define the (joint, unnormalized) model

$$\tilde{p}_t(\tilde{\boldsymbol{y}}, \sigma \mid \tilde{\boldsymbol{x}}) \ \propto \ \exp(I(\sigma)).$$

Then the Bayes posterior is exactly the "soft assignment" used in the main text:

$$p(\sigma \mid \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \frac{\exp(I(\sigma))}{\sum_{\sigma' \in S_N} \exp(I(\sigma'))} =: q(\sigma \mid \boldsymbol{x}, \boldsymbol{y}).$$

Consequently (Fisher's identity / mixture-score identity),

$$\nabla_{\boldsymbol{y}} \log \tilde{p}_t(\tilde{\boldsymbol{y}} \mid \tilde{\boldsymbol{x}}) = \mathbb{E}_{\sigma \sim q(\sigma \mid \boldsymbol{x}, \boldsymbol{y})} \big[\nabla_{\boldsymbol{y}} I(\sigma)\big],$$

which is the permutation-symmetrized score target.

ELBO view. For any distribution $r(\sigma)$ on S_N, Jensen's inequality gives

$$\begin{aligned} \log \tilde{p}_t(\tilde{\boldsymbol{y}} \mid \tilde{\boldsymbol{x}}) &= \log \sum_{\sigma} \exp(I(\sigma)) + \text{const} \\ &= \log \mathbb{E}_{\sigma \sim r} \left[\frac{\exp(I(\sigma))}{r(\sigma)} \right] + \text{const} \\ &\geq \mathbb{E}_r[I(\sigma)] + H(r) + \text{const} \ =: \ \mathrm{ELBO}(r), \end{aligned}$$

and one can write the exact decomposition

$$\log \tilde{p}_t(\tilde{\boldsymbol{y}} \mid \tilde{\boldsymbol{x}}) = \mathrm{ELBO}(r) + D_{\mathrm{KL}}\big(r(\sigma) \,\|\, q(\sigma \mid \boldsymbol{x}, \boldsymbol{y})\big).$$

Thus maximizing the evidence on the quotient is equivalent to maximizing this ELBO, and the optimal variational distribution is the posterior $r^\star = q(\sigma \mid \boldsymbol{x}, \boldsymbol{y})$. Our training objective uses the corresponding posterior expectation of the Euclidean score; in practice we approximate $\mathbb{E}_q[\cdot]$ by MCMC samples of σ.