STIEFEL FLOW MATCHING FOR MOMENT-CONSTRAINED STRUCTURE ELUCIDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular structure elucidation is a critical step in understanding chemical phenomena, with applications in identifying molecules in natural products, lab syntheses, forensic samples, and the interstellar medium. We consider the task of elucidating a molecule's 3D structure given only its molecular formula and moments of inertia, motivated by the ability of rotational spectroscopy to precisely measure these moments. While existing generative models can conditionally sample 3D structures with approximately correct moments, this soft conditioning fails to leverage the many digits of precision afforded by experimental rotational spectroscopy. To address this, we first show that the space of n-atom point clouds with a fixed set of moments of inertia is embedded in the Stiefel manifold $\mathrm{St}(n, 4)$. We then propose *Stiefel Flow Matching* as a generative model for elucidating 3D structure under exact moment constraints. Additionally, we learn simpler and shorter flows by finding approximate solutions for optimal transport on the Stiefel manifold. Empirically, satisfying moment constraints exactly allows Stiefel Flow Matching to achieve higher success rates and faster sampling than Euclidean diffusion models, even on high-dimensional manifolds corresponding to large molecules in the GEOM dataset.

1 INTRODUCTION

Elucidating the structure of unknown molecules is a ubiquitous task in chemistry, important for analyzing environmental samples (Moneta et al., 2023), identifying novel drugs (Sonstrom et al., 2023), and determining potential building blocks of life in the interstellar medium (McGuire et al., 2016). The challenge is to aggregate information from multiple sources of analytical data to unambiguously determine a molecule's structure. Rotational spectroscopy holds a unique capacity to provide precise measurements about a molecule's moments of inertia, which in turn has routinely provided the highest quality gas-phase 3D structures attainable from experiment (Domingos et al., 2020). As the moments of inertia are trivial to compute for a given 3D structure, structure elucidation with rotational spectroscopy usually proceeds by confirming whether a *known* structure's moments matches with experimentally measured moments (Lee & McCarthy, 2019; McCarthy et al., 2020). However, this approach is limited only to molecules whose structures are known *a priori*. This fundamentally limits applications related to identifying molecules that have not been catalogued, such as novel natural products, undiscovered molecules, and key reactive intermediate species that cannot be easily isolated (Womack et al., 2015).

To overcome this, we apply generative modeling to structure elucidation by inferring candidate 3D structures based on the moments and molecular formula *alone*. By themselves, moments of inertia only describe the spread of the molecule's mass in three directions. Going from moments to 3D structure ($3 \to 3n$ values) is therefore a heavily underconstrained inverse problem. Nevertheless, deep generative models such as diffusion (Ho et al., 2020; Song et al., 2020) and flow matching (Lipman et al., 2023; Liu et al., 2022) have shown promise in solving various inverse problems (Song et al., 2022; Chung et al., 2023; Song et al., 2023a). Previous work has trained a Euclidean diffusion model to sample 3D structures conditioned on a given set of moments of inertia (Cheng et al., 2024).

However, analytic formulas for the moments of inertia are sufficiently simple that we can define a feasible space where these constraints are always exactly satisfied. Such precise adherence to moments can potentially leverage the many digits of precision provided by rotational spectroscopy

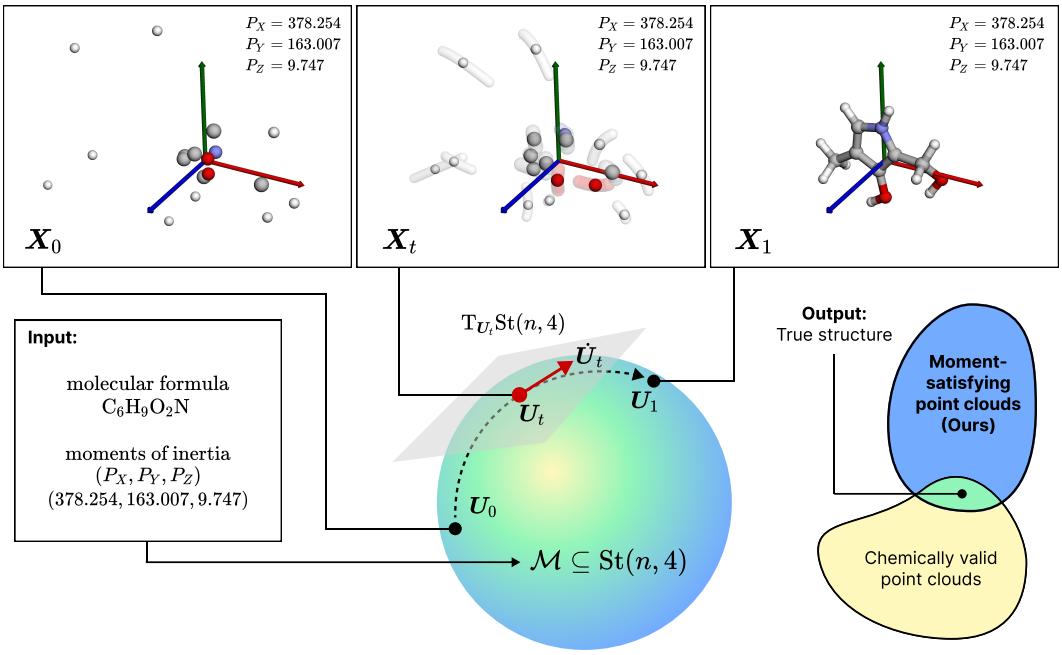

Figure 1: Stiefel Flow Matching learns to elucidate 3D molecular structure from moments and molecular formula alone by transforming uniform Stiefel noise $\boldsymbol{X}_0$ into *valid* molecular structures $\boldsymbol{X}_1$. Generative modelling on the Stiefel manifold $\mathrm{St}(n, 4)$ guarantees that samples always have the correct moments of inertia, which allows the network to focus only on generating chemically stable structures. Within the intersection of these spaces lies the true 3D structure.

(Shipman et al., 2011) in order to constrain the space of plausible structures. We first show that $\mathcal{M}$, the set of n-atom point clouds with fixed moments of inertia, is embedded in the Stiefel manifold $\mathrm{St}(n, 4)$. We then propose *Stiefel Flow Matching* as a generative model on the Stiefel manifold for solving the moment-constrained structure elucidation problem (Figure 1). Stiefel Flow Matching builds on ideas from Riemannian flow matching (Chen & Lipman, 2024) with equivariant optimal transport (Klein et al., 2023; Song et al., 2023c), which simplifies and shortens generation paths.

Concretely, our contributions are:

1. We propose the task of moment-constrained structure elucidation as a challenging generative modelling problem on the Stiefel manifold.

2. To solve this problem, we present *Stiefel Flow Matching*, a Riemannian flow matching approach. Furthermore, we formulate an objective for equivariant optimal transport on the Stiefel manifold, which obtains shorter and simpler flows.

3. Stiefel Flow Matching predicts 3D structure with greater success rate and lower cost than Euclidean diffusion models on both the QM9 (Ramakrishnan et al., 2014) and GEOM (Axelrod & Gomez-Bombarelli, 2022) datasets.

2 Background and Approach

We consider a 3D molecule as a point cloud of n atoms with atomic numbers $\boldsymbol{a} \in \mathbb{N}^n$ and 3D coordinates $\boldsymbol{X} = (x_i, y_i, z_i)_{i=1}^N \in \mathbb{R}^{n\times 3}$. We also refer to $\boldsymbol{X}$ as the molecule's *3D structure*. Molecules have translational, rotational, and permutation symmetry. However, as $\boldsymbol{a}$ and $\boldsymbol{X}$ are stored on a computer, they necessarily have a node ordering and orientation. Therefore, when we refer to *structure*, we really mean the equivalence class containing $\boldsymbol{X}$ under these symmetries.

2.1 PROBLEM STATEMENT

We are given a molecule's molecular formula and moments of inertia, and wish to infer the 3D structure of the molecule. The molecular formula provides us with the number of atoms n, atomic numbers $\boldsymbol{a} \in \mathbb{N}^n$, and atomic masses $\boldsymbol{m} \in \mathbb{R}^n$. The moments of inertia[1] (P_X, P_Y, P_Z) are three nonnegative numbers that summarize the mass distribution of the point cloud. They are defined as the eigenvalues of the *planar dyadic* $\boldsymbol{P} \in \mathbb{R}^{3\times 3}$ (Kraitchman, 1953), calculated from masses $\boldsymbol{m}$ and the 3D atomic coordinates $\boldsymbol{X} = (x_i, y_i, z_i)_{i=1}^N \in \mathbb{R}^{n\times 3}$ as

$$\boldsymbol{P} = \boldsymbol{X}^\top (\operatorname{diag} \boldsymbol{m}) \boldsymbol{X} = \sum_{i=1}^{n} \begin{pmatrix} m_i x_i^2 & m_i x_i y_i & m_i x_i z_i \\ m_i y_i x_i & m_i y_i^2 & m_i y_i z_i \\ m_i z_i x_i & m_i z_i y_i & m_i z_i^2 \end{pmatrix}. \tag{1}$$

This definition assumes a coordinate system whose origin is the weighted center of mass of the molecule, which gives the constraint $\boldsymbol{m}^\top \boldsymbol{X} = \boldsymbol{0}$. We assume that $P_X > P_Y > P_Z > 0$, which eliminates rare edge cases (Appendix A.2). Note that $\boldsymbol{P}$ is symmetric and positive semidefinite and can be thought of as the mass-weighted covariance matrix of the point cloud, as in principal component analysis (PCA).

Diagonalizing $\boldsymbol{P}$ yields three eigenvectors, referred to as the *principal axes of rotation*, which orient a molecule in a canonical representation up to sign-flips. The principal axes correspond to directions which "explain" the most molecular mass, in analogy to PCA. The *principal axis system* is then the coordinate system whose origin is the center-of-mass and whose axes are the principal axes of rotation. We fix our coordinate system to be the principal axis system *by construction*, which gives the following constraints on $\boldsymbol{X}$:

$$\begin{aligned} P_X &= \sum_{i=1}^{n} m_i x_i^2, \quad 0 = \sum_{i=1}^{n} m_i y_i z_i, \quad 0 = \sum_{i=1}^{n} m_i x_i, \\ P_Y &= \sum_{i=1}^{n} m_i y_i^2, \quad 0 = \sum_{i=1}^{n} m_i x_i z_i, \quad 0 = \sum_{i=1}^{n} m_i y_i, \\ P_Z &= \sum_{i=1}^{n} m_i z_i^2, \quad 0 = \sum_{i=1}^{n} m_i x_i y_i, \quad 0 = \sum_{i=1}^{n} m_i z_i. \end{aligned} \tag{2}$$

These constraints also canonicalize a 3D structure up to sign-flips of the axes (e.g. $x \mapsto -x$). The center-of-mass constraint removes translational degrees of freedom, while the off-diagonal constraints remove rotational degrees of freedom.

The goal of molecular identification from moments and molecular formula is to find all molecular structures $\boldsymbol{X}$ which are consistent with these constraints *and* are thermodynamically *stable*, i.e., local minima of potential energy surfaces. Structures which satisfy these criteria can then be compared to experimental measurements from rotational spectroscopy.

2.2 THE FEASIBLE SPACE OF MOMENT-CONSTRAINED STRUCTURES

The *Stiefel manifold* is the set of orthonormal $n \times p$ matrices, defined as

$$\operatorname{St}(n, p) = \{\boldsymbol{U} \in \mathbb{R}^{n\times p} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_p\}, \tag{3}$$

where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, and $n \geq p$. Elements of $\operatorname{St}(n, p)$ can be thought of as the first p columns of some n-dimensional (improper) rotation matrix, or can be thought of as orthonormal p-frames residing in n-dimensional space.

Moment-constrained structures $\boldsymbol{X}$ can then be mapped into $\operatorname{St}(n, 4)$ by scaling rows by masses and columns by moments. Letting $M = \sum_{i=1}^{n} m_i$ be the total mass, consider the following construction:

$$\boldsymbol{U} = \begin{pmatrix} \sqrt{\frac{m_1}{P_X}} x_1 & \sqrt{\frac{m_1}{P_Y}} y_1 & \sqrt{\frac{m_1}{P_Z}} z_1 & \sqrt{\frac{m_1}{M}} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\frac{m_n}{P_X}} x_n & \sqrt{\frac{m_n}{P_Y}} y_n & \sqrt{\frac{m_n}{P_Z}} z_n & \sqrt{\frac{m_n}{M}} \end{pmatrix} \in \mathbb{R}^{n\times 4}. \tag{4}$$

[1]This overview is a slight simplification of rotational spectroscopy. Full details are outlined in Appendix A.

It can be verified by inspection and comparison to the Equation 2 constraints that the columns of $\boldsymbol{U}$ are orthonormal. That is, $\boldsymbol{U} \in \mathrm{St}(n, 4)$. However, we cannot yet freely convert between $\boldsymbol{X}$ and $\boldsymbol{U}$: the last column of $\boldsymbol{U}$ is not free as it must equal the unit mass vector $\hat{\boldsymbol{m}} = \sqrt{\boldsymbol{m}/M}$ to satisfy the zero center-of-mass constraint. To convert an arbitrary $\boldsymbol{U} \in \mathrm{St}(n, 4)$ to an $\boldsymbol{X}$ satisfying all constraints, we first apply a rigid n-dimensional rotation $\boldsymbol{R}$ to $\boldsymbol{U}$ so that its last column is aligned to $\hat{\boldsymbol{m}}$ (Appendix B.5), before finally unscaling the rows and columns of $\boldsymbol{U}$.

The feasible space $\mathcal{M}$ of moment-constrained structures is therefore the subset of $\mathrm{St}(n, 4)$ whose last column is fixed to $\hat{\boldsymbol{m}}$, i.e.,

$$\mathcal{M} = \{\boldsymbol{U} \in \mathrm{St}(n, 4) \mid \boldsymbol{U}_{\cdot,4} = \hat{\boldsymbol{m}}\}. \tag{5}$$

In fact, we show in Appendix B.7 that $\mathcal{M}$ is a totally geodesic submanifold of $\mathrm{St}(n, 4)$, which means that geodesics between points in $\mathcal{M}$ stay in $\mathcal{M}$. The first three columns of elements in $\mathcal{M}$ form the intersection between $\mathrm{St}(n, 3)$ and the orthogonal complement to $\mathrm{span}(\hat{\boldsymbol{m}})$, which is in turn equivalent to $\mathrm{St}(n-1, 3)$. Hence, the dimension of $\mathcal{M}$ is $3n - 9$, which corresponds to removing 3 translational, 3 rotational, and 3 moment degrees of freedom. Going forward, we assume that the molecule of interest contains $n \geq 5$ atoms, so that we always deal with Stiefel manifolds of rectangular matrices.

2.3 Navigating the Stiefel Manifold

The Stiefel manifold provides rich structure for navigating the feasible space of molecular structures. As a *manifold*, it is locally Euclidean but globally curved. This means that every point $\boldsymbol{U} \in \mathrm{St}(n, p)$ is attached a vector space called its *tangent space* $\mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, p)$. For the Stiefel manifold, these tangent spaces are given as

$$\mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, p) = \{\Delta \in \mathbb{R}^{n \times p} \mid \boldsymbol{U}^\top \Delta + \Delta^\top \boldsymbol{U} = \boldsymbol{0}\}. \tag{6}$$

Then, equipping every tangent space with an inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{U}}$ turns $\mathrm{St}(n, p)$ into a *Riemannian manifold*, giving rise to notions of angles and distances. The collection of inner products for each tangent space is called the *Riemannian metric*. One such metric for the Stiefel manifold is the canonical metric (Edelman et al., 1998),

$$\langle \Delta, \tilde{\Delta} \rangle_{\boldsymbol{U}} = \mathrm{trace}\, \Delta^\top \left(\boldsymbol{I}_n - \tfrac{1}{2}\boldsymbol{U}\boldsymbol{U}^\top\right) \tilde{\Delta}, \tag{7}$$

which we exclusively use for this work. The canonical metric induces a norm $||\Delta||_{\boldsymbol{U}} = \sqrt{\langle \Delta, \Delta \rangle_{\boldsymbol{U}}}$ on each tangent space, allowing us to define the length of a curve $\gamma \colon [0, 1] \to \mathrm{St}(n, p)$ as $L(\gamma) = \int_0^1 ||\dot{\gamma}(t)||_{\gamma(t)} dt$. Curves that are locally length-minimizing are called *geodesics*, providing a notion of "straight lines" for efficiently navigating around the manifold. Geodesics are defined by their starting point and initial velocity. Indeed, the exponential map $\exp_{\boldsymbol{U}}(\Delta)$ takes in a starting point $\boldsymbol{U} \in \mathrm{St}(n, p)$ and an initial velocity $\Delta \in \mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, p)$, and outputs the final manifold point after following this geodesic for unit time. The exponential map is locally invertible, which gives the existence of the logarithmic map $\log_{\boldsymbol{U}}(\tilde{\boldsymbol{U}})$. The logarithmic map takes in a starting point $\boldsymbol{U}$ and a target point $\tilde{\boldsymbol{U}}$, and outputs the tangent vector needed to travel from $\boldsymbol{U}$ to $\tilde{\boldsymbol{U}}$. Algorithms for computing exponential and logarithmic maps under the canonical metric for the Stiefel manifold are given in Appendix B.3. In addition, points which are not on the manifold can be projected onto the manifold (Appendix B.6), which can turn a Euclidean generative model into a Stiefel generative model.

3 Stiefel Flow Matching

Having shown that the feasible space of moment-constrained structures is a Stiefel manifold, we can now formulate the problem of moment-constrained structure elucidation as an *unconstrained generative modeling problem on the Stiefel manifold*. An attractive approach to this is flow matching (FM), which trains a network as a time-dependent velocity field u_t that transforms samples from a prior noise distribution $x_0 \sim p_0(x_0)$ into samples which approximately match the data distribution $x_1 \sim p_1(x_1) \approx p_{\text{data}}(x_1)$ (Lipman et al., 2023). The integration of the velocity field over time is then a *continuous normalizing flow* ψ_t, which generates marginal probability densities p_t over time by the pushforward operation $p_t = [\psi_t]_{\#}(p_0)$. In practice, this is realized by sampling initial conditions $x_0 \sim p_0(x_0)$ and evolving them from time 0 to t according to the ODE $\frac{dx}{dt} = u_t(x)$ (Appendix C.4). The goal of training is to approximate this u_t using a neural network $v_\theta(t, x)$ parameterized by θ.

Flow matching is readily generalized to distributions on Riemannian manifolds (Chen & Lipman, 2024). When closed-form geodesics are available, Riemannian flow matching provides a simulation-free training objective for learning u_t, called Riemannian conditional flow matching,

$$\mathcal{L}_{\text{RCFM}}(\theta) = \mathbb{E}_{t\sim U(0,1),\, p(\boldsymbol{U}_0),\, p(\boldsymbol{U}_1)} \left[||v_\theta(t, \boldsymbol{U}_t) - \dot{\boldsymbol{U}}_t||^2_{\boldsymbol{U}_t} \right], \tag{8}$$

To compute this loss, we require (1) sampling $\boldsymbol{U}_0 \sim p(\boldsymbol{U}_0)$ from a prior noise distribution, (2) geodesic interpolation between $\boldsymbol{U}_0$ and $\boldsymbol{U}_1$ to get $\boldsymbol{U}_t$, (3) computing the time derivative of the interpolant $\dot{\boldsymbol{U}}_t$, and (4) evaluating the norm $||\cdot||_{\boldsymbol{U}_t}$.

To sample uniformly from the feasible space $\mathcal{M}$, we can sample $\boldsymbol{U}$ uniformly from $\mathrm{St}(n, 4)$ (Appendix B.4) and then rigidly rotate $\boldsymbol{U}$ so that its last column aligns with the unit mass vector $\hat{\boldsymbol{m}}$ (Appendix B.5). Then, geodesics can be computed from the exponential (Edelman et al., 1998) and logarithmic (Zimmermann & Hüper, 2022) maps, which are efficient to compute for $\mathrm{St}(n, 4)$,

$$\boldsymbol{U}_t = \exp_{\boldsymbol{U}_0}(t \log_{\boldsymbol{U}_0}(\boldsymbol{U}_1)), \tag{9}$$

making our training objective simulation-free. Once we have the interpolant $\boldsymbol{U}_t$, which will be the input to the neural network, we can then calculate the network's target $\dot{\boldsymbol{U}}_t$. Instead of autodifferentiation, which introduces unnecessary overhead, we compute $\dot{\boldsymbol{U}}_t$ using the logarithmic map,

$$\dot{\boldsymbol{U}}_t = \frac{1}{1-t} \log_{\boldsymbol{U}_t}(\boldsymbol{U}_1), \tag{10}$$

observing that $\dot{\boldsymbol{U}}_t$ is a unit length tangent vector along the geodesic from $\boldsymbol{U}_t$ to $\boldsymbol{U}_1$. Finally, we compute the norm of $v_\theta(t, \boldsymbol{U}_t) - \dot{\boldsymbol{U}}_t$ following Appendix B.2. Additionally, Appendix Theorem 4 shows how we can compute the logarithm in $\mathrm{St}(n, 3)$, rather than $\mathrm{St}(n, 4)$, which slightly saves time.

3.1 Reflection and Permutation Equivariance

As mentioned earlier, we set the coordinate axes as the principal axes of rotation by construction. This canonicalizes the 3D structure, removing translational symmetries and reducing the rotational symmetries to sign-flip symmetries of the eigenvectors of $\boldsymbol{P}$, which are the coordinate axes (Puny et al., 2022; Duval et al., 2023). Hence, the flow needs to be equivariant with respect to sign-flips of the coordinate axes, which we call "reflection-equivariance" for brevity (Lim et al., 2023; Cheng et al., 2024). In addition, because 3D structure is invariant under node order permutations, the learned velocity should be equivariant to permutations. Together, the explicit equivariance constraints on the network are given as $v_\theta(t, \boldsymbol{\Pi}\boldsymbol{U}_t\boldsymbol{R}) = \boldsymbol{\Pi} v_\theta(t, \boldsymbol{U}_t)\boldsymbol{R}$, for all node permutations $\boldsymbol{\Pi}$ and reflections $\boldsymbol{R} = \mathrm{diag}(\pm 1, \pm 1, \pm 1, 1)$. We satisfy these constraints with a reflection-equivariant graph neural network architecture described in Appendix C.1.

3.2 Equivariant Optimal Transport

Flow matching learns a velocity field which transports samples from the noise distribution to the data distribution, but there is no guarantee that samples will follow paths γ that are optimal with respect to transport cost $L(\gamma)$. Optimal paths are desired because they afford more efficient training and faster generation (Pooladian et al., 2023; Tong et al., 2024). During training, the model should be encouraged to learn dynamics which follow paths with lower transport cost. Since molecules have permutation and rotational symmetries, optimal paths should connect structures whose equivalence classes are close to each other. This corresponds to finding optimal node permutations and rotations to align noise samples $\boldsymbol{X}_0$ to data samples $\boldsymbol{X}_1$. In contrast to the Euclidean case (Klein et al., 2023; Song et al., 2023c), only reflections are needed, as opposed to rotations, because the coordinate axes are already fixed in place by the principal axes of rotation, enforced by the orthogonality condition.

In addition, transport cost on $\mathrm{St}(n, 4)$ must be measured using the Riemannian distance. Thus, the optimal transport map from $\boldsymbol{U}_0$ to $\boldsymbol{U}_1$ minimizes the following cost over atom-type-preserving node permutations $\boldsymbol{\Pi}$ (i.e., $\boldsymbol{\Pi}\boldsymbol{a} = \boldsymbol{a}$) and reflections $\boldsymbol{R} = \mathrm{diag}(\pm 1, \pm 1, \pm 1, 1)$:

$$c(\boldsymbol{U}_0, \boldsymbol{U}_1) = ||\log_{\boldsymbol{\Pi}\boldsymbol{U}_0\boldsymbol{R}}(\boldsymbol{U}_1)||_{\boldsymbol{\Pi}\boldsymbol{U}_0\boldsymbol{R}}. \tag{11}$$

While searching for optimal alignments, we do not compute the full Stiefel logarithm and instead approximately calculate distance using only one iteration of the inner loop described in Appendix

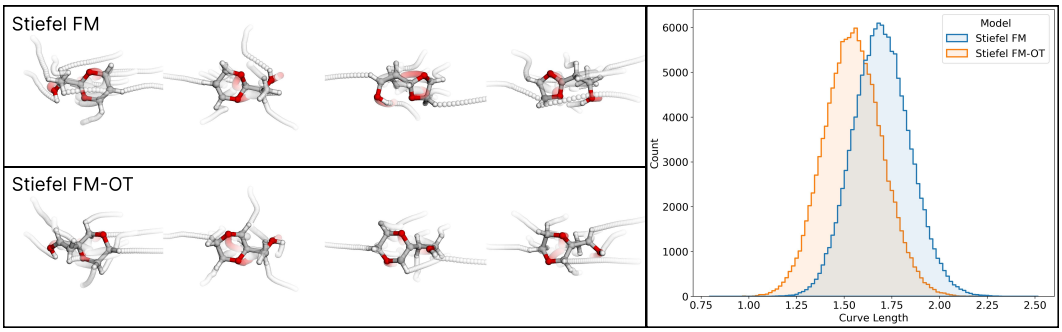

Figure 2: (*Left*) Learned sampling trajectories for Stiefel FM and Stiefel FM-OT on QM9. Each column begins generation from the same noise. (*Right*) Histogram of curve lengths of all QM9 sampling trajectories for Stiefel FM and Stiefel FM-OT. Permutation and reflection alignment lead to simpler and shorter paths.

Algorithm 2, and justified in Appendix Figure 11 . This approximate distance is then heuristically optimized over atom permutations and reflections with a greedy random local search. Appendix Algorithm 4 outlines this procedure, which samples several $\boldsymbol{U}$ matrices for each reflection to approximately identify the best reflection, and then performs a local search of random atom type-preserving index swaps (Appendix Figure 9).

4 Experiments

We evaluate Euclidean diffusion models and Stiefel Flow Matching on the QM9 and GEOM datasets. For each example, the model takes in moments and molecular formula and produces $K = 10$ samples.

Datasets. For QM9 (Ramakrishnan et al., 2014), we use the conformers provided by the GEOM dataset. We abbreviate GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022) as GEOM. We use the same training, validation, and test splits as Cheng et al. (2024), except that examples which are unstable or have less than 5 atoms are removed, which drops 76 examples from QM9 and drops no examples from GEOM. QM9 has train/val/test splits of 104265/13056/13033 molecules, while GEOM has splits of 233625/29203/29203 molecules, or 5537598/29203/29203 conformers. We only predict 3D structures for the lowest-energy conformers of GEOM. This reflects experimental reality, as the lowest-energy conformer typically has the highest proportion in the population after cooling by supersonic jet expansion, assuming thermodynamic equilibrium (Ruoff et al., 1990).

Structure elucidation. The only measure of a model's success is its ability to generate the correct 3D structure $\boldsymbol{X}$ with high accuracy at least once, where accuracy is measured by root-mean-squared-deviation (RMSD) of the predicted coordinates to the ground truth. High accuracy is needed because the only confirming evidence available to us is (1) agreement with moments and (2) thermodynamic stability by quantum chemistry. For this reason, we use stringent thresholds of RMSD < 0.25 Å and RMSD < 0.10 Å to ensure that the predicted structure is in the same potential energy basin as the true structure. Success is reported if the minimum RMSD over $K = 10$ generated samples satisfies these thresholds. The success rate is the percentage of the test set whose generated samples has a *minimum* RMSD which satisfies each threshold. Error bars are standard errors of the mean.

Evaluating RMSD is nontrivial because the generated and ground truth structure must first be aligned under same-atom-type permutations and reflections. We use the same RMSD procedure as Cheng et al. (2024): We first align node permutations by solving a linear assignment problem whose cost matrix is squared Euclidean distance, and repeat this for all 8 reflections, taking the minimum. Then we compute RMSD between the ground truth coordinates and the aligned coordinates. This is similar to the alignment procedure used in (Klein et al., 2023; Song et al., 2023c).

Auxiliary metrics. To characterize the samples generated by each model, we report additional metrics. In contrast to success rate, which checks the minimum RMSD of generated samples for each example, these auxiliary metrics are averaged over *all* generated samples (except diversity). *Error* measures how much the generated structure violates the moment constraints. If $\hat{\boldsymbol{P}}$ is the computed

Table 1: Experimental results on QM9. Stiefel FM shows no violation of moment constraints as shown in the **Error** metrics, and has the highest success rate for structure elucidation, with the lowest computational cost.

Method	% < RMSD ↑ 0.25 Å	0.10 Å	Error ↓	Valid ↑	Stable ↓	Diverse ↑	NFE ↓
Stiefel Random	0.00±0.00	0.00±0.00	0.00	0.061	nan	2.640	0
KREED	11.22±0.28	9.55±0.26	5.18	0.878	−1.335	1.429	1000
KREED-XL	13.65±0.30	10.94±0.27	3.64	0.933	−1.048	0.870	1000
KREED-XL-DPS	12.36±0.29	9.40±0.26	1.33	0.744	−0.826	1.060	1000
KREED-XL-proj	13.67±0.30	10.93±0.27	0.00	0.924	−0.905	0.871	1000
Stiefel FM	**15.17±0.31**	**13.82±0.30**	0.00	0.882	−1.125	1.040	200
Stiefel FM-OT	13.99±0.30	12.68±0.29	0.00	0.835	−1.039	1.045	200

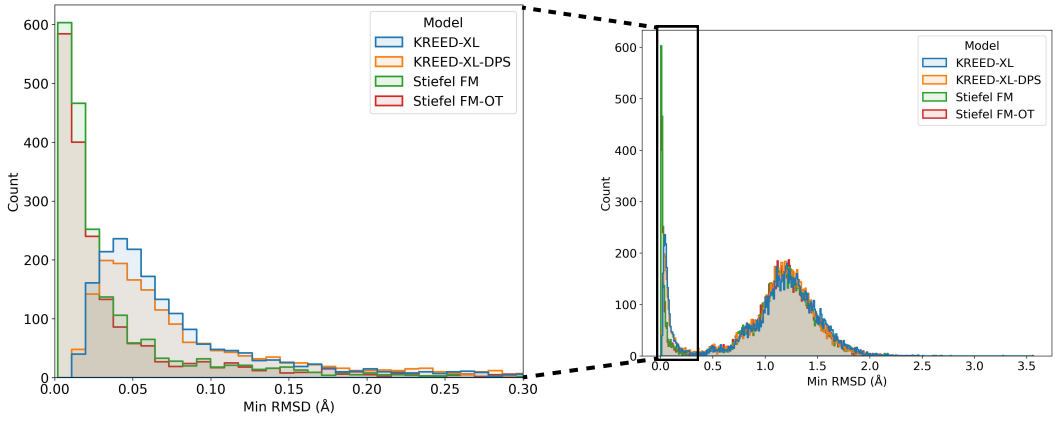

Figure 3: Histograms of minimum RMSD for predicted QM9 examples show two distinct clusters for RMSD. The 0.25Å threshold captures molecular structures that are useful for structure elucidation.

planar dyadic of the generated structure, this is computed as $\frac{1}{\sqrt{6}}||\text{triu}(\hat{\boldsymbol{P}} - \boldsymbol{P})||_2$, where triu takes the upper triangular part of the matrix (contains 6 elements). *Validity* is a heuristic check based on bond detection with `rdDetermineBonds.DetermineConnectivity` (Landrum, 2013; Kim & Kim, 2015). *Stability* is the log norm of the gradient of energy with respect to coordinates, as reported by the `xtb` quantum chemistry program (Bannwarth et al., 2019). Stable structures should have a gradient norm close to zero (log norm very negative), assuming the structure is at a local minimum and not a saddle point. *Diversity* is calculated as the average pairwise RMSD of all generated samples of a single example. *NFE* is the number of function evaluations used during generation, and measures computational cost. We do not set boldface for these metrics because they do not correspond directly to success criteria.

Baselines. Given the novelty of the problem, the number of available baselines is limited. We compare the performance of Stiefel FM to KREED (Cheng et al., 2024). KREED is a reflection-equivariant diffusion model trained to generate 3D structure conditioned on molecular formula and moments of inertia, and is a specialization of E(3)-equivariant approaches like EDM (Hoogeboom et al., 2022). Since our architecture for Stiefel FM is much larger (parameter-wise) than the model architecture used in KREED, and because KREED is tailored for a slightly different task, we also train another reflection-equivariant diffusion model with an identical neural network architecture to Stiefel Flow Matching, which we label as KREED-XL. The planar dyadic is computable at every step of the generation process, which means that this task can be treated as a nonlinear inverse problem: On top of KREED-XL, we apply Diffusion Posterior Sampling (DPS) (Chung et al., 2023), which guides generation with an additional drift term for minimizing the planar dyadic error. As a simple baseline which exactly satisfies moment constraints, we report performance for uniform random

Table 2: Experimental results on GEOM. Stiefel FM generates few valid structures on its own due to the increased difficulty of manifold-constrained generative modelling. When adjusted to generate the same number of valid molecules, Stiefel FM-OT (filter) obtains the highest success rate, without surpassing baselines in computational cost.

Method	% < RMSD ↑ 0.25 Å	0.10 Å	Error ↓	Valid ↑	Stable ↓	Diverse ↑	NFE ↓
Stiefel Random	0.00±0.00	0.00±0.00	0.00	0.000	nan	4.104	0
KREED	0.05±0.01	0.02±0.01	58.36	0.353	−0.583	2.297	1000
KREED-XL	3.54±0.11	2.02±0.08	30.71	0.907	−0.900	2.190	1000
KREED-XL-proj	3.54±0.11	2.04±0.08	0.00	0.904	−0.752	2.188	1000
Stiefel FM	2.17±0.08	1.24±0.07	0.00	0.388	−0.066	2.212	200
Stiefel FM-OT	2.45±0.09	1.49±0.07	0.00	0.376	−0.002	2.195	200
Stiefel FM (filter)	3.57±0.11	2.06±0.08	0.00	0.889	−0.437	2.208	600
Stiefel FM-OT (filter)	**3.79±0.11**	**2.34±0.09**	0.00	0.869	−0.352	2.193	600

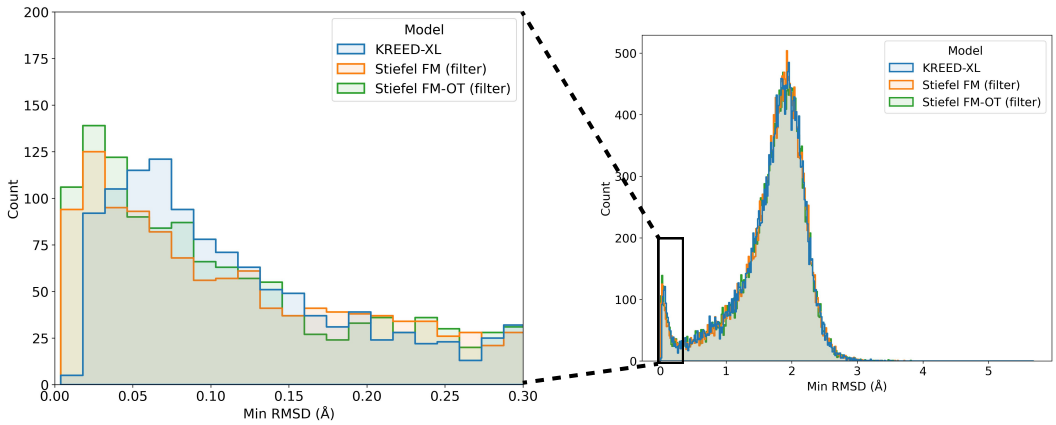

Figure 4: Histograms of minimum RMSD for predicted GEOM examples.

sampling on the Stiefel manifold. We also report the results of KREED-XL after projecting samples to the feasible manifold. Relevant hyperparameters for all methods are provided in Appendix C.3.

Results. Table 1 shows our experimental results on QM9, reporting Stiefel FM with and without optimal transport (OT). We note that the average number of atoms in QM9 is 18, meaning that on average the model must infer $3(18) - 9 = 39$ values from 3 moments. We find that Stiefel FM can generate the correct structure with a greater success rate than all Euclidean diffusion models. We also see that incorporating the analytic formula of the moments via DPS does improve agreement with the moments, but at the cost of accuracy. In contrast, Stiefel FM does not suffer from this tradeoff. Projecting samples onto the manifold does not change success rate, because the projection leaves correct structures untouched, while only distorting incorrect structures. In addition, Stiefel FM uses only 20% of the computation used by diffusion models. Unconstrained diffusion models can produce more valid and stable structures, though they may not necessarily generate the correct structure. Figure 3 reveals that when Stiefel FM's predictions are correct, it is likely to be extremely accurate, achieving RMSD even below 0.05 Å. Training with equivariant optimal transport helps learn simpler generation paths and reduces the average curve length of generation trajectories from 1.696 to 1.547 (Figure 2). Interestingly, training with optimal transport slightly reduces the success rate of Stiefel FM on QM9, though this trend is reversed for GEOM. Appendix Table 5 reports additional trials experimenting with optimal transport, logit-normal timestep sampling (Esser et al., 2024), and stochasticity (Bose et al., 2024).

The task of structure elucidation poses an even harder challenge on GEOM, with an average of 46 atoms, giving a task of inferring $3(46) - 9 = 129$ values. Naively, one would expect that this problem is hopelessly underconstrained. Nevertheless, Table 2 shows that both diffusion models

and flow matching can obtain a nontrivial success rate for structure elucidation. However, the low accuracy, validity, and stability of Stiefel FM(-OT) suggests that the model is underfitting the dataset, even as KREED-XL is able to generate numerous valid samples. We observe that less than half of the structures generated by Stiefel FM are valid, which aligns with the fact that *there are fewer valid structures on the feasible manifold $\mathcal{M}$ than in regular Euclidean space*. This suggests that a fairer evaluation utilizing the strength of the manifold constraint should generate a similar number of *valid* structures for each model. This is justified by the fact that validity does not use ground truth labels and can be computed at only nominal cost. Therefore, we generate 30 samples for Stiefel FM and Stiefel FM-OT before filtering to retain up to $K = 10$ valid samples. Note that the combined computational cost of generating 3x as many samples (3 x 200 NFE) is still lower than KREED-XL (1000 NFE). After filtering, Stiefel FM-OT obtains a similar validity rate to KREED-XL, but obtains the highest success rate for structure elucidation. Even after filtering, the mediocre stability of generated samples suggests that success rate can be improved further, though at the slightly higher cost of quantum chemistry calculations. Now, we see that optimal transport does help Stiefel FM to predict accurate structures, while also reducing average curve length from 1.421 to 1.344 (Appendix Figure 8). However, biasing generation by selecting for valid samples seems to also select for longer generation trajectories. We find that generation paths that land on the correct structure are usually longer than generation paths that land on incorrect structures (Appendix Figure 7) This may be explained by the fact that initial points are sampled anywhere uniformly on the manifold, but for success they must end up on the single true structure. In contrast, there are many incorrect structures all over the manifold, which may end up on average closer to random initial points.

It should be noted that one should expect success rates to be of this magnitude for solving a heavily underconstrained problem. This is simply due to the fact that there exist many stable structures with the same molecular formula and very similar moments of inertia. Indeed, trained models usually generate realistic-looking molecules – see generated examples in Appendix Figure 5 and Figure 6. Furthermore, only 10 to 30 samples were queried for each molecule, but an actual structure elucidation campaign would have a much larger compute budget for generating thousands of samples. As a highlight, our results show that it is actually possible at 0.25 Å resolution to elucidate 27.4% of the test set of QM9 (3580/13033) when combining KREED-XL, Stiefel FM, and Stiefel FM-OT; and 7.9% of the test set of GEOM (2297/29203), when combining KREED-XL, Stiefel FM (filter), and Stiefel FM-OT (filter).

5 Related Work

Generative models for 3D molecules. While research in applying generative models for 3D structure, only few works have applied these methods to the task of structure elucidation (Cheng et al., 2024). Adjacent work in applying diffusion and flow matching models to 3D molecules include molecular generation (Hoogeboom et al., 2022; Song et al., 2023c), conformer search (Jing et al., 2022; Xu et al., 2022), docking (Corso et al., 2022), biomolecular assembly (Abramson et al., 2024) and Boltzmann generators (Klein et al., 2023). Generative models on Riemannian manifolds have also been applied to protein design (Bose et al., 2024; Yim et al., 2023) and crystal structure prediction (Jiao et al., 2024). Recent work has applied generative modelling to predict 3D structure from powder X-ray diffraction patterns (Lai et al., 2024; Riesel et al., 2024).

Deep learning for molecular identification using rotational spectroscopy. A limited number of works have provided parts of the overall puzzle needed for structure elucidation using rotational spectroscopy. Zaleski & Prozument (2018) propose RAINet as a forward modeling approach for assigning rotational spectra, where a set of peaks is fed into a classifier, that categorizes the spectra to an appropriate multilayer pereception that outputs spectroscopic parameters, including moments. McCarthy & Lee (2020) adopt a "mixture-of-experts" approach that maps spectroscopic parameters and approximate molecular formula into a set of complementary experimental observables and SMILES strings, though with limited success. Cheng et al. (2024) present a Euclidean diffusion model KREED for determining 3D structure from moments, molecular formula, and unsigned substitution coordinates, the latter of which is also measurable from rotational spectroscopy, but can be difficult and expensive to obtain. Most recently, Schwarting et al. (2024) provide a thorough analysis into the inverse problem, examining the frequency with which different molecules have moments of inertia that are very close in value. Stiefel Flow Matching could disambiguate these structures by providing a diversity of structures that satisfy moment constraints exactly.

Statistics on the Stiefel manifold. While the Stiefel manifold is often studied in the context of optimization (Absil et al., 2008; Chen et al., 2021; Kong et al., 2022), a number of works study probability distributions on the Stiefel manifold (Chakraborty & Vemuri, 2019; Chikuse, 1990). One distribution on the Stiefel manifold is called the *matrix von Mises-Fisher distribution* or *matrix Langevin distribution* (Pal et al., 2020; Chikuse, 2003; Jupp & Mardia, 1979). Wang & Solo (2020) propose a particle filtering algorithm on the Stiefel manifold, with the first application of optimal transport on the Stiefel manifold. Yataka et al. (2023) propose a continuous normalizing flow on the Grassmann manifold, which is closely related to the Stiefel manifold.

6 Conclusion

We propose Stiefel Flow Matching, a Riemannian generative model for generating samples subject to exact orthogonality constraints, and apply it to the challenging inverse problem of structure elucidation from moments of inertia and molecular formula. Empirically, Stiefel Flow Matching achieves a higher success rate than other Euclidean diffusion approaches. Satisfying the constraints exactly will enable future advances in Riemannian generative modelling to directly transfer to generating more stable molecules, without needing to consider agreement to the moments.

6.1 Future Research Directions

Improving Stiefel generative models. Riemannian flow matching empirically shows degradation compared to Riemannian diffusion (Lou et al., 2024; Zhu et al., 2024). This has been attributed to two pathologies of Riemannian flow matching for compact manifolds: (1) the geodesic-based velocity field is discontinuous at the cut locus (Lou et al., 2024; Zhu et al., 2024), and (2) the probability density has a shrinking support (Stark et al., 2024; Holderrieth et al., 2024). These pathologies may explain the difficulty of Stiefel FM in fitting GEOM, and motivate the development of alternative probability paths for Stiefel flow matching, such as Stiefel diffusion (De Bortoli et al., 2022), or flows which asymptotically land on the Stiefel manifold (Ablin & Peyré, 2022; Gao et al., 2022). Training a Stiefel diffusion model with the denoising score matching objective requires the heat kernel as training targets for the neural network, which can be computed using techniques by Azangulov et al. (2022) and Lou et al. (2024). Alternatively, the matrix Langevin distribution may be amenable to modelling with Star-Shaped Denoising Diffusion Probabilistic Models (Okhotin et al., 2024). Future work can also explore varying the metric used on the Stiefel manifold, since the Stiefel manifold admits a 1-parameter family of metrics generalizing the canonical metric (Hüper et al., 2021). These metrics have efficient numerical algorithms for the exponential and logarithm (Mataigne et al., 2024). Sample-time advances in flow matching, such as corrector sampling (Gat et al., 2024) or enhancing the flow with a jump process (Holderrieth et al., 2024), are also orthogonal avenues for improvement.

A key limitation of Stiefel Flow Matching is the requirement of molecular formula as input. Recent approaches in discrete flow matching (Campbell et al., 2024; Gat et al., 2024) could enable a multimodal flow to simultaneously vary both continuous atom positions and discrete atom types. The jump processes of generator matching (Holderrieth et al., 2024) are particularly natural for this problem, as they could allow the flow to jump between Stiefel manifolds of different sizes.

Modeling on real-world chemical data. Since we require only minimal information in the moments and molecular formula, another direction is to incorporate other conditioning information, such as energy and force information, fragments of 2D graphs, dipole moments, or other sources of analytical chemistry data. We can do so through MCMC sampling (Du et al., 2023), guidance (Song et al., 2023a;b; Mardani et al., 2023), or diversity sampling approaches (Corso et al., 2024). This presents an opportunity in mass spectrometry, as the 3D structural information provided by the moments can distinguish molecules which have the same mass. Controllable diversity can also be leveraged here since, in a molecule identification campaign, we can eliminate candidates once we know they are not correct (by calculating the rovibrational correction and comparing to experiment). We then want to sample structures which are different from these eliminated candidates.

Additional applications of Stiefel Flow Matching. Stiefel generative models can also be applied to other domains of orthogonality constrained data, such as molecular orbitals (Mrovec & Berger, 2021; Aoto & da Silva, 2021), orthogonal neural network weights (Kong et al., 2022), and covariance matrices in neural data (Nejatbakhsh et al., 2024).

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.

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

Yuri Alexandre Aoto and Márcio Fabiano da Silva. An optimisation on the grassmannian with applications to quantum chemistry. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 8(1), 2021.

Simon Axelrod and Rafael Gomez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):1–14, 2022.

Iskander Azangulov, Andrei Smolensky, Alexander Terenin, and Viacheslav Borovitskiy. Stationary kernels and gaussian processes on lie groups and their homogeneous spaces i: the compact case. *arXiv preprint arXiv:2208.14960*, 2022.

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15 (3):1652–1671, 2019.

Thomas Bendokat, Ralf Zimmermann, and P-A Absil. A grassmann manifold handbook: Basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(1):1–51, 2024.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=kJFIH23hXb.

Gordon G Brown, Brian C Dian, Kevin O Douglass, Scott M Geyer, and Brooks H Pate. The rotational spectrum of epifluorohydrin measured by chirped-pulse fourier transform microwave spectroscopy. *Journal of Molecular Spectroscopy*, 238(2):200–212, 2006.

Gordon G Brown, Brian C Dian, Kevin O Douglass, Scott M Geyer, Steven T Shipman, and Brooks H Pate. A broadband fourier transform microwave spectrometer based on chirped pulse excitation. *Review of Scientific Instruments*, 79(5), 2008.

Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *arXiv preprint arXiv:2402.04997*, 2024.

Rudrasis Chakraborty and Baba C Vemuri. Statistics on the stiefel manifold: theory and applications. 2019.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=g7ohDlTITL.

Shixiang Chen, Alfredo Garcia, Mingyi Hong, and Shahin Shahrampour. Decentralized riemannian gradient descent on the stiefel manifold. In *International Conference on Machine Learning*, pp. 1594–1605. PMLR, 2021.

Austin H Cheng, Alston Lo, Santiago Miret, Brooks H Pate, and Alán Aspuru-Guzik. Determining 3d structure from molecular formula and isotopologue rotational spectra in natural abundance with reflection-equivariant diffusion. *The Journal of Chemical Physics*, 160(12), 2024.

Yasuko Chikuse. Distributions of orientations on stiefel manifolds. *Journal of multivariate analysis*, 33(2):247–264, 1990.

Yasuko Chikuse. Concentrated matrix langevin distributions. *Journal of Multivariate Analysis*, 85(2): 375–394, 2003.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Gabriele Corso, Yilun Xu, Valentin De Bortoli, Regina Barzilay, and Tommi S. Jaakkola. Particle guidance: non-i.i.d. diverse sampling with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=KqbCvIFBY7.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

J. Demaison, James E. Boggs, and Attila G. Császár (eds.). *Equilibrium Molecular Structures: From Spectroscopy to Quantum Chemistry*. CRC Press, Boca Raton, 2011. ISBN 978-1-4398-1132-0.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Sander Dieleman, Laurent Sartran, Arman Roshannai, Nikolay Savinov, Yaroslav Ganin, Pierre H Richemond, Arnaud Doucet, Robin Strudel, Chris Dyer, Conor Durkan, et al. Continuous diffusion for categorical data. *arXiv preprint arXiv:2211.15089*, 2022.

Sérgio R. Domingos, Cristóbal Pérez, Mark D. Marshall, Helen O. Leung, and Melanie Schnell. Assessing the performance of rotational spectroscopy in chiral analysis. *Chemical Science*, 11(40): 10863–10870, October 2020. ISSN 2041-6539. doi: 10.1039/D0SC03752D.

Yilun Du, Conor Durkan, Robin Strudel, Joshua B Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl. Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and mcmc. In *International conference on machine learning*, pp. 8489–8510. PMLR, 2023.

Alexandre Agm Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D Malliaros, Yoshua Bengio, and David Rolnick. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pp. 9013–9033. PMLR, 2023.

Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. *arXiv preprint arXiv:2403.03206*, 2024.

Matthias Fey and Jan Eric Lenssen. Fast Graph Representation Learning with PyTorch Geometric, May 2019. URL https://github.com/pyg-team/pytorch_geometric.

Bin Gao, Simon Vary, Pierre Ablin, and P-A Absil. Optimization flows landing on the stiefel manifold. *IFAC-PapersOnLine*, 55(30):25–30, 2022.

Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky TQ Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete flow matching. *arXiv preprint arXiv:2407.15595*, 2024.

Walter Gordy, Robert L Cook, and Arnold Weissberger. *Microwave molecular spectra*, volume 18. Wiley New York, 1984. URL https://app.knovel.com/kn/resources/kpMMSE0001/toc.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky TQ Chen, and Yaron Lipman. Generator matching: Generative modeling with arbitrary markov processes. *arXiv preprint arXiv:2410.20587*, 2024.

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.

Knut Hüper, Irina Markina, and Fátima Silva Leite. A lagrangian approach to extremal curves on stiefel manifolds. AIMS, 2021.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36, 2024.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35:24240–24253, 2022.

Peter E Jupp and Kanti V Mardia. Maximum likelihood estimators for the matrix von mises-fisher and bingham distributions. *The Annals of Statistics*, 7(3):599–606, 1979.

Yeonjoon Kim and Woo Youn Kim. Universal structure conversion method for organic molecules: from atomic connectivity to three-dimensional geometry. *Bulletin of the Korean Chemical Society*, 36(7):1769–1777, 2015.

Leon Klein, Andreas Krämer, and Frank Noe. Equivariant flow matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=eLH2NFOO1B.

Lingkai Kong, Yuqing Wang, and Molei Tao. Momentum stiefel optimizer, with applications to suitably-orthogonal attention, and optimal transport. *arXiv preprint arXiv:2205.14173*, 2022.

J Kraitchman. Determination of molecular structure from microwave spectroscopic data. *American Journal of Physics*, 21(1):17–24, 1953.

Qingsi Lai, Lin Yao, Zhifeng Gao, Siyuan Liu, Hongshuai Wang, Shuqi Lu, Di He, Liwei Wang, Cheng Wang, and Guolin Ke. End-to-end crystal structure prediction from powder x-ray diffraction. *arXiv preprint arXiv:2401.03862*, 2024.

Greg Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013.

John M. Lee. *Smooth Manifolds*, pp. 1–29. Springer New York, New York, NY, 2003. ISBN 978-0-387-21752-9. doi: 10.1007/978-0-387-21752-9_1. URL https://doi.org/10.1007/978-0-387-21752-9_1.

Kin Long Kelvin Lee and Michael McCarthy. Study of Benzene Fragmentation, Isomerization, and Growth Using Microwave Spectroscopy. *The Journal of Physical Chemistry Letters*, 10(10): 2408–2413, April 2019. ISSN 1948-7185. doi: 10.1021/acs.jpclett.9b00586.

Kin Long Kelvin Lee and Michael McCarthy. Bayesian Analysis of Theoretical Rotational Constants from Low-Cost Electronic Structure Methods. *The Journal of Physical Chemistry A*, 124(5): 898–910, February 2020. ISSN 1089-5639. doi: 10.1021/acs.jpca.9b09982.

Derek Lim, Joshua Robinson, Stefanie Jegelka, and Haggai Maron. Expressive sign equivariant networks for spectral geometric learning. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=UWd4ysACo4.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Aaron Lou, Minkai Xu, Adam Farris, and Stefano Ermon. Scaling riemannian diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. *arXiv preprint arXiv:2305.04391*, 2023.

Alan G Marshall and Christopher L Hendrickson. High-resolution mass spectrometers. *Annu. Rev. Anal. Chem.*, 1:579–599, 2008.

Simon Mataigne, Ralf Zimmermann, and Nina Miolane. An efficient algorithm for the riemannian logarithm on the stiefel manifold for a family of riemannian metrics. *arXiv preprint arXiv:2403.11730*, 2024.

Michael McCarthy and Kin Long Kelvin Lee. Molecule Identification with Rotational Spectroscopy and Probabilistic Deep Learning. *The Journal of Physical Chemistry A*, 124(15):3002–3017, April 2020. ISSN 1089-5639, 1520-5215. doi: 10.1021/acs.jpca.0c01376.

Michael C. McCarthy, Kin Long Kelvin Lee, P. Brandon Carroll, Jessica P. Porterfield, P. Bryan Changala, James H. Thorpe, and John F. Stanton. Exhaustive Product Analysis of Three Benzene Discharges by Microwave Spectroscopy. *The Journal of Physical Chemistry A*, 124(25):5170–5181, June 2020. ISSN 1089-5639, 1520-5215. doi: 10.1021/acs.jpca.0c02919.

Brett A McGuire. 2018 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules. *The Astrophysical Journal Supplement Series*, 239(2):17, 2018.

Brett A. McGuire, P. Brandon Carroll, Ryan A. Loomis, Ian A. Finneran, Philip R. Jewell, Anthony J. Remijan, and Geoffrey A. Blake. Discovery of the interstellar chiral molecule propylene oxide (CH3CHCH2O). *Science*, 352(6292):1449–1452, June 2016. doi: 10.1126/science.aae0328.

Benedetta Giannelli Moneta, Sara Elsa Aita, Elena Barbaro, Anna Laura Capriotti, Andrea Cerrato, Aldo Laganà, Carmela Maria Montone, Susy Piovesana, Federico Scoto, Carlo Barbante, et al. Untargeted analysis of environmental contaminants in surface snow samples of svalbard islands by liquid chromatography-high resolution mass spectrometry. *Science of the Total Environment*, 858: 159709, 2023.

Martin Mrovec and JA Berger. A diagonalization-free optimization algorithm for solving kohn–sham equations of closed-shell molecules. *Journal of Computational Chemistry*, 42(7):492–504, 2021.

Amin Nejatbakhsh, Isabel Garon, and Alex Williams. Estimating noise correlations across continuous conditions with wishart processes. *Advances in Neural Information Processing Systems*, 36, 2024.

Andrey Okhotin, Dmitry Molchanov, Arkhipkin Vladimir, Grigory Bartosh, Viktor Ohanesian, Aibek Alanov, and Dmitry P Vetrov. Star-shaped denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 36, 2024.

Subhadip Pal, Subhajit Sengupta, Riten Mitra, and Arunava Banerjee. Conjugate priors and posterior inference for the matrix langevin distribution on the stiefel manifold. 2020.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28100–28127. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/pooladian23a.html.

Omri Puny, Matan Atzmon, Edward J. Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=zIUyj55nXR.

Cristina Puzzarini and John F Stanton. Connections between the accuracy of rotational constants and equilibrium molecular structures. *Physical Chemistry Chemical Physics*, 25(3):1421–1429, 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.

Eric A Riesel, Tsach Mackey, Hamed Nilforoshan, Minkai Xu, Catherine K Badding, Alison B Altman, Jure Leskovec, and Danna E Freedman. Crystal structure determination from powder diffraction patterns with generative machine learning. *Journal of the American Chemical Society*, 2024.

RS Ruoff, TD Klots, T Emilsson, and HS Gutowsky. Relaxation of conformers and isomers in seeded supersonic jets of inert gases. *The Journal of chemical physics*, 93(5):3142–3150, 1990.

Marcus Schwarting, Nathan A. Seifert, Michael J. Davis, Ben Blaiszik, Ian Foster, and Kirill Prozument. Twins in rotational spectroscopy: Does a rotational spectrum uniquely identify a molecule?, April 2024.

Steven T Shipman, Justin L Neill, Richard D Suenram, Matt T Muckle, and Brooks H Pate. Structure determination of strawberry aldehyde by broadband microwave spectroscopy: Conformational stabilization by dispersive interactions. *The Journal of Physical Chemistry Letters*, 2(5):443–448, 2011.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023a. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pp. 32483–32498. PMLR, 2023b.

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.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=vaRCHVj0uGI.

Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant flow matching with hybrid probability transport for 3d molecule generation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023c.

Reilly E. Sonstrom, Zoua Pa Vang, Haley N. Scolati, Justin L. Neill, Brooks H. Pate, and Joseph R. Clark. Rapid Enantiomeric Excess Measurements of Enantioisotopomers by Molecular Rotational Resonance Spectroscopy. *Organic Process Research & Development*, 27(7):1185–1197, July 2023. ISSN 1083-6160. doi: 10.1021/acs.oprd.3c00028.

Jean-Thibaut Spaniol, Kin Long Kelvin Lee, Olivier Pirali, Cristina Puzzarini, and Marie-Aline Martin-Drumel. A rotational investigation of the three isomeric forms of cyanoethynylbenzene (HCC-C 6 H 4 -CN): Benchmarking experiments and calculations using the "Lego brick" approach. *Physical Chemistry Chemical Physics*, 25(8):6397–6405, 2023. doi: 10.1039/D2CP04825F.

Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to dna sequence design. *arXiv preprint arXiv:2402.05841*, 2024.

Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=CD9Snc73AW. Expert Certification.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Natalaja Vogt, Jürgen Vogt, and Jean Demaison. Accuracy of the rotational constants. *Journal of molecular structure*, 988(1-3):119–127, 2011.

Zhichao Wang and Victor Solo. Particle filtering on the stiefel manifold with optimal transport. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 4111–4116. IEEE, 2020.

Caroline C Womack, Marie-Aline Martin-Drumel, Gordon G Brown, Robert W Field, and Michael C McCarthy. Observation of the simplest criegee intermediate ch2oo in the gas-phase ozonolysis of ethylene. *Science advances*, 1(2):e1400105, 2015.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Ryoma Yataka, Kazuki Hirashima, and Masashi Shiraishi. Grassmann manifold flows for stable shape generation. *Advances in Neural Information Processing Systems*, 36:72377–72411, 2023.

Lia Yeh, Lincoln Satterthwaite, and David Patterson. Automated, context-free assignment of asymmetric rotor microwave spectra. *The Journal of chemical physics*, 150(20), 2019.

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.

Daniel P. Zaleski and Kirill Prozument. Automated assignment of rotational spectra using artificial neural networks. *The Journal of Chemical Physics*, 149(10):104106, September 2018. ISSN 0021-9606. doi: 10.1063/1.5037715.

Yuchen Zhu, Tianrong Chen, Lingkai Kong, Evangelos A Theodorou, and Molei Tao. Trivialized momentum facilitates diffusion generative modeling on lie groups. *arXiv preprint arXiv:2405.16381*, 2024.

Ralf Zimmermann and Knut Hüper. Computing the riemannian logarithm on the stiefel manifold: Metrics, methods, and performance. *SIAM Journal on Matrix Analysis and Applications*, 43(2): 953–980, 2022.

A ROTATIONAL SPECTROSCOPY & MOLECULAR STRUCTURE

A.1 NOTATION

(P_X, P_Y, P_Z) are really called the *planar moments of inertia*. We refer to them as the moments of inertia as shorthand. The actual *moments of inertia* (I_A, I_B, I_C) are uniquely related to the planar moments by a simple linear transformation (Kraitchman, 1953).

A.2 INERTIA EDGE CASES

We assume that the molecule of interest has moments of inertia $P_X > P_Y > P_Z > 0$, or in other words, a nonplanar asymmetric rotor. This assumption holds for the vast majority of molecules. We now discuss edge cases, such as perfectly symmetric, planar, or linear molecules. It is worth noting that, owing to their rarity and symmetries, these edge cases have significant overlap with the set of molecules that have already been studied (McGuire, 2018).

When structures have two equal eigenvalues ($P_Y = P_Z$) in their inertia matrix, there is no longer a unique choice of these two principal axes. These axes now sweep out a plane of possibilities, and numerical diagonalization will arbitrarily pick two orthogonal axes from this plane. But, making an arbitrary choice does not break the mapping between $\boldsymbol{X}$ and $\boldsymbol{U}$ in Equation (4). The only issue is that to respect this additional symmetry, the flow should be invariant to in-plane rotoreflections of the molecule. However, these examples are so rare that they can be ignored: 76 examples in QM9 and 8 examples in GEOM. If needed, this symmetry can be handled using data augmentation.

Stiefel Flow Matching cannot handle *exactly* planar ($P_Z = 0$) and *exactly* linear ($P_Y = P_Z = 0$) molecules due to divide by zero in Equation (4). The vast majority of "planar" molecules are actually slightly nonplanar and therefore pose no issue. In rare cases where molecules are truly planar, Stiefel Flow Matching can be reformulated for 1 and 2 dimensions.

A.3 EXPERIMENTAL WORKFLOW

For an unknown molecule, we assume that its molecular formula can be measured by high-resolution mass spectrometry (Marshall & Hendrickson, 2008) and that its moments can be measured from rotational spectroscopy (Gordy et al., 1984).

At a high level, rotational spectroscopy observes how molecules freely rotate in the gas phase. The rotation of molecules is *quantized*, giving rise to a discrete set of rotational states. Molecules can absorb or emit radiation at *characteristic* wavelengths to transition between these energy levels. Rotational spectroscopy measures the energies of these transitions. A broadband microwave spectrometer can simultaneously measure thousands of these transitions, producing a spectrum of many sharp peaks (Brown et al., 2006; 2008). Rotational transition energies are in the microwave to terahertz region, which is why rotational spectroscopy is also known as microwave spectroscopy.

For an asymmetric rigid molecule, and neglecting effects like centrifugal distortion and hyperfine structure, the molecule's rotational energy levels are essentially determined by three unique *rotational constants*, $A(BC), A > B > C$. Rotational constants are inversely proportional to the principal moments of inertia, i.e., $A = \frac{\hbar^2}{2I_A}$. Each energy level is also indexed by quantum numbers. Transitions are the differences in these energy levels. Transition strength is also affected by the molecule's dipole moment.

Once a spectrum is measured, the rotational spectroscopist is tasked with assigning each transition to its quantum numbers and ultimately assigning rotational constants $A(BC)$. Spectral assignment is a challenging problem tackled in other works (Zaleski & Prozument, 2018; Yeh et al., 2019). Rotational constants can then simply be inverted to obtain effective moments of inertia.

Using these effective moments, a spectroscopist can now search for the true structure using any of the models developed in this work. In an actual structure elucidation campaign, only a few targets are considered, which permits querying $K > 1000$ samples for each target, and also leaves enough computational resources to evaluate every generated sample using quantum chemistry.

A.4 EXPERIMENTAL PRECISION

While the proposed method generates structures which satisfy the moment constraints *exactly*, and while rotational spectroscopy can measure experimental rotational constants to many digits of precision (Vogt et al., 2011), it is unfortunate that the experimental rotational constants do not directly translate to moments of inertia. This is because molecules are not perfectly rigid: Experiment observes properties that have been *vibrationally* averaged, including the rotational constants. Vibrational averaging is distinguished from conformational fluctuations such as torsions, which can be frozen out by cooling molecules to their ground vibrational state. However, even in the ground vibrational state, a molecule is still vibrating due to zero-point energy. As a result, the experimental rotational constants $A(BC)_0$ are proportional to $\langle \frac{1}{r^2} \rangle$, whereas equilibrium rotational constants $A(BC)_e$ are proportional to $\frac{1}{\langle r \rangle^2}$, where $\langle \cdot \rangle$ denotes a vibrational average. Structures in QM9 and GEOM have been geometry optimized to reach equilibrium structures $\langle r \rangle$. Equilibrium rotational constants are on the order of 1% relative error of the experimental rotational constants (Vogt et al., 2011; Puzzarini & Stanton, 2023).

Typically, experimental rotational constants $A(BC)_0$ can be corrected into equilibrium rotational constants $A(BC)_e$ by a rovibrational calculation. Experimental and equilibrium rotational constants are related by a perturbative expansion (Demaison et al., 2011)

$$A(BC)_0 = A(BC)_e - \frac{1}{2}\sum_i \alpha_i^{A(BC)} + \frac{1}{8}\sum_{ij} \gamma_{ij}^{A(BC)} + \cdots, \tag{12}$$

where i, j index normal modes of vibration ($i, j \in \{1, 2, \ldots, 3n-6\}$ for n atoms), and α and γ are first and second order interaction constants that couple rotational and vibrational motions together. As a Taylor series, α is generally much larger than γ, so γ is usually neglected. The rovibrational correction α can be calculated using an electronic structure method with a good compromise between computational time and accuracy (Puzzarini & Stanton, 2023; Spaniol et al., 2023). However, this calculation requires knowing the structure in the first place.

Alternatively, experimental rotational constants can be corrected to equilibrium rotational constants by simple semi-empirical scaling factors under marginalized Bayesian uncertainty (Lee & McCarthy, 2020).

But, exact moment constraints do allow application in the following sense: if one were to *a priori* guess the equilibrium moments correctly, one could then verify whether they are indeed correct by generating structures, calculating their rovibrational corrections, and then checking their agreement to the experimental rotational constants. Therefore, given experimental rotational constants $A(BC)_0$, one can use this verification procedure in a fine-grid search for the true equilibrium rotational constants $A(BC)_e$. This is feasible since $A(BC)_e$ consist of only 3 numbers. Experimental precision in $A(BC)_0$ is maintained up to the precision in computing α. To verify that a structure is the true structure, we must know that (1) its $A(BC)_e$ and α match $A(BC)_0$ and (2) gradient norm is near 0.

B THE STIEFEL MANIFOLD

In this section, we discuss various facts about the Stiefel manifold and computations thereon. We refer readers to (Lee, 2003; Edelman et al., 1998; Bendokat et al., 2024) for further details.

The Stiefel manifold $\mathrm{St}(n, p)$ is the set of orthonormal p-frames in $\mathbb{R}^n$:

$$\mathrm{St}(n, p) = \{\boldsymbol{U} \in \mathbb{R}^{n \times p} \mid \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_p\}. \tag{13}$$

$\mathrm{St}(n, p)$ is a manifold of dimension $np - \frac{1}{2}p(p+1)$. Elements of $\mathrm{St}(n, p)$ can be thought of as the first p columns of an $n \times n$ orthogonal matrix. In fact, $\mathrm{St}(n, p)$ generalizes some well-known spaces. For example, $\mathrm{St}(n, n)$ is the orthogonal group $\mathrm{O}(n)$, and $\mathrm{St}(n, n-1)$ is diffeomorphic to the special orthogonal group $\mathrm{SO}(n)$, and $\mathrm{St}(n, 1)$ recovers the unit n-sphere S^{n-1}.

B.1 TANGENT SPACE

The tangent space of $\mathrm{St}(n, p)$ at a point $\boldsymbol{U}$ is identified with the subspace:

$$\mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, p) = \{\Delta \in \mathbb{R}^{n \times p} \mid \boldsymbol{U}^\top \Delta + \Delta^\top \boldsymbol{U} = \boldsymbol{0}\}, \tag{14}$$

In other words, this is the set of $n \times p$ matrices Δ for which $\boldsymbol{U}^\top \Delta$ is skew-symmetric.

B.2 CANONICAL METRIC

Smoothly equipping every tangent space with an inner product turns $\mathrm{St}(n,p)$ into a Riemannian manifold. Herein, we will consider the canonical metric:

$$\langle \Delta, \tilde{\Delta} \rangle_{\boldsymbol{U}} = \operatorname{trace} \Delta^\top \left(\boldsymbol{I}_n - \tfrac{1}{2} \boldsymbol{U}\boldsymbol{U}^\top \right) \tilde{\Delta}, \tag{15}$$

Any inner product induces a norm by $||\Delta||_{\boldsymbol{U}} = \sqrt{\langle \Delta, \Delta \rangle_{\boldsymbol{U}}}$. For the canonical metric, the induced norm is equivalent to:

$$||\Delta||_{\boldsymbol{U}}^2 = \operatorname{trace} \Delta^\top \Delta - \tfrac{1}{2} \operatorname{trace}(\boldsymbol{U}^\top \Delta)^\top (\boldsymbol{U}^\top \Delta) = ||\Delta||_F^2 - \tfrac{1}{2} ||\boldsymbol{U}^\top \Delta||_F^2, \tag{16}$$

where $||\cdot||_F$ is the Frobenius norm. Computing the canonical norm in this way is much faster for the nodewise-batching approach used by PyTorch Geometric (Fey & Lenssen, 2019).

B.3 EXPONENTIAL AND LOGARITHM

The Stiefel exponential can be computed by the algorithm presented by Edelman et al. (1998), who give the closed-form expression for a geodesic γ with initial conditions $\gamma(0) = \boldsymbol{U}$ and $\dot{\gamma}(0) = \Delta$. Here, we reproduce the algorithm for computing $\exp_{\boldsymbol{U}}(\Delta) = \gamma(1)$.

Algorithm 1 Computing a Stiefel geodesic $\gamma(t)$. (Edelman et al., 1998)

Require: Base point $\boldsymbol{U} \in \mathrm{St}(n,p)$, initial direction $\Delta \in \mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n,p)$, time $t \in \mathbb{R}$

$\boldsymbol{Q}\boldsymbol{R} \leftarrow (\boldsymbol{I}_n - \boldsymbol{U}\boldsymbol{U}^\top)\Delta$, where $\boldsymbol{Q} \in \mathbb{R}^{n \times p}$ and $\boldsymbol{R} \in \mathbb{R}^{p \times p}$ ▷ QR decomposition

$\boldsymbol{A} \leftarrow \begin{pmatrix} \boldsymbol{U}^\top \Delta & -\boldsymbol{R}^\top \\ \boldsymbol{R} & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{2p \times 2p}$

$\left(\begin{array}{c|c} \boldsymbol{M}(t) & \cdots \\ \hline \boldsymbol{N}(t) & \cdots \end{array} \right) \leftarrow \exp_m(t\boldsymbol{A})$, where $\boldsymbol{M}(t), \boldsymbol{N}(t) \in \mathbb{R}^{p \times p}$ ▷ take submatrices

return $\boldsymbol{U}\boldsymbol{M}(t) + \boldsymbol{Q}\boldsymbol{N}(t) \in \mathrm{St}(n,p)$

An efficient algorithm for computing the Stiefel logarithm is reproduced and simplified here from Zimmermann & Hüper (2022). We implement the logarithm in C++ for speed on the CPU, as it must be called on every fetch of a data example. This does not present a bottleneck: computing logarithmic maps takes an average of 0.1 ms for molecules in both QM9 and GEOM. The logarithm has a cost of $O(np^2)$. The thin QR decomposition and initial matrix multiplications have a cost of $O(np^2)$, while the inner loop's Schur decomposition, Sylvester solve, matrix multiplications, and matrix exponential have a cost of $O(p^3)$.

Algorithm 2 Computing the Stiefel logarithm. (Zimmermann & Hüper, 2022)

Require: Base point $\boldsymbol{U} \in \mathrm{St}(n,p)$, target point $\tilde{\boldsymbol{U}} \in \mathrm{St}(n,p)$

$\boldsymbol{M} \leftarrow \boldsymbol{U}^\top \tilde{\boldsymbol{U}} \in \mathbb{R}^{p \times p}$

$\boldsymbol{Q}\boldsymbol{N} \leftarrow \tilde{\boldsymbol{U}} - \boldsymbol{U}\boldsymbol{M} \in \mathbb{R}^{n \times p}$ ▷ thin QR

$\boldsymbol{O}\boldsymbol{R} \leftarrow \begin{pmatrix} \boldsymbol{M} \\ \boldsymbol{N} \end{pmatrix} \in \mathbb{R}^{2p \times p}$, $\boldsymbol{O} \in \mathbb{R}^{2p \times 2p}$ ▷ compute orthogonal completion via full QR

$\boldsymbol{V} \leftarrow \begin{pmatrix} \boldsymbol{M} & \boldsymbol{O}_{\cdot,p:2p} \\ \boldsymbol{N} & \end{pmatrix} \in \mathrm{SO}(2p)$ ▷ flip sign if $\det < 0$

for $k = 1, \ldots, 20$ **do**

  $\begin{pmatrix} \boldsymbol{A} & -\boldsymbol{B}^\top \\ \boldsymbol{B} & \boldsymbol{C} \end{pmatrix} \leftarrow \log_m(\boldsymbol{V})$ ▷ use Schur matrix logarithm

  $\boldsymbol{S} \leftarrow \frac{1}{12}\boldsymbol{B}\boldsymbol{B}^\top - \frac{1}{2}\boldsymbol{I}_p$

  solve $\boldsymbol{C} = \boldsymbol{S}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}\boldsymbol{S}$ for $\boldsymbol{\Gamma} \in \mathbb{R}^{p \times p}$ ▷ symmetric Sylvester equation

  $\boldsymbol{\Phi} \leftarrow \exp_m(\boldsymbol{\Gamma}) \in \mathbb{R}^{p \times p}$ ▷ matrix exponential, or use Cayley transform

  $\boldsymbol{V}_{\cdot,p:2p} \leftarrow \boldsymbol{V}_{\cdot,p:2p}\boldsymbol{\Phi}$ ▷ rotate last p columns of $\boldsymbol{V}$

return $\boldsymbol{U}\boldsymbol{A} + \boldsymbol{Q}\boldsymbol{B} \in \mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n,p)$

B.4 UNIFORM SAMPLING

To sample uniformly on $\mathrm{St}(n, p)$ with respect to the Haar measure, we can compute $\boldsymbol{U} = \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1/2}$ for a random matrix $\boldsymbol{Z} \in \mathbb{R}^{n \times p}$ whose elements are drawn i.i.d. from a standard Gaussian (i.e. $\boldsymbol{Z} =$ `randn(n, p)`).

B.5 HOUSEHOLDER REFLECTIONS

To align the final column of a matrix $\boldsymbol{U} \in \mathrm{St}(n, p)$ to a unit vector $\boldsymbol{y}$, we can rotate its columns under a transformation $\boldsymbol{R} \in \mathrm{SO}(n)$ such that $\boldsymbol{R}\boldsymbol{U}_{:,4} = \boldsymbol{y}$. Let $\boldsymbol{U}_{:,4} = \boldsymbol{x}$. For any unit vector $\boldsymbol{v} \in \mathbb{R}^3$, the Householder matrix $\boldsymbol{H}(\boldsymbol{v}) = \boldsymbol{I}_n - 2\boldsymbol{v}\boldsymbol{v}^\top$ has determinant -1. Then,

$$\boldsymbol{R} = \boldsymbol{H}(\boldsymbol{y})\boldsymbol{H}\left(\frac{\boldsymbol{x} + \boldsymbol{y}}{||\boldsymbol{x} + \boldsymbol{y}||}\right) \in \mathrm{SO}(n) \tag{17}$$

is our desired rotation. In particular, we can verify that

$$\boldsymbol{H}\left(\frac{\boldsymbol{x} + \boldsymbol{y}}{||\boldsymbol{x} + \boldsymbol{y}||}\right)\boldsymbol{x} = \boldsymbol{x} - \frac{(2 + 2\boldsymbol{x}^\top \boldsymbol{y})}{||\boldsymbol{x} + \boldsymbol{y}||^2}(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{x} - (\boldsymbol{x} + \boldsymbol{y}) = -\boldsymbol{y} \tag{18}$$

so that $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{H}(\boldsymbol{y})(-\boldsymbol{y}) = \boldsymbol{y}$, as desired.

B.6 ORTHOGONAL PROJECTIONS

The orthogonal projection of a point $\boldsymbol{U}_0 \in \mathbb{R}^{n \times p}$ onto the $\mathrm{St}(n, p)$ is a special case of the well-known orthogonal Procrustes problem:

$$\operatorname*{arg\,min}_{\boldsymbol{U}} ||\boldsymbol{U}\boldsymbol{I}_p - \boldsymbol{U}_0||_F, \quad \text{subject to } \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_p. \tag{19}$$

Letting $\boldsymbol{U}_0 = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^\top$ under a singular value decomposition, the solution to this problem is $\boldsymbol{A}\boldsymbol{B}^\top$. The orthogonal projection of a point $\boldsymbol{Z} \in \mathbb{R}^{n \times p}$ onto the tangent space $\mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, p)$ is computed by $\boldsymbol{Z} - \boldsymbol{U}\,\mathrm{sym}(\boldsymbol{U}^\top \boldsymbol{Z})$, where $\mathrm{sym}(\boldsymbol{A}) = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^\top)$ is a symmetrization operation.

B.7 THE ZERO CENTER-OF-MASS SUBMANIFOLD

For $n \geq 5$, we are interested in the subset of $\mathrm{St}(n, 4)$ obtained by fixing the final column to a fixed unit vector $\boldsymbol{a} \in S^{n-1}$:

$$\mathcal{M} = \{\boldsymbol{U} \in \mathrm{St}(n, p) \mid \boldsymbol{U}_{:,4} = \boldsymbol{a}\}. \tag{20}$$

Note that $S = \pi^{-1}(\{\boldsymbol{a}\})$ is a level set of the projection map:

$$\pi \colon \mathrm{St}(n, 4) \to S^{n-1}, \quad \boldsymbol{U} \mapsto \boldsymbol{U}_{:,4}. \tag{21}$$

In fact, π is a smooth surjective map of constant rank, so it is a submersion by the global rank theorem. Hence, $\mathcal{M}$ is an embedded submanifold by the submersion level set theorem. The tangent vectors to $\mathcal{M}$ are exactly those whose final column is zero, since

$$\mathrm{T}_{\boldsymbol{U}}\mathcal{M} = \ker \mathrm{d}\pi_{\boldsymbol{U}} = \{\Delta \in \mathrm{T}_{\boldsymbol{U}}\mathrm{St}(n, 4) \mid \Delta_{:,4} = \boldsymbol{0}\}. \tag{22}$$

If $\mathcal{M}$ further inherits the canonical metric from $\mathrm{St}(n, 4)$, then it becomes a Riemannian manifold. Note that $\mathcal{M}$ is homeomorphic to $\mathrm{St}(n-1, 3)$, so it is connected and compact. Hence, it is geodesically complete and any two points in $\mathcal{M}$ can be connected with a length-minimizing geodesic on $\mathcal{M}$. Theorem 1 shows that $\mathcal{M}$ is totally geodesic, i.e., any geodesic on $\mathcal{M}$ is a geodesic on $\mathrm{St}(n, 4)$. This allows us to perform simpler computations in the ambient space. Theorem 5 computes projections of arbitrary matrices onto $\mathrm{T}_{\boldsymbol{U}}\mathcal{M}$.

Theorem 1. *As defined above, $\mathcal{M}$ is totally geodesic.*

Proof. A sufficient condition is that for any $\boldsymbol{U} \in \mathcal{M}$ and $\Delta \in \mathrm{T}_{\boldsymbol{U}}\mathcal{M}$, the geodesic γ with initial conditions $\gamma(0) = \boldsymbol{U}$ and $\dot{\gamma}(0) = \Delta$ stays within $\mathcal{M}$. Fortunately, γ is given in closed-form by Edelman et al. (1998) in Algorithm 1. Since $\Delta_{:,4} = \boldsymbol{0}$, the fourth columns of $\boldsymbol{R}$ and $\boldsymbol{U}^\top\Delta$ are zero. Since $\boldsymbol{U}^\top\Delta$ is skew-symmetric by Equation 14, its fourth row is also zero. Then, the fourth row and

column of $\boldsymbol{A}$ and $\exp_m(t\boldsymbol{A})$ are zero, but $\exp_m(t\boldsymbol{A})_{4,4} = 1$. It follows that $\boldsymbol{M}(t)_{\cdot,4} = (0,0,0,1)$ and $\boldsymbol{N}(t)_{\cdot,4} = \boldsymbol{0}$, so that

$$\gamma(t)_{\cdot,4} = \boldsymbol{U}\left(\boldsymbol{M}(t)_{\cdot,4}\right) + \boldsymbol{Q}\left(\boldsymbol{N}(t)_{\cdot,4}\right) = \boldsymbol{U}_{\cdot,4} = \boldsymbol{a}. \tag{23}$$

Hence, $\gamma(t) \in \mathcal{M}$ as desired. □

We now provide some lemmas which are useful for proving that the Stiefel logarithm on $\mathcal{M}$ can be computed using $\mathrm{St}(n,3)$ rather than $\mathrm{St}(n,4)$ (Theorem 4).

Lemma 2. *Let $\boldsymbol{X}$ be a square matrix whose i^{th} row and column are 0. Positive powers of $\boldsymbol{X}$ retain zeros in the i^{th} row and column. Suppose*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{0} & \boldsymbol{F} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \boldsymbol{G} & \boldsymbol{0} & \boldsymbol{H} \end{pmatrix}, \qquad \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}^k = \begin{pmatrix} \hat{\boldsymbol{E}} & \hat{\boldsymbol{F}} \\ \hat{\boldsymbol{G}} & \hat{\boldsymbol{H}} \end{pmatrix}. \tag{24}$$

Then we have

$$\boldsymbol{X}^k = \begin{pmatrix} \hat{\boldsymbol{E}} & \boldsymbol{0} & \hat{\boldsymbol{F}} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \hat{\boldsymbol{G}} & \boldsymbol{0} & \hat{\boldsymbol{H}} \end{pmatrix}. \tag{25}$$

Proof. We show this by induction over k. It is clear this is true for $k = 1$. Then, we assume $\boldsymbol{X}^k = \begin{pmatrix} \hat{\boldsymbol{E}} & \boldsymbol{0} & \hat{\boldsymbol{F}} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \hat{\boldsymbol{G}} & \boldsymbol{0} & \hat{\boldsymbol{H}} \end{pmatrix}$ and aim to show for $k+1$. We know that

$$\begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}^{k+1} = \begin{pmatrix} \hat{\boldsymbol{E}} & \hat{\boldsymbol{F}} \\ \hat{\boldsymbol{G}} & \hat{\boldsymbol{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{E}}\boldsymbol{E} + \hat{\boldsymbol{F}}\boldsymbol{G} & \hat{\boldsymbol{E}}\boldsymbol{F} + \hat{\boldsymbol{F}}\boldsymbol{H} \\ \hat{\boldsymbol{G}}\boldsymbol{E} + \hat{\boldsymbol{H}}\boldsymbol{G} & \hat{\boldsymbol{G}}\boldsymbol{F} + \hat{\boldsymbol{H}}\boldsymbol{H} \end{pmatrix}. \tag{26}$$

At the same time, we have that

$$\boldsymbol{X}^{k+1} = \begin{pmatrix} \hat{\boldsymbol{E}} & \boldsymbol{0} & \hat{\boldsymbol{F}} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \hat{\boldsymbol{G}} & \boldsymbol{0} & \hat{\boldsymbol{H}} \end{pmatrix} \begin{pmatrix} \boldsymbol{E} & \boldsymbol{0} & \boldsymbol{F} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \boldsymbol{G} & \boldsymbol{0} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{E}}\boldsymbol{E} + \hat{\boldsymbol{F}}\boldsymbol{G} & \boldsymbol{0} & \hat{\boldsymbol{E}}\boldsymbol{F} + \hat{\boldsymbol{F}}\boldsymbol{H} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \hat{\boldsymbol{G}}\boldsymbol{E} + \hat{\boldsymbol{H}}\boldsymbol{G} & \boldsymbol{0} & \hat{\boldsymbol{G}}\boldsymbol{F} + \hat{\boldsymbol{H}}\boldsymbol{H} \end{pmatrix}, \tag{27}$$

which completes the proof. □

Lemma 3. *Let $\boldsymbol{X}$ be a square matrix whose i^{th} row and column are 0, and consider the matrix exponential of its nonzero portions:*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{0} & \boldsymbol{F} \\ \boldsymbol{0} & 0 & \boldsymbol{0} \\ \boldsymbol{G} & \boldsymbol{0} & \boldsymbol{H} \end{pmatrix}, \qquad \exp_m \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{E}} & \bar{\boldsymbol{F}} \\ \bar{\boldsymbol{G}} & \bar{\boldsymbol{H}} \end{pmatrix}. \tag{28}$$

Then we have

$$\exp_m(\boldsymbol{X}) = \begin{pmatrix} \bar{\boldsymbol{E}} & \boldsymbol{0} & \bar{\boldsymbol{F}} \\ \boldsymbol{0} & 1 & \boldsymbol{0} \\ \bar{\boldsymbol{G}} & \boldsymbol{0} & \bar{\boldsymbol{H}} \end{pmatrix}. \tag{29}$$

Proof. The matrix exponential is given by

$$\exp_m(\boldsymbol{X}) = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{X}^k. \tag{30}$$

By Lemma 2, all non-identity terms of this series are identical to terms in the series for $\exp_m \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}$ but with zeros inserted in the i^{th} row and column. The identity term then contributes the extra 1 in the main diagonal. □

Theorem 4. *Let $\boldsymbol{U} \in \mathcal{M}$, and let $\tilde{\boldsymbol{U}}$ be its first three columns so that $\boldsymbol{U} = \begin{pmatrix}\tilde{\boldsymbol{U}} & \boldsymbol{a}\end{pmatrix}$.*

The exponential on $\mathcal{M}$ can be computed as the exponential on $\mathrm{St}(n,3)$ after discarding the last column of $\Delta = \begin{pmatrix}\tilde{\Delta} & \boldsymbol{0}\end{pmatrix}$.

$$\exp_{\boldsymbol{U}}(\Delta) = \begin{pmatrix}\exp_{\tilde{\boldsymbol{U}}}(\tilde{\Delta}) & \boldsymbol{a}\end{pmatrix} \tag{31}$$

Similarly, the logarithm on $\mathcal{M}$ can be computed as the logarithm on $\mathrm{St}(n,3)$ followed by concatenating a zero column:

$$\log_{\boldsymbol{U}_0}(\boldsymbol{U}_1) = \begin{pmatrix}\log_{\tilde{\boldsymbol{U}}_0}(\tilde{\boldsymbol{U}}_1) & \boldsymbol{0}\end{pmatrix} \tag{32}$$

Proof. We go line-by-line through Algorithm 1 to show that the Stiefel exponential on $\mathcal{M}$ is equivalent to the exponential for $\mathrm{St}(n,3)$.

The last column of $(\boldsymbol{I}_n - \boldsymbol{U}\boldsymbol{U}^\top)\Delta$ is 0 because the last column of Δ is 0. Then,

$$\boldsymbol{Q}\boldsymbol{R} = (\boldsymbol{I}_n - \boldsymbol{U}\boldsymbol{U}^\top)\Delta = \begin{pmatrix}\tilde{\boldsymbol{Q}} & \boldsymbol{q}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{R}} & \boldsymbol{0}\\ \boldsymbol{0} & 0\end{pmatrix}, \tag{33}$$

where $\boldsymbol{q}$ is some orthogonal vector to the 3 columns of $\tilde{\boldsymbol{Q}}$, and $\tilde{\boldsymbol{R}}$ is the first 3 columns and rows of $\boldsymbol{R}$. This corresponds to the QR decomposition $\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{R}} = (\boldsymbol{I}_n - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^\top)\tilde{\Delta}$.

In addition, we have that $\boldsymbol{U}^\top\Delta = \begin{pmatrix}\tilde{\boldsymbol{U}}^\top\tilde{\Delta} & \boldsymbol{0}\\ \boldsymbol{0} & 0\end{pmatrix}$.

Thus, the block matrix $\boldsymbol{A}$ in Algorithm 1 is given as

$$\begin{pmatrix}\tilde{\boldsymbol{U}}^\top\tilde{\Delta} & \boldsymbol{0} & -\tilde{\boldsymbol{R}}^\top & \boldsymbol{0}\\ \boldsymbol{0} & 0 & \boldsymbol{0} & 0\\ \tilde{\boldsymbol{R}} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & 0 & \boldsymbol{0} & 0\end{pmatrix}. \tag{34}$$

By Lemma 3, the matrix exponential of $\boldsymbol{A}$ containing extra rows and columns of zeros is equivalent to the matrix exponential with these rows and columns of zeros removed.

Therefore,

$$\exp_m(t\boldsymbol{A}) = \left(\begin{array}{c|c}\boldsymbol{M}(t) & \cdots\\ \hline \boldsymbol{N}(t) & \cdots\end{array}\right) = \left(\begin{array}{cc|c}\tilde{\boldsymbol{M}}(t) & \boldsymbol{0} & \cdots\\ \boldsymbol{0} & 1 & \cdots\\ \hline \tilde{\boldsymbol{N}}(t) & \boldsymbol{0} & \cdots\\ \boldsymbol{0} & 0 & \cdots\end{array}\right). \tag{35}$$

Finally, the output of Algorithm 1 is

$$\begin{aligned}\boldsymbol{U}\boldsymbol{M}(t) + \boldsymbol{Q}\boldsymbol{N}(t) &= \begin{pmatrix}\tilde{\boldsymbol{U}} & \boldsymbol{a}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{M}}(t) & \boldsymbol{0}\\ \boldsymbol{0} & 1\end{pmatrix} + \begin{pmatrix}\tilde{\boldsymbol{Q}} & \boldsymbol{q}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{N}}(t) & \boldsymbol{0}\\ \boldsymbol{0} & 0\end{pmatrix} && (36)\\ &= \begin{pmatrix}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{M}}(t) & \boldsymbol{a}\end{pmatrix} + \begin{pmatrix}\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{N}}(t) & \boldsymbol{0}\end{pmatrix} && (37)\\ &= \begin{pmatrix}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{M}}(t) + \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{N}}(t) & \boldsymbol{a}\end{pmatrix} && (38)\\ &= \begin{pmatrix}\exp_{\tilde{\boldsymbol{U}}}(\tilde{\Delta}) & \boldsymbol{a}\end{pmatrix}. && (39)\end{aligned}$$

Since the Stiefel exponential is locally invertible, this also shows that the Stiefel logarithm can be computed using $\tilde{\boldsymbol{U}}_0$ and $\tilde{\boldsymbol{U}}_1$.

□

Theorem 5. *Let $\boldsymbol{U} \in \mathcal{M}$, and let $\tilde{\boldsymbol{U}}$ be its first three columns so that $\boldsymbol{U} = [\tilde{\boldsymbol{U}}\ \boldsymbol{a}]$. Given $\boldsymbol{Z} \in \mathbb{R}^{n\times 3}$, the minimum-norm projection of $[\boldsymbol{Z}\ \boldsymbol{0}]$ onto $\mathrm{T}_{\boldsymbol{U}}\mathcal{M}$ is given by $[\pi(\boldsymbol{Z})\ \boldsymbol{0}]$, where*

$$\begin{aligned}\pi(\boldsymbol{Z}) &= \tilde{\boldsymbol{U}}\,\mathrm{skew}(\tilde{\boldsymbol{U}}^\top\boldsymbol{Z}) + (\boldsymbol{I}_n - \boldsymbol{U}\boldsymbol{U}^\top)\boldsymbol{Z} \in \mathbb{R}^{n\times 3} && (40)\\ &= \tilde{\boldsymbol{U}}\,\mathrm{skew}(\tilde{\boldsymbol{U}}^\top\boldsymbol{Z}) + (\boldsymbol{I}_n - \boldsymbol{U}\boldsymbol{U}^\top)\boldsymbol{Z} \in \mathbb{R}^{n\times 3} && (41)\end{aligned}$$

Note there is no tilde in the second term.

Proof. We can rewrite

$$\mathrm{T}_{\boldsymbol{U}}\mathcal{M} = \{[\tilde{\Delta}\ \mathbf{0}] \mid \tilde{\boldsymbol{U}}^\top\tilde{\Delta} + \tilde{\Delta}^\top\tilde{\boldsymbol{U}} = \mathbf{0},\ \tilde{\Delta}^\top\boldsymbol{a} = \mathbf{0}\} \tag{42}$$

$$= \{[\tilde{\Delta}\ \mathbf{0}] \mid \tilde{\Delta} \in \mathrm{T}_{\tilde{\boldsymbol{U}}}\mathrm{St}(n,3) \cap \mathrm{span}(\boldsymbol{a})^\perp\}, \tag{43}$$

Edelman et al. (1998) give an orthogonal projection of $\boldsymbol{Z}$ onto $\mathrm{T}_{\tilde{\boldsymbol{U}}}\mathrm{St}(n,3)$ as:

$$\pi_1(\boldsymbol{Z}) = \tilde{\boldsymbol{U}}\,\mathrm{skew}(\tilde{\boldsymbol{U}}^\top\boldsymbol{Z}) + (\boldsymbol{I}_n - \tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}^\top)\boldsymbol{Z}, \tag{44}$$

where $\mathrm{skew}(\boldsymbol{A}) = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^\top)$. The orthogonal projection of $\boldsymbol{Z}$ onto $\mathrm{span}(\boldsymbol{a})^\perp$ is

$$\pi_2(\boldsymbol{Z}) = (\boldsymbol{I}_n - \boldsymbol{a}\boldsymbol{a}^\top)\boldsymbol{Z}. \tag{45}$$

We can check that π_1 and π_2 commute with $\pi = \pi_1 \circ \pi_2 = \pi_2 \circ \pi_1$. Hence, π is an orthogonal projection onto $\mathrm{T}_{\tilde{\boldsymbol{U}}}\mathrm{St}(n,3) \cap \mathrm{span}(\boldsymbol{a})^\perp$, as desired. □

C Experimental Details

C.1 Architecture

Explicitly, the neural network takes in moments (P_X, P_Y, P_Z), time t, atom types $\boldsymbol{a}$, and coordinates $\boldsymbol{X}$, and outputs a Stiefel tangent vector. Moments are embedded using sinusoidal features with wavelength geometrically spaced from 0.0001 to 10,000. Time is similarly embedded but with a wavelength range from 0.001 to 1. We use a reflection-equivariant network. Note that given a reflection-*invariant* function f, the mapping $\boldsymbol{X} \mapsto \mathrm{sign}(\boldsymbol{X}) \odot f(\boldsymbol{X})$ is reflection-equivariant, where $\mathrm{sign}(\boldsymbol{X})$ gives the element-wise signs of $\boldsymbol{X}$, with $\mathrm{sign}(0) = 0$. Thus, the problem is reduced to constructing a network that is reflection-invariant with respect to the input coordinates $\boldsymbol{X}$.

As input to our network, we begin by featurizing the molecule in a reflection-invariant manner. We obtain invariant node features $\boldsymbol{h}_i \in \mathbb{R}^{d_{\text{node}}}$ by using the molecule's unsigned coordinates and atom types, and edge features $\boldsymbol{e}_{ij} \in \mathbb{R}^{d_{\text{edge}}}$ are computed from the unsigned differences between pairs of atomic coordinates. These features are passed through a Transformer backbone (Vaswani et al., 2017). We use the PreLN layout with an adaptive version of LayerNorm (Dieleman et al., 2022; Dhariwal & Nichol, 2021) that conditions on the timestep and molecule's moments. In addition, the attention module is replaced with a message-passing block that jointly updates the node and edge features:

$$(\boldsymbol{v}_{ij}, a_{ij}, \boldsymbol{e}'_{ij}) \leftarrow \mathrm{MLP}(\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{e}_{ij}), \quad \text{for all } i, j, \tag{46}$$

$$\boldsymbol{e}_{ij} \leftarrow \boldsymbol{e}_{ij} + \boldsymbol{e}'_{ij}, \quad \text{for all } i, j, \tag{47}$$

$$\boldsymbol{y}_i \leftarrow \sum_{j=1}^{n} \left(\frac{\exp(a_{ij})}{\sum_{k=1}^{n} \exp(a_{ik})} \right) \boldsymbol{v}_{ij}, \quad \text{for all } i, \tag{48}$$

$$\boldsymbol{h}_i \leftarrow \boldsymbol{h}_i + \mathrm{Linear}(\boldsymbol{y}_i) \quad \text{for all } i. \tag{49}$$

The last two equations are reminiscent of self-attention in Transformers, and along the same lines, we use a multi-headed extension of them. Across all experiments, we use $d_{\text{node}} = 768$ and $d_{\text{edge}} = 192$. We use 16 Transformer-like blocks, 12 update heads, SiLU/Swish activations, and a $4\times$ expansion in each block's feed-forward module. In total, our model has 154M trainable parameters.

This architecture is distinct from the pretrained network of KREED (Cheng et al., 2024), which was tailored for predicting 3D structure given molecular formula, moments of inertia, and *substitution coordinates*. KREED was trained using random dropout of input substitution coordinates. For QM9, the model sometimes observed examples with no substitution coordinates during training. However, for GEOM, the model always received at least some substitution coordinates during training. Therefore, KREED is operating out-of-distribution when provided with no substitution coordinates at all.

C.2 Stochasticity

Diversity is important for identifying unknown molecules outside the training set. Therefore, we experiment with adding stochasticity to the dynamics of the flow, leading individual paths to be stochastic. To do so during training, we apply an exponential map to a Gaussian variable with noise scale $\gamma \cdot \frac{\cos(\pi t)+1}{2}$ in the tangent space of the interpolant before calculating $\dot{\boldsymbol{U}}$ again. During sampling, we add Gaussian noise of the same noise schedule to the tangent vector at every time step.

Table 3: General training and sampling hyperparameters.

	Hyperparameter	QM9	GEOM
Training	Epochs	1000	60
	Effective batch size	1024	96
	Optimizer	AdamW	AdamW
	Learning rate	10^{-4}	10^{-4}
	Learning rate warmup steps	2000	2000
	Weight decay	0.01	0.01
	Gradient clipping	yes	yes
	EMA decay	0.9995	0.9995
KREED	Timesteps	1000	1000
	Schedule	polynomial	polynomial
Stiefel FM	Timesteps	200	200

Table 4: Training and sampling hyperparameters for Stiefel Flow Matching.

Dataset	Model	Timestep sampling	OT	stochasticity γ
QM9	Stiefel FM	uniform	no	0.00
	Stiefel FM-OT	uniform	yes	0.00
	Stiefel FM-OT-stoch	uniform	yes	0.10
	Stiefel FM-ln	logit-normal	no	0.00
	Stiefel FM-ln-OT	logit-normal	yes	0.00
GEOM	Stiefel FM	uniform	no	0.00
	Stiefel FM-OT	uniform	yes	0.00

Table 5: Extended ablation study for QM9.

Method	% < RMSD ↑ 0.25 Å	% < RMSD ↑ 0.10 Å	Error ↓	Valid ↑	Stable ↓	Diverse ↑	NFE ↓
Stiefel FM	**15.17 ± 0.31**	**13.82 ± 0.30**	0.00	0.882	-1.125	1.040	200
Stiefel FM-OT	13.99 ± 0.30	12.68 ± 0.29	0.00	0.835	-1.039	1.045	200
Stiefel FM-stoch	**15.13 ± 0.31**	**13.83 ± 0.30**	0.00	0.877	-1.116	1.045	500
Stiefel FM-ln	**15.74 ± 0.32**	11.45 ± 0.28	0.00	0.880	-0.600	0.982	200
Stiefel FM-ln-OT	14.90 ± 0.31	12.45 ± 0.29	0.00	0.875	-0.687	1.026	200

Method	Dataset	Training (min / epoch)	Training (it / s)	Sampling (seconds / K=10 samples)
KREED-XL	QM9	1.02	6.7	13.9
Stiefel FM	QM9	1.32	5.1	2.9
Stiefel FM-OT	QM9	3.48	1.9	2.9
KREED-XL	GEOM	225.6	17.0	71.3
Stiefel FM	GEOM	224.4	17.1	15.0
Stiefel FM-OT	GEOM	229.8	16.7	15.0

Table 6: Training and sampling timing on QM9 and GEOM. A CPU bottleneck of computing logarithms and optimal transport exists for QM9 due to a large batch size of 256, but this CPU bottleneck disappears for GEOM due to a small batch size of 24.

C.3 Training

Models were trained on 4 NVIDIA A100 40GB GPUs. Tables 3 and 4 gives our hyperparameters. We use the adaptive gradient clipping strategy from Hoogeboom et al. (2022).

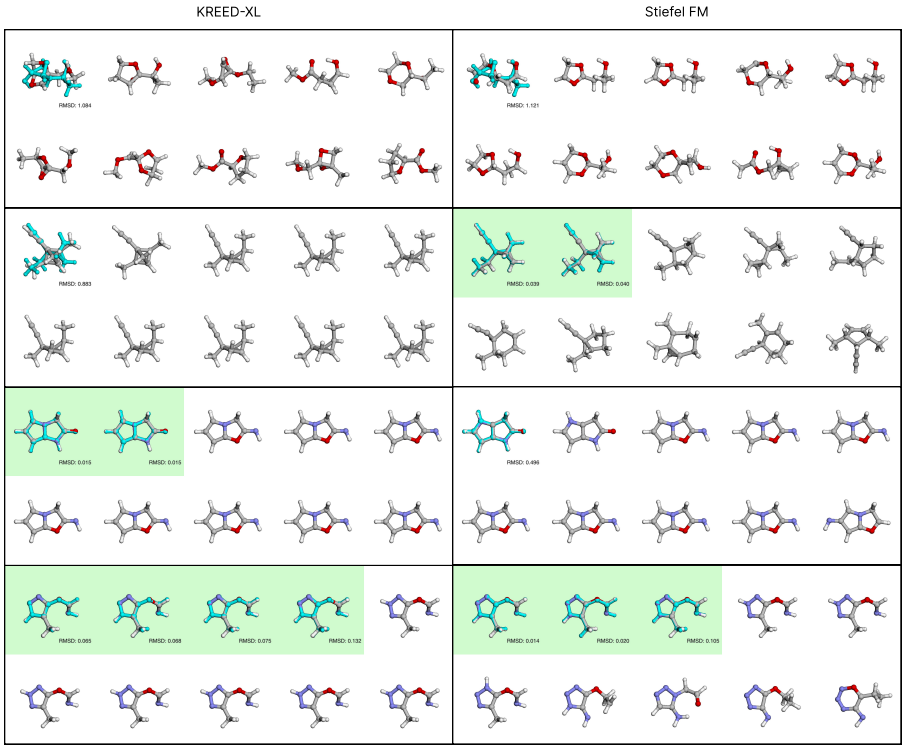

Figure 5: Selected QM9 examples. Best viewed zoomed in. Examples are sorted by RMSD to ground truth, which is shown in cyan. Green panels indicate meeting the threshold of 0.25 Å.

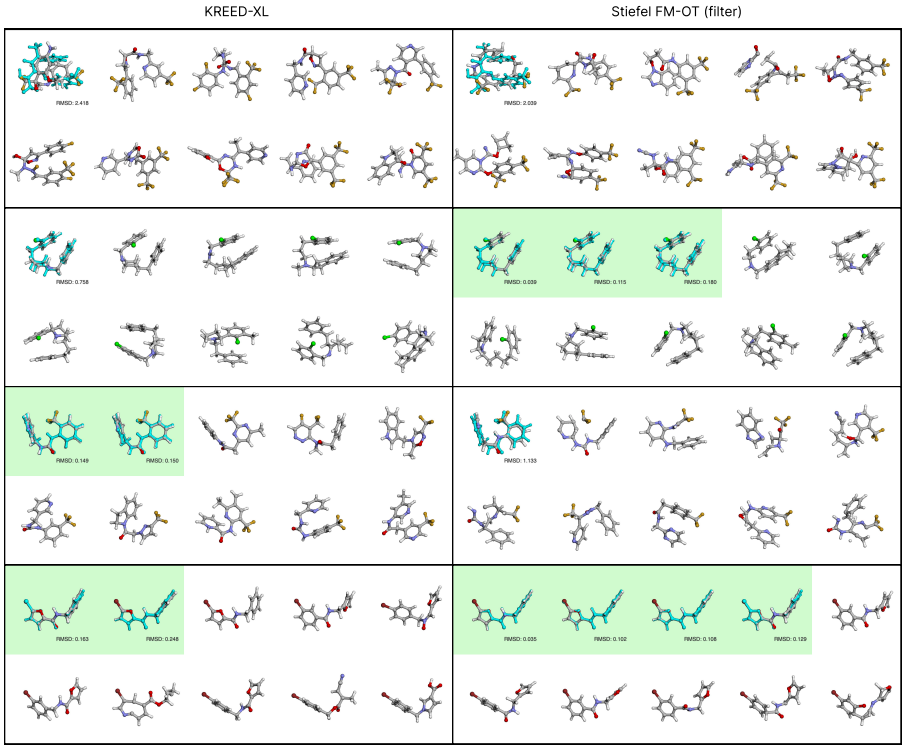

Figure 6: Selected GEOM examples. Best viewed zoomed in. Examples are sorted by RMSD to ground truth, which is shown in cyan. Green panels indicate meeting the threshold of 0.25 Å.

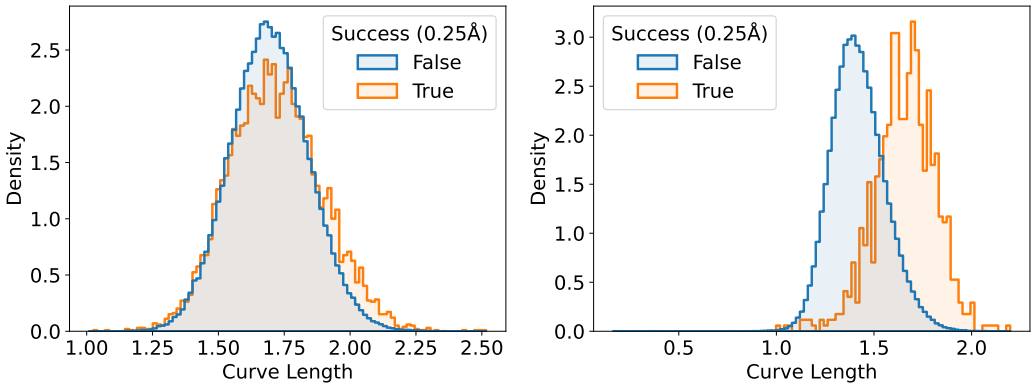

Figure 7: Normalized histogram of curve lengths of success (0.25 Å) and failure examples of Stiefel FM on the QM9 (left) and GEOM (right) datasets. Success cases are more often the result of longer generation paths.

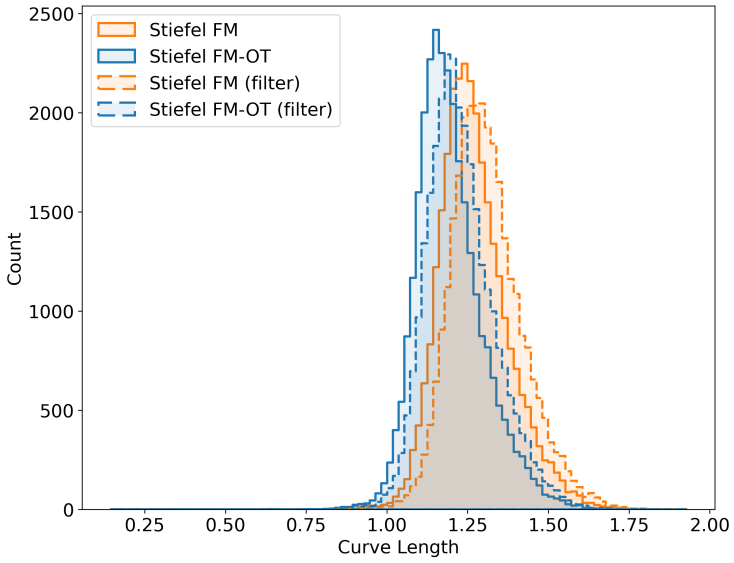

Figure 8: Equivariant optimal transport shortens generation trajectories for GEOM. However, generating more samples and filtering by validity has a slight bias towards *longer* generation trajectories.

C.4 Sampling

During sampling, we sample uniformly from $\mathcal{M}$ and iteratively query the trained model for a tangent vector Δ at every step. If stochasticity is turned on, Gaussian noise with scale $\gamma \cdot \frac{\cos(\pi t)+1}{2}$ is added to this tangent vector. At every step, $\boldsymbol{U}_t$ is projected onto the manifold (Appendix B.6) before projecting the tangent vector to the tangent space on the manifold $\mathrm{T}_{\boldsymbol{U}_t}\mathrm{St}(n,4)$ (Theorem 5). Integration proceeds by applying $\exp_{\boldsymbol{U}_t}$ to this tangent vector, scaled by dt.

Algorithm 3 Sampling under Stiefel Flow Matching.

Require: Size n, atom types $\boldsymbol{a}$, moments $P_{XYZ} = (P_X, P_Y, P_Z)$, stochasticity γ, timesteps T

$\boldsymbol{U} \sim \mathrm{Uniform}(\mathrm{St}(n, 4))$ using Appendix B.4
Project $\boldsymbol{U}$ to $\mathcal{M}$ (defined in Equation 5) using Appendix B.5
$t \leftarrow 0$
$t_\Delta \leftarrow 1/T$
for step in $1, \ldots, T$ **do**
    $\boldsymbol{X} \leftarrow \boldsymbol{X}(\boldsymbol{U}, \boldsymbol{m}, P_{XYZ})$ by inverting Equation 4
    $\Delta \leftarrow v_\theta(t, \boldsymbol{X}, P_{XYZ}, \boldsymbol{a}) \in \mathbb{R}^{n\times 3}$
    if step $< T$ **then**
        $\Delta \leftarrow t_\Delta \Delta + \frac{1}{2}\gamma\sqrt{t_\Delta}(\cos(\pi t) + 1)\boldsymbol{\varepsilon}$, where $\varepsilon_{ij} \sim \mathcal{N}(0, 1)$
    else
        $\Delta \leftarrow t_\Delta \Delta$
    Project Δ to $\mathrm{T}_{\boldsymbol{U}}\mathcal{M}$ using Theorem 5
    $\boldsymbol{U} \leftarrow \exp_{\boldsymbol{U}}(\Delta)$
    $t \leftarrow t + t_\Delta$
return $\boldsymbol{X} \leftarrow \boldsymbol{X}(\boldsymbol{U}, \boldsymbol{m}, P_{XYZ})$ by inverting Equation 4

D GREEDY RANDOM OPTIMAL ASSIGNMENT ALGORITHM

The computational cost of computing the optimal transport map is $O(np^2)$, as it relies on a fixed number of computations of the logarithm.

Algorithm 4 Heuristic alignment algorithm.

Require: $\boldsymbol{U}_0, \boldsymbol{U}_1 \in \mathrm{St}(n, p)$, atom types $\boldsymbol{a}$, number of restarts R, local search budget L

$c^* \leftarrow \infty$ ▷ lowest cost seen so far
for reflections $\boldsymbol{R} \in \{\mathrm{diag}(\boldsymbol{s}) \mid \boldsymbol{s} \in \{-1, +1\}^3\}$ **do** ▷ find a good reflection
    for $k = 0, 1, \ldots, R$ **do**
        sample a random atom-type-preserving node permutation $\boldsymbol{\Pi}$
        $\boldsymbol{U}_{\text{cand}} \leftarrow \boldsymbol{\Pi}\boldsymbol{U}_0\boldsymbol{R}^\top$
        $c \leftarrow \tilde{d}(\boldsymbol{U}_{\text{cand}}, \boldsymbol{U}_1)$, ▷ approximate distance
        if $c < c^*$ **then**
            $c^* \leftarrow c$
            $\boldsymbol{U}_{\text{best}} \leftarrow \boldsymbol{U}_{\text{cand}}$
for $k = 0, 1, \ldots, L$ **do** ▷ local search over node swaps
    sample (i, j) such that $\boldsymbol{a}_i = \boldsymbol{a}_j$, without repeating pairs
    $\boldsymbol{U}_{\text{cand}} \leftarrow \mathrm{swap}(\boldsymbol{U}_{\text{best}}, i, j)$
    $c \leftarrow \tilde{d}(\boldsymbol{U}_{\text{cand}}, \boldsymbol{U}_1)$, ▷ approximate distance
    if $c < c^*$ **then**
        $c^* \leftarrow c$
        $\boldsymbol{U}_{\text{best}} \leftarrow \boldsymbol{U}_{\text{cand}}$
return $\boldsymbol{U}_{\text{best}}$

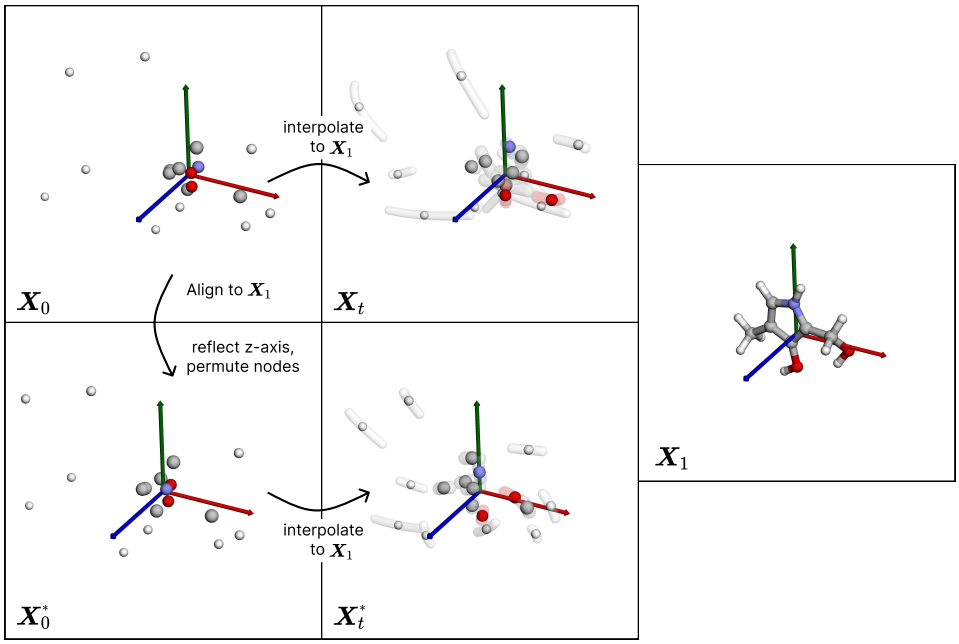

Figure 9: Equivariant optimal transport aligns noise samples $\boldsymbol{X}_0$ to training examples $\boldsymbol{X}_1$ over reflections and permutations, leading to smoother and shorter paths shown to the model during training.

E STIEFEL LOGARITHM EMPIRICAL ANALYSIS

We sample 100k training examples from QM9 or GEOM to be used as $\boldsymbol{U}_1$ and for each example sample one random point $\boldsymbol{U}_0$. We compute the true logarithm $\Delta = \log_{\boldsymbol{U}_0}(\boldsymbol{U}_1)$ using a large number of iterations and compare it to the 20-iteration truncated logarithm. We record error as the infinity norm of the difference between the true and approximate logarithms. We set the convergence threshold to be 1e-6.

For QM9, the 20-iteration logarithm converges 97.9% of the time (median 9 iterations to converge), and the median error in case of nonconvergence is 2.4e-4. The Spearman correlation between the 1-iteration approximate distance and the true distance is $\rho = 0.88$. For GEOM, the 20-iteration logarithm converges 99.7% of the time (median 9 iterations to converge), and the median error in case of nonconvergence is 8.8e-5. The Spearman correlation between the 1-iteration approximate distance and the true distance is $\rho = 0.82$.

These results empirically validate that the 1-iteration approximate distance used in Algorithm 4 is an upper bound on the true distance, and that it is a valid heuristic which generally maintains the same relative ordering as the true Stiefel distance.

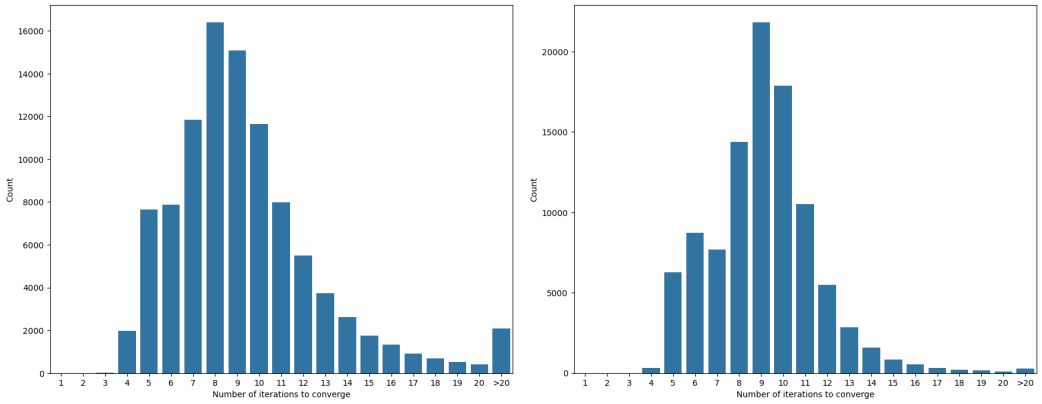

Figure 10: Histograms of the number of inner iterations of the Stiefel logarithm required to converge to an error of 1e-6 for 100k random training examples of QM9 (left) and GEOM (right).

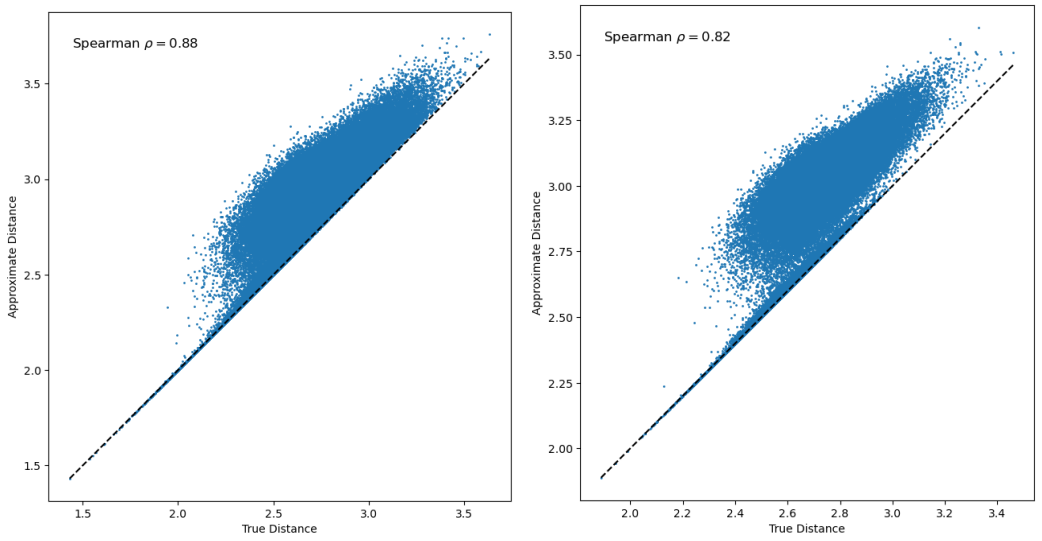

Figure 11: Parity plots comparing the 1-iteration approximate Stiefel distance to the true Stiefel distance of QM9 (left) and GEOM (right).