arXiv:2602.15022v1 [cs.LG] 16 Feb 2026

通过规范化重新思考具有对称性的扩散模型及其在分子图生成中的应用

Cai Zhou [*1] **Zijie Chen** [*2] **Zian Li** [3] **Jike Wang** [2] **Kaiyi Jiang** [1 4] **Pan Li** [5] **Rose Yu** [6] **Muhan Zhang** [3] **Stephen Bates** [1] **Tommi Jaakkola** [1]

摘要

化学与科学中的许多生成任务涉及对群对称性（例如置换和旋转）不变的分布。一种常见策略通过架构约束来强制不变性与等变性，例如等变去噪器和不变先验。本文以另一种规范化视角挑战这一传统：首先将每个样本映射为具有规范姿态或顺序的轨道代表元，在典范切片上训练无约束（非等变）的扩散或流模型，最后在生成时采样随机对称变换以恢复不变分布。基于形式化的商空间视角，本工作通过证明以下内容，为规范扩散提供了全面理论：(i) 典范生成模型相对于不变目标的正确性、通用性与更强表达能力；(ii) 规范化通过消除由群混合引起的扩散分数复杂度，并降低流匹配中的条件方差，从而加速训练。我们进一步表明，对齐先验与最优传输同规范化互补作用，并能进一步提升训练效率。我们将该框架实例化为 $S_n \times SE(3)$ 对称性下的分子图生成。通过利用基于几何谱的规范化以及温和的位置编码，规范扩散在三维分子生成任务中显著优于等变基线，且计算量相近甚至更少。此外，借助新颖架构 *Canon*，CanonFlow 在具有挑战性的 GEOM-DRUG 数据集上取得了最先进的性能，并且在少步生成中优势依然显著。

1. 引言

生成建模从根本上植根于数据的几何结构。在计算机视觉和自然语言处理（NLP）等领域中，数据呈现出特定的对称性——例如，图像中的物体具有平移不变性，而文本的语义则依赖于顺序。近期的扩散（Song 与 Ermon，2019年；Ho 等人，2020；Song 等人，2021年）以及基于流的（Liu 等人，2022年；Lipman 等人，2022年；Albergo 等人，2023年）生成模型已在图像（Dhariwal 与 Nichol，2021年；Rombach 等人，2022年）、视频（Ho 等人，2022年；Singer 等人，2022年）、文本（Li 等人，2022年；Gong 等人，2022年）和生物分子（Watson 等人，2023年；Corso 等人，2023年）等方面取得了巨大成功。然而，至关重要的是，图像等模态并非对所有变换都完全不变：倒置的风景或颠倒的句子通常会失去其语义有效性。因此，这些领域中最先进的扩散模型会显式地打破对称性：它们利用固定网格拓扑或注入位置编码（PE），将生成过程锚定到规范方向。

相比之下，分子生成作用于一个在根本上不同的几何空间，该空间由置换与欧几里得对称性的直积定义，$S_N \times SE(3)$。不同于自然图像，旋转后的分子或重新索引的分子图都表示完全相同的物理实体。分子生成模型通常通过构建等变架构和/或不变先验来施加这些约束（Hoogeboom 等人，2022年；Xu 等人，2022年；Tian 等人，2024），从而使学习到的向量场（得分/速度）遵循群作用，并使生成分布与对称性保持一致。尽管这种方法具有理论依据，但通常会带来显著的架构和计算开销（例如等变层、张量代数），还可能掩盖另一个挑战：对称性会产生潜在的“规范”歧义，因此中间噪声状态可能对应于多个等价的群变换构型。这种复杂的混合分布特性会导致“轨迹交叉”和冲突梯度，因为模型难以确定应从对称噪声向量生成哪一种有效方向，从而使学习到的动力学不够直线。

受这一观察的启发，我们认为，图像生成的效率

[*]同等贡献 [1]麻省理工学院 [2]浙江大学 [3]北京大学 [4]普林斯顿大学 [5]佐治亚理工学院 [6]加利福尼亚大学圣迭戈分校。通信作者：CaiZhou <caiz428@mit.edu>。

预印本。*2026*年*2*月*17*日。

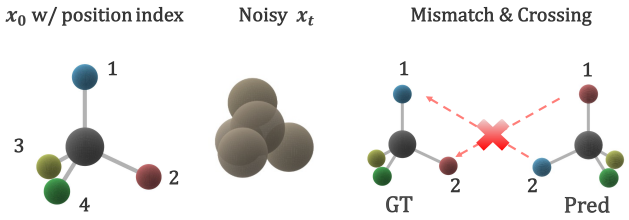


图 *1.* 规范扩散的动机：在扩散模型中，按位置进行监督会使等价分子在对称性下产生失配损失。

恰恰源于其不具备完全不变性。我们提出通过**规范化**将这一优势迁移到分子建模中。在这项工作中，我们质疑等变生成模型的必要性和有效性，并提出一种新颖的规范扩散框架，从理论上分析这一对称性破缺过程如何加速扩散训练，同时确保相较于不变或等变目标的有效性。通过将分子的商空间映射到唯一的典范截面——在训练过程中显式打破对称性——我们使噪声分布与数据分布相互对齐。这有效解决了轨迹失配问题，将复杂的等变生成任务转化为沿典范流形进行的简化传输问题。

我们的出发点是流匹配恒等式：不可约回归误差等于给定中间状态时端点位移的期望条件方差。在第 3 节中，我们证明，在群对齐提升构造下，这个条件方差可分解为两个非负分量：切片内项反映了典范切片上真实的传输难度；对称性歧义项则完全源于对潜在群元进行边缘化。规范化通过固定规范（即在每个轨道上选取一个典范代表元）消除了对称性歧义项；而类似最优传输的耦合则通过使耦合更接近类 Monge，并使轨迹更接近直线传输，来针对切片内项。

基于上述分析，在第4 节中，我们提出了一种实用的对称性感知分子生成流程，将规范化器与灵活的非等变骨干网络相结合。我们通过联合固定置换规范和旋转规范来对分子进行规范化（例如，为S_N 采用Fiedler向量排序，为 $SO(3)$采用秩锚定标架），然后在典范空间中使用无间隙先验和规范条件直接训练扩散/流模型。这为完全等变生成建模提供了一种简单而强大的替代方案：它减少了对称性诱导的歧义，借助更直的学习传输提高了少步准确率，并释放了通用Transformer/图神经网络在大规模分子生成中的表达能力和计算优势。此外，我们还开发了一种名为 ***Canon*** （图6）的新颖架构，它显式地融入并细化额外原子隐状态中的规范信息，从而实现典范性感知去噪。

从实验结果来看，规范化与任何（等变或非等变）基础模型兼容。第 5 节表明，SemlaFlow 等基线（Irwin 等人，2024）的规范化模型在包括 QM9 和 GEOM-DRUG 在内的热门无条件三维分子生成基准上取得了显著更好的结果，同时计算开销完全可以忽略不计。此外，借助“Canon”架构，*CanonFlow* 在分子稳定性和有效性指标方面达到了最先进水平，以较大幅度超越了所有基线。值得注意的是，规范化模型展现出更快的训练收敛速度，以及更高质量的少步采样结果。

2. 预备知识

对称数据的规范化。 我们在一个可测空间 $\mathcal{M}$上处理结构化对象。在分子生成中，本文始终使用以下状态：

$$\mathbf{Z} = (\mathbf{X}, \mathbf{H}, \mathbf{A}) \in \mathcal{M},$$

其中，$\mathbf{X} \in \mathbb{R}^{N\times 3}$ 表示坐标；$\mathbf{H} \in \mathbb{R}^{N\times d_h}$ 表示原子特征，其中每个元素由若干 d_h标量组合而成，例如原子类型 $\{1,\dots,K\}^N$ 和形式电荷；$\mathbf{A} \in \mathbb{R}^{N\times N\times d_e}$ 对化学键/边特征进行编码。

令群作用 $\mathcal{G}$ 作用于 $\mathcal{M}$，其定义为

$$g\cdot\mathbf{Z} = (R,\ t,\ \pi)\cdot(\mathbf{X},\mathbf{H},\mathbf{A}) := (\pi(\mathbf{X})R^\top + \mathbf{1}t^\top,\ \pi(\mathbf{H}),\ \pi(\mathbf{A})), \tag{1}$$

其中（R，t）$\in SE(3)$，且 $\pi \in S_N$（节点及邻接关系的同步置换）。

定义 2.1（轨道与商空间）。 $\mathbf{Z}$ 的轨道为$\mathcal{O}$（$\mathbf{Z}$）$= \{g\cdot \mathbf{Z}:\ g \in \mathcal{G}\}$。商空间为$\mathcal{M}/\mathcal{G} = \{\mathcal{O}(\mathbf{Z}):\ \mathbf{Z}\in\mathcal{M}\}$。

定义 2.2（不变性）。 若概率测度 μ 在 $\mathcal{M}$上满足 $\mathcal{G}$不变，则称其为 $g\#\,\mu = \mu$，对所有 $g \in \mathcal{G}$均成立。

现在，我们将规范化 引入为轨道选择。

定义 2.3（规范化映射与切片）。 若一个可测映射 Ψ：$\mathcal{M}\to\mathcal{M}$ 满足 (i)Ψ（$\mathbf{Z}$）$\in\mathcal{O}$（$\mathbf{Z}$），且 (ii) Ψ（$g\cdot\mathbf{Z}$）$=\Psi$（$\mathbf{Z}$），对所有 $g\in\mathcal{G}$ 均成立，则称其为规范化映射 。其像 S：$=\Psi(\mathcal{M})$ 是一个典范切片。

注：2.4（稳定子与非唯一性）。如果 $\mathbf{Z}$ 具有非平凡稳定子 $\mathrm{Stab}(\mathbf{Z}) = \{g:\ g\cdot\mathbf{Z} = \mathbf{Z}\}$（例如图自同构或对称几何体），则规范代表可能并不唯一。这本质上会导致不连续性或多值选择；加权/概率标架是一种解决方法（Dym 等人，2024）。

幸运的是，我们有以下假设；对于含噪真实世界三维分子而言，该假设通常是合理的（精确对称性的测度为零）。

假设 2.5（几乎处处自由作用）。 在 p_0 条件下，稳定器几乎必然是平凡的。

存在一个有限且平移不变的测度（哈尔测度 λ），在紧致拓扑群上可以将其归一化，使总测度为 1。对于置换 S_N，Haar 测度是均匀的；对于 $SO(3)$，Haar 测度是均匀旋转测度；但对于平移，Haar 测度并非有限。因此，遵循先前的工作（Hoogeboom 等人，2022年；Li 等人，2025），在实践中我们通过中心化（例如，质心中心化）去除平移，从而得到一个紧致对称群 $SO(3) \times S_N$。

假设 2.6（中心化表示）。 在全文中，我们假设 $\mathbf{X}$ 是零中心的，从而去除全局平移；剩余的群是紧致的：$\mathcal{G} = SO(3) \times S_N$，其 Haar 概率测度为 λ。

扩散模型与流匹配模型。 我们给出连续时间表述（离散时间去噪扩散概率模型（Ho 等人，2020）和连续时间马尔可夫链变体也类似）。

设 $\mathbf{Z}_0 \sim p_0$ 为数据。我们首先介绍基于得分的扩散（Song 等人，2021年）。考虑在 $\mathcal{M}$（例如 $\mathbf{X}$ 等连续部分）的环境欧几里得嵌入空间上定义的前向加噪随机微分方程，标准的方差保持随机微分方程为

$$\mathrm{d}\mathbf{Z}_t = -\frac{1}{2}\beta(t)\mathbf{Z}_t\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}W_t, \tag{2}$$

其边缘分布为 p_t，得分为 $s_\star(\mathbf{Z}\ t) = \nabla_{\mathbf{Z}} \log p_t(\mathbf{Z})$，并由参数化得分网络学习。

一些近期生成模型基于等价的流匹配视角构建，通过学习时变向量场 $v_t(\mathbf{Z})$，经由常微分方程将先验 p_1 传输至数据 p_0：

$$\frac{\mathrm{d}\mathbf{Z}_t}{\mathrm{d}t} = v_t(\mathbf{Z}_t), \qquad \mathbf{Z}_1 \sim p_1, \quad \mathbf{Z}_0 \sim p_0. \tag{3}$$

在条件流匹配中，指定一个耦合 $\gamma(\mathbf{Z}_0,\mathbf{Z}_1)$ 以及一个微观条件路径 $\mathbf{Z}_t = \Phi_t(\mathbf{Z}_0,\mathbf{Z}_1)$，其条件向量场为 $u_t(\cdot \mid \mathbf{Z}_0, \mathbf{Z}_1)$。最优边缘场为条件期望：

$$v_t(\mathbf{Z}) = \mathbb{E}_\gamma\big[u_t(\mathbf{Z} \mid \mathbf{Z}_0, \mathbf{Z}_1) \mid \mathbf{Z}_t = \mathbf{Z}\big]. \tag{4}$$

相比之下，类别部分（例如 $\mathbf{A}$，$\mathbf{H}$）通常采用适当的离散扩散进行建模，先验通常为均匀状态、吸收状态或掩码状态。

3. 理解具有对称性的扩散模型中的规范化

基于形式化的商空间视角，我们在本节给出规范扩散的完整理论。我们证明：**(i)** 通过切片分布与 Haar 随机化得到的不变测度因子分解定理，以及规范化生成模型相对于不变目标的通用性与更强表达能力（第3.1节）；**(ii)** 由群混合诱导、且可通过规范化消除的扩散得分复杂度的显式表达式，以及流匹配的条件方差分解，表明规范化能够降低条件方差，从而加速扩散/流匹配训练（第3.2节）。完整证明及更多讨论见附录 B。

3.1. 典范生成模型诱导通用不变分布

本小节形式化说明，为何在典范切片上学习并采用 Haar 随机化进行采样，不仅足以表示任何不变目标（第3.1.1节），而且在配合非等变去噪骨干网络时也更具表达能力（第3.1.2节）。

3.1.1. 典范参数化对不变目标与等变目标的通用性

首先，我们证明不变分布经由典范切片分解。回顾可知，Haar 概率 λ 对于 $S_N \times SO(3)$ 始终是均匀的。

定理 3.1（不变测度的分解；已知）。假设*2.5*与 *2.6* 成立。设 μ 为任意 $\mathcal{G}$-不变概率测度，定义在 $\mathcal{M}$上。设 Ψ 为一个轨道代表映射，其在 μ-几乎必然意义下有定义，并设 $\nu = \Psi\#\mu$为 $S = \Psi(\mathcal{M})$ 上的切片分布。则

$$\mu = \int_S \left(\int_{\mathcal{G}} \delta_{g\cdot\mathbf{Z}}\,\mathrm{d}\lambda(g) \right) \mathrm{d}\nu(\mathbf{Z}). \tag{5}$$

~ ~等价地，若 $\mathbf{Z} \sim \nu$, $g \sim \lambda$ 相互独立，则 $g \cdot \mathbf{Z} \sim \mu$.**推论 3.2**(切片建模的充分性)**.** 要对任意不变目标分布 μ进行建模，只需对切片分布ν进行建模；通过 *Haar* 随机化即可恢复不变性。

我们提出，规范化是一种用于构造等变/不变函数的通用技术，无需等变骨干网络，并将 Kaba 等人(2023年) 的方法加以推广：

命题 3.3 (规范化参数化在不变和等变目标上的通用性)**.** 设 $\mathcal{G}$ 在紧集 $K \subset \mathbb{R}^d$ 上连续作用。假设我们有一个（可测的）规范化映射 $\Psi : K \to K$，如定理 *2.3*所述，以及一个（可测的）规范映射 $\kappa : K \to \mathcal{G}$ 满足：

$$\Psi(g \cdot x) = \Psi(x),\ \kappa(g \cdot x) = g\kappa(x),\ x = \kappa(x) \cdot \Psi(x) \tag{6}$$

考虑如下参数化

$$\phi(x) = \kappa(x) \cdot f(\Psi(x)), \tag{7}$$

其中，f 是 $\Psi(K)$上的通用逼近器。于是，ϕ是 K 上连续 $\mathcal{G}$等变函数的通用逼近器，而 $f \circ \Psi$ 则是 K上连续 $\mathcal{G}$不变函数的通用逼近器。

3.1.2. 通过对称性破缺增强表达能力

我们现在说明，借助非等变模型进行对称性破缺，在实践中可以获得更强的表达能力。尽管不变分布的真实得分具有等变性，但在去噪器内部强制施加等变性可能会**限制架构选择**，从而限制表达能力，同时增加不必要的计算开销(Yan 等人，2023年)。值得注意的是，非等变模型比其等变对应模型具有更强的表达能力（Zhang 等人，2021年；Zhou 等人，2023a）。然而，要恢复不变输出，需要采用框架平均或关系池化，其复杂度与群的规模成正比。这促使我们采用规范化扩散：它使用一种对称破缺规范（规范顺序/位姿），在切片上利用更强的非等变骨干网络，然后通过随机化恢复不变性。由于无需遍历并进行平均，规范化的非等变模型可以在计算量相近或更低的情况下实现更强的表达能力。更多详情见附录B.1.2。

3.2. 规范化加速扩散训练

在本小节中，我们从两个互补的角度进行分析：得分混合复杂度（第3.2.1节）和流匹配条件方差（第3.2.2节），严格阐明规范化如何降低对称诱导方差，并加速扩散模型或流模型的训练。我们进一步表明，对齐典范先验和最优传输还能进一步降低切片内方差（第3.2.3节）。

3.2.1. 扩散得分中的混合结构

为简化起见，考虑一个有限群 $\mathcal{G}$，它包含 M 个元素，并在 $\mathbb{R}^d$ 上正交地作用，且 $g_m \in M$ 表示第 m个元素。设 q 为一个支撑于规范区域 S上的切片密度。定义不变混合分布

$$p(z_t) = \frac{1}{M}\sum_{m=1}^{M} q(g_m^{-1}\cdot z_t). \tag{8}$$

其中 z_t 指噪声状态；对于无限群，求和可以替换为积分。假设在需要处 q 可微且为正。根据链式法则和正交性，得分为

$$\begin{aligned}\nabla \log p(z_t) &= \sum_{m=1}^{M} w_m(z_t)\, g_m \cdot \nabla \log q(g_m^{-1}\cdot z_t),\\ w_m(z_t) &:= \frac{q(g_m^{-1}\cdot z_t)}{\sum_{j=1}^{M} q(g_j^{-1}\cdot z_t)}.\end{aligned} \tag{9}$$

多个群副本重叠时，其责任度$w_m(x_t)$ 会迅速变化。由此产生的尖锐得分场需要更强的模型容量和更小的步长，才能确保反向积分稳定。如（Yan 等人，2023年）所述，不变训练会导致模态增多，并产生类似混合分布的最优去噪得分（高斯混合模型类比）。相比之下，规范化消除了混合分布：切片上只有一个分量（$M = 1$），从而平滑了得分景观，简化了去噪得分网络的训练。

3.2.2. 通过消除对称性歧义降低条件流方差

我们在此形式化陈述核心结果：规范化通过消除对称性歧义，降低流匹配训练中的条件方差。

不失一般性，考虑线性路径设定，此时条件向量场为常量：

$$\begin{aligned}\mathbf{Z}_t &= (1-t)\mathbf{Z}_0 + t\mathbf{Z}_1, \qquad (\mathbf{Z}_0,\mathbf{Z}_1)\sim\gamma,\\ u_t(\mathbf{Z}_t \mid \mathbf{Z}_0,\mathbf{Z}_1) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{Z}_t = \mathbf{Z}_1 - \mathbf{Z}_0.\end{aligned} \tag{10}$$

最优边缘场为 $v_t(\mathbf{Z}) = \mathbb{E}[\mathbf{Z}_1 - \mathbf{Z}_0 \mid \mathbf{Z}_t = \mathbf{Z}]$。令 $\hat{v}$ 为基于 $(t,\mathbf{Z}_t)$ 对 $v_t(\mathbf{Z}_t)$ 的任意可测预测器。那么可达到的最小均方误差（等价地，不可约流匹配误差）为条件方差：

$$\inf_{\hat{v}} \mathbb{E}\big[\,\|\hat{v}(t,\mathbf{Z}_t) - (\mathbf{Z}_1 - \mathbf{Z}_0)\|^2\,\big] = \mathbb{E}\big[\mathrm{Var}(\mathbf{Z}_1 - \mathbf{Z}_0 \mid t, \mathbf{Z}_t)\big].$$

我们旨在理解 $\mathrm{Var}(Z_1 - Z_0|t, Z_t)$ 何时较大，尤其是在存在对称性的情况下。考虑以下对称性混合数据模型。设紧群$\mathcal{G}$ 在线性且正交地作用于 $\mathbb{R}^d$ （因此 $\|g\cdot x\| = \|x\|$）。假设数据分布 p_0 具有 $\mathcal{G}$-不变性：

$$\begin{aligned}Z_0 \overset{d}{=} G\cdot\widetilde{Z}_0 \sim p_0,\ G\sim\lambda\ (\text{Haar}),\ \widetilde{Z}_0\sim q_0,\\ \widetilde{Z}_0 = \Psi(Z_0),\ q_0 = \Psi_{\#}p_0,\end{aligned} \tag{11}$$

其中，Ψ 是一个（可测的）规范化器，将每个轨道映射到切片上的一个代表元。直观地说，如果不进行规范化，后验分布 $p(Z_0 \mid Z_t)$在潜在对称元素 G上可能呈多峰，从而增大条件方差。

比较以下两种训练范式：**(i)规范切片训练**（商空间），在规范化后的~数据 $Z_0 := \Psi(Z_0) \sim q_0$ 上训练，采用切片先验 $Z_1 \sim q_1$ 和一个~耦合 $\tilde{\gamma}(Z_0, Z_1)$；**(ii) 不变混合训练**（环境空间），直接在不变数据 $Z_0 \sim p_0$ 上训练，采用环境先验 $Z_1 \sim p_1$ 和一个耦合 $\gamma(Z_0,\ Z_1)$。

为将切片耦合与不变的环境空间耦合联系起来，我们定义给定 $\tilde{\gamma}$的群对齐提升：

$$\begin{aligned}G\sim\lambda,\ (\widetilde{Z}_0,\widetilde{Z}_1)\sim\tilde{\gamma}\ \text{independent},\\ (Z_0,Z_1) := (G\cdot\widetilde{Z}_0,\ G\cdot\widetilde{Z}_1)\sim\gamma^{\text{lift}}.\end{aligned} \tag{12}$$

我们强调，这种耦合对应关系（即不变训练使用诱导的群对齐提升作为~环境耦合）在实践中是合理的。当 $Z_1 \sim$

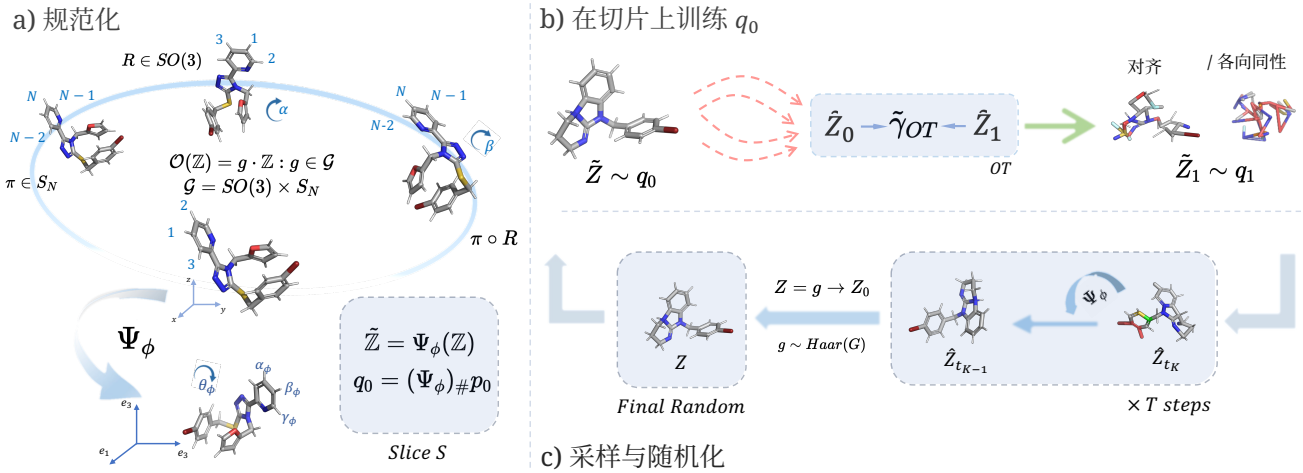


图 2。 我们的规范化生成流程概览。(a) 规范化：根据 $\mathcal{G} = SO(3) \times S_N$，将分子 Z 映射为切片代表元 $\tilde{Z} = \Psi_\phi(Z)$，从而得到 $q_0 = (\Psi_\phi)_{\#} p_0$。(b) 训练：在具有切片先验 q_1 的 $\tilde{Z}_0 \sim q_0$ 上训练扩散/流模型（可选地使用最优传输耦合进行对齐配对）。(c) 采样：在切片上生成样本（可选地使用投影规范采样），然后应用 $g \sim \text{Haar}(\mathcal{G})$ 以获得不变样本 $Z = g \rightarrow \hat{Z}_0$。

$\mathcal{N}(0, I)$ 且切片耦合是独立乘积$\tilde{\gamma} := q_0 \otimes q_1$，那么环境先验 $Z_1 \sim p_1 = \mathcal{N}(0, I)$ 和不变耦合便可还原广泛采用的乘积耦合（不使用最优传输）$\gamma_{mix} := p_0 \otimes p_1$；更多详情请参阅命题 B.11 。然而，如果$\tilde{}$ $\tilde{}$ $\tilde{}$具有群感知特性，或者 Z_1 和 Z_1 彼此并非 Z_0 独立，则通常不成立。更多详情见附录B.2.2。

备注 3.4。 下面的所有方差分解都在提升耦合下 γ^{lift} 进行比较（即分离出对 G 进行边缘化与对 G 进行条件化的影响）。

我们将在提升耦合下给出以下精确的方差分解与条件方差比较。

定理 3.5（群对齐提升下的方差分解）。 假设 $\mathcal{G}$ 正交作用。在群$\tilde{}$ $\tilde{}$对齐的提升耦合下，$G \sim \lambda$ 独立于 (Z_0, Z_1) $\tilde{}$ $\tilde{}$和 $(Z_0, Z_1) = (G \cdot Z_0, G \cdot Z_1)$，令 $Z_t = (1-t) Z_0 + t Z_1\tilde{}$ $\tilde{}$ $\tilde{}$以及 $Z_t = (1-t) Z_0 + t Z_1$。令 $U := Z_1 - Z_0 = G \cdot \Delta\tilde{}$ $\tilde{}$以及 $\Delta := Z_1 - Z_0$，则有

$$\begin{aligned} \mathrm{Var}(U \mid Z_t) = & \underbrace{\mathbb{E}\Big[\mathrm{Var}(\Delta \mid \widetilde{Z}_t) \mid Z_t\Big]}_{\textit{within-slice difficulty}} \\ & + \underbrace{\mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t)}_{\textit{symmetry ambiguity} \geq 0}. \end{aligned} \tag{13}$$

因此，未经规范化的不变训练会导致更大的条件方差：

$$\mathbb{E}\big[\mathrm{Var}(U \mid Z_t)\big] \;\geq\; \mathbb{E}\big[\mathrm{Var}(\Delta \mid \widetilde{Z}_t)\big]. \tag{14}$$

具体而言，规范化消除了由对称性歧义导致的第二项；只要给定 Z_t 后，G 仍存在歧义，通常就会有严格不等式条件均值漂移取决于 G；具体的后验碰撞界见附录 B.2.2。

注 3.6（只有非等变模型能从规范化中获益）。 值得注意的是，定理 3.5 仅给出了不可约误差的下界，但并未表明所有模型在实践中都能达到这一下界。我们强调，实践中只有非等变模型能从对称性歧义消除中获益，因为规范化和规范条件扩展了有效假设类，这与我们在第 3.1.2 节中的分析相呼应。相比之下，等变模型在规范化后的环境贝叶斯风险与不变训练时相同。更多细节请参见附录中的命题 B.14 。

3.2.3. 利用对齐典范先验和蒙日耦合进行切片内简化

定理3.5 揭示了不可约误差的两个不同来源：切片内难度和对称性歧义。规范化降低了第二项，但第一项仍然取决于切片耦合$\tilde{}$ $\tilde{}$ $\tilde{\gamma}(Z_0, Z_1)$，当切片耦合远离直线输运时，该项仍可能很大。为此，我们提出针对第一项的技术： 对齐典范先验和 最优传输。

典范切片的对齐先验原理。 我们从$\tilde{}$先验的选择 $Z_1 \sim q_1$出发。考虑这样一种情况：如果保留一个未对齐的简单先验（例如 $q_1 = \mathcal{N}(0, I)$ 而q_0 在典范切片上远离中心/各向同性，或具有低维性），那么切片耦合仍然很难处理，定理3.5 中的第一项可能仍然很大；闭式切片内条件方差的一个示例见命题 B.18。因此，我们提出两种简单而有效的方法来构造**对齐的典范先验**：(i)在特定函数类中定义可学习先验

（例如，具有可学习均值和方差的高斯分布）；(ii)将 q_0 进行KL投影到特定函数类上（例如矩匹配高斯分布）。例如，在高斯规范先验中，矩匹配高斯分布为 $\sim$ $\sim q_1^\star \sim \mathcal{N}(\mathbb{E}_{q_0}[Z_0],$ 协方差$_{q_0}(Z_0))$（命题B.19）。

最优传输和近莫恩日耦合互为补充。 给定规范数据 q_0 和先验$\sim$ $\sim$ $\sim q_1$，切片条件方差 $\mathbb{E}[\mathrm{Var}(Z_1 - Z_0 \mid Z_t)]$仍取决于耦合 $\tilde{\gamma}$。在有利的条件下，例如满足标准正则性条件的类 Monge（近似）最优传输（OT），这一项可以很小。

值得注意的是，已有一些工作（Ir- win 等人，2024）应用了最优传输（OT）。现在，我们正式阐明 OT 在对称空间中的有效性，无论是否应用规范化。无论是否应用规范化，OT 通常都能减少轨迹交叉和条件方差，但在存在对称性的情况下，OT 解并不唯一。平均化会产生对角线不变的最优耦合，但可能破坏确定性（使莫恩日映射变成混合分布）。幸运的是，规范化通过提供一个规范来稳定解，并进一步惠及将规范条件作为输入的非等变模型，正如附录B.2.2中所示。另一种做法是，规范化通过将数据-噪声对对齐到同一规范中，充当一种“隐式 OT”。

备注 3.7（规范化与最优传输互为补充）。 在存在对称性的情况下，规范化消除了群歧义，而最优传输降低了切片内难度。通过规范化，最优传输解可能更接近唯一解或类 Monge 解（图4）。

训练-采样一致性。假设网络以辅助变量 C （例如“规范秩”位置编码、帧 ID，或规范化器的任意确定性/随机输出）为条件。训练随后会隐式地以联合分布 $\tilde{\pi}_t(c, z)$ 下的条件场 $\tilde{v}(t, z, c)$ 为目标，该联合分布由训练流程诱导而来。如果推理时联合分布 $\tilde{\pi}_1^{\mathrm{inf}}(C, \widetilde{Z}_1)$ 等于训练时联合分布$\sim\tilde{\pi}_1^{\mathrm{tr}}(C, Z_1)$，我们称其为起始时不存在条件失配。采用以下任一方案：（i）对齐典范先验；或（ii）各向同性先验和训练时最优传输，均可保持训练-采样一致性。事实上，训练轨迹在图3 有力地支持了我们的主张：规范化（尤其是与对齐典范先验结合时）能显著加快训练。

4. 用于分子图生成的规范扩散

基于第 3 节所奠定的理论基础，我们现在在联合对称群 $\mathcal{G} = S_N \times SO(3)$（原子索引置换和全局旋转）下实例化规范分子图扩散框架。

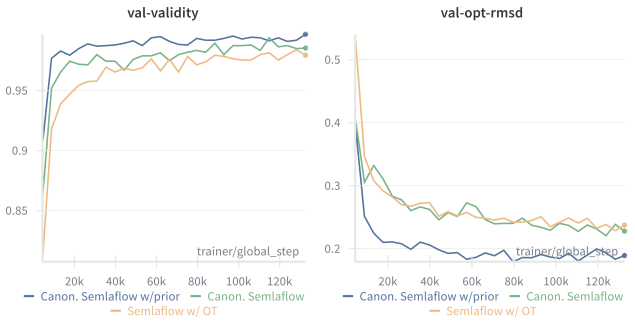


图3. 规范化模型的训练轨迹（学习到的输运动力学的可视化）。

4.1. 分子图规范化

关键的设计选择在于构建一个兼具几何原理依据和计算可行性的规范化器。我们将完整的对称群分解，并依次处理置换对称性和旋转对称性：

$$\Psi_\phi(\mathbf{Z}) = \Psi_\phi^{(\mathrm{rot})}\left(\Psi_\phi^{(\mathrm{perm})}(\mathbf{Z})\right), \tag{15}$$

其中 $\Psi_\phi^{(\mathrm{perm})}$ 确立了规范原子排序，$\Psi_\phi^{(\mathrm{rot})}$ 固定了全局参考系。

我们研究**通过几何谱进行规范化**，以协调的方式同时打破置换对称性和旋转对称性。Fiedler向量，即对应于图的拉普拉斯矩阵第二小特征值（代数连通度）的特征向量，以自然的方式确定规范原子排序。具体而言，给定一个分子，我们根据成对距离（例如，$W_{ij} = k(\|\mathbf{X}_i - \mathbf{X}_j\|)$，用于径向核 k）构建对称核矩阵 $\mathbf{W}(\mathbf{X},\mathbf{A}) \in \mathbb{R}^{N\times N}$，以处理三维几何，形成归一化几何拉普拉斯

$$\mathbf{L}(\mathbf{X}, \mathbf{A}) = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{W}), \quad \mathbf{D} = \mathrm{diag}(\mathbf{W}\mathbf{1}). \tag{16}$$

$\mathbf{L}$ 仅取决于距离，并且具有旋转不变性。随后，我们使用Fiedler向量 $\mathbf{u_2} \in \mathbb{R}^N$ 来定义规范顺序

$$\pi^\star(\mathbf{Z}) := \mathrm{argsort}(\mathbf{u_2}) \in S_N. \tag{17}$$

根据拉普拉斯矩阵的标准质量-弹簧解释，这对应于最低频率的非刚性振动模态（Knyazev，2018）。该排序产生经典的谱二分，并提供了一种有原则的序列化方法，使强连通子结构保持连续（Pothen 等，1990）。在分子语境中，谱排序充当一种几何感知线性化。

最后，我们还可以选择在应用 $\pi^\star$ 后，以与排序一致的方式定义全局 $SO(3)$ 规范。我们按规范顺序选取一小组锚原子（例如，两个极端原子定义主轴，第三个锚原子确定平面法向），并将中心化坐标旋转到

所得的右手正交标架中。由此可在联合 $S_N \times SO(3)$ 作用下得到一个确定性的代表元；完整算法见算法2。

此外，还有其他现有的规范化方法，主要针对抽象图中的置换（Zhao 等人，2024；Ma 等人，2023年；Dong 等人，2024），而非几何图；现有的最优传输方法则大多针对坐标。我们的方法始终以稳定的方式考虑联合群。不同的规范化方法可能导致不同的稳定性和实际性能，但我们的实验发现，新颖的几何谱规范化方法优于其他方法（附录 C.1）。

4.2. 规范扩散与 CanonFlow

典范性感知去噪架构。我们详细介绍两种架构调整，以便在规范扩散中使用非等变去噪模型：(i) 仅采用位置编码，即保留原始模型架构，只将模型的条件设为典范秩嵌入 $C = \varphi(r_i)$ 和归一化秩$r_i = (i-1)/(N-1)$，以适应不同的分子尺寸；(ii) Canon 架构，这是我们新设计的架构，它显式地将原子典范嵌入作为可学习的隐藏状态 C $\in \mathbb{R}^{N\times d_c}$，并使其在每一层与 X、H、A 的嵌入交互（图 6）；更多细节见附录 C.2。我们用“Canon. BM”指代第一种类型，其余设置均与基础模型“BM”一致，只有 PE 除外；“CanonFlow”特指采用 Canon 架构的流匹配模型，其余设置与 SemlaFlow 对齐。

在典范切片上训练。训练过程采用标准的扩散或流匹配方法，在典范切片 $\widetilde{\mathbf{Z}} = \Psi_\phi(\mathbf{Z})$ 和切片分布$q_0 = (\Psi_\phi)_{\#}p_0$上进行。为方便说明，我们以流匹配框架为例，扩散模型也可类似地处理。给定规范化数据 $\widetilde{\mathbf{Z}}_0 \sim q_0$ 与（可选的对齐）先验 $\widetilde{\mathbf{Z}}_1 \sim q_1$ 之间的耦合 $\widetilde{\gamma}(\widetilde{\mathbf{Z}}_0,\widetilde{\mathbf{Z}}_1)$，我们针对 $t \sim \mathrm{Unif}[0,1]$构造插值路径 $\widetilde{\mathbf{Z}}_t = \Phi_t(\widetilde{\mathbf{Z}}_0,\widetilde{\mathbf{Z}}_1)$。模型 m_θ 学习预测条件速度场：

$$\min_\theta \mathbb{E}\Big[\big\|m_\theta(\widetilde{\mathbf{Z}}_t, t; C) - u_t(\widetilde{\mathbf{Z}}_t \mid \widetilde{\mathbf{Z}}_0, \widetilde{\mathbf{Z}}_1)\big\|^2\Big], \qquad (18)$$

其中 $u_t = \mathrm{d}\Phi_t/\mathrm{d}t$ 包括连续特征和离散特征的速度。如第 3.2.3 节所述，我们也可以选择采用对齐先验，或在训练初期启用最优传输耦合，并随训练轮数逐步退火其使用程度（称为 最优传输退火），以稳定优化过程，同时避免对对齐配对过度专门化。

采样并恢复不变性。从切片先验 $\widetilde{\mathbf{Z}}_1 \sim q_1$ （与训练中使用的边缘先验相同）出发，对学习到的动力学进行积分，以获得切片上的规范**样本** $\widetilde{\mathbf{Z}}_0$；即使训练时使用了最优传输匹配，推理时也不需要进行最优传输匹配。将采样得到的规范分布记为 S，随后只需通过采样 $\tilde{\mu}$来恢复不变性，$g \sim \lambda$ $\mathbf{Z} \sim \tilde{\mu}$ 并输出$g\cdot\widetilde{\mathbf{Z}}$，由此得到的随机化分布 $\mu := \int_{\mathcal{G}}(g\cdot)_{\#}\tilde{\mu}\,\mathrm{d}\lambda(g)$具有不变性（命题C.1）。

利用规范性的改进技术。训练与推理之间存在一个微妙的差异：训练期间，规范秩直接根据 $\Psi_\phi(\mathbf{Z})$ 精确计算，而推理时，模型从噪声生成数据，无法获取真实秩。此外，遵循（Dym 等人，2024）的思路，对于 S_N 和 $SO(d)$，连续规范化是不可能的。为了解决这一问题并缩小训练-测试差距，我们将秩放宽为连续值，并在训练期间向秩特征注入小幅扰动；此外，还可选用一个辅助头，从中间表示中自行估计秩。在采样时，模型可以使用固定的规范条件，也可以自适应地重新估计秩，以对齐带噪数据与规范条件；我们将后者称为投影规范采样 （PCS）。受（Li 等人， 2025）启发，我们还会在训练期间以概率p_{drop} 随机丢弃规范条件，并在推理时启用无分类器引导（CFG）（Ho 与 Salimans， 2022年），对条件生成与无条件生成进行外推。更多训练与采样技术详见附录C。

5. 实验

实验设置。我们在两个广泛采用的无条件三维分子生成基准测试上评估规范化方法：QM9（Ramakrishnan 等人，2014年）和 GEOM-DRUG（Axelrod & Gomez-Bombarelli，2022）。QM9 由相对较小的分子组成，而 GEOM-DRUG 提供了更具挑战性的测试平台，能够更好地区分不同生成模型的能力。我们采用与先前工作相同的训练/验证/测试集划分（Vignac 等人，2023；Le 等人，2023）。遵循（Irwin 等人，2024），为提高计算效率，我们从 GEOM-DRUG 训练集中排除原子数超过 72 个的分子，这约占数据的 1%；验证集和测试集则保持完整。评估时，我们从测试集的经验分布中采样分子大小，并通过对学习到的流动力学进行数值积分，生成相应数量的原子。所有报告的指标均基于 3 次独立运行生成的 10k 个分子计算得出。除非另有说明，我们采用 SemlaFlow（Irwin 等人，2024）中的设置和超参数，并以其作为基础模型。默认情况下，规范化在 S_N 群上进行；规范化 $S_N \times SO(3)$ 的结果以及更多消融实验见附录D。

表 *1.* 典范化 SemlaFlow 在 QM9 数据集上的性能。原子稳定性、分子稳定性和有效性以百分比表示；Opt-RMSD 以 ˚A 表示。

方法	原子稳定性 ↑	分子稳定性↑	有效性 ↑	Opt-RMSD ↓	函数评估次数 ↓
EDM	98.7	82.0	91.9	–	1000
GCDM	98.7	85.7	94.8	–	1000
MUDiff	98.3	89.9	95.3	–	1000
FlowMol	99.7	96.2	97.3	–	100
MiDi	99.8	97.5	97.9	–	500
EQGAT-diff	$99.9_{\pm 0.0}$	$98.7_{\pm 0.18}$	$99.0_{\pm 0.16}$	–	500
Semlaflow（含最优传输）	$99.9_{\pm 0.0}$	$99.56_{\pm 0.07}$	$99.31_{\pm 0.08}$	$0.23_{\pm 0.0}$	100
规范化SemlaFlow，采用最优传输退火 + PCS（我们的方法）	$99.9_{\pm 0.0}$	$99.62_{\pm 0.08}$	$99.49_{\pm 0.06}$	$0.21_{\pm 0.0}$	100
规范化SemlaFlow（带先验） + PCS（我们的方法）	$99.9_{\pm 0.0}$	$\mathbf{99.64}_{\pm 0.04}$	$\mathbf{99.58}_{\pm 0.06}$	$\mathbf{0.17}_{\pm 0.0}$	100

表 *2.* 典范化SemlaFlow和CanonFlow在GEOM-DRUG 数据集上的表现。报告了原子稳定性、分子稳定性、有效性、唯一性和新颖性，数值均为百分比。

方法	原子稳定性 ↑	分子稳定性↑	有效性 ↑	唯一性 ↑	新颖性 ↑	函数评估次数 ↓
FlowMol	99.0	67.5	51.2	-	-	100
MiDi	99.8	91.6	77.8	100.0	100.0	500
EQGAT-diff	$99.8_{\pm 0.0}$	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	$100.0_{\pm 0.0}$	$99.9_{\pm 0.07}$	500
Semlaflow（含最优传输）	$99.8_{\pm 0.0}$	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	$100.0_{\pm 0.0}$	$99.6_{\pm 0.03}$	100
规范化SemlaFlow 最优传输退火（我们的方法）	$99.8_{\pm 0.0}$	$98.1_{\pm 0.03}$	$95.0_{\pm 0.20}$	$100.0_{\pm 0.0}$	$99.7_{\pm 0.02}$	100
CanonFlow + 最优传输退火（我们的方法）	$\mathbf{99.9}_{\pm 0.0}$	$\mathbf{98.4}_{\pm 0.02}$	$\mathbf{95.9}_{\pm 0.08}$	$100.0_{\pm 0.0}$	$99.7_{\pm 0.01}$	100

主要结果。 将规范扩散框架应用于 SemlaFlow 和 CanonFlow 后，我们在两个数据集的大多数指标上都取得了最先进的结果。

表 1 展示了QM9上的结果。尽管所有方法的原子稳定性都接近饱和，但我们的规范化变体在分子稳定性和有效性方面均优于SemlaFlow。值得注意的是，Canon. SemlaFlow（先验 + PCS）取得了最低的Opt-RMSD（0.17 ˚A），与基线（0.23 ˚A）相比降低了26%，这表明规范化能够引导模型生成更接近能量最小化构象的几何结构。此外，我们在图 3 中报告了QM9上的训练动态。可以观察到，与基线相比，在相同训练轮数下，采用规范化的模型在验证集上实现了显著更高的有效性和更低的Opt-RMSD，而对齐典范先验进一步扩大了这一优势，验证了规范化能够加速训练收敛。

表 2 展示了更具挑战性的 GEOM- DRUG 数据集上的结果。该数据集包含规模更大、结构更多样的类药分子，能够更好地区分模型的能力。在此基准测试中，我们的规范化方法相较于基线取得了明显提升。具体而言，规范化SemlaFlow（使用 CFG 训练）在有效性和分子稳定性方面均持续优于 SemlaFlow 基线。值得注意的是，我们的 CanonFlow 在原子稳定性、分子稳定性和有效性这几项最重要的指标上均达到了最先进的 性能，大幅超越此前所有模型（分子稳定性比 SemlaFlow 基线高1.1%，有效性高2.0%）。

表*3*。 GEOM-DRUG 上少步生成的比较（训练和采样时均不使用 CFG）。采样时间以生成 1000 个分子的平均秒数衡量。Canonical SemlaFlow 使用更少的采样步数，便显著优于基线，而采样时间仅略有增加。

模型	分子稳定性↑	有效性 ↑	采样时间↓	函数评估次数 ↓
EQGAT-diff	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	2293.0	500
SemlaFlow_{50} 搭配最优传输	$97.0_{\pm 0.21}$	$93.9_{\pm 0.12}$	49.8	50
规范化SemlaFlow_{50}	$97.5_{\pm 0.14}$	$94.9_{\pm 0.07}$	50.1	50
SemlaFlow_{100}，搭配最优传输	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	99.3	100
规范化SemlaFlow_{100}	$97.9_{\pm 0.09}$	$94.9_{\pm 0.18}$	99.7	100
Data	100.0	100.0	–	–

我们还在表 3 中报告了少步生成的性能。即使仅使用 50 步（仅为以往大多数模型步数的十分之一），规范化模型的优势仍然显著。值得注意的是，我们的方法几乎不会增加计算或采样时间开销。这进一步验证了规范条件化在采样过程中所提供的引导信号既高效又有效。

6. 结论

我们从规范化与传输的角度研究了基于扩散/流的生成建模中的对称性。对于表示对置换和刚体运动具有不变性的分子数据，我们证明，对称性可能在中间噪声水平诱发潜在的群歧义，增大条件方差，并降低少步求解器的可靠性。规范化提供了一种有原则的规范固定方法：它将每个轨道压缩到一个唯一的切片

代表元，从而消除对称性诱导的歧义项，并使得能够使用具有稳定、固定索引语义（例如基于排序的输入）的高表达力非等变骨干网络。训练过程中，采用更好的对齐先验和近 Monge 耦合，可以缓解剩余的切片内难度。总体而言，这些结果阐明了规范化与最优传输在何时、为何互补，如何避免投影或条件化诱导的不匹配，以及如何设计实用、可扩展的分子对称性感知生成器，在保持不变性的同时提升少步生成质量。规范扩散在多个常用的几何分子生成基准测试指标上取得最先进结果，并在训练收敛方面具有显著优势。

致谢

Cai Zhou 谨向加州大学圣迭戈分校的 Yusu Wang 教授、麻省理工学院和慕尼黑工业大学的 Stefanie Jegelka 教授，以及来自北京大学的 Xingang Peng 和 Xiyuan Wang 致以诚挚谢意，感谢他们进行的富有建设性的讨论。

影响声明

本文旨在推动机器学习、生成模型和分子设计领域的发展。我们的工作可能会产生一些社会影响，但我们认为没有任何影响需要在此特别指出。

参考文献

Albergo, M. S.、Boffi, N. M. 和 Vanden-Eijnden, E. 随机插值：流与扩散的统一框架。*arXiv* 预印本 *arXiv:2303.08797*，2023年。Anonymous。商空间扩散模型。载于第十四届国际学习表征会议，2026年。网址：https://openreview.net/forum?id=3JPAkwSVc4。Axelrod, S. 和 Gomez-Bombarelli, R. Geom：用于属性预测和分子生成的能量标注分子构象。*Scientific Data*，9(1):185，2022年。Bo, D.、Shi, C.、Wang, L. 和 Liao, R. Specformer：谱图神经网络与 Transformer 的结合。*ArXiv*，abs/2303.01028，2023年。Cai, C.、Hy, T. S.、Yu, R. 和 Wang, Y. 论 mpnn 与图Transformer 之间的联系。载于国际机器学习会议，第 3408–3430 页。PMLR，2023年。Campbell, A.、Benton, J.、Bortoli, V. D.、Rainforth, T.、Deligiannidis, G. 和 Doucet, A. 离散去噪模型的连续时间框架。载于 Oh, A. H.、Agarwal, A.、Belgrave, D. 和 Cho, K.（编），神经信息处理系统进展，2022年。网址：https://openreview.net/forum?id=DmT862YAieY。

Cao, N. D. 和 Kipf, T. MolGAN：用于小分子图的隐式生成模型，2022年。网址 https://arxiv.org/abs/1805.11973。

Chen, X., He, J., Han, X., 和 Liu, L.-P. 通过离散扩散建模实现高效且受度数引导的图生成。*arXiv* 预印本 *arXiv:2305.04111*，2023年。

Corso, G., St”ark, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock：用于分子对接的扩散步骤、扭转与转向。载于 国际学习表征会议，2023年。Dhariwal, P. 和 Nichol, A. 扩散模型在图像合成方面胜过生成对抗网络。载于 神经信息处理系统进展，34:8780–8794，2021年。Dong, Z., Zhang, M., Payne, P., Province, M., Cruchaga, C., Zhao, T., Li, F., 和 Chen, Y. 重新审视图规范化在具有稳定性的图表示学习中的作用。载于国际学习表征会议，第2024卷，第30797–30817页，2024。Dwivedi, V. P., Luu, A. T., Laurent, T., Bengio, Y., 和 Bresson, X. 具有可学习结构与位置表示的图神经网络。载于 *ArXiv*，abs/2110.07875，2021年。Dym, N., Lawrence, H., 和 Siegel, J. W. 等变标架与连续规范化的不可能性。载于 *arXiv* 预印本 *arXiv:2402.16077*，2024。Feng, W., Wang, L., Lin, Z., Zhu, Y., Wang, H., Dong, J., Bai, R., Wang, H., Zhou, J., Peng, W., Huang, B., 和 Zhou, W. 通过语言模型在口袋中生成三维分子，2023年。网址 https://arxiv.org/abs/2305.10133。Garcia Satorras, V., Hoogeboom, E., Fuchs, F., Posner, I., 和 Welling, M. E(n) 等变归一化流。神经信息处理系统进展，34: 4181–4192，2021年。Gebauer, N., Gastegger, M., 和 Sch¨utt, K. 面向分子定向发现的三维点集对称性适配生成。载于 神经信息处理系统进展，32，2019年。Gong, S., Li, M., Feng, J., Wu, Z., 和 Kong, L. DiffuSeq：使用扩散模型进行序列到序列文本生成。载于 *arXiv* 预印本 *arXiv:2210.08933*，2022年。

Hassan, M., Shenoy, N., Lee, J., Stark, H., Thaler, S., 和 Beaini, D. Et-flow：用于分子构象生成的等变流匹配，2024。网址 https://arxiv.org/abs/2410.22388。

Ho, J. 和 Salimans, T. 无分类器扩散引导。*arXiv* 预印本 *arXiv:2207.12598*，2022 年。

Ho, J.、Jain, A. 和 Abbeel, P. 去噪扩散概率模型。*ArXiv*，abs/2006.11239，2020。

Ho, J.、Salimans, T.、Gritsenko, A.、Chan, W.、Norouzi, M. 和 Fleet, D. J. 视频扩散模型。神经信息处理系统进展，35:8633–8646，2022年。Hong, H.、Lin, W. 和 Tan, K. C. 通过联合几何最优传输加速三维分子生成，2025。网址：https://arxiv.org/abs/2405.15252。

Hoogeboom, E.、Satorras, V. G.、Vignac, C. 和 Welling, M. 面向三维分子生成的等变扩散。发表于国际机器学习会议，第8867–8887页。PMLR，2022年。Hou, X.、Zhu, T.、Ren, M.、Bu, D.、Gao, X.、Zhang, C. 和 Sun, S. 通过流匹配和最优传输改进分子图生成，2024。网址：https://arxiv.org/abs/2411.05676。Hua, C.、Luan, S.、Xu, M.、Ying, R.、Fu, J.、Ermon, S. 和 Precup, D. MUDiff：用于完整分子生成的统一扩散，2024。网址：https://arxiv.org/abs/2304.14621。

Huang, H.、Sun, L.、Du, B. 和 Lv, W. 学习联合二维和三维扩散模型以实现完整分子生成，2023。网址：https://arxiv.org/abs/2305.12347。

Huang, Y.、Peng, X.、Ma, J. 和 Zhang, M. 通过 I^2-GNNs 增强图神经网络的环计数能力。*ArXiv*论文，文章编号：arXiv:2210.13978，2022年10月。doi：10.48550/arXiv.2210.13978。Irwin, R.、Tibo, A.、Janet, J. P. 和 Olsson, S. SemlaFlow——利用潜在注意力和等变流匹配实现高效的三维分子生成。*ArXiv* 预印本 *arXiv:2406.07266*，2024。Jang, Y.、Kim, D. 和 Ahn, S. 基于 k^2-树的分层图生成。发表于*I CML 2023*结构化概率推断研讨会 {\&} 生成建模，2023。Jiang, K.、Cui, J.、Dong, X. 和 Toni, L. 用于图生成的Bures-Wasserstein流匹配，2025。网址：https://arxiv.org/abs/2506.14020。

Jin, W.、Barzilay, R. 和 Jaakkola, T. 用于分子图生成的连接树变分自编码器，2019年。网址：https://arxiv.org/abs/1802.04364。

Kaba, S.-O.、Mondal, A. K.、Zhang, Y.、Bengio, Y. 和 Ravanbakhsh, S. 通过学习式典范化函数实现等变性。载于第*40*届国际机器学习会议论文集，第202卷，收录于机器学习研究论文集，第15546–15566页。PMLR，2023年7月23–29日。网址：https://proceedings.mlr.press/v202/kaba23a.html。

Kim, J., Nguyen, T. D., Min, S., Cho, S., Lee, M., Lee, H. 和 Hong, S. 纯 Transformer 是强大的图学习器。*ArXiv*，abs/2207.02505，2022年。Knyazev, A. V. 关于有符号图的谱划分，2018。网址 https://arxiv.org/abs/1701.01394。

Kornilov, N., Mokrov, P., Gasnikov, A. 和 Korotin, A. 最优流匹配：一步学习直线轨迹，2024。网址https://arxiv.org/abs/2403.13117。

Lawrence, H., Portilheiro, V., Zhang, Y. 和 Kaba, S.- O. 通过概率对称性破缺改进等变网络。*arXiv* 预印本 *arXiv:2503.21985*，2025年。Le, T., Cremer, J., No´e, F., Clevert, D.-A. 和 Sch¨utt, K. 探索用于从头生成 3D 分子的等变扩散生成模型设计空间。*arXiv* 预印本 *arXiv:2309.17296*，2023年。Lee, J., Kim, S., Moon, S., Kim, H. 和 Kim, W. Y. Fragfm：通过片段级离散流匹配实现高效分子生成的分层框架，2025。网址 https://arxiv.org/abs/2502.15805。Li, X., Thickstun, J., Gulrajani, I., Liang, P. S. 和 Hashimoto, T. B. Diffusion-LM 改进了可控文本生成。神经信息处理系统进展，35:4328–4343，2022年。

Li, Z., Wang, X., Huang, Y. 和 Zhang, M. 距离矩阵足以用于几何深度学习吗？载于神经信息处理系统进展，第 36 卷，第 37413–37447 页，2023年。Li, Z., Zhou, C., Wang, X., Kang, S. 和 Zhang, M. 不变几何深度学习模型的完备性。*arXiv* 预印本 *arXiv:2402.04836*，2024。Li, Z., Zhou, C., Wang, X., Peng, X. 和 Zhang, M. 几何表示条件改进等变分子生成。载于论文集

第 *42* 届国际机器学习大会，第 267 卷，机器学习研究论文集，第 36921–36953 页。PMLR，2025年7月13–19日。网址 https://proceedings.mlr.press/v267/li25dz.html。

Lim, D., Robinson, J., Zhao, L., Smidt, T., Sra,S., Maron, H., 和 Jegelka, S. 用于谱图表示学习的符号与基不变网络。*arXiv* 预印本 *arXiv:2202.13013*，2022年。

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., 和 Le, M. 用于生成建模的流匹配。*arXiv* 预印本 *arXiv:2210.02747*，2022年。

Liu, X., Gong, C., 和 Liu, Q. 流动笔直且快速：通过 rectifiedflow 学习生成和迁移数据。*arXiv* 预印本 *arXiv:2209.03003*，2022年。

Luo, T., Mo, Z., 和 Pan, S. J. 通过谱扩散实现快速图生成。*IEEE*模式分析与机器智能汇刊，2023年。

Luo, Y. 和 Ji, S. 从零开始生成三维分子几何的自回归流模型。载于 国际学习表征会议（*ICLR*），2022年。Luo, Y.、Yan, K. 和 Ji, S. Graphdf：用于分子图生成的离散流模型，2021年。网址 https://arxiv.org/abs/2102.01189。

Ma, J.、Wang, Y. 和 Wang, Y. 拉普拉斯规范化：一种实现符号与基底不变谱嵌入的极简方法。*ArXiv*预印本 *arXiv:2310.18716*，2023年。

Maron, H.、Ben-Hamu, H.、Serviansky, H. 和 Lip- man, Y. 可证明具有强大表达能力的图网络。*ArXiv*，abs/1905.11136，2019年。Morris, C.、Ritzert, M.、Fey, M.、Hamilton, W. L.、Lenssen, J. E.、Rattan, G. 和 Grohe, M. 魏斯费勒-莱曼走向神经网络：高阶图神经网络。载于 *AAAI*人工智能会议论文集，第4602–4609页，2019年。

Morris, C.、Rattan, G. 和 Mutzel, P. Weisfeiler 和 le- man 转向稀疏：迈向可扩展的高阶图嵌入。神经信息处理系统进展，33:21824–21840，2020。

Pothen, A.、Simon, H. D. 和 Liou, K.-P. 使用图的特征向量对稀疏矩阵进行划分。*SIAM*矩阵分析与应用杂志，11(3):430–452，1990。doi: 10.1137/0611030。

Qin, Y.、Madeira, M.、Thanou, D. 和 Frossard, P. De-fog：用于图生成的离散流匹配。*arXiv* 预印本 *arXiv:2410.04263*，2024。

Ramakrishnan, R.、Dral, P. O.、Rupp, M. 和 Von Lilienfeld, O. A. 13.4万种分子的量子化学结构与性质。*Scientific Data*，1(1):1–7，2014年。Rampasek, L.、Galkin, M.、Dwivedi, V. P.、Luu, A. T.、Wolf, G. 和 Beaini, D. 通用、强大且可扩展的图 Transformer方案。*ArXiv*，abs/2205.12454，2022年。Rombach, R.、Blattmann, A.、Lorenz, D.、Esser, P. 和 Ommer, B. 利用潜在扩散模型进行高分辨率图像合成。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第10684–10695页，2022年。

Satorras, V. G.、Hoogeboom, E. 和 Welling, M. E (n) 等变图神经网络。载于国际机器学习会议，第 9323–9332 页。PMLR，2021年。

Shi, C.、Xu, M.、Zhu, Z.、Zhang, W.、Zhang, M. 和 Tang, J. Graphaf：一种用于分子图生成的基于流的自回归模型，2020。网址 https://arxiv.org/abs/2001.09382。

Shirzad, H., Velingker, A., Venkatachalam, B., Sutherland, D. J. 和 Sinop, A. K. Exphormer：面向图的稀疏变换器。*ArXiv*，abs/2303.06147，2023年。

Simonovsky, M. 和 Komodakis, N. Graphvae：使用变分自编码器生成小型图，2018。URLhttps://arxiv.org/abs/1802.03480。

Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O. 等。Make-a-video：无需文本-视频数据的文本生成视频。*arXiv* 预印本 *arXiv:2209.14792*，2022年。

Song, Y. 和 Ermon, S. 通过估计数据分布的梯度进行生成建模。*ArXiv*，abs/1907.05600，2019年。

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S. 和 Poole, B. 通过随机微分方程进行基于得分的生成建模。在 国际学习表征会议，2021年。网址：https://openreview.net/forum?id=PxTIG12RRHS。Song, Y., Gong, J., Xu, M., Cao, Z., Lan, Y., Ermon, S., Zhou, H. 和 Ma, W.-Y. 通过混合概率传输实现用于三维分子生成的等变流匹配。神经信息处理系统进展，36，2023年。

Tahmasebi, B. 和 Jegelka, S. 规范化的泛化界：与群平均的比较研究。发表于第十三届国际会议

关于表征学习，2025。网址 https://openreview.net/forum?id=n0lXaskyk5。

Tian, Q.、Xu, Y.、Yang, Y.、Wang, Z.、Liu, Z.、Yan, P. 和 Li, X. Equiflow：采用最优传输的等变条件流匹配，用于三维分子构象预测，2024年。网址 https://arxiv.org/abs/2412.11082。

Tong, A.、Fatras, K.、Malkin, N.、Huguet, G.、Zhang, Y.、Rector-Brooks, J.、Wolf, G. 和 Bengio, Y. 使用小批量最优传输改进并泛化基于流的生成模型，2024。网址 https://arxiv.org/abs/2302.00482。

Vignac, C.、Krawczuk, I.、Siraudin, A.、Wang, B.、Cevher, V. 和 Frossard, P. Digress：用于图生成的离散去噪扩散。*ArXiv* 预印本 *ArXiv:2209.14734*，2022年。

Vignac, C., Osman, N., Toni, L. 和 Frossard, P. MiDi：用于分子生成的混合图与三维去噪扩散。载于欧洲机器学习与数据库知识发现联合会议，第 560–576 页。Springer，2023年。

Wang, J., Luo, H., Qin, R., Wang, M., Fang, M., Zhang, O., Gou, Q., Su, Q., Shen, C., You, Z., Wan, X., Liu, L., Hsieh, C.-Y., Hou, T. 和 Kang, Y. 使用仅使用词元的大语言模型进行基于三维分子口袋的生成。*ChemRxiv*，2024(0819)，2024。doi: 10.26434/chemrxiv-2024-0ckgt-v2。URL h ttps://chemrxiv.org/doi/abs/10.26434/chemrxiv-2024-0ckgt-v2。

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F. 等。利用 RFdiffusion 从头设计蛋白质结构与功能。《自然》，620:1089–1100，2023年。

Wu, L., Gong, C., Liu, X., Ye, M. 和 Liu, Q. 利用信息性先验桥进行基于扩散的分子生成。神经信息处理系统进展，35: 36533–36545，2022年。

Xu, K., Hu, W., Leskovec, J. 和 Jegelka, S. 图神经网络的表达能力有多强？*ArXiv*，abs/1810.00826，2018。

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S. 和 Tang, J. Geodiff：用于分子构象生成的几何扩散模型。*arXiv* 预印本 *arXiv:2203.02923*，2022年。

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., 和 Leskovec, J. 用于 3D 分子生-成的几何潜扩散模型。载于国际机器学习会议，第 38592–38610 页。PMLR，2023年。

Yan, Q., Liang, Z., Song, Y., Liao, R., 和 Wang, L. Swingnn：重新思考图生成扩-散模型中的置换不变性。*arXiv* 预印本 *arXiv:2307.01646*，2023年。

You, J., Gomes-Selman, J. M., Ying, R., 和 Leskovec, J. 身份感知图神经网络。载于*AAAI*人工智能会议，2021年。

You, Y., Zhou, R., Park, J., Xu, H., Tian, C., Wang, Z., 和 Shen, Y. 潜三维图扩散。载于第十二届国际学习表征会议，2023年。

Zhang, M., Li, P., Xia, Y., Wang, K., 和 Jin, L. 标签技巧：使用图神经网络进行多节点表示学习的理论。神经信息处理系统进展，34：9061–9073，2021年。

Zhao, L., Ding, X., 和 Akoglu, L. PARD：用于图生成的置换不变自回归扩散。神经信息处理系统进展，37：7156–7184，2024。

Zhou, C., Wang, X., 和 Zhang, M. 从关系池化到子图图神经网络：一个增强图神经网络表达能力的通用框架。载于机器学习第*40*届国际会议论文集，机器学习研究论文集第202卷，第42742–42768页。PMLR，2023年a。

Zhou, C., Wang, X., 和 Zhang, M. 利用单纯复形上的随机游走提升图神经网络性能。载于神经信息处理系统进展，第36卷，第16172–16206页，2023年b。

Zhou, C., Wang, X., 和 Zhang, M. 利用潜图扩散统一图上的生成与预测。神经信息处理系统进展，第37卷，2024a。

Zhou, C.、Yu, R. 与 Wang, Y. 论高阶图变换器的理论表达能力与设计空间。载于第*27*届国际人工智能与统计会议论文集，机器学习研究论文集第238卷，第2179–2187页。PMLR，2024年5月2日至4日b。

附录

A. 相关工作

生成模型中的对称性。生成建模中许多受关注的数据集在本质上仅在群作用意义下被定义——例如三维空间中的旋转与平移，或集合与图中的置换——因此，一个有原则的生成器应当尊重这些对称性，以避免伪模态以及对等价构型的重复计数。一种主流方法是通过强制群等变性（从而使诱导分布具有不变性）将对称性构建到架构中，例如 E(n)/SE(3) 等变消息传递网络，以及它们在面向几何数据的等变流与扩散／基于分数的模型中的应用（Satorras 等人，2021年；Garcia Satorras 等人，2021年；Hoogeboom 等人，2022年）。系统性研究进一步阐明了在此类几何对称性下等变生成建模的设计空间与实际权衡（Le 等人，2023；Lawrence 等人，2025）。另一种与架构无关的视角通过在群轨道结构上操作来实现对称性，例如借助群平均或规范化／商表示，从而能够使用表达能力强的非等变骨干网络，同时仍得到等变或不变模型；学习式典范化已被证明提供了一种由典范代表元构造等变函数的通用机制（Kaba 等人，2023年）。然而，规范化并非“免费”：对于常见群，存在根本性的连续性／稳定性障碍，这促使人们采用软化或平均化构造，例如加权标架，以缓解对称构型附近的不连续性（Dym 等人，2024）。相关理论分析了规范化或群平均在何种情况下具有更好的统计行为，并突出了不同的泛化机制（Tahmasebi 与 Jegelka，2025）。作为对这些视角的补充，近期关于图扩散的工作认为，强制严格不变性会在对称变换上诱导类混合目标，从而使学习更加困难，并提出在采样时间进行事后群随机化，以在不约束训练模型的情况下恢复不变性（Yan等人，2023）。最后，同期的研究方向明确地在商空间上表述扩散，以避免学习由对称性引入的冗余自由度，进一步凸显了轨道空间视角对于高效生成式学习的价值 (Anonymous, 2026年)。

最优传输。 在连续时间生成建模中，最优传输（OT）常被用作一种原则，用于在简单的基分布与数据分布之间选择成本更低、结构更清晰的概率路径。流匹配中一种广泛采用的选择是 OT 位移插值，它产生更接近最小成本传输的轨迹，并且在经验上能带来更快的收敛和更少的采样步数（Lipman 等人，2022年）。与之相关的“轨迹拉直”观点，例如整流流，可以理解为逐步将学习到的动力学转变为更直的类传输路径，从而进一步提升数值稳定性，并减少推理时所需的积分步数（Liu 等人，2022年；Tong 等人，2024；Kornilov 等人，2024）。

在分子生成中，受最优传输启发的目标函数被更明确地采用，以便在几何对称性下实现高质量的少步生成。对于二维分子图生成，MolGAN（Cao 和 Kipf，2022年）是最早成功使用 Wasserstein-1 距离来稳定分子图生成器训练的方法之一。后续工作，如 BWFlow（Jiang 等人，2025），也展示了最优传输在二维分子设计中的有效性。对于三维分子生成，多种方法已证明，最优传输不仅可作为理论视角，也可作为快速、高保真分子合成的实用设计工具（Song 等人，2023年；Tian 等人，2024；Hong 等人，2025）。

分子生成模型。早期研究主要通过基于字符串或图的参数化方式生成离散分子结构，包括能够显式构建化学有效图的结构化潜变量模型（Jin et al., 2019; Simonovsky & Komodakis, 2018; Jang et al., 2023）。近期一项重要研究方向是在离散图属性上使用扩散过程。离散去噪扩散模型直接处理类别属性，并已在分子基准数据集上展现出强劲的规模化性能（Vignac et al., 2022）。与此同时，基于流的方法为分子图生成提供了基于似然或受传输启发的替代形式（Luo et al., 2021; Shi et al., 2020）；离散流匹配则在保持较高生成质量的同时，进一步提升了采样灵活性和效率（Qin et al., 2024; Hou et al., 2024; Lee et al., 2025; Chen et al., 2023; Luo et al., 2023）。

通过逐步放置原子并保持几何一致性，采用序列式和自回归方案直接在三维空间中生成分子的方法已有探索（Gebauer 等人，2019年；Luo 和 Ji，2022年；Feng 等人，2023年；Wang 等人，2024）。扩散模型通过学习对受扰坐标进行去噪，并采用旨在满足欧几里得对称性的架构，已成为三维分子生成的主流范式（Hoogeboom 等人，2022年）。与之密切相关的条件设定，例如给定固定分子图进行构象生成，也受益于几何扩散和流匹配形式（Xu 等人，2022年；Hassan 等人，2024；Hong 等人，2025；Wu 等人，2022年）。

越来越多的研究致力于联合生成离散分子图和连续三维几何结构，以避免脆弱的事后键推断，并更好地将化学结构与空间排列相结合。MiDi（Vignac 等人，2023年）提出一种混合离散—连续扩散方法，以端到端可微的方式同时生成分子图和构象。后续研究进一步探索了对全面分子表征进行联合扩散（Huang 等人，2023年；Hua 等人，2024；Irwin 等人，2024）。

此外，还有针对分子二维、三维或联合表示开发的潜在生成模型，包括 GeoLDM（Xu 等人，2023年）、LGD（Zhou 等人，2024a）和 LDM-3DG（You 等人，2023年）。最近，GeoRCG（Li 等人，2025）进一步采用了两阶段生成：首先生成分子表示，然后利用几何表示来引导分子生成。我们强调，在温和条件下，我们的分析和所提出的方法同样适用于这些潜在空间／表示空间生成模型，并且有效。

B. 省略的证明

本节给出所有省略的证明，并对主要见于第3节的理论结果进行更详细的讨论。

B.1. 第3.1节的证明

本小节证明了 th　　典范化生成模型能够诱导通用不变分布，并优于
那些等变基线　　直接在不变数据上训练的模型。

B.1.1. 证明第3.1.1 节

我们首先证明典范切片建模的正确性和充分性。

定理 B.1（不变测度的分解；正文中的定理3.1 ）。 假设 *2.5* 和 *2.6* 成立。设μ 是任意一个 $\mathcal{G}$-不变概率测度，定义在 $\mathcal{M}$上。设 Ψ 是一个几乎处处定义的轨道代表映射 μ，并设 $\nu = \Psi\#\mu$ 是定义在 $S = \Psi(\mathcal{M})$ 上的切片分布。则

$$\mu = \int_S \left(\int_{\mathcal{G}} \delta_{g\cdot \mathbf{z}} \, \mathrm{d}\lambda(g) \right) \mathrm{d}\nu(\mathbf{Z}). \tag{19}$$

˜　　˜等价地，若 Z ∼ ν, g ∼ λ 相互独立，则 g ·
Z ∼ µ。

˜证明。这是一个已知结果，此处列出以求完整。设 Z ∼ µ 并定义 Z = Ψ(Z) ∈ S。根据轨道˜代表元的定义，存在 g(Z) ∈ G，使得 Z = g(Z) · Z。根据 2.5， g(Z) 几乎处处唯一。任取可测的 $A \subseteq \mathcal{G}$。对于任意 $h \in \mathcal{G}$，由 μ 的不变性可得

$$\mathbb{P}(g(\mathbf{Z}) \in A \mid \tilde{\mathbf{Z}}) = \mathbb{P}(h\, g(\mathbf{Z}) \in A \mid \tilde{\mathbf{Z}}),$$

˜因此，给定Z 时， g(Z)的条件分布是左不变的。根据紧致 G上的 Haar 概率测度的唯一性，该条件分布为 λ。因此，

$$\begin{aligned} \mu(B) &= \mathbb{E}[\mathbf{1}\{g(\mathbf{Z}) \cdot \tilde{\mathbf{Z}} \in B\}] = \mathbb{E}\Big[\int_{\mathcal{G}} \mathbf{1}\{g \cdot \tilde{\mathbf{Z}} \in B\} \, \mathrm{d}\lambda(g) \Big] \\ &= \int_S \Big(\int_{\mathcal{G}} \delta_{g\cdot \mathbf{z}}(B) \, \mathrm{d}\lambda(g) \Big) \mathrm{d}\nu(\mathbf{Z}), \end{aligned} \tag{20}$$

这就是（5）。 □

推论 B.2（切片建模的充分性；正文中的推论 3.2 ）。 要对任意不变目标分布 μ进行建模，只需对切片分布 ν进行建模；通过 *Haar* 随机化即可恢复不变性。

接下来，我们将展示典范化参数化对于不变函数和等变函数的通用性。

命题 B.3 (等变参数化的通用性；正文中的命题3.3 **)**。 设 $\mathcal{G}$ 在紧集 $K \subset \mathbb{R}^d$上连续作用。假设我们有一个（可测的）规范化映射 $\Psi : K \to K$，如定理 *2.3*，以及一个（可测的）规范映射 $\kappa : K \to \mathcal{G}$，使得

$$\Psi(g \cdot x) = \Psi(x), \; \kappa(g \cdot x) = g\kappa(x), \; x = \kappa(x) \cdot \Psi(x) \tag{21}$$

(定义在一个全测度集上；在自由作用下，κ几乎处处唯一）。考虑参数化 n

$$\phi(x) = \kappa(x) \cdot f(\Psi(x)), \tag{22}$$

其中 f 是 $\Psi(K)$上的通用逼近器。则 ϕ 是 K上连续 $\mathcal{G}$-等变函数的通用逼近器，且 $f \circ \Psi$ 对 K上的连续 $\mathcal{G}$-不变函数是通用的。

证明。 设 $F\colon K \to \mathbb{R}^d$连续且 $\mathcal{G}$-等变： $F(g \cdot x) = g \cdot F(x)$。将其在切片 S上的限制$= \Psi(K)$定义为 $F_S(z) := F(z)$, 其中 $z \in S$。对任意 $x \in K$，记 $x = \kappa(x) \cdot \Psi(x)$。由等变性,

$$F(x) = F(\kappa(x) \cdot \Psi(x)) = \kappa(x) \cdot F(\Psi(x)) = \kappa(x) \cdot F_S(\Psi(x)). \tag{23}$$

Since f 在 S上是通用的，对任意 $\varepsilon > 0$，存在 f 使得$\sup_{z \in S} \|f(z) - F_S(z)\| < \varepsilon$。则

$$\sup_{x \in K} \|\phi(x) - F(x)\| \leq \sup_{z \in S} \|f(z) - F_S(z)\| < \varepsilon, \tag{24}$$

利用群作用保范数这一事实。 □

B.1.2. 证明：第3.1.2节

现在，我们展示具有高表达能力的（可能是）非等变网络的典范化生成模型，相较于采用等变骨干网络的标准生成模型所具备的优越性。

等变模型与非等变模型的表达能力。 众所周知，就区分同构图而言，基于消息传递的图神经网络的表达能力受到抽象图上的一阶 Weisfeiler–Lehman（1-WL）算法的限制（Xu 等人，2018）。由此直接导致了它们无法计数子结构，例如环（Huang 等人，2022年），甚至无法进行链接预测（Zhang 等人，2021年）。由于 Transformer 与 MPNNs 等价（Cai 等人，2023），类似问题同样存在于 Transformer 中；在处理三维几何图时，等变图神经网络（包括等变图神经网络 EGNNs）也存在这些问题（Satorras 等人，2021年）（Li 等人，2023；2024）。目前有一些主流方法可以克服表达能力方面的缺陷：

1. 高阶架构：高阶图神经网络（Morris 等人，2019年；2020；Maron 等人，2019年）和高阶变换器（Kim 等人，2022；Zhou 等人，2024b）在 k-WL 框架内外都具有可证明更强的表达能力，但计算成本呈指数级增长；例如，SwinGNN（Yan 等人，2023年）采用与3-WL 等价的 PPGN（Maron 等人，2019年）作为去噪主干网络。然而，我们认为，更强的表达能力并不一定要依赖复杂度呈指数增长的高阶网络：如下所述，对称性破缺也能提升表达能力。

2. 对称性破缺：通过有意引入非对称性，可以证明底层网络的表达能力得到增强。一种广泛采用的技术是用额外的身份特征对节点进行“标记”（You 等人，2021年；Zhang 等人，2021年；Huang 等人，2022；Zhou 等人，2023a），且表达能力会随着身份标识数量的增加而单调提升。不幸的是，这些方法需要借助关系池化或平均化来产生不变输出，这又会引入群阶数量级的复杂度。规范化则提供了一种无需付出高昂代价进行遍历和平均化操作，即可生成适当且稳定输出的方法（Dong 等人，2024）。

3. 位置编码与结构编码：基于图结构的 PE 显著提升了神经网络骨干的表达能力和实际性能（Rampasek 等人，2022年）。常见的 PE 包括基于空间结构的 PE（例如 RWSE（Dwivedi 等人，2021年）和 RRWP（Shirzad 等人，2023年）），以及基于谱的 PE（利用图拉普拉斯算子的特征值和特征向量（Ma 等人，2023年；Lim 等人，2022年；Bo 等人，2023年）或霍奇拉普拉斯算子（Zhou 等人，2023b））。然而，与图像或文本中基于词元绝对索引的固定位置编码不同，在图上计算上述 PE 依赖于真实值结构，而扩散生成中无法获得该结构。一些生成模型（如 DeFoG（Qin 等人，2024））利用从当前噪声图中估计得到的 PE；我们认为这可能并非最优选择。相比之下，我们的规范化秩也可视为空间和谱位置编码，却几乎不会带来额外计算开销，也不存在训练-测试差距。

使用规范化的非等变骨干网络可证明能降低总变差误差。 紧接着，我们展示了将非等变去噪网络与规范化相结合所具有的卓越表达能力。

假设 B.4 （等变模型的生成建模误差）。 对于任意目标不变分布 μ，假设存在参数 θ(eq)，使得等变生成模型 $\hat{\nu}_{\theta(\text{equi})}$ 满足 TV($\hat{\mu}_{\theta(\text{eq})}, \mu$) $< \varepsilon$。

因此我们得出结论：规范化的非等变模型的生成建模误差可能更小。

定理 B.5 （规范化模型在 TV 距离上匹配（或优于）等变基线）。 在假设 *2.5* 和*2.6*下，令μ 为定义在 $\mathcal{G}$上的任意 $\mathcal{M}$ -不变目标分布，并令 $\nu := \Psi_{\#}\mu$ 为其定义在 $S = \Psi(\mathcal{M})$上的规范切片分布。假设存在一个等变基线（定义在环境空间中），其生成的不变模型分布为$\hat{\mu}_{\theta(\text{eq})}$，且满足

$$\mathrm{TV}\big(\hat{\mu}_{\theta(\text{eq})}, \mu\big) < \varepsilon. \tag{25}$$

令 $\hat{\nu}_{\theta(\text{eq})} := \Psi_{\#}\hat{\mu}_{\theta(\text{eq})}$ 为诱导的切片分布。若切片模型类 $\{\hat{\nu}_{\theta(\text{free})}\}$ 在 S 上的表达能力至少相当于以下意义：

$$\inf_{\theta(\text{free})} \mathrm{TV}\big(\hat{\nu}_{\theta(\text{free})}, \nu\big) \;\le\; \mathrm{TV}\big(\hat{\nu}_{\theta(\text{eq})}, \nu\big), \tag{26}$$

那么存在 θ（自由），使得 *Haar* 随机化模型

$$\hat{\mu}_{\theta(\text{free})} := \int_S \left(\int_{\mathcal{G}} \delta_{g\cdot \mathbf{z}} \, \mathrm{d}\lambda(g) \right) \mathrm{d}\hat{\nu}_{\theta(\text{free})}(\mathbf{Z}) \tag{27}$$

满足 总变差$\big(\hat{\mu}_{\theta(\text{free})}, \mu\big) < \varepsilon$。

证明。 首先，推前不会增加总变差（数据处理不等式）。因此

$$\mathrm{TV}(\hat{\nu}_{\theta(\text{eq})}, \nu) = \mathrm{TV}\big(\Psi_{\#}\hat{\mu}_{\theta(\text{eq})}, \Psi_{\#}\mu\big) \le \mathrm{TV}\big(\hat{\mu}_{\theta(\text{eq})}, \mu\big) < \varepsilon. \tag{28}$$

根据（26）（注意，根据上文讨论，这通常成立），选取 θ（自由）使得 总变差($\hat{\nu}_{\theta(\text{free})}, \nu$) ≤总变差($\hat{\nu}_{\theta(\text{eq})}, \nu$) $< \varepsilon$。最后，Haar 随机化是一个马尔可夫核 T （参见定理 3.1 ），因此是总变差不扩张的：

$$\mathrm{TV}\big(T(\hat{\nu}_{\theta(\text{free})}), T(\nu)\big) \le \mathrm{TV}\big(\hat{\nu}_{\theta(\text{free})}, \nu\big). \tag{29}$$

由于 T（ν） $= \mu$（定理 3.1），我们得出 总变差($\hat{\mu}_{\theta(\text{free})}, \mu$) $< \varepsilon$。 □

B.2. 第 3.2 节的证明

在本小节中，我们将详细分析并完整证明关于通过规范化加速扩散模型和流模型训练的结果。

B.2.1. 证明第节 3.2.1

本部分主要讨论基于得分的扩散建模，说明不变数据中群对称性的存在会导致得分函数具有混合结构。

命题 B.6 （对称性混合分布的得分）。假设在需要处 q 可微且为正。则

$$\nabla \log p(x) = \sum_{m=1}^{M} w_m(x)\, g_m \cdot \nabla \log q(g_m^{-1}\cdot x), \quad w_m(x) := \frac{q(g_m^{-1}\cdot x)}{\sum_{j=1}^{M} q(g_j^{-1}\cdot x)}. \tag{30}$$

证明。对 $\log g_m \cdot \nabla q\ g_m^{-1}\cdot x$ = ( )

$\log(\frac{1}{M}\sum_m q(g_m^{-1}\cdot x))$ 求导，并应用链式法则。正交性意味着 $\nabla_x q(g_m^{-1}\cdot x)$ ( )。 □

B.2.2. 证明第节 3.2.2

这篇论文 rt 强调了我们的核心结果：规范训练降低了流匹配中的条件方差 g.

流匹配中的不可约误差。首先，我们证明一个标准引理：不可约流匹配误差就是条件方差。

引理 B.7 (不可约流匹配误差 = 条件方差)。 设 $\hat{v}$ 是从$(t,\mathbf{Z}_t)$预测 $v_t(\mathbf{Z}_t)$ 的任意可测预测器。那么，可达到的最小均方误差满足

$$\inf_{\hat{v}} \mathbb{E}\big[\,\|\hat{v}(t,\mathbf{Z}_t)-(\mathbf{Z}_1-\mathbf{Z}_0)\|^2\,\big] = \mathbb{E}\big[\mathrm{Var}(\mathbf{Z}_1-\mathbf{Z}_0 \mid t,\mathbf{Z}_t)\big]. \tag{31}$$

证明。 这是标准的 L^2 回归恒等式：条件期望 $\mathbb{E}[\mathbf{Z}_1-\mathbf{Z}_0 \mid t,\mathbf{Z}_t]$ 是唯一的最小化者，且最小风险等于条件方差的期望（全方差公式）。 □

推论 B.8 (确定性耦合导致不可约误差为零)。 如果$\mathbf{Z}_1-\mathbf{Z}_0$ 是$(t,\mathbf{Z}_t)$的（几乎处处）确定性函数，那么右侧为 0；原则上，流可以以任意小的误差进行学习（受近似和优化的限制）。

下面我们给出一个关于这种不可约误差的具体示例，说明在存在多群歧义时，可达到的最小均方误差损失严格大于零。

通过后验碰撞界量化对称性歧义 我们现在通过以下方式强化方差分解视角：**(i)** 给出定理 对称性歧义项的精确恒等式（见定理 3.5）；**(ii)** 仅根据潜在对称性的后验 $\varrho(g \mid Z_t)$推导下界。

在群对齐提升耦合下，将潜在对称性的后验分布写作 $\varrho(g \mid z) := \mathbb{P}(G=g \mid Z_t=z)$（有限群情形；对于紧群，将 $\varrho(\cdot \mid z)$ 解释为相对于 Haar 测度的后验密度）。
我们首先证明，歧义项可以写成成对均值分离恒等式。以下恒等式是标准结果（通过独立同分布副本表示方差），但将其应用于此处可为对称性歧义项提供明确的几何刻画。

引理 B.9 （对称性歧义表示为成对条件均值分离）。定义群条件均值漂移

$$m_g(z) := \mathbb{E}[U \mid Z_t=z,\ G=g], \qquad m(z) := \mathbb{E}[U \mid Z_t=z]. \tag{32}$$

那么，定理3.5中的歧义项对每个 z，

$$\mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t=z) = \mathbb{E}_{g\sim\varrho(\cdot|z)}\big[\|m_g(z)-m(z)\|^2\big] \tag{33}$$

$$= \frac{1}{2}\,\mathbb{E}_{g,g'\overset{i.i.d.}{\sim}\varrho(\cdot|z)}\big[\|m_g(z)-m_{g'}(z)\|^2\big]. \tag{34}$$

证明。 式（33）是取值为 $m_G(z)$ 的离散随机变量条件方差的定义。式（34）由以下关于独立同分布副本的标准恒等式 $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}\|X-X'\|^2$ 推出 $(X,\ X')$（在给定 $Z_t=z$ 的条件下应用）。 □

在此给出后验不确定性（碰撞概率）的下界。引理 B.9表明，只要后验分布 $\varrho(g \mid Z_t)$ 将概率质量分散到条件均值漂移存在显著差异的群元素上，歧义就会很大。

命题 B.10 （后验碰撞下界）。 固定 z，并设 $\mathcal{G}_\Delta(z) \subseteq \mathcal{G}$ 为任意子集，使得对所有 $g \in \mathcal{G}_\Delta(z)$，

$$\|m_g(z)-m(z)\| \geq \Delta(z). \tag{35}$$

Then

$$\mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t=z) \ \geq\ \Delta(z)^2\ \mathbb{P}\big(G \in \mathcal{G}_\Delta(z) \mid Z_t=z\big). \tag{36}$$

特别地，若存在两个不相交的子集 $\mathcal{A}(z)$ $\mathcal{B}(z)$，使得对所有 $g \in \mathcal{A}(z)$ 和$g' \in \mathcal{B}(z)$，均有 $\|m_g(z) - m_{g'}(z)\| \geq \Delta(z)$，则

$$\mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t=z) \ \geq\ \frac{\Delta(z)^2}{2}\,\Big(1-\sum_g \varrho(g \mid z)^2\Big), \tag{37}$$

其中 $1-\sum_g \varrho(g \mid z)^2$ 是后验碰撞概率的补数。

证明。第一个界可直接由 (33) 得出，只保留 $g \in \mathcal{G}_\Delta(z)$，并在此处使用 $\|m_g(z) - m(z)\|^2 \geq \Delta(z)^2$。对于第二个界，使用 (34)，并在 $(g, g') \in \mathcal{A}(z) \times \mathcal{B}(z) \cup \mathcal{B}(z) \times \mathcal{A}(z)$ 时，将成对距离下界为 $\Delta(z)$。然后

$$\mathbb{P}(g \neq g' \mid z) = 1 - \sum_g \varrho(g \mid z)^2, \tag{38}$$

因此，所述不等式成立（相差因子 $1/2$，根据（34））。 □

碰撞项 $1 - \sum_g \varrho(g \mid Z_t)^2$ 为 0，当且仅当 G 几乎处处由 Z_t 决定。因此，除非条件均值漂移 $m_g(Z_t)$ 在不同的 g 之间一致，否则只要 G 仍存在不确定性，对称性歧义就会产生严格为正的方差分量。

关于提升耦合的讨论。一个微妙但重要的要点是，（12）中的群对齐提升在 G 之间共享相同的潜变量 G，涉及 (Z_0, Z_1)，这原则上可能会引入依赖性。然而，对于标准各向同性高斯分布先验和正交群作用（包括 S_N 置换和 $SO(3)$ 旋转），这种依赖性会消失。

命题 B.11（乘积切片耦合的提升等于环境空间乘积耦合）。假设 $\mathcal{G}$ 在 $\mathbb{R}^d$ 上正交作用。令 $G \sim \lambda$ 和 $(Z_0, Z_1) \sim q_0 \otimes \mathcal{N}(0, I)$ 相互独立，并定义 $(Z_0, Z_1) = (G \cdot Z_0, G \cdot Z_1)$。那么 *(i)* $Z_1 \sim \mathcal{N}(0, I)$ 和 $Z_1 \perp G$，以及 *(ii)* $Z_1 \perp Z_0$。此外，如果不变数据分布满足解体关系 $p_0 = \int (g\cdot)_{\#} q_0 \, d\lambda(g)$，那么提升后的环境耦合满足

$$(Z_0, Z_1) \sim p_0 \otimes \mathcal{N}(0, I). \tag{39}$$

证明。对于任意可测的 B 和任意 g，$\mathcal{N}(0, I)$ 的正交不变性意味着 $\mathbb{P}(Z_1 \in B \mid G = g) = \mathbb{P}(g \cdot Z_1 \in B) = \mathbb{P}(Z_1 \in B)$，其不依赖于 g，因此 $Z_1 \perp G$ 和 $Z_1 \sim \mathcal{N}(0, I)$。对于任意可测的 A，B，

$$\mathbb{P}(Z_0 \in A, Z_1 \in B) = \mathbb{E}\Big[\mathbf{1}_{\{Z_0 \in A\}} \mathbb{P}(Z_1 \in B \mid G, \widetilde{Z}_0)\Big] = \mathbb{P}(\widetilde{Z}_1 \in B)\, \mathbb{P}(Z_0 \in A), \tag{40}$$

由于 Z_1 与 (G, Z_0) 独立，且 $\mathbb{P}(g \cdot Z_1 \in B)$ 关于 g 为常数。因此 $Z_1 \perp Z_0$。若进一步通过所述解体得到 $Z_0 \sim p_0$，则联合分布为 $p_0 \otimes \mathcal{N}(0, I)$。 □

因此，常用的环境空间乘积耦合 $\gamma_{\text{mix}} = p_0 \otimes \mathcal{N}(0, I)$ 可视为采用乘积耦合和各向同性高斯分布切片先验的切片训练这一特例，因此两种范式的条件方差可直接比较。在此情形下，规范化不会改变耦合；它改变的是表示形式，从而使定理 3.5 中的群歧义项在切片上变得明确且可避免。因此，我们的理论涵盖了大多数无需最优传输的实际训练方案。

然而，如果 Z_1 是群感知的（例如，在我们的实现中采用学习得到的先验或群统计相关先验），或者 Z_1 与 Z_0 不独立（例如，在规范化切片上进行最优传输），那么通常这一结论并不成立。特别是，如果切片先验不是各向同性的（例如，一个对齐高斯分布 $\mathcal{N}(\mu, \Sigma)$，其中某个 g 满足 $g\Sigma g^\top \neq \Sigma$），那么从分布上看，$g \cdot \widetilde{Z}_1$ 依赖于 g，因此 Z_1 不再独立于 G，提升耦合通常也不是乘积耦合。同样，如果切片耦合 $\tilde{\gamma}$ 是非乘积的（例如，最优传输/蒙日），那么提升耦合仍然是非乘积的。

规范化通过消除对称性歧义来降低条件方差。我们现在证明规范化流匹配优势的核心结果。

定理 *B.12*（群对齐提升下的方差分解；正文中的定理 *3.5*）。假设 $\mathcal{G}$ 正交地作用。在群对齐提升耦合 $G \sim \lambda$ 独立于 (Z_0, Z_1) 和 $(Z_0, Z_1) = (G \cdot Z_0, G \cdot Z_1)$ 的条件下，令 $Z_t = (1-t)Z_0 + tZ_1$ 且 $\widetilde{Z}_t = (1-t)Z_0 + tZ_1$。令 $U := Z_1 - Z_0 = G \cdot \Delta$ 和 $\Delta := Z_1 - Z_0$，则有

$$\mathrm{Var}(U \mid Z_t) = \underbrace{\mathbb{E}\Big[\mathrm{Var}(\Delta \mid \widetilde{Z}_t) \mid Z_t\Big]}_{\textit{within-slice difficulty}} + \underbrace{\mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t)}_{\textit{symmetry ambiguity} \geq 0}. \tag{41}$$

因此，

$$\mathbb{E}\big[\mathrm{Var}(U \mid Z_t)\big] \;\geq\; \mathbb{E}\big[\mathrm{Var}(\Delta \mid \widetilde{Z}_t)\big]. \tag{42}$$

Proof。 对潜在对称变量 G 应用标准的全方差公式：

$$\mathrm{Var}(U \mid Z_t) = \mathbb{E}[\mathrm{Var}(U \mid Z_t, G) \mid Z_t] + \mathrm{Var}(\mathbb{E}[U \mid Z_t, G] \mid Z_t)\,. \tag{43}$$

第二项非负。对于第一项，在给定 $(Z_t,\ G)$ 的条件下，$\widetilde{Z}_t = G^{-1} \cdot Z_t$ 是确定的。在提升~耦合下，$(\Delta,\ Z_t)$ 与 G 相互独立，而正交性意味着 $\mathrm{Var}(G \cdot X \mid \cdot) = \mathrm{Var}(X \mid \cdot)$。因此，

$$\mathrm{Var}(U \mid Z_t, G) = \mathrm{Var}(G \cdot \Delta \mid \widetilde{Z}_t, G) = \mathrm{Var}(\Delta \mid \widetilde{Z}_t). \tag{44}$$

代入 式 (43) 得到式 (13)。对 Z_t 取期望并利用塔式性质，可得式 (14) . □

定理 3.5 中的下界是“切片内”条件方差；两者之间的差距恰好是由于不知道是哪一个对称元素生成了观测而产生的额外不确定性。这正式阐明了这一直觉：对称性会导致群元素的后验多模态性，从而增大条件方差，并为粗略求解器带来不可约误差下限。

备注 B.13（此处比较的内容及不比较的内容）。 上述引理在相同的提升耦合下，比较“边缘化与不边缘化 G”。它并不意味着，在耦合方式任意选择的情况下，规范化训练范式普遍优于其他任何范式。其价值在于分离出一个完全由对称性歧义导致、且在切片上不存在的非负方差分量。然而，正如命题 B.11所述，提升耦合很容易还原大多数实际流匹配模型（不使用 OT）所采用的标准乘积高斯噪声。

接下来我们将进一步解释备注 3.6。固定一个耦合 γ 定义在 $(Z_0,\ Z_1)$ 上，并定义线性插值

$$Z_t = (1-t)Z_0 + tZ_1, \qquad U := Z_1 - Z_0. \tag{45}$$

对于任意预测器 $v\colon \mathbb{R}^d \to \mathbb{R}^d$，时刻 t 的总体平方损失为

$$\mathcal{L}_t(v) := \mathbb{E}\big[\|v(Z_t) - U\|^2\big]. \tag{46}$$

众所周知， 贝叶斯回归函数为 $v^\star(z) = \mathbb{E}[U \mid Z_t = z]$，贝叶斯风险等于条件方差 ：

$$\inf_v \mathcal{L}_t(v) = \mathbb{E}\big[\mathrm{Var}(U \mid Z_t)\big]. \tag{47}$$

规范化不会降低等变模型的环境贝叶斯风险。设紧群 $\mathcal{G}$ 正交地作用于 $\mathbb{R}^d$（例如置换和旋转）。设 γ 是对角不变的：

$$(g, g)_{\#}\gamma = \gamma \quad \forall g \in \mathcal{G}, \tag{48}$$

这对于环境空间乘积耦合 $p_0 \otimes \mathcal{N}(0,\ I)$ 在 p_0 为 $\mathcal{G}$-不变时尤其成立，对于对称化最优传输耦合也成立。

命题 B.14（对角不变性下贝叶斯回归函数的等变性）。 假设 $\mathcal{G}$ 正交作用，且 γ 具有对角不变性。那么贝叶斯回归函数 $v^\star(z) = \mathbb{E}[U \mid Z_t = z]$ 是 $\mathcal{G}$-等变的：

$$v^\star(g \cdot z) = g \cdot v^\star(z) \qquad \forall g \in \mathcal{G}. \tag{49}$$

因此，将范围限制为等变函数类 $\mathcal{H}_{\mathrm{eq}} := \{v\colon \ v(g \cdot z) = g \cdot v(z)\}$ 不会 增加最小总体损失：

$$\inf_{v \in \mathcal{H}_{\mathrm{eq}}} \mathcal{L}_t(v) = \inf_v \mathcal{L}_t(v) = \mathbb{E}[\mathrm{Var}(U \mid Z_t)]. \tag{50}$$

证明。 对角不变性意味着 $(Z_t, U) \overset{d}{=} (g \cdot Z_t, g \cdot U)$。因此，对于任意可测集 A,

$$\mathbb{E}[U \mathbf{1}_{\{Z_t \in A\}}] = \mathbb{E}[g \cdot U\ \mathbf{1}_{\{g \cdot Z_t \in A\}}] = g \cdot \mathbb{E}[U \mathbf{1}_{\{Z_t \in g^{-1}A\}}], \tag{51}$$

根据条件期望的拉东–尼科迪姆刻画，可得 $v^\star(g \cdot z) = g \cdot v^\star(z)$。由于 $v^\star \in \mathcal{H}_{\mathrm{eq}}$，下确界相等。

□

注释 B.15。 如果保留 相同的环境回归问题 (47)，那么对输入进行规范化并不会改善等变模型类的总体最优值，因为贝叶斯回归函数本身已经是等变的（命题B.14）。

规范条件拓展了非等变模型的有效假设类。 当模型在架构上不 具备等变性时，规范化最为有用。直观而言，环境空间中的非等变模型必须从数据中学习对称平均。规范化和规范条件提供了一种规范 （例如，规范化器 Ψ 以及相关联的群元素 $\kappa(\cdot)$），使非等变模型能够通过显式公式实现等变行为。

我们假设存在一个精确的规范化器模型，即规范化器提供 $\Psi:\ \mathbb{R}^d \to S$ (切片) 和 $\kappa:\ \mathbb{R}^d \to \mathcal{G}$，使得 $\Psi(x) = \kappa(x)^{-1} \cdot x$ 在全测度集上成立（考虑稳定子的影响）。

命题 B.16 （典范条件提升实现等变函数）。 设 $\mathcal{A}$ 是切片 S 上的任意函数类（不必具有等变性）。使用典范条件定义诱导的环境类：

$$\mathcal{H}_{\text{lift}}(\mathcal{A}) := \big\{\, h(x) = \kappa(x) \cdot a(\Psi(x)) \ :\ a \in \mathcal{A} \,\big\}. \tag{52}$$

于是，每个 $h \in \mathcal{H}_{\text{lift}}(\mathcal{A})$ 都是 $\mathcal{G}$-等变的（在 Ψ、κ 一致的集合上）。此外，如果 $\mathcal{A}$ 在 S 上具有普适性，那么 $\mathcal{H}_{\text{lift}}(\mathcal{A})$ 可以在 $\mathbb{R}^d$ 上近似任意等变目标函数，但需考虑通常的稳定子限制。

证明。等变性由 $\Psi(g \cdot x) = \Psi(x)$ 和 $\kappa(g \cdot x) = g\kappa(x)$ 推出（在自由作用情形下）：

$$h(g \cdot x) = \kappa(g \cdot x) \cdot a(\Psi(g \cdot x)) = g\kappa(x) \cdot a(\Psi(x)) = g \cdot h(x). \tag{53}$$

通用性得以继承，因为任何等变函数都由其在切片上的限制所确定。 □

注释B.17（为何这对非等变网络来说是真正的“捷径”）。一个只能看到x 的通用非等变网络，必须隐式地学习等变结构。提供（$\Psi(x)$ $\kappa(x)$）（或规范秩／帧 ID 等离散代理），即可像命题B.16中那样，通过构造实现等变性。这可以降低逼近复杂度并改善优化效果。

总而言之，切片训练并不声称能够降低环境贝叶斯风险（47）。实际能够达到的经验贝叶斯风险仍取决于架构，而且对于等变模型而言无法进一步降低（命题B.14）。~相反，它在切片上学习一个不同的回归问题（根据 Δ 预测 Z_t），其不可约方差是定理3.5 中的第一项，并且可以证明不大于环境中的对应方差。这正是规范化能够让非等变模型更容易训练、采样所需步数更少的原因。s

B.2.3. 第 3.2.3 节的证明 3.2.3

T本部分为对齐切片先验、最优传输耦合和训练—测试一致性提供补充 y.

规范化消除了对称性歧义项，但剩余的切片内条件方差 $\mathbb{E}[\mathrm{Var}(\Delta \mid Z_t)]$ 取决于切片先验 q_1 和切片耦合 $\widetilde{\gamma}$。第3.2.3节说明，规范化后剩余的~不可约项是切片内难度 $\mathbb{E}[\mathrm{Var}(\Delta \mid Z_t)]$, ，它取决于 (i) 切片先验 q_1 和 (ii) 切片耦合 $\widetilde{\gamma}$ (定理 3.5)。本附录三级小节汇集了三项简短的技术补充：(i) 给出闭式表达式，说明未对齐的高斯先验如何在乘积耦合下增大切片内方差；(ii) 解释最优传输／近莫恩日耦合为何互为补充，却可能因对称性而不唯一；(iii) 给出一条最简一致性陈述，阐明训练时最优传输何时与推理兼容，以及条件化何时会造成不匹配。

乘积耦合下的高斯错配：切片内方差的精确公式。 规范化消除了对称性歧义项，但即使在切片上，如果先验与切片几何对齐不佳，条件方差仍可能很大。最简单（也很常见）的基线是乘积耦合 $\widetilde{\gamma} = q_0 \otimes q_1$。以下命题给出了线性路径（整流流）设定下切片内条件协方差的精确表达式，说明 Σ_0 与 Σ_1 之间的不匹配如何决定不可约误差。

命题 B.18 （独立高斯分布的闭式切片内条件方差）。 假设切片数据和切片先验是相互独立的高斯分布：

$$\widetilde{Z}_0 \sim \mathcal{N}(\mu_0, \Sigma_0), \qquad \widetilde{Z}_1 \sim \mathcal{N}(\mu_1, \Sigma_1), \qquad \widetilde{Z}_0 \perp \widetilde{Z}_1.$$

~ ~ ~ ~ ~令 $Z_t = (1-t)Z_0 + tZ_1$ 且 $\Delta = Z_1 - Z_0$。则

$$\mathrm{Cov}(\Delta \mid \widetilde{Z}_t) = \frac{1}{t^2}\,\mathrm{Cov}(\widetilde{Z}_0 \mid \widetilde{Z}_t), \tag{54}$$

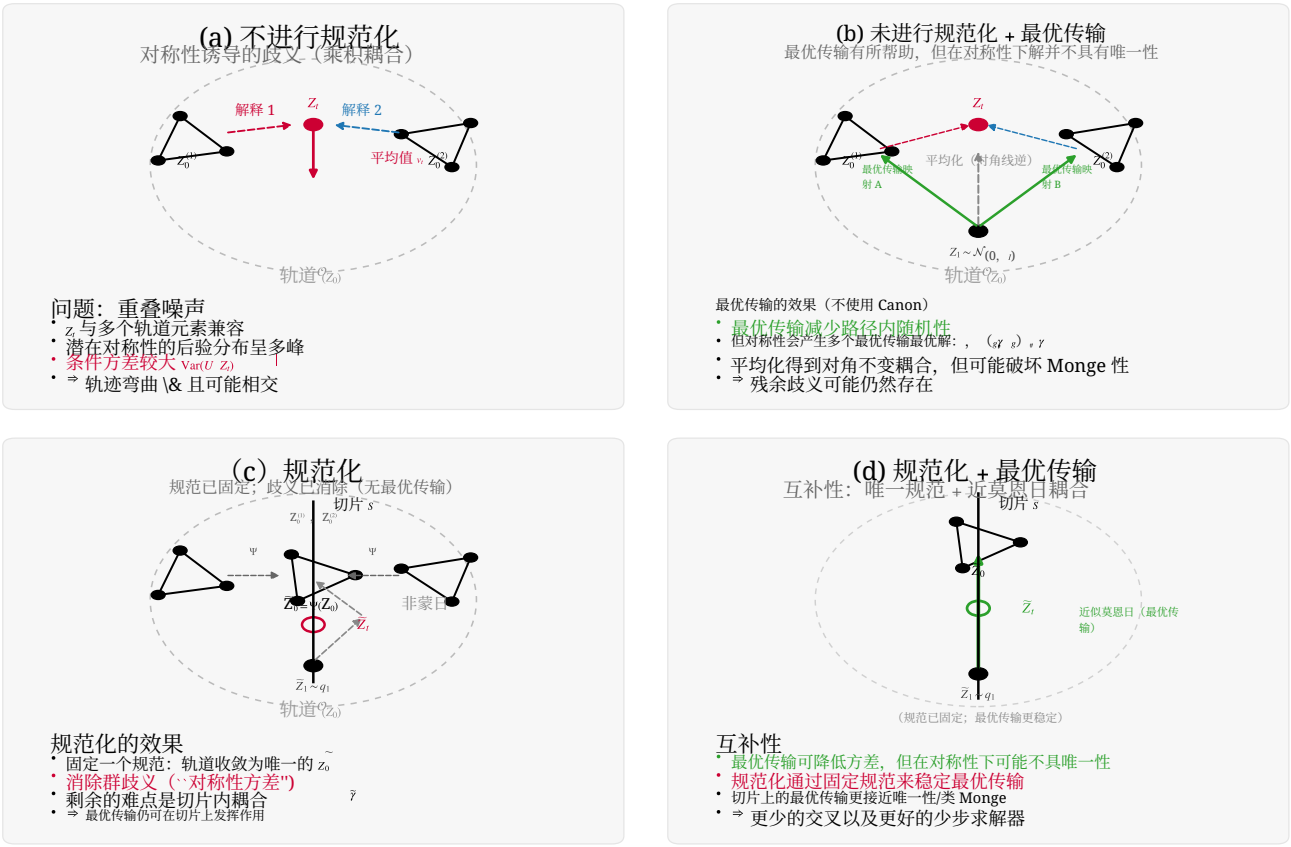


图 *4.* 最优传输与规范化互补作用。无论是否进行规范化，最优传输都可以降低条件方差，但在存在对称性时，最优传输解通常不唯一（若 γ 最优，则 $(g,g)_{\#}\gamma$ 同样最优）；平均化可得到对角不变最优解，但可能破坏 Monge 性。规范化固定一个规范（将每条轨道坍缩为唯一的切片代表元），消除群歧义并稳定切片上的最优传输。

且条件协方差具有闭式表达

$$\mathrm{Cov}(\widetilde{Z}_0 \mid \widetilde{Z}_t) = \Sigma_0 - (1-t)^2\,\Sigma_0\Big((1-t)^2\Sigma_0 + t^2\Sigma_1\Big)^{-1}\Sigma_0. \tag{55}$$

因此，流匹配中的切片内不可约误差为

$$\mathbb{E}\big[\mathrm{Var}(\Delta \mid \widetilde{Z}_t)\big] = \mathrm{tr}\left(\frac{1}{t^2}\Sigma_0 - \frac{(1-t)^2}{t^2}\,\Sigma_0\Big((1-t)^2\Sigma_0 + t^2\Sigma_1\Big)^{-1}\Sigma_0\right). \tag{56}$$

证明。恒等式 $\Delta = (\widetilde{Z}_t - \widetilde{Z}_0)/t$ 是代数恒等式，因此 $\mathrm{Cov}(\Delta \mid \widetilde{Z}_t) = \frac{1}{t^2}\mathrm{Cov}(\widetilde{Z}_0 \mid \widetilde{Z}_t)$。由于 $(\widetilde{Z}_0, \widetilde{Z}_t)$ 服从联合高斯分布，且 $\mathrm{Cov}(\widetilde{Z}_0, \widetilde{Z}_t) = (1-t)\Sigma_0$，$\mathrm{Cov}(\widetilde{Z}_t) = (1-t)^2\Sigma_0 + t^2\Sigma_1$，由标准高斯条件公式可得 (55)。取迹即可得到最终表达式。 □

因此，若使用各向同性先验 $q_1 = \mathcal{N}(0, I)$，而 q_0 呈强各向异性（Σ_0 中条件数很大），则上述迹在 Σ_0 主导 I 的方向上仍然较大，从而形式化地说明：即使经过规范化，一个未对齐的简单先验仍可能降低少步准确率。一种实用的补救方法是在易于处理的 $\widetilde{\quad}$ $\widetilde{\quad}$族中选择对齐先验，例如矩匹配高斯分布 $q_1^\star = \mathcal{N}(\mathbb{E}_{q_0}[Z_0], \mathrm{Cov}_{q_0}(Z_0))$（到高斯分布上的 KL投影），或学习得到的先验。虽然可以学习规范先验 q_1，但使用 q_0 的 KL 投影是一种简单而有效的方法。该计算仅发生在预处理中，不会给训练带来任何额外开销。例如，在高斯分布中，矩匹配高斯分布是 q_0 的 KL投影：

命题 B.19（矩匹配高斯分布是 KL 最优高斯近似）。 设 q_0 为 $\mathbb{R}^d$ 上具有有限二阶矩的任意分布。考虑族 $\mathcal{N}(\mu, \Sigma)$，其中 $\Sigma \succ 0$。$\mathrm{KL}(q_0 \| \mathcal{N}(\mu, \Sigma))$ 的极小化元为

$$\mu^\star = \mathbb{E}_{q_0}[\widetilde{Z}_0], \qquad \Sigma^\star = \mathrm{Cov}_{q_0}(\widetilde{Z}_0). \tag{57}$$

证明。 这是标准的指数族投影：高斯对数密度是充分统计量（x， $xx^\top$）的仿射函数，因此最小化 KL($q_0 \| \mathcal{N}(\mu$， Σ)) 可得到矩匹配。 □

最优传输/近莫恩日耦合：与规范化互为补充，但对称性可能破坏唯一性。 给定$\widetilde{\ }(q_0$， $q_1)$，选择 $\widetilde{\gamma}$ 接近最优传输耦合的耦合，可使传输更加确定性（类 Monge），减小 $\mathbb{E}[\mathrm{Var}(\Delta \mid Z_t)]$并使轨迹更平直（从而有利于少步求解器）。然而，在具有对称边缘分布和对称代价的环境空间中，最优传输解通常并不唯一：若 γ 是最优解，那么 $(g$， $g)_{\#}\gamma$ 也同样最优。对 g 进行平均化可得到对角不变最优解，但可能会破坏 Monge 性，将确定性映射变成混合分布。规范化固定了一种规范，因此能够稳定 最优传输：在切片上，由对称性引起的退化有所减轻，而且经验上最优传输解更接近唯一/类 Monge。这正是规范化与最优传输互为补充的确切含义：规范化消除了群歧义（定理 3.5），而最优传输则针对剩余的切片内难度。图 4 对此作了说明。

训练时最优传输与推理兼容；失配源于条件变量。一个常见的疑问是，训练时使用最优传输是否意味着推理时也必须使用最优传输。在流匹配／整流流中，最优传输会改变训练耦合 $\widetilde{\gamma}$（进而改变监督信号），但推理只需要从边缘先验 q_1 中采样，并对学习到的动力学进行积分。特别是，如果总体向量场定义良好且足够规整，那么在给定正确向量场的情况下，采样无需耦合$\widetilde{\ }$：抽取 $Z_1 \sim q_1$，并对学习到的常微分方程／随机微分方程进行积分，即可获得来自 q_0的样本；推理时不需要成对的端点。

实际的训练–测试失配风险源于模型以辅助变量 C 为条件（例如“规范秩”$\widetilde{\ }$位置编码、帧 ID，或规范化器的任何确定性／随机输出）。设 $\widetilde{\pi}_1^{\mathrm{tr}}(C$， $Z_1)$为起始时刻的训练时联合分布， $\widetilde{\pi}_1^{\mathrm{inf}}(C$， $\widetilde{Z}_1)$为推理时联合分布。避免失配的一个充分条件是：

$$\widetilde{\pi}_1^{\mathrm{tr}}(C, \widetilde{Z}_1) = \widetilde{\pi}_1^{\mathrm{inf}}(C, \widetilde{Z}_1). \tag{58}$$

特别地，如果 C 是规范坐标系中固有的确定性索引（例如，坐标索引 i 是 规范秩），那么无论选择 q_1为何，该等式都成立；反之，如果推理阶段以不同于训练阶段的方式，根据采样噪声计算 C，则失配不可避免。最后请注意，在推理阶段规范化噪声 （例如，$\widetilde{\ }$设置 $Z_1 := \Psi(\varepsilon)$，其中 $\varepsilon \sim \mathcal{N}(0, I)$）通常会诱导出复杂的切片先验 $\Psi_{\#}\mathcal{N}(0, I)$；除非训练时使用相同的诱导先验，否则这会造成原本可以避免的先验失配，并可能增大切片内方差。

C. 实现细节

C.1. 规范化方法详述

C.1.1. 几何谱排序

我们定义一个确定性的规范化函数 $\kappa : \mathcal{M} \to \mathcal{M}$，它将每个分子图映射到其在 $S_N \times SE(3)$ 作用下的轨道中的唯一代表元。关键在于利用由三维原子坐标构造的几何拉普拉斯；其 Fiedler向量以连续且旋转不变的方式编码了分子连接结构。

几何拉普拉斯的构造与规范置换。 几何拉普拉斯的构造以及（带符号的）Fiedler向量的计算，可以简洁地概括为算法1中的伪算法。

规范排序则为： $\pi^* = \mathrm{argsort}(u_2)$ 规范化分子为 $\kappa(\mathcal{M}) = \pi^*(\mathcal{M})$。Fiedler向量近似刻画了图拉普拉斯算子的基本振荡模式，从而提供一种“从核心到外围”的排序。Fiedler值相近的原子往往在空间上接近且结构等价，这诱导出一种从分子核心到官能团的自然生成顺序；见图5 中的说明性示例。

规范 $SO(3)$ 框架。 上述过程确定了 S_N上的规范顺序，并以此为基础进一步（可选地）使用算法处理 $SO(3)$对称性2。

算法1 几何拉普拉斯与带符号Fiedler向量

输入：坐标 X $=$(x1, . . . , x$_N$) $\in \mathbb{R}^{N\times 3}$，键集 E。计算平均键长 $\bar{d}_{\text{bond}} \leftarrow \frac{1}{|E|}\sum_{(i,j)\in E}\|\mathrm{x}_i - \mathrm{x}_j\|$。设置带宽 $\sigma^2 \leftarrow 4\,\bar{d}^2_{\text{bond}}$。对于 $i = 1, \ldots, N$ 执行 对于 $j = 1, \ldots, N$ 执行 如果 $i = j$ 则 $W_{ij} \leftarrow 0$。否则$W_{ij} \leftarrow \exp\left(-\|\mathrm{x}_i - \mathrm{x}_j\|^2/(2\sigma^2)\right)$。结束 if 结束 for 结束 for 度矩阵 D $\leftarrow$ diag($\sum_j W_{ij}$)。随机游走拉普拉斯 $L_{\text{rw}} \leftarrow$ D^{-1}(D $-$ W)。计算 L_{rw} 的特征对： $0 = \lambda_1 \leq \lambda_2 \leq \cdots$，其特征向量为 $\{u_k\}$

Fiedler向量 $u_2 \leftarrow$是 λ_2的特征向量。质心 $\bar{\mathbf{x}} \leftarrow \frac{1}{N}\sum_i \mathbf{x}_i$，平均半径为 $\bar{d} \leftarrow \frac{1}{N}\sum_i \|\mathbf{x}_i - \bar{\mathbf{x}}\|$。根据质心方向约定确定符号：$u_2 \leftarrow$ 符号$\left(\sum_i u_{2,\ i}(\|\mathbf{x}_i - \bar{\mathbf{x}}\| - \bar{d})\right) \cdot u_2$。**输出：**带符号Fiedler向量 u_2。

C.1.2. 替代排序

图规范化是一个受到广泛研究的课题，也存在其他已有的规范化方法。然而，其中大多数是针对抽象图中的置换定义的（Zhao等人，2024；Ma等人，2023年；Dong等人，2024），而非针对几何图。

我们还实现了两种更简单的排序方法，用于消融实验：

1. 结构多跳（受 PARD（Zhao 等人，2024）启发）：通过加权多跳度进行迭代度剥离 w_K（v）$= \sum_{k=1}^{K} d_k$（v）$\cdot N^{K-k}$，其中 d_k（v）统计恰好相隔 k 跳的节点。
2. 原子序编号：按照原子序数进行优先级排序（例如，优先排列重原子或稀有原子，氢最后排列）。

然而，我们通过实验发现，我们提出的结构感知几何谱排序优于所有这些方法，验证了其有效性。

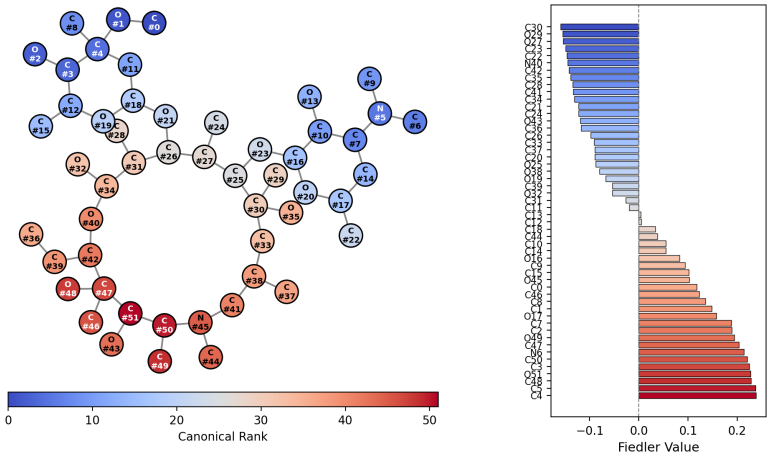
算法2 谱 $SO(3)$ 规范化

要求： 原子坐标 $\mathbf{X} \in \mathbb{R}^{N\times 3}$ 和带符号Fiedler向量 v（由算法 1计算得出）**确保：** 规范化坐标$\mathbf{X}'$

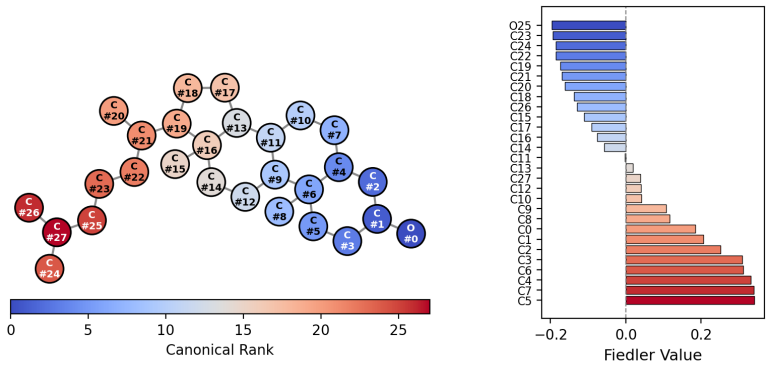
1：应用谱排序 $\pi^* =$ argsort(v) 并强制符号 $\sum_i v_i^3 > 0$2：令头索引 h 的排序为 0，尾索引 t 的排序为 $N-1$3：选择锚点索引

$$a = \operatorname{argmax}_{k\in[\lfloor N/3\rfloor,\lfloor 2N/3\rfloor)} \|(\mathbf{x}_k - \mathbf{x}_h) \times (\mathbf{x}_t - \mathbf{x}_h)\|$$

4：计算纵轴 e$_1 \leftarrow$ x$\|$ $_{t-}^{t-}$ ^{h}x x x$_h\|$5：计算平面法向 n $\leftarrow$ e$_1$ $\times$(x$_a$ $-$x$_h$) 6：归一化法向量 e$_3 \leftarrow$ n $\|$n$\|$7：构造右手基 $\overline{\mathrm{e}_2} \leftarrow$ e$_3$ $\times$e$_1$8：组装旋转矩阵 R $\leftarrow$ [e$_1$,e$_2$,e$_3$]$^\top$9：对 X$'$ $\leftarrow$ (X $-$ $\bar{}$X)R$^\top$进行中心化和旋转



(a) 阿奇霉素（高对称性／大环）：谱排序以核心为中心，将大环集中在谱中心附近，并将外围糖基团映射至序列末端。



(b) 胆固醇（低对称性／各向异性）：谱排序沿主轴展开，从极性头部经过刚性四环核心，延伸至柔性烃尾，同时保留局部连通性。

图5. 谱规范化通过（带符号的）Fiedler向量进行规范化，可为两者生成稳定且保持局部性的原子排序 y
对称和弱对称分子（蓝色：起点，红色：终点），展示了在多种对称性条件下的稳健表现 s.

C.2. 规范扩散与 CanonFlow

我们介绍 规范扩散与流匹配的其他算法和架构设计。

C.2.1. 规范扩散与流匹配的详细信息

我们已在第 4.2 节介绍了规范扩散或流匹配的整体框架。现在，我们将在本部分提供更多细节。

规范位置编码。 为了明确打破置换等变性，我们将由规范秩得到的位置信息注入模型架构。对于每个秩为 i、归一化秩为 $r_i = \text{rank}_i/N \in [0,1$ 的原子，我们计算正弦位置编码：

$$\text{PE}(r, 2k) = \sin\left(\frac{r \cdot M}{10000^{2k/d_{\text{pe}}}}\right), \qquad \text{PE}(r, 2k+1) = \cos\left(\frac{r \cdot M}{10000^{2k/d_{\text{pe}}}}\right) \tag{59}$$

其中 $M = 10000$ 是最大位置尺度，d_{pe} 是编码维度。该编码在初始投影之前与原子特征拼接。这在架构层面完全打破了置换等变性，使模型能够依据原子的结构角色区分原子。请注意，此位置编码既可应用于我们新颖的 Canon 架构，也可应用于其他现有架构，以构建一个（非等变）规范性兼容版本。位置编码丢弃 机制会替换

$\mathbf{R}$ 以概率p_{drop} 在训练期间被替换为学习得到的“伪位置编码”嵌入，该嵌入用于针对规范条件的无分类器引导。

自适应位置相关先验。 标准流匹配从均匀先验 p_0 中采样噪声。我们将其替换为一种位置相关先验，用以编码规范位置与原子类型之间的经验相关性：

$$p_0^{\text{eff}}(c \mid r) = \beta_r \cdot p_{\text{prior}}(c) + (1 - \beta_r) \cdot p_r(c \mid r) \tag{60}$$

其中 $r = r_i/N$ 是相对位置，β_r 是混合系数（对分类数据通常为 0.1），p_{prior} 是先验分布，例如 p_{uniform}，而 $p_r(\cdot|r)$ 是从训练集估计得到的经验分布。

对于 先验估计，我们将 $r \in [0, 1)$ 离散化为 K 个箱并计算：

$$P(c \mid B_k) = \frac{\text{Count}(c, k) + \epsilon}{\sum_{c'}(\text{Count}(c', k) + \epsilon)} \tag{61}$$

对于分类数据；对连续坐标，我们使用以数据集中的均值和方差为充分统计量的高斯分布。在运行时，我们对连续的 r 使用相邻箱之间的线性插值：

$$p_r(c|r) = (1 - \delta) \cdot P(c|B_{k_0}) + \delta \cdot P(c|B_{k_1}) \tag{62}$$

其中 $k_0 = \lfloor r \cdot K \rfloor$，$k_1 = \min(k_0 + 1, K - 1)$，以及 $\delta = r \cdot K - k_0$。

C.2.2. CANON 架构

由（Irwin 等人，2024）提出的 Semla 架构在每一层维护与 $\mathbf{X}$、$\mathbf{H}$ 相对应的两个隐藏状态。在我们的 *Canon* 架构中，我们额外引入第三个隐藏状态，通过与其他原子特征和键特征交互，在每一层更新并细化规范信息。图 6 总结了 Canon 架构。

输入处理。 我们实现了一个三流分子动力学网络 **Canon**，它用显式的规范秩流以及可选的规范位置编码（PE）增强类似 Semla 的等变消息传递主干网络。具体而言，在扩散/流时间 t，该模型的输入为

$$\mathbf{X}_t \in \mathbb{R}^{B\times N\times 3}, \qquad \mathbf{H}_t \in \mathbb{R}^{B\times N\times d_h}, \qquad \mathbf{R} \in \mathbb{R}^{B\times N},$$

其中 $\mathbf{X}_t$ 为坐标，$\mathbf{H}_t$ 为节点级不变特征（例如原子特征、可选条件化），$\mathbf{R}$ 为规范秩（或其归一化变体）。位置编码的处理与上文的一般描述 ˆ ˆ 相同。可选的自条件化提供先前的估计 $\mathbf{X}_0$, ，以及可选的先前秩估计 $\mathbf{R}$。在实现中，自条件化通过堆叠（$\mathbf{X}_t$，$\hat{\mathbf{X}}_0$），并经由线性映射Π_{CS}投影到 K 个坐标集来实现：

$$\mathbf{CS}_t \;=\; \Pi_{CS}\big([\mathbf{X}_t, \hat{\mathbf{X}}_0]\big) \in \mathbb{R}^{B\times K\times N\times 3}, \tag{63}$$

并对填充原子应用掩码。

我们将大小嵌入 Emb(|V|) 和可选的规范 PE 附加到不变特征上，并将它们投影到主干宽度 d：

$$\widetilde{\mathbf{H}}_t \;=\; \phi_{\text{in}}\Big([\mathbf{H}_t, t, \text{Emb}(|V|), \text{PE}(\mathbf{R})]\Big) \in \mathbb{R}^{B\times N\times d}, \tag{64}$$

其中，ϕ_{in} 是一个具有 SiLU 非线性的两层多层感知机。

CanonDynamics：具有节点/坐标/秩流的消息传递层。主干网络（CanonDynamics）堆叠 L 个消息传递层，并可选择使用边特征。每一层都会更新三个相互耦合的状态：

$$(\mathbf{CS}, \mathbf{H}, \mathbf{R}) \mapsto (\mathbf{CS}', \mathbf{H}', \mathbf{R}'),$$

在每个边消息层中，我们使用*(i)* 投影节点特征、*(ii)* 投影秩特征以及*(iii)*由坐标集点积导出的几何特征来构造成对消息。具体而言，在归一化之后我们计算

$$\mathbf{G}_{ij}^{(k)} \;=\; \langle \mathbf{CS}_i^{(k)},\, \mathbf{CS}_j^{(k)} \rangle, \qquad k = 1, \ldots, K, \tag{65}$$

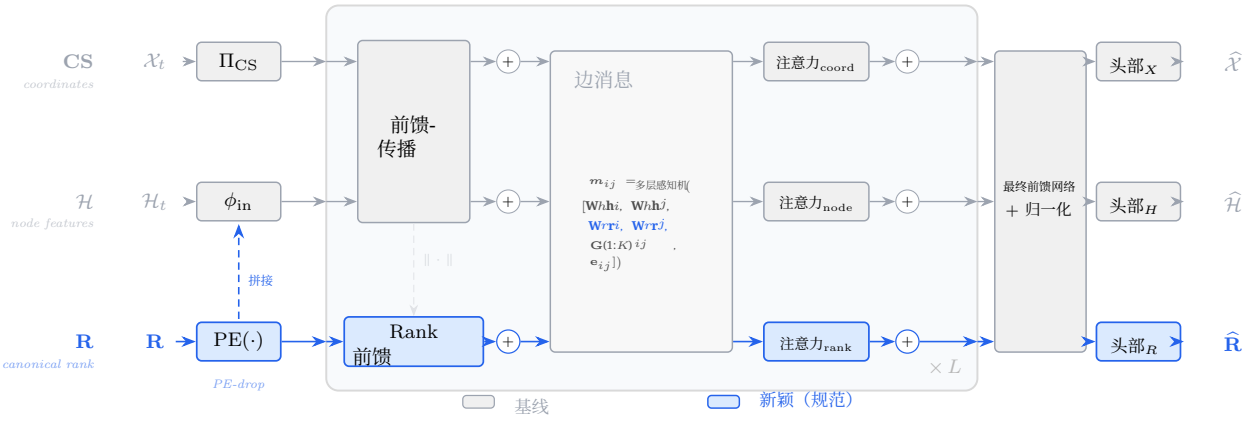


图 6. Canon 架构概览。三个并行流——坐标（**坐标流**）、节点特征（$\mathcal{H}$）和规范秩（$\mathbf{R}$）——通过 L 个消息传递层进行更新，这些层由前馈块、成对边消息和基于注意力的聚合组成。灰色模块表示继承自 Semla 的组件；蓝色模块表示 Canon 引入的新颖规范秩组件。

并拼接

$$\boldsymbol{m}_{ij} \;=\; \mathrm{MLP}\Big(\,[\mathbf{W}_h\boldsymbol{h}_i,\; \mathbf{W}_h\boldsymbol{h}_j,\; \mathbf{W}_r\boldsymbol{r}_i,\; \mathbf{W}_r\boldsymbol{r}_j,\; \mathbf{G}_{ij}^{(1:K)},\; \boldsymbol{e}_{ij}]\,\Big), \tag{66}$$

其中， $\boldsymbol{e}_{ij}$ 是可选的边特征输入，而 $\mathbf{W}_h$,$\mathbf{W}_r$ 是可学习的投影权重。

接下来，我们采用基于注意力的更新来处理节点/坐标/秩特征。消息张量被拆分为三个通道（另加可选的边输出）：

$$\boldsymbol{m}_{ij} \;=\; \big(\boldsymbol{m}_{ij}^{\text{node}},\; \boldsymbol{m}_{ij}^{\text{coord}},\; \boldsymbol{m}_{ij}^{\text{rank}},\; \boldsymbol{m}_{ij}^{\text{edge}}\big), \tag{67}$$

这些通道用于驱动三个注意力模块：

$$\mathbf{H} \leftarrow \mathbf{H} + \mathrm{Attn}_{\text{node}}(\mathbf{H}, \boldsymbol{m}^{\text{node}}), \quad \mathbf{CS} \leftarrow \mathbf{CS} + \mathrm{Attn}_{\text{coord}}(\mathbf{CS}, \boldsymbol{m}^{\text{coord}}), \quad \mathbf{R} \leftarrow \mathbf{R} + \mathrm{Attn}_{\text{rank}}(\mathbf{R}, \boldsymbol{m}^{\text{rank}}), \tag{68}$$

交错穿插着前馈残差块，这些残差块由用于（$\mathbf{H}$，**坐标流**）和$\mathbf{R}$的等变或不变多层感知机构成。

输出头 **s.** 经过 L 层后，Canon 通过归一化**linear head**生成坐标更新 + :

$$\widehat{\mathbf{X}} \;=\; \mathrm{Head}_X\big(\mathrm{Norm}(\mathbf{CS})\big) \in \mathbb{R}^{B\times N\times 3}, \tag{69}$$

并通过线性分类器基于最终节点特征预测原子类型和电荷逻辑值。如果启用，还会通过线性头和最小-最大归一化为每个节点预测一个秩分数，归一化至 $[0,1]$。可选的边特征由键精化模块进行精化，并投影为边类型逻辑值。

C.3. 训练

鲁棒性用秩噪声。 为缩小训练-推理差距（因为采样期间无法获取真实秩），我们在训练期间向规范秩添加噪声：

$$r_{\text{noised}} = r + \mathcal{N}(0, \sigma_r^2)\cdot(1-t) \tag{70}$$

噪声随 t 衰减，以避免分子接近重建完成时产生较大的扰动。

辅助秩预测。 模型还可以选择将规范秩作为辅助任务进行预测：

$$\hat{r} = \sigma\big(\mathrm{MLP}(h_i^{(L)})\big) \in [0,1] \tag{71}$$

对原子进行最小-最大归一化，并使用训练损失：

$$\mathcal{L}_{\text{rank}} = \lambda_r \cdot \mathbb{E}\left[\|r^* - \hat{r}\|^2\right] \tag{72}$$

这使得推理过程中能够进行动态秩估计。

算法 3 无间隙典范训练，可选用最优传输耦合（推理时无需最优传输）**要求：** 数据集 $(G,\ X)$；对称群 $G \in \{S_N,\ S_N \times SO(3)\}$；规范化器 Ψ_G，生成按排序排列的典范~坐标 X。1: **群设置：** 相应地选择 Ψ_G 和规范空间先验 q_1。~2: 对每个训练样本仅进行一次典范化： $Z_0 = \mathrm{vec}(\Psi(X))$；将秩视为固定的坐标索引。3: 在典范空间中选择边缘先验 q_1 （对齐或各向同性）。4: **对于**每次训练迭代**执行**5: 采样小批量 $\{\tilde{z}_0^i\} \sim q_0$ 和噪声小批量 $\{\tilde{z}_1^j\} \sim q_1$。6: 构造耦合：

- 乘积：配对（$\tilde{z}_0^i,\ \tilde{z}_1^i$）；或者
- 最优传输／Sinkhorn：计算使 Π 最小化$\sum_{i,\ j} \Pi_{ij} \|\tilde{z}_0^i - \tilde{z}_1^j\|^2$ 的传输计划，然后从 Π 中采样配对。

7: 使用固定／预测的秩嵌入，在规范坐标上训练流／扩散模型 位置编码(i)。8: **结束循环**

最优传输退火。 我们发现，噪声与目标分子之间的最优传输（OT）匹配会降低流方差，但可能限制泛化能力。我们实现了**OT退火**：训练期间使用 OT 的概率线性递减：

$$p_{\mathrm{OT}}(\mathrm{epoch}) = \max\left(0, 1 - \frac{\mathrm{epoch}}{\max \mathrm{epochs}}\right) \tag{73}$$

这使模型能够在训练早期受益于最优传输对齐，同时在后期学会处理任意的噪声—目标配对。

训练目标。 总体训练损失结合了坐标回归、离散特征的类别交叉熵，以及可选的秩预测：

$$\mathcal{L}_{\mathrm{coord}} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathcal{X}}_{1,i} - \mathcal{X}_{1,i}\|^2, \qquad \mathcal{L}_{\mathrm{type}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{CE}(\hat{\mathbf{h}}_{1,i}, \mathbf{h}_{1,i}) \tag{74}$$

$$\mathcal{L}_{\mathrm{bond}} = \frac{1}{N^2}\sum_{i,j} \mathrm{CE}(\hat{\mathbf{E}}_{1,ij}, \mathbf{E}_{1,ij}), \qquad \mathcal{L}_{\mathrm{rank}} = \frac{1}{N}\sum_{i=1}^{N} (\hat{r}_i - r_i^*)^2 \tag{75}$$

$$\mathcal{L} = \mathcal{L}_{\mathrm{coord}} + \lambda_{\mathrm{type}}\mathcal{L}_{\mathrm{type}} + \lambda_{\mathrm{bond}}\mathcal{L}_{\mathrm{bond}} + \lambda_{\mathrm{charge}}\mathcal{L}_{\mathrm{charge}} + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{rank}} \tag{76}$$

其中默认权重为 $\lambda_{\mathrm{type}} = 0.2$，$\lambda_{\mathrm{bond}} = \lambda_{\mathrm{charge}} = 1.0$ 和 $\lambda_{\mathrm{rank}} = 0.1$。根据（Irwin 等人，2024），我们为$\lambda_{\mathrm{bond}} = 0.5$ QM9 数据集设定该值。
完整的训练过程可总结为算法 3。

C.4. 采样

通过 Haar 随机化恢复不变性。 关键观察是，规范样本本身虽然并非从不变分布中抽取，但可以通过一个简单的后处理步骤提升为不变分布。由定理 3.1可直接得到以下定理。该定理将（Yan 等人，2023年）中的观察推广到任意紧群，并为不变分布的采样提供理论保证：

命题 C.1（事后随机化可实现不变采样）。 设 $\tilde{\mu}$ 是 $\mathcal{M}$ 上的任意分布。定义随机化分布

$$\mu := \int_{\mathcal{G}} (g\cdot)\#\tilde{\mu}\, \mathrm{d}\lambda(g). \tag{77}$$

~那么 μ 具有 $\mathcal{G}$不变性。此外，对 $g \sim \lambda$ 和 $\mathbf{Z} \sim \tilde{\mu}$ 进行采样，并输出 $g \cdot \mathbf{Z}$，即可得到 μ.

算法 4 统一少步规范采样（机制 A/B + 选项）要求：对称群 $G \in \{S_N,\ S_N \times SO(3)\}$；训练好的模型 v_θ；基线先验 q_1；步时间 $1 = t_K > \cdots > t_0 = 0$。要求：采样机制标志：机制 $\in$ {A、B}。要求：选项：(i) 对齐先验；(ii) PCS（若启用，则每一步后应用 Ψ；要求机制 $=$ B）。1：注意（秩条件化）：如果将排序用作条件/位置编码，则将其视为固定坐标索引，不要根据噪声重新计算。2：（可选的先验选择）：可选择用替代先验替换 q_1。~3：在固定规范坐标系中采样 $Z_{t_K} \sim q_1$。4：对于 $k = K, \ldots, 1$，执行~ ~5：$Z_{t_{k-1}} \leftarrow$ Step($Z_{t_k}, v_\theta, t_k \to t_{k-1}$) {常微分方程/随机微分方程求解器}6：如果机制 $=$ B，则~ ~7：$Z_{t_{k-1}} \leftarrow \Psi($ $Z_{t_{k-1}}$) 8：否则 9：注意：在机制 A 的采样过程中，不要重新规范化／重新排序／投影；保持排序索引固定。10：结束 if 11：结束循环~12：输出 Z_{t_0}。~13：可选的不变性恢复：采样 $g \sim \lambda(G)$ 并输出 $g \cdot Z_{t_0}$。

证明。对于任意 $h \in \mathcal{G}$,

$$(h\cdot)\#\mu = \int_{\mathcal{G}} (hg\cdot)\#\tilde{\mu}\,\mathrm{d}\lambda(g) = \int_{\mathcal{G}} (g'\cdot)\#\tilde{\mu}\,\mathrm{d}\lambda(g') = \mu, \tag{78}$$

其中我们进行了变量替换 $g' = hg$ 并利用了 Haar 的左不变性。 □

备注 C.2（规范化与随机化的比较）。 仅靠随机化可以强制实现不变性，但无法简化学习问题。规范化强制进行一种规范选择，能够大幅减少由对称性引起的多模态性/方差，从而加快训练收敛；随后，随机化则在最后恢复不变性。因此，将规范化训练与事后随机化相结合，能够兼得两者之长：高效学习和有理论保证的不变生成。

动态秩估计。 由于采样可以使用固定规范秩作为条件来进行，因此秩估计的动态更新完全是可选的。

一种实用的少步稳定器是将中间状态投影回切片（算法4中的机制 B）。特别是，为了保持秩条件与部分含噪样本的一致性，我们允许采用投影到切片的采样方式，并将其称为投影规范采样（PCS），详见算法 5。请注意，在保持位置编码秩不变的同时将数据投影到典范切片，在概念上等价于保持数据的原始顺序不变，但更新规范秩以使二者对齐。

在没有真实值的情况下，我们支持两种策略来估计算法 5 中的规范秩：

1. 预测模式：直接使用模型输出的秩预测 $\hat{}r$ 进行扭曲。
2. 规范化模式：根据中间预测定期重新计算规范秩：将当前的原子类型预测离散化：$\hat{a}_i = \mathrm{argmax}(\mathrm{atomics}_i)$。然后，对预测分子应用相同的规范化算法。每隔 K 步更新一次秩，当 $t \geq T$时。

我们通过实验发现，预测模式始终表现更好，这可能是因为规范化算法在中间噪声状态下不稳定；相比之下，自预测将这一过程融入模型学习，从而提升鲁棒性和表达能力。

值得注意的是，如果禁用动态秩估计（算法 4 中的机制 A），模型始终会根据规范秩条件执行“条件生成”任务，这也是一种有效且不存在训练-测试差距的任务。

完整的采样过程可总结为算法 4。

算法 5 投影规范采样（PCS）

要求： 规范化器 Ψ_ϕ；训练得到的切片模型（得分或向量场） m_θ；时间网格 $1 = t_K > \cdots > t_0 = 0$。
˜1：初始化$\mathbf{Z}_{t_K} \sim q_1$。 2：**对于** $k = K$，.....，1 **执行**⌢˜ ˜3：在切片上执行一次反向步：$\mathbf{Z}$ $\leftarrow$ Step($\mathbf{Z}_{t_k}$， m_θ， $t_k \to t_{k-1}$)。⌢˜ ˜4：投影：$\mathbf{Z}_{t_{k-1}} \leftarrow \Psi_\phi(\mathbf{Z})$。 5：**循环结束**ˆ ˜6：输出规范样本$\mathbf{Z}_0 = \mathbf{Z}_{t_0}$，并可选择通过 $g \sim \lambda$ 进行随机化。

D. 实验细节和其他结果

D.1. 实验细节

D.1.1. 评估指标

我们使用以下指标评估生成的分子。设 $\mathcal{M} = \{m_i\}_{i=1}^N$ 表示生成分子的集合，设 $\tilde{m}_i$ 表示 m_i经 RDKit 力场优化后的构象。

- **有效性**。 通过 RDKit 化学有效性检查的生成分子所占比例。
- **原子稳定性**。 显式价处于预定义允许范围内的生成原子所占比例。
- **分子稳定性**。 其中每个 原子都稳定的生成分子所占比例。
- **唯一性**。 在可转换为规范 SMILES 的分子中，不同 SMILES 字符串所占比例。
- **新颖性**。 在具有有效规范 SMILES 的分子中，不存在于训练集中的分子所占比例。
- **Opt-RMSD**。 生成构象与优化后构象之间的均方根偏差均值：

$$\text{Opt-RMSD} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \text{RMSD}(m_i,\, \tilde{m}_i),$$

其中， $\mathcal{I} = \{i$： 优化成功且均方根偏差为有限值$\}$。这量化了生成的三维几何与局部能量极小值之间的偏差。

- **函数评估次数**。 即常微分方程求解器在采样过程中所需的神经网络前向传播总次数。函数评估次数是一种与硬件无关的生成成本衡量指标。
- **采样时间**。 生成完整评估集（N=1，000 个分子）所需的实际运行时间).

D.1.2. 超参数。

我们的绝大多数超参数都完全遵循 Irwin 等人(2024) 对 Canonical SemlaFlow 和 CanonFlow 的设定。所有骨干网络均由 L=12 个等变消息传递层（d_{model}=384、 d_{msg}=128、 d_{edge}=128、 H=32 个注意力头、 S=64 个坐标集，并对每个集合进行长度归一化）组成，输入端和输出端各有两个边感知层。维度为 64 的可学习分子大小嵌入会与每个原子的输入特征拼接。所有输出头（原子类型、电荷、键类型以及可选的规范排序头）均为采用 SiLU 激活函数的两层多层感知机。默认采用自条件化：训练期间，模型以0.5 的概率先在条件化输入置零的情况下生成初步预测，再将其作为额外上下文反馈；推理期间，则在每个积分步都应用自条件化。

模型在 GEOM-Drugs 上训练 200 轮，在 QM9 上训练 300 轮，使用 Adam（lr=3×10^{-4}、AMSGrad、不使用权重衰减），采用线性预热（GEOM-Drugs 为 10 000 步，QM9 为 2 000 步），之后保持学习率恒定；全局范数为 1.0 时进行梯度裁剪，并采用成本预算为 4096、线性成本缩放的基于分桶的动态批处理。对模型参数维护指数移动平均（EMA，衰减率为 0.999），并将其用于所有推理。所有训练均以 FP32 精度进行。在单张 NVIDIA A100 80GB GPU 上，QM9 的总训练时间约为 10 小时，GEOM-DRUG 的总训练时间约为 28 小时。

表 4. 在 GEOM-DRUG 上使用我们新颖的 Canon 架构的 CanonFlow。规范化在 S_N 上进行。

模型	分子稳定性 ↑	有效性 ↑	NFE
EQGAT-diff	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	500
SemlaFlow_{50}，使用最优传输	$97.0_{\pm 0.21}$	$93.9_{\pm 0.12}$	50
CanonFlow_{50}	$\mathbf{98.0}_{\pm 0.08}$	$\mathbf{95.4}_{\pm 0.09}$	50
SemlaFlow_{100} 含最优传输	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	100
CanonFlow_{100}	$\mathbf{98.4}_{\pm 0.02}$	$\mathbf{95.9}_{\pm 0.08}$	100
Data	100.0	100.0	–

坐标采用线性插值，$\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\,\mathbf{x}_1$，并添加高斯噪声（$\sigma$=0.2）。原子类型和键类型使用均匀采样类别插值（Campbell 等人，2022年）。插值时间从$\text{Beta}(2,\ 1)$。坐标对可选择采用等变最优传输（Kabsch 对齐 + 匈牙利匹配）。启用无分类器引导（CFG）时，位置编码的 dropout 率默认为 p_{drop}=0.1。

D.2. 补充结果

除了目前为止　　在第 5 节介绍的主要结果外，我们还在此提供了更全面的消融实验　　e.

使用 CanonFlow 实现出色的少步生成。 表 4展示了我们的 CanonFlow（采用 OT退火训练）的少样本生成结果。它始终优于未采用规范化的 SemlaFlow 基线，即使仅使用 50 步，优势依然显著。这进一步支持了我们的观点：规范化不仅能提高生成模型的性能上限，还能通过条件提供强有力的引导信号，从而促进少步生成。

表 5. Canonical 语义 laFlow 在 GEOM-DRUG 数据集上使用无分类器引导（CFG）。在 S 上进行规范化　　$_N$.

模型	CFG 缩放系数	分子稳定性 ↑	有效性 ↑	NFE
EQGAT-diff	–	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	500
SemlaFlow_{50}	–	$97.0_{\pm 0.21}$	$93.9_{\pm 0.12}$	50
规范化SemlaFlow_{50}	1.0	$\mathbf{97.6}_{\pm 0.13}$	$94.6_{\pm 0.32}$	50
规范化SemlaFlow_{50}	1.5	$97.4_{\pm 0.07}$	$\mathbf{94.7}_{\pm 0.12}$	50
规范化SemlaFlow_{50}	2.0	$96.7_{\pm 0.04}$	$94.5_{\pm 0.03}$	50
SemlaFlow_{100}	–	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	100
规范化SemlaFlow_{100}	1.0	$\mathbf{98.1}_{\pm 0.03}$	$95.0_{\pm 0.20}$	100
规范化SemlaFlow_{100}	1.5	$97.8_{\pm 0.03}$	$95.0_{\pm 0.09}$	100
规范化SemlaFlow_{100}	2.0	$97.5_{\pm 0.08}$	$\mathbf{95.1}_{\pm 0.07}$	100
Data	–	100.0	100.0	–

表6. 规范化 SemlaFlo　　在 GEOM-DRUG 数据集上使用无分类器引导（CFG）的 w。规范化在 $S_N \times SO$(3 上进行　　).

模型	CFG 缩放系数	分子稳定性 ↑	有效性 ↑	NFE
EQGAT-diff	–	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	500
SemlaFlow_{100}	–	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	100
规范化SemlaFlow_{100}	1.0	$\mathbf{97.9}_{\pm 0.09}$	$93.8_{\pm 0.30}$	100
规范化SemlaFlow_{100}	1.5	$97.7_{\pm 0.06}$	$94.2_{\pm 0.23}$	100
规范化SemlaFlow_{100}	2.0	$97.6_{\pm 0.09}$	$\mathbf{94.4}_{\pm 0.05}$	100
Data	–	100.0	100.0	–

无分类器引导的效果。 我们进一步通过 CFG 消融实验评估采样质量。为了全面评估规范化对称群的影响，我们报告了规范化 SemlaFlow 在以下两种设置下的表现： S_N （采用 OT退火训练），见表 5；以及 $S_N \times SO(3)$ （未采用 OT 训练），见表 6。在适当范围内，较大的 CFG 缩放系数往往能提高有效性指标，但会略微降低分子稳定性。有趣的是，在 $S_N \times SO(3)$ 规范化设置下，提升更为显著，因为预测更加依赖规范信息，而将条件捷径和无条件基线外推有助于利用这些信息。

对齐先验和最优传输的消融研究。在表7 中，我们全面研究了对齐先验和最优传输分别应用于等变与非等变骨干网络时的效果。我们纳入了 S_N 对称群，并评估了不同的采样步数。值得注意的是，如果没有对齐先验，采用各向同性先验的等变 SemlaFlow（不使用 PE）将与基线完全相同；换言之，对齐先验也是我们典范生成模型的重要组成部分，即使等变模型也能从更优的先验-数据耦合中获益，而无需借助 PE。

表 7. 在 GEOM-DRUG 上使用对齐先验的 Canonical SemlaFlow。在 S_N 上进行规范化，并研究使用或不使用 OT 的等变与非等变骨干网络。请注意，先验是根据不同的群 G 一致计算的

群	模型	PE	OT	分子稳定性 ↑	有效性 ↑	NFE
	EQGAT-diff	-	-	$93.4_{\pm 0.21}$	$94.6_{\pm 0.24}$	500
	SemlaFlow_{20}	-	等变	$95.3_{\pm 0.14}$	$93.0_{\pm 0.10}$	20
	SemlaFlow_{50}	-	等变	$97.0_{\pm 0.21}$	$93.9_{\pm 0.12}$	50
	SemlaFlow_{100}	-	等变	$97.3_{\pm 0.08}$	$93.9_{\pm 0.19}$	100
S_N	规范化SemlaFlow_{20}	-	等变	$95.1_{\pm 0.09}$	$93.5_{\pm 0.17}$	20
S_N	规范化SemlaFlow_{50}	-	等变	$97.3_{\pm 0.08}$	$94.4_{\pm 0.02}$	50
S_N	规范化SemlaFlow_{100}	-	等变	$97.5_{\pm 0.10}$	$94.2_{\pm 0.11}$	100
$S_N + PE$	规范化SemlaFlow_{20}	True	等变	$92.6_{\pm 0.18}$	$91.7_{\pm 0.20}$	20
$S_N + PE$	规范化SemlaFlow_{50}	True	等变	$96.3_{\pm 0.03}$	$93.9_{\pm 0.07}$	50
$S_N + PE$	规范化SemlaFlow_{100}	True	等变	$97.5_{\pm 0.10}$	$94.2_{\pm 0.20}$	100
	Data	100.0	100.0	–		

E. 规范扩散示例

本节展示了一些 在 GEOM-Drugs 上训练的典范化 SemlaFlow 生成的样本。这些样本由 d
随机生成的样本，但我们在必要时对它们进行了旋转 以便于可视化。

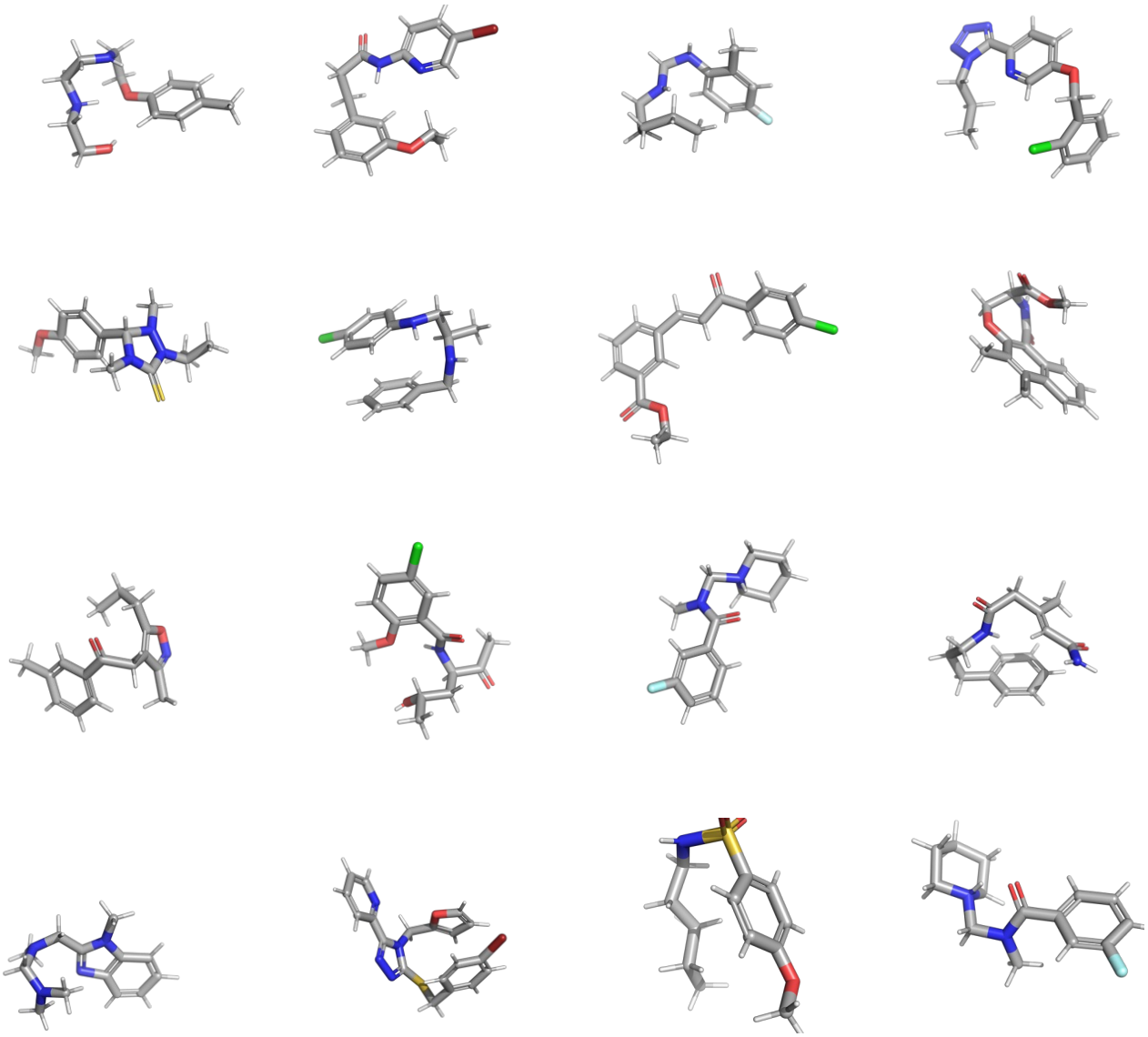


图 7. 在 GEOM Drugs 上训练的典范化SemlaFlow模型生成的随机样本。这些样本通过 100 个常微微分方程积分步生成。