arXiv:2602.03118v1 [math.NA] 3 Feb 2026

学习等变模型的数据增强高成本

Henri Klinteb¨ack[1],Christoph Ortner[1],和 Lior Silberman[1]

[1]1984数学路，温哥华，不列颠哥伦比亚大学数学系， klintebaeck@math.ubc.ca

2026年2月4日

摘要

根据诺特定理，哈密顿系统中存在连续对称性等价于存在守恒量；然而，数据驱动模型并不总是显式地强制施加这些对称性。关于将对称性编码进模型架构是否为最优方法，至今仍有争议。另一种相 competing 的方法是通过数据增强来实现近似对称性。本文研究了两种旨在改善此类近似方案对称性的方法：一种基于紧李群上的哈尔测度的求积公式，该群编码了所关注的连续对称性；另一种则基于对该哈尔测度进行随机采样。我们通过理论分析和实证结果表明，求积增强能使多项式模型精确保留对称性，而随机增强的对称化误差仅以平方根速度收敛。

1 引言

将机器学习（ML）应用于开发复杂物理系统的代理模型，是一个活跃的研究领域。这源于物理系统的内在特性：它们往往极其复杂，传统代理模型难以充分描述。围绕机器学习代理的丰富研究领域包括用于分子动力学的量子精确力场 [1], 、粗粒化分子动力学 [2], ，以及更一般地从数据中学习动力系统 [3]。

本研究的重点是对称性在机器学习模型架构中的作用。这些对称性在物理、化学和生物世界中无处不在，无论是在原子势、分子结构还是蛋白质相互作用中。处理机器学习模型中对称性的两种相互竞争的方法是：(i) 使用一个通用模型

该模型不感知对称性，但在学习过程中对训练数据进行数据增强 [4, 5]；(ii) 在模型中精确强制已知对称性 [6, 7]。每种可能性都可以通过多种现有技术实现。这两种方法中哪一种更可取，以及如何最优地实现它们，仍是一个持续的争论（参见前述参考文献）。一般而言，在模型设计中强制结构与从数据中学习结构之间取得平衡，仍然是科学机器学习中的一项挑战。我们工作的目的是严格分析两种用于学习已知模型对称性的自然数据增强方法，并将它们相互对比，同时与具有内置对称性的类似模型进行对比。

支持将对称性硬编码到模型架构中的一个常见论点是，这样可以减少模型参数的数量。这通常是正确的，但在深度学习领域，参数数量很少足以衡量模型的性能或鲁棒性。训练动态的性质通常更为重要，而且越来越多的证据表明，无对称性约束模型可能确实更容易训练，并且在分布外数据上具有同等的鲁棒性和准确性 [8, 9]。这已经推动了高效无约束模型的发展，使其能够处理几何优化等静态任务 [10, 11]

我们关注保持对称性，重点在于保留模型所描述的重要定性物理性质。根据诺特定理，对称性与守恒量有着内在联系；存在连续对称性，当且仅当运动存在第一积分。例如，平移不变性会带来线性动量守恒，而旋转不变性会带来角动量守恒。在许多应用中，我们希望数值格式能够保持这些守恒量。比如，这促成了几何积分的发展，尤其是辛积分 [12]。辛积分器是一种积分器，其时间步进操作是辛同胚，因此模拟系统的任意第一积分都将得到保持 [13]，前提是哈密顿量可分。

在第 2 节中，我们将说明，在代理模型中仅具有近似对称性，会普遍对第一积分的守恒产生重大影响。我们尤其关注 SO(n) 下的不变性；在二维和三维中，与之相关的守恒量是角动量。我们将回顾相关的既有理论，该理论说明——除非系统是完全可积系统——在这些守恒量的时间演化过程中会观察到指数漂移。这促使我们需要让代理模型精确地或以极高精度保持守恒量。

在第3节中，我们将介绍并研究两种不同的数据增强方法，以改善不受对称性约束的近似方案的对称性。具体而言，一种方法是通过求积公式对紧李群上的哈尔测度进行数据增强，该紧李群的表示对应于我们关注的连续对称性；另一种方法则是对该哈尔测度进行随机采样。我们将严格证明这两种方案的收敛性（及其收敛速率）。最后，在第4节中，我们将通过更多的实证分析来阐释我们的理论

并进一步详细说明数据增强方案和近似方案的性能，具体采用一个最小二乘玩具模型——我们将考察 S^1 和 S^2上的三个粒子这一最简单的非平凡情形。在第3节中，我们将介绍并研究两种不同的数据增强方法，以改善不受对称性约束的近似方案的对称性。具体而言，一种方法是通过求积公式对紧李群上的哈尔测度进行数据增强，该紧李群的表示对应于我们关注的连续对称性；另一种方法则是对该哈尔测度进行随机采样。我们将严格证明这两种方案的收敛性（及其收敛速率）。

最后，在第4节中，我们提供数值测试结果，以验证我们的严谨理论，并进一步展示数据增强方案的性能细节。

2 角动量的近似守恒

本节回顾有关辛积分器的近似守恒性质，以及当系统仅近似对称时这些性质如何失效的研究结果。 $H = T + V$表示可分哈密顿量，$(p,\ q)$表示相空间中的正则坐标。本节对关键思想进行形式化讨论；技术细节请参见 [12] 。

如果哈密顿系统的数值积分器与流形结构相容，则称其为辛积分器；参见 [12, 14]。辛积分器的典型例子是我们在数值应用中选用的二阶速度 Verlet 积分器 [15]。虽然通过该积分方案得到的动力学并不精确，但它对应于一个接近原哈密顿量的哈密顿量，称为影子哈密顿量 H_{eff}。对于任意阶数为 p、步长为 h的辛积分器

$$H_{\mathrm{eff}} = H + O(h^p). \tag{1}$$

H_{eff} 的一种构造可见于 [13]。这意味着，任何同时与 T 和 V 对易的量，也与 H_{eff} 对易，因为其所有高阶项都是 T 和 V 的对易子。实际上，这一结果对第一积分的守恒具有重要意义，因为 J 是第一积分，当且仅当 $\{J, H\} = 0$。因此， H 和 H_{eff} 具有相同的守恒量。

我们关注旋转不变势 V 的情形，此时角动量在哈密顿动力学下守恒。更一般地，我们可以考虑连续对称性。

现在假设，在近似过程中（见后文），精确不变性遭到破坏，但所得哈密顿量仍近似不变，即 $H = H_{\mathrm{I}} + \varepsilon H_{\mathrm{P}}$，其中 H_{I} 具有连续对称性， $\varepsilon > 0$ 而 H_{P} 是一个哈密顿量。在这种情况下，无法保证近似的守恒性质。两种情形最能说明预期行为：轨迹要么保持受控，要么系统中存在一定程度的混沌。一般情形属于后一种；即使在解析性假设下，我们也观察到误差呈指数增长。这是意料之中的，因为众所周知，如果一个系统并非完全可积，那么它很可能存在混沌 [16–20]。

尽管辛积分器不会改变这种行为，但它们确实能确保守恒量得到正确保持。为观察这一点，我们将 $X_{\mathrm{V}} = X_{\mathrm{i}} + \epsilon X_{\mathrm{p}}$代入 X_{eff}的表达式，其中 X_{i} 和 X_{p} 分别是 H_{i}和 H_{p}的刘维尔算子 [13, 14] 。由此可得

$$X_{\mathrm{eff}} = X_{\mathrm{T}} + X_{\mathrm{i}} + \epsilon X_{\mathrm{p}} + \frac{h}{2}[X_{\mathrm{T}}, X_{\mathrm{i}}] + \frac{h\epsilon}{2}[X_{\mathrm{T}}, X_{\mathrm{p}}] + \cdots \tag{2}$$

$$J \circ \Phi_t = e^{tX_{\mathrm{eff}}} J \tag{3}$$

$$= e^{t\epsilon X_{\mathrm{eff,p}}} e^{t(X_{\mathrm{T}}+X_{\mathrm{i}})} J \tag{4}$$

$$= e^{t\epsilon X_{\mathrm{eff,p}}} J, \tag{5}$$

其中 Φ_t 是哈密顿流，也就是满足 $\Phi_t(p(0)\ q(0)) = (p(t)\ q(t))$ 的映射。我们指出，$e^{t\epsilon X_{\mathrm{eff,\ p}}} J = J$ 当且仅当 X_{p} 与 J对易。

指数偏离是普遍情况，而可积系统中则会呈现更为刚性的行为；这类系统具有 n（构型空间的维数）个函数独立的运动常数。KAM理论 [21] 表明，在完全可积系统中，足够无理的非退化环面在扰动下仍能存续，这意味着守恒量在受扰动的系统中将近似守恒。第十一章[12]对此作了说明。

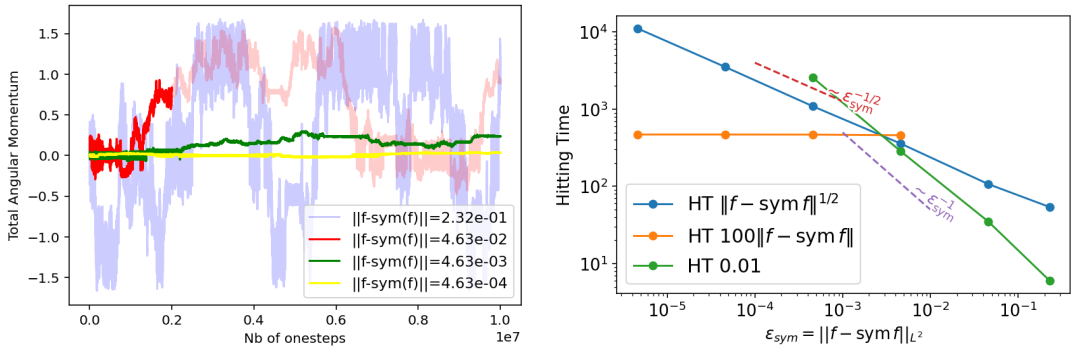
我们通过以下模拟来展示这一行为。我们考虑圆 S^1 上的 3 个粒子，它们处于势函数 $U = F + \epsilon G$作用下，其中 F 是旋转不变的解析函数，而G 则选为解析但不具有旋转不变性的函数。根据诺特定理，势函数具有连续对称性等价于存在第一积分；旋转不变性对应于角动量守恒。初始条件设定为初始总角动量为零。我们针对 ϵ的不同取值进行模拟。为了比较近似对称的势函数，我们定义了对称化误差。

定义 1。设 G是一个紧李群， H 是 G上的哈尔测度。设 f 是一个函数 $f\colon\ V \to \mathbb{C}$，其中 G 作用于向量空间 V。设 对称化 是相对于 L^2 内积投影到 G 不变函数空间上的正交投影。于是，我们将对称化误差定义为

$$\epsilon_{\mathrm{sym}}(f) \stackrel{def}{=} ||f - \mathrm{sym}(f)||_{L^2}. \tag{6}$$

当 f 是 G-不变的时，$\epsilon_{\mathrm{sym}}(f) = 0$。范数的选择通常会产生重大影响。然而，在我们的应用中，我们只考虑非常光滑的函数类，使得 L^2 范数和 L^∞ 范数在适当的常数因子下保持可比。

图1展示了对于近似旋转不变势，角动量守恒会迅速失效。因此，若希望保持守恒量，要么应研究完全可积系统，要么采用能够精确保留对称性的近似方案。



(a) 由3个粒子组成的系统在 S^1 （先前所述的势）上的总角动量时间演化。(b) 不同误差目标在 ε下的击中时间。击中时间对应于第一个时间步 t_n，使得 $|J(t_n)| \geq \varepsilon$。模拟共运行1000万个时间步。

图1: 所有计算均使用步长为 $dt = 0.05$的Verlet积分器完成 .

总之，我们指出，除完全可积性这一极为特殊的情况外，所采用的数值格式中的哈密顿量应尽可能精确地保留对称性。事实上，除 KAM理论之外，只有精确对称的哈密顿量才能确保模拟动力学的输出始终保留所需的第一积分。

2.1 几何优化

前述讨论表明，必须将对称性保持在极高精度之内，以防止重要守恒定律的丧失以及由此导致的（可能是灾难性的）分子动力学模拟失败。此类失败是否发生取决于所进行的具体模拟，例如，角动量守恒在气体、液体或固体中的重要性各不相同。

此外，这并不妨碍不保持对称性的模型很好地适用于其他模拟任务。例如，在“几何优化”（能量极小值、鞍点等的计算）中，我们通常寻求 V的平衡点，即求解非线性系统

$$\nabla V(\bar{q}) = 0.$$

现在假设存在一个近似模型 V_ϵ，使得 $|V - V_\epsilon| \leq \epsilon^2$，并且 $\|\nabla V - \nabla V_\epsilon\|_2 = \epsilon$ 在平衡点 $\bar{q}$的某个邻域内。在对该平衡点的稳定性作出适当且自然的附加假设后，反函数定理保证存在一个近似平衡 $\bar{q}_\epsilon$，使得

$$\nabla V_\epsilon(\bar{q}_\epsilon) = 0, \qquad \|\bar{q} - \bar{q}_\epsilon\|_2 \leq C\epsilon \quad \text{and} \quad |V(\bar{q}) - V_\epsilon(\bar{q}_\epsilon)| \leq C\epsilon^2.$$

得出这一结论无需了解 V_ϵ的对称性。因此，精确保留（或以高精度保留）对称性的要求必然高度依赖于具体任务。这与在具有挑战性的基准任务 [11]中观察到的大量实证结果一致。

3 数据增强分析

我们首先总结本文其余部分所使用的记号。 $\mathbb{N}$ 是包含0的自然数。S^d、 $d \in \mathbb{N} \setminus \{0\}$ 是 d-球面， S^N、^{d}是其 N重积。我们用 $L^2(S^N,^d)$表示复值可测平方可积函数的空间。$R = (\mathbf{r}_j)_{j=1}^N \in S^N$、^{d} 是一般 d-球面上的一组点；对于$d = 1$的特殊情形，我们用 $\Theta = (\theta_1, \cdots, \theta_N) \in S^N$,1表示代表单位圆上点的一组角度。符号$\mathbf{l}$、$\mathbf{m}$、$\mathbf{k}$表示多重指标，即整数元组，例如$\mathbf{k} = (k_1, \cdots, k_N) \in \mathbb{Z}^N$。 $Q \in \mathrm{SO}(d+1)$是一个旋转。 Q在 Θ或 R上的作用记为 $Q \cdot \Theta$ 或 $Q \cdot R$，表示逐元素旋转，即

$$Q \cdot R = (Q\mathbf{r}_1, \cdots, Q\mathbf{r}_N). \tag{7}$$

3.1 多项式回归

我们研究逼近（或“学习”）旋转不变势f的不同方法： $S^N, ^d \to \mathbb{C}$。我们将重点讨论使用多项式张量积基进行的逼近。其元素为：

$$\varphi_{\mathbf{k}}(\theta_1, \cdots, \theta_N) = \prod_{j=1}^{N} e^{ik_j\theta_j}, \qquad \text{for } d = 1, \tag{8}$$

$$\varphi_{\mathbf{l},\mathbf{m}}(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \prod_{i=1}^{N} Y_{l_i}^{m_i}(\mathbf{r}_i), \qquad \text{for } d = 2, \tag{9}$$

其中， Y_l^m是复球谐函数：对于 $l \in \mathbb{N}$, $-l \le m \le l$, $(\theta, \varphi) \in S^2$，它们定义为

$$Y_l^m(\theta, \varphi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos(\theta)) e^{im\varphi},$$

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{\frac{m}{2}} \frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l.$$

在实践中，我们从不使用这个定义，因为我们依赖数值稳定的递归求值算法 [22]。这些函数是 $\mathbb{R}^{d+1}$ 的笛卡尔坐标中的多项式。我们将

$$V_K^{N,d} \subset L^2(S^{N,d}) \tag{10}$$

S^N 上（以笛卡尔坐标表示）的多项式函数子空间，^{d} 总次数至多为K的，具体而言，

$$V_K^{N,1} := \mathrm{Span}\left\{\varphi_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^N, \|\mathbf{k}\|_1 \le K\right\}, \quad \text{and} \tag{11}$$

$$V_K^{N,2} := \mathrm{Span}\left\{\varphi_{\mathbf{l},\mathbf{m}} : \mathbf{l} \in \mathbb{N}^N, \mathbf{m} \in \mathbb{Z}^N, |m_i| \le l_i, \|\mathbf{l}\|_1 \le K\right\}. \tag{12}$$

$\varphi_{\mathbf{lm}}$ 的总次数取为 $\|\mathbf{l}\|_1$，以确保 $V^N{}_{,K}{}^2$ 在旋转下保持封闭（[23]；另见 § 3.2）。众所周知，$\bigcup_K V^N{}_{,K}{}^d$ ——所有多项式构成的空间——在L^2 [24]中稠密。

为方便记号，我们对基元素任意指定一个顺序，$\mathbf{k}_j$ 或 $(\mathbf{l}_j,\mathbf{m}_j)$，并分别将 $\varphi_j = \varphi_{\mathbf{k}_j}$ 和 $\varphi_j = \varphi_{\mathbf{l}_j\mathbf{m}_j}$对应起来。于是，任意多项式 $P \in V^N{}_{,K}{}^d$ 都可以写成如下形式

$$P_\beta = \sum_j \beta_j \varphi_j, \tag{13}$$

参数为 β_j。

我们采用最小二乘近似方案（LSQ）来构造逼近元。设 $(R_i, f(R_i))_{i=1}^n$ 为相对于某个分布抽取的独立同分布数据（详见第4节）。据此，我们构造设计矩阵 A, $A_{ij} = \varphi_j(R_i)$，其次数为 K，以及数据向量 Y，其元素为 $Y_i = f(R_i)$。然后通过求解线性最小二乘问题来确定最小二乘参数

$$\beta = \mathrm{argmin}_\beta L(\beta), \qquad \text{where} \qquad L(\beta) = \frac{1}{2}\|A\beta - Y\|^2. \tag{14}$$

众所周知，在数据充足的假设下，最小二乘法会逼近 L^2 在 $V^N{}_{,K}{}^d$ [25] 上的投影。然而，关于所估计多项式的对称性，并未提供任何额外信息，因此，先验地 无法确定其偏离精确不变性的程度是否会导致模拟出现严重的定性失效（参见图1）。

3.2 不变子空间中的回归

可以说，学习不变函数的典型方法是将逼近限制在 $S^{N,d}$ 上总次数至多为 K 的旋转不变多项式子空间中：

$$B_K^{N,d} := \left\{P \in V_K^{N,d} : P \circ Q = P, \forall Q \in \mathrm{SO}(d+1)\right\} \subset V_K^{N,d}, \tag{15}$$

其中，$V^N{}_{,K}{}^d$ （定义见式（10））是 $S^{N,d}$ 上总次数至多为 K 的多项式空间，旋转按元素逐一作用，如式（7）所示。

引理 1. 令 H 为 $G = SO(d+1)$ 上的归一化的哈 *Haar*测度，并令 $f \in L^2(S^{N,d})$。那么，

$$\mathrm{sym}(f) = \int_G f \circ Q H(dQ). \tag{16}$$

此外，对称化误差 $\varepsilon_{\mathrm{sym}}(f)$ 满足

$$\varepsilon_{\mathrm{sym}}(f) \overset{def}{=} \|f - \mathrm{sym}(f)\|_{L^2} \leq \|f - \bar{f}\|_{L^2}$$

¯对于任意 G-不变函数 f。

引理的证明是标准的；为方便读者，我们将其列出 e.

证明。对于一个 L^2函数 f，我们将 $\pi(f)$ 定义为式 (16) 中的群平均。我们将证明，它是到 L^2_G - 旋转不变 L^2 函数空间上的 L^2 正交投影。首先注意， $\pi(f)$ 是旋转不变的，因为与旋转复合只会使积分发生平移：令 $Q' \in G$，则

$$\pi(f) \circ Q' = \int_G f \circ (QQ')H(dQ) = \pi(f). \tag{17}$$

因此，$\mathrm{Im}(\pi) \subset L^2_G$。通过变量替换，不难看出 $\pi(\pi(f)) = \pi(f)$，并且¯ ¯ ¯当 f 为旋转不变时， $\pi(f) = f$，换言之， π 是到不变函数空间上的投影。我们验证它是自伴的，由此即可得出它是正交投影算子。对于任意函数 $g \in L^2$,

$$\begin{aligned}
\langle \pi(f), g \rangle &= \left\langle \int_G (f \circ Q)H(dQ), g \right\rangle && (18)\\
&= \int_G \langle f \circ Q, g \rangle H(dQ) && (19)\\
&= \int_G \left\langle f, g \circ Q^{-1} \right\rangle H(dQ) && (20)\\
&= \int_G \langle f, g \circ Q \rangle H(dQ) && (21)\\
&= \langle f, \pi(g) \rangle . && (22)
\end{aligned}$$

因此， $\pi\colon L^2 \to L^2_G$ 是 L^2-正交投影算子，由此得出 $\pi =$ 对称化。该不等式随即成立，因为 $\mathrm{sym}(f) = \mathrm{argmin}_{\bar{f} \in L^2_G}(\|f - \bar{f}\|)$。这一定义良好，因为我们声称 L^2_G是闭集。确实，令 $\rho\colon G \to U(L^2)$ 为我们的酉表示——也就是说，它是一个函数与旋转的上述复合（参见 (7)）。那么，

$$L^2_G = \{f \in L^2 \colon \rho(Q)f = f, \forall Q \in G\} = \bigcap_{Q \in G} \ker(\rho(Q) - \mathrm{Id}). \tag{23}$$

结论成立，因为任意多个闭集的交仍为闭集。 □

利用引理 1，我们可以刻画旋转不变子空间。对于 $d = 1$,

$$B_K^{N,1} = \mathrm{Span}\left\{ \varphi_{\mathbf{k}} \,:\, \varphi_{\mathbf{k}} \in V_K^{N,1}, \sum_{i=1}^{N} k_i = 0 \right\}. \tag{24}$$

这可以通过对 $V^{N,}{}_{K}{}^{1}$ 的基元素进行对称化来证明。

$$\begin{aligned}\mathrm{sym}(\varphi_{\mathbf{k}}) &= \frac{1}{2\pi}\int_0^{2\pi}\prod_{j=1}^{N} e^{ik_j(\theta_j+\alpha)}d\alpha \\ &= \frac{1}{2\pi}\int_0^{2\pi} e^{i\alpha\sum_j k_j}\prod_{j=1}^{N} e^{ik_j\theta_j}d\alpha = \begin{cases}\varphi_{\mathbf{k}}, & \sum_i k_i = 0,\\ 0, & \text{otherwise.}\end{cases}\end{aligned} \tag{25}$$

旋转不变子空间 $V^{N}{}_{,K}{}^{2}$ 的基有多种自然选择。可能的方法包括直接应用（16），或使用克莱布施–戈尔丹系数递归分解不可约表示对的张量积。无论采用哪种方法，最终都会得到不变基函数

$$\varphi_{\mathbf{l},\mu} = \sum_{\mathbf{m}\in\mathcal{M}_{\mathbf{l}}} \mathcal{C}^{\mathbf{l}\mu}_{\mathbf{m}}\varphi_{\mathbf{lm}}, \tag{26}$$

其中，μ 编码了所有可能的不变耦合，而 $\mathcal{C}^{\mathbf{l}\mu}_{\mathbf{m}}$ 是广义的（高阶）克莱布施–戈尔丹系数。详情请参见 [23]和 [26]。

我们所有的数值测试均使用 $N = 3$ 个粒子进行；在这种情况下，$B^{3}{}_{,L}{}^{2}$ 的基为

$$\tilde{\varphi}_{\mathbf{l}}(R) = \sum_{\substack{\boldsymbol{\mu}\in\mathcal{M}_{\mathbf{l}}\\ \mu_1+\mu_2+\mu_3=0}} C^{l_3,0}_{l_1,0,l_2,0}C^{l_3,\mu_1+\mu_2}_{l_1,\mu_1,l_2,\mu_2}C^{00}_{l_3,0,l_3,0}C^{00}_{l_3,-\mu_3,l_3,\mu_3}\prod_{i=1}^{3}Y^{m_i}_{l_i}(\mathbf{r}_i), \tag{27}$$

$$\text{where} \quad |l_1 - l_2| \le l_3 \le l_1 + l_2. \tag{28}$$

现在我们可以定义不变设计矩阵 $(A_{\mathrm{I}})_{ij} = \tilde{\varphi}_j(x_i)$，其中我们同样指定了 $B^{N}{}_{,K}{}^{d}$ 的基元素顺序。不变回归问题为

$$\bar{\beta}_{\mathrm{I}} = \mathrm{argmin}_{\beta_{\mathrm{I}}} L_{\mathrm{I}}(\beta_{\mathrm{I}}) \quad \text{where} \quad L_{\mathrm{I}}(\beta_{\mathrm{I}}) = \frac{1}{2}\|A_{\mathrm{I}}\beta_{\mathrm{I}} - Y\|^2 . \tag{29}$$

在架构中强制施加结构这一想法在机器学习界十分普遍。例如，物理信息神经网络 [27]是一类神经网络，它们强制其输出成为它们试图学习的方程的解。就对称性而言，在机器学习模型中强制施加此类约束会得到所谓的等变神经网络（ENNs） [28]。它们被广泛用于研究原子势 [6, 26]，或其他对称性普遍存在的领域，例如格点规范理论 [29]。应用并不局限于物理学，例如，等变神经网络也被用于图像处理 [30]，因为相差某个刚体变换的两幅图像通常被认为是相同的。

3.3 数据增强

强制施加对称性总是会使学习得到的模型具有精确的对称性。然而有趣的是，这未必总是最优选择。事实上，近期研究表明，强制施加对称性可能会损害性能 [31]，或使训练更加困难 [9]。为避免这些可能的问题，人们或许希望选择另一种学习对称性的方式，例如通过数据增强。对于 $T \in \mathbb{N}$，令 $(w_t, Q_t)_{t=1}^{T}$ 为一组实权重与旋转。我们将在下文说明它们是如何得到的。于是，对称性增强的线性最小二乘问题为

$$\beta_{\text{aug}} = \arg\min L_{\text{aug}}(\beta) \qquad \text{where} \qquad L_{\text{aug}}(\beta) = \frac{1}{2}\sum_{t=1}^{T} w_t \big\|(A \circ Q_t)\beta - Y\big\|^2, \tag{30}$$

其中，$A \circ Q_t$ 表示在点 $Q_t \cdot R$、$t = 1$、.....、n处计算得到的设计矩阵。记号 $A \circ Q$ 源于这样一个事实：我们可以将设计矩阵视为从数据空间映射到矩阵空间的函数。将数据旋转 Q_t 后得到的设计矩阵，相当于与 Q_t进行复合。请注意，Y 无需旋转，因为目标函数被选为旋转不变函数。

这一思路源于上述观察：对于旋转不变函数 f，其在一点 R处的值在整个轨道 $G \cdot R$上保持不变。我们介绍两种数据增强方法：对 G进行随机采样和最优采样，分别带来近似的平方根衰减和精确不变性。一个重要的观察是，这种增强方式并不等同于对解进行简单的对称化。增强最小二乘系统（参见式（30））的输出具有非线性依赖关系，因此必须仔细分析增强所产生的影响。

众所周知，V^{N,K^d} 在旋转下保持封闭。也就是说，对于任意的 $v \in V^{N,K^d}$ 以及$Q \in \mathrm{SO}(d+1)$，$v \circ Q \in V^{N,K^d}$。此外，这一作用在多项式空间上是线性的。它们通过（酉）Wigner D-矩阵 [23] 联系在一起，记作D。若考虑球谐函数 Y_l^m，则其经 Q 旋转后可写为

$$Y_l^m \circ Q = \sum_{m'=-l}^{l} Y_l^{m'} D^l_{m'm}(Q). \tag{31}$$

Le引理 2. 对于任意 $Q \in \mathrm{SO}(d+1)$，存在一个酉矩阵 $\mathcal{D}(Q)$ 使得

$$A \circ Q = A\mathcal{D}(Q). \tag{32}$$

证明。 我们可以将维格纳 D 矩阵的概念推广到多项式张量基，因为表示的张量积仍然是表示。此外，这些矩阵仍为酉矩阵，因为紧李群的有限维表示都是酉表示 [24]。因此，

我们知道，在给定 $Q \in \mathrm{SO}(d+1)$ 的情况下，存在 $\mathcal{D}(Q)$，使得对所有 $v \in V^{N,\ K}{}^{d}$，$v \circ Q = v\mathcal{D}(Q)$ 均成立。由于设计矩阵是此类向量的拼接，上述 $\mathcal{D}$ 满足所需性质。

□

这些矩阵 $\mathcal{D}$ 可以视为广义 D-矩阵。有关详情请参阅 [23, 26]，不过为便于说明并加深直观理解，我们将对圆上的粒子情形（$d = 1$）进行明确计算。回顾一下，基元素 $\varphi_{\mathbf{k}}$ 由三角多项式的乘积（9）给出。对其施加角度为 α 的旋转 Q_α，得到

$$Q_\alpha \cdot \varphi_{\mathbf{k}}(\Theta) = \prod_{j=1}^{N} e^{ik_j(\theta_j+\alpha)} \tag{33}$$

$$= e^{i\alpha\sum_j k_j}\phi_{\mathbf{k}}(\Theta). \tag{34}$$

因此，重新排列基以实现分解 $V^{N,\ K}{}^{1} = B^{N,\ K}{}^{1} \oplus (B^{N,\ K}{}^{1})^{\perp}$ 后，$\mathcal{D}(Q_\alpha)$ 具有块对角形式：

$$\mathcal{D}(Q_\alpha) = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & \mathrm{D}_{\alpha,\mathrm{N}} \end{pmatrix}, \tag{35}$$

其中 $\mathrm{D}^{\ \alpha,\ \mathrm{N}}$ 是一个对角矩阵，其对应于 $\varphi_{\mathbf{k}}$ 的行上的元素为 $e^{i\alpha\sum_j k_j}$。

为简化记法，我们引入 $D_t := \mathcal{D}(Q_t)$，那么 $\nabla_\beta L_{\mathrm{aug}}(\beta)$ 可直接表示为

$$\nabla_\beta L_{\mathrm{aug}}(\beta) = \sum_t w_t D_t^*(A^*AD_t\beta - A^*Y), \tag{36}$$

由此得到正规方程

$$\left(\sum_t w_t D_t^* A^* A D_t\right)\beta_{\mathrm{aug}} = \sum_t w_t D_t A^* Y. \tag{37}$$

要理解数据增强的影响，我们需要分析（37）。由于 D_t 矩阵以非线性方式出现，这项任务并非易事。我们首先按照上文针对 $d = 1$ 的说明，将 $V^{N,\ K}{}^{d} = B^{N,\ K}{}^{d} \oplus (B^{N,\ K}{}^{d})^{\perp}$ 进行分解。我们可以选取一个与该分解相容的新标准正交基。稍作符号滥用后，可得到新的设计矩阵 A、参数 β 和 D 矩阵 D_t，它们具有分块结构，其中 I，N 分别表示不变部分和非不变部分：

$$A = \begin{pmatrix} A_{\mathrm{I}} & A_{\mathrm{N}} \end{pmatrix},\ A^* = \begin{pmatrix} A_{\mathrm{I}}^* \\ A_{\mathrm{N}}^* \end{pmatrix},\ D_t = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & D_{t,\mathrm{N}} \end{pmatrix},\ \beta = \begin{pmatrix} \beta_{\mathrm{I}} \\ \beta_{\mathrm{N}} \end{pmatrix}. \tag{38}$$

在这种分解中，D_t 的形状随之确定，因为旋转平凡地作用于 $B^{N,\ K}{}^{d}$ 这意味着 $(B^{N,\ K}{}^{d})^{\perp}$ 在该作用下也必须是封闭的。事实上，由一个旋转及其逆 QQ^{-1} 所实施的作用应当是平凡作用。

注 1. 我们可以看到

$$\varepsilon_{\mathrm{sym}}(\beta) = ||\beta_{\mathrm{N}}||_2. \tag{39}$$

对于给定的 t，我们得到

$$D_t^* A^* A D_t = \begin{pmatrix} A_{\mathrm{I}}^* A_{\mathrm{I}} & A_{\mathrm{I}}^* A_{\mathrm{N}} D_{t,\mathrm{N}} \\ D_{t,\mathrm{N}}^* A_{\mathrm{N}}^* A_{\mathrm{I}} & D_{t,\mathrm{N}}^* A_{\mathrm{N}}^* A_{\mathrm{N}} D_{t,\mathrm{N}} \end{pmatrix}. \tag{40}$$

最后，利用求积公式的权重之和等于积分空间的测度这一事实，我们得到

$$\sum_t w_t D_t^* A^* A D_t = \begin{pmatrix} A_{\mathrm{I}}^* A_{\mathrm{I}} & \sum_i w_t A_{\mathrm{I}}^* A_{\mathrm{N}} D_{t,\mathrm{N}} \\ \sum w_t D_{t,\mathrm{N}}^* A_{\mathrm{N}}^* A_{\mathrm{I}} & \sum_i w_t D_{t,\mathrm{N}}^* A_{\mathrm{N}}^* A_{\mathrm{N}} D_{t,\mathrm{N}} \end{pmatrix}. \tag{41}$$

在接下来的计算中，我们将（41）写成分块形式

$$\begin{pmatrix} B & C \\ C^* & D \end{pmatrix}, \tag{42}$$

并假设原始的、未经增强的设计矩阵 A 具有满列秩。在我们的数值应用中， A 满秩等价于数据点的数量多于参数的数量。

引理 3. 如果 A 满秩，那么 A_{I} 和 A_{N} 也满秩。

由此可知，存在常数 c_0， $c_1 > 0$ ，使得对于任意向量 $u \in \mathbb{R}^{n_{\mathrm{N}}}$ （n_{N} 是 A_{N} 的列数）以及任意旋转 Q,

$$c_0\|u\| \le \|A_{\mathrm{N}} u\| \le c_1\|u\| \qquad \text{and} \tag{43}$$

$$c_0\|u\| \le \|A_{\mathrm{N}} D(Q) u\| \le c_1\|u\|, \tag{44}$$

其中第二行是由 $D(Q)$的酉性得出的。

引理 4. 设 β_{I} 表示不变最小二乘问题(29)的解，则

$$S\beta_{\mathrm{N}} = \bar{D}_T^* A_{\mathrm{N}}^* \big(Y - A_{\mathrm{I}} \bar{\beta}_{\mathrm{I}}\big), \tag{45}$$

其中 $S = D - C^* B^{-1} C$ 是分块 $A_{\mathrm{I}}^* A_{\mathrm{I}}$ (41) 的舒尔补 *[32]*，且 $D_T = \sum_t w_t D_{t,\ \mathrm{N}}$。

S 是对称正定的，其谱的上下界由 A的奇异值决定，因此，

$$c_2\|\beta_{\mathrm{N}}\| \le \big\|\bar{D}_T^* A_{\mathrm{N}}^* \big(Y - A_{\mathrm{I}} \bar{\beta}_{\mathrm{I}}\big)\big\| \le c_3\|\beta_{\mathrm{N}}\|, \tag{46}$$

其中 c_2、 c_3 仅取决于 A 的奇异值。

证明。 我们将矩阵写成式 (41) 中的形式。由于 A_{I} 满秩（参见引理 1）， B 可逆。因此，其舒尔补可逆，我们得到 β_{N} 的如下表达式

$$S\beta_{\mathrm{N}} = \overline{D}_T^* A_{\mathrm{N}}^* Y - C^* B^{-1} C A_{\mathrm{I}}^* Y \tag{47}$$

$$= \overline{D}_T^* A_{\mathrm{N}} (Y - A_{\mathrm{I}} A_{\mathrm{I}}^{\dagger} Y). \tag{48}$$

† 符号表示 Moore-Penrose 伪逆。由于 A_{I} 满秩， $A^{\dagger}{}_{\mathrm{I}} Y$ 是不变最小二乘问题 (29) 的解。最后，根据 S 的表达式可知，它是正定的，因此 (46) 成立。

□

结合引理 4 和注 1，可得以下上界:

$$\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}}) \leq \frac{1}{c_2} \|A_{\mathrm{N}} \overline{D}_T\|_{\mathrm{op}} \|A_{\mathrm{I}} \bar{\beta}_{\mathrm{I}} - Y\|, \tag{49}$$

$\|\cdot\|_{\mathrm{op}}$ 表示由 2-范数诱导的算子范数。

3.3.1 求积数据增强

由 (41) 可知，若 $(w_t,\ Q_t)_{t=1}^T$ 是 SO($d+1$) 的求积公式，则增强后的系统将变为分块对角形式，因为求积公式能够精确积分次数较低的多项式。

引理 5。 对于 $V^{N,}{}_{K}{}^{d}$ 中的一个最小二乘问题，引理 *2* 中矩阵 $\mathcal{D}(Q)$ 的各个元素是关于 Q 的 K 次多项式。

这是一个标准事实； $\mathcal{D}$ 是 SO($d+1$) [33]的有限维表示。

定理 1。 假设 $(w_t、\ Q_t)_t$ 是一个精度为 $n \geq K$ 的求积公式，则在 $V^{N}{}_{、K}{}^{d}$ 上的增强最小二乘（参见 (30)）逼近是严格不变的。也就是说，设 P_{A} 为所得的多项式，则

$$\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}}) = 0 \tag{50}$$

证明。 由 (49) 中得到的上界可知，$D_T^* = 0$ 是 $\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}}) = 0$的必要条件。根据求积公式的性质和引理 5，结论得证。

□

注 2. 引理 (49)中的上界表明，在求积公式的次数足够高之前，并不能保证$\varepsilon_{\mathrm{sym}}$ 会有所改善；这一点在 $d = 1$ (35)的情形下尤为明显。然而，仍然可以将求积增强不足的情况视为随机增强的一种特殊情形，我们将在下一小节中对此进行研究。

精度	点数
2	11
4	43
7	168
9	300
14	960
21	12648
51	207576

表1：SO(3)求积公式的大小

注 3. 定理 *1* 的条件将得到与不变最小二乘问题 (29)相同的逼近元。这可以从以下事实推得：(41) 中的非对角项积分后恰好为 0 ，且右下角块可逆（因为 A_{N}满秩，且$D_{t,\ \mathrm{N}}$具有酉性）。

为了从求积增广法中受益，必须获得适当的求积公式。对于 SO(2)，将其与 $[0, 2\pi)$ 视为同一对象后，我们得到，精度为 $n-1$ 的等权求积公式由 $\left\{\frac{1}{n}, \frac{2\pi t}{n}\right.$ 给出： $\left.t \in [n-1]\right\}$ [34]。对于 SO(3)，文献 [35]，中已经发展出了求积公式，我们在所有数值测试中都使用了这些公式。这些公式无论是在获取还是使用方面，计算成本都显著更高。我们通过表1对此进行说明，其中给出了某些求积公式 [35, 36] 的大小。求积公式大小的快速增长无法改变。

引理6。 对于 *SO(3)* 上任何次数为 n 的求积公式，都存在一个常数 c_1 ，使得该公式至少包含 $c_1 n^3$ 个点。

证明。 设 V_n 为所有最高权至多为 n 的矩阵系数的表示所张成的空间。根据角动量加法定理，若 $f, g \in V_{n/2}$ 则 $fg \in V_n$。

现设 I 为一个积分泛函，且（$\mathbf{P}\ w$）为一个 N 点求积泛函，我们将其记为 J。然后假设它们在 V_n上一致。特别地，它们在 $V_{n/2}$上也一致，并且由于它们是线性泛函，它们在 $V_{n/2}$ 上诱导出一个内积，且它们是满秩的。此外，根据构造， J 的秩至多为 N。并且 $N \geq V_{n/2} > n^3$。 □

3.3.2 随机数据增强

逼近这一对称化过程的常见方法是通过随机采样的旋转 [37, 38] 进行增强。我们证明，这会导致 $\varepsilon_{\mathrm{sym}}$ 中接近 $T^{-1/2}$ 的衰减。

定理 2. 设 A 为设计矩阵。设 Q_1、Q_2、$\cdots$ 为从 $\mathrm{SO}(d+1)$ 中独立同分布均匀抽取的元素。对于固定的 $T \in \mathbb{N}$，设 $(T^{-1}, Q_t)_{t=1}^{T}$ 为增广参数。设 P_{A} 为

多项式有界　由增广最小二乘近似得到（参见 (30)）。那么，对于任意 $\epsilon > 0$ ，

$$\frac{\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}})\sqrt{T}}{\ln(T)^{1/2+\epsilon}} \xrightarrow{a.s.} 0 \tag{51}$$

证明。 我们研究随机变量

$$\overline{D}_T = \frac{1}{T}\sum_t D_{t,\mathrm{N}}. \tag{52}$$

我们知道 $\mathbb{E}_{Q_t}(D_{t,\ \mathrm{N}}) = 0$。因此，对于 $D_{t,\ \mathrm{N}}$ d_i 的任意元素， $\mathbb{E}(d_i) = 0$ 且 $\mathbb{E}(d_i^2) < \infty$ ，因为它是定义在紧集上的连续函数。因此，如果 $d_T^i = T^{-1}\sum_t d_t^i$ 是 D_T 的一个元素，根据 [39], 的定理 2.5.$\overline{11}$，对于任意 $\epsilon > 0$,

$$\frac{\overline{d}_T\sqrt{T}}{\ln(T)^{1/2+\epsilon}} \xrightarrow{a.s.} 0. \tag{53}$$

此外，如果 $\overline{D}_T$ 是一个大小固定为 $m \times m$ 的方阵，那么其弗罗贝尼乌斯范数满足

$$||\overline{D}_T||_F \leq m \max_{kl} |\overline{d}_{Tkl}|. \tag{54}$$

最大值取遍 D_T 的所有元素。此外，模运算是连续的，因此

$$\frac{\|\overline{D}_T\|_F\sqrt{T}}{\ln(T)^{1/2+\epsilon}} \xrightarrow{a.s.} 0. \tag{55}$$

令 $\|\cdot\|$ 表示由 2-范数诱导的算子范数。由于 $\|\cdot\|_F \geq \|\cdot\|$，上述收敛性质对 $\|\cdot\|$ 成立。利用算子范数的次乘性，我们得到

$$\frac{\|A_{\mathrm{N}}\overline{D}_T^*\|\sqrt{T}}{\ln(T)^{1/2+\epsilon}} \xrightarrow{a.s.} 0. \tag{56}$$

将这一结果应用于（49），即可得到所需结果

$$\frac{||\beta_{\mathrm{N}}||_2\sqrt{T}}{\ln(T)^{1/2+\epsilon}} \xrightarrow{a.s.} 0. \qquad \square$$

注 4. 若改用 *[39]* 中的定理 *8.5.2*，则可证明存在 $\lambda > 0$ 使得

$$\limsup_{T\to\infty} \frac{\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}})\sqrt{T}}{\sqrt{\ln(\ln(T))}} = \lambda \ a.s. \tag{57}$$

因此，尽管收敛速度快于 $\sqrt{\ln(T)/T}$，该速率接近最优　　*l.*

注 5. 尽管如此，如果我们放宽收敛要求，仅要求依概率收敛，则可以将收敛速率改进为 $T^{-1/2}h(T)$，其中 $h(T)$ 是一个在无穷远处无界的可测函数。事实上，存在 $\lambda > 0$，使得对所有$\delta > 0$

$$\mathbb{E}(\varepsilon_{\mathrm{sym}}(P_{\mathrm{A}})) \leq \frac{\lambda}{\sqrt{T}}, \tag{58}$$

结论由马尔可夫不等式即可得出。

（*58*）的证明概要。我们使用 (49) 并取期望，得到

$$c_2\mathbb{E}(\|\beta_N\|_2) \leq \mathbb{E}(\|\overline{D}_T\|)\gamma, \tag{59}$$

其中 γ 是常数（相对于 Q_t）。此外，

$$\left\|\frac{1}{T}\sum_t D_{t,\mathrm{N}}\right\|_F^2 = \frac{1}{T^2}\left(\|D_{t,\mathrm{N}}\|_F^2 + 2\sum_{k\neq l}\mathrm{Tr}(D_{k,\mathrm{N}}^* D_{l,\mathrm{N}})\right). \tag{60}$$

取期望后，我们得到，

$$\mathbb{E}(\|\overline{D}_T\|_F^2) = \frac{1}{T^2}\left(\sum_t \mathbb{E}(\|D_{t,\mathrm{N}}\|_F^2) + 2\sum_{k\neq l}\mathrm{Tr}(\mathbb{E}(D_{k,\mathrm{N}}^* D_{l,\mathrm{N}}))\right) \tag{61}$$

$$= \frac{1}{T^2}\left(\sum_t \mathbb{E}(\|D_{t,\mathrm{N}}\|_F^2) + 2\sum_{k\neq l}\mathrm{Tr}(\mathbb{E}(D_{k,\mathrm{N}}^*)\mathbb{E}(D_{l,\mathrm{N}}))\right) \tag{62}$$

$$= \frac{1}{T}\mathbb{E}(\|D_{1,\mathrm{N}}\|_F^2) \stackrel{\mathrm{def}}{=} \frac{\mu}{T}. \tag{63}$$

设 $\mu = \mathbb{E}(\|D_{1,\mathrm{N}}\|_F^2)$。由于这是一个连续函数在紧集上的实值积分，因此该量定义良好。由詹森不等式、 $\|\cdot\|_F \geq \|\cdot\|_2$ 这一事实以及注 1，可得所需结果。

□

虽然随机增强易于实现，但其渐近误差率的下界为 $T^{-1/2}$；鉴于第 2 节中的讨论，这一结果并不充分。

4 数值测试

我们通过一系列数值测试来阐释上一节提出的理论。在这些测试中，我们选择 $d \in \{2, 3\}$ 和 $N = 3$，即由三个二维或三维粒子组成的系统。选择 $N = 3$ 对应于最简单的非平凡不变基。事实上，对于 $N = 1$，唯一的旋转不变函数是常数函数，

而对于 $N = 2$，任意一个粒子的位置都能确定整个系统。只有从$N = 3$ 开始，我们才会观察到非平凡的耦合。

将测试限制在相对较少的粒子上，也是因为在规模大得多的点云模型中，总会应用降维技术，把多粒子相互作用降至中等相关阶数。例如，非常成功的MACE模型 [6] 仅采用两层 $N = 3$ 来对复杂的原子间相互作用进行建模。

分子动力学中，尤其是针对材料系统，另一种常见做法是将粒子旋转到某个参考系（自然晶胞或超胞）中。因此，我们也考虑非均匀分布。然而，这些分布具有如下性质：当它们在SO($d + 1$)上取平均时，会变成均匀分布。除非另有说明，我们仅在拟合模型时改变分布，误差将使用均匀分布计算。我们的模型不是置换不变的，因此在对非均匀分布进行采样时，受约束粒子的选择在整个测试过程中保持固定。我们本可以使基具有置换不变性，但选择不这样做，因为我们希望专注于旋转不变性这一具体情况。

4.1 二维结果

	粒子 1	粒子2	粒子3
UUU	U	U	U
δUU	δ	U	U
$\delta^* UU$	M	U	U

表 2：所用测度为每行中所有测度的乘积。 U 是均匀测度， δ 是在 0 处的狄拉克测度， δ^* 是集中于 0 处的冯·米塞斯测度。本文选择了 $\kappa = 100$。它旨在作为狄拉克测度的平滑化。

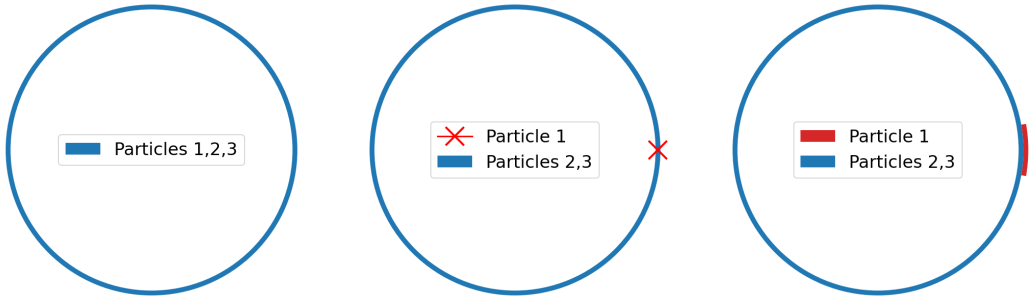
二维测试所用的分布总结于表 2，并在图 2 中进行了说明。

我们选择一个高次数多项式作为目标函数 $f \in B^{3,\ 1\ 30}$，即其系数以速率 $\alpha = 2$ 发生指数衰减，

$$f(\Theta) = \sum_{\mathbf{k}\in\mathbb{Z}^N, \|\mathbf{k}\|_1 \le \tilde{K}} c_{\mathbf{k}} e^{-\alpha||\mathbf{k}||_1} e^{\mathbf{k}\cdot\Theta}, \tag{64}$$

其中 $c_{\mathbf{k}}$ 从 $[-1, 1]$中均匀抽取。指数衰减是单变量情形下解析性的自然推广。根据假设的衰减性质，很容易看出存在 $C > 0$ ，使得

$$||f - f_K||_{L_2} \le Ce^{-\alpha K}, \tag{65}$$



(a) UUU (b) δUU (c) $\delta^* UU$

图 2：表 2 中所述分布的示意图

.

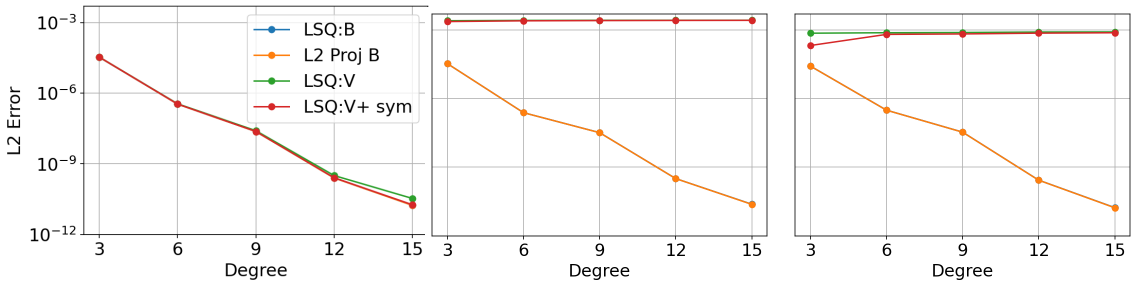
其中 $f_K \in B^{3,1}_K$ 是 f的次数为 K 的截断（L^2-正交投影）。经过充分训练后，所有最小二乘方法都应达到这一近似误差率。

图 3 展示了在不同分布下进行采样时，我们的最小二乘方案的性能（参见表 2）；旋转不变最小二乘（29）始终实现最优近似误差。对于使用 UUU 样本的全基最小二乘（14）而言，情况也是如此（见子图 3a）。对于更高次数，由于我们仅使用了大小为 8000 的训练集，该近似误差并非完全最优。不过，我们可以看到，对称化将该误差改善至最优值。子图 3b 表明，对于从 δUU采样的数据，由于缺少关于粒子 1 的信息，拟合完整基函数是不可能的。图 3c 表明，当对粒子 1 的约束略有放宽时，情况仍然如此。

在图 4 中我们可以看到，求积增强最小二乘（30）对所有采样分布都实现了精确的对称性学习。事实上，在求积公式的次数超过最小二乘系统的次数后，便会立即实现精确对称性。我们注意到，该数值实验也说明了注 2，因为在求积阶数超过逼近阶数之前，ε_{sym} 并没有明显改善。如果求积增强对最小二乘逼近中所有较低次数的项进行了对称化，那么由于多项式系数的衰减，我们预期对称化误差会呈指数衰减——这由注 1 可知。

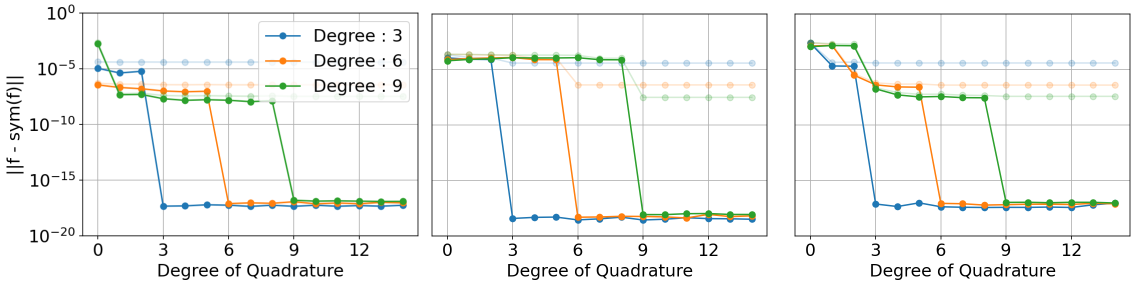
在定理 1 和 2 的证明中，为了简化起见，我们假设设计矩阵是满秩的。然而，对于采样δUU时的全基最小二乘 (14)，情况并非如此。而在采样 $\delta^* UU$ 时，尽管得到的设计矩阵是满秩的，但其条件很差。例如，对于次数为 9、学习集中有 800 个数据点的逼近，最小奇异值的量级为 10^{-6}。这说明了正则化的必要性。不过，数据增强能改善系统的条件数，而求积增强仍能实现精确的对称性学习。图 5 展示了这一点。

在图 6 中，我们将训练数据集缩减至 100 个数据点。本实验



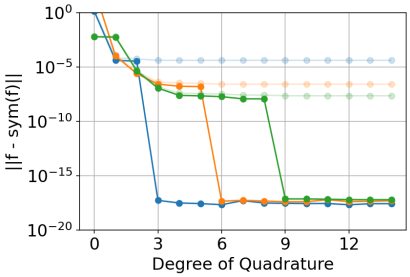
(a) UUU (b) δUU (c) $\delta^* UU$

图 3：不同分布的逼近率。我们的样本量为8000，测试集大小为2000。$\delta^* UU$ 的设计矩阵被发现是病态的，因此最小二乘需要正则化，我们通过选择截断奇异值 $10^{-4.5}$来实现。该截断值被选为使近似误差最小的值。三幅图共用相同的y轴。请注意，蓝色、绿色和红色曲线无法区分。



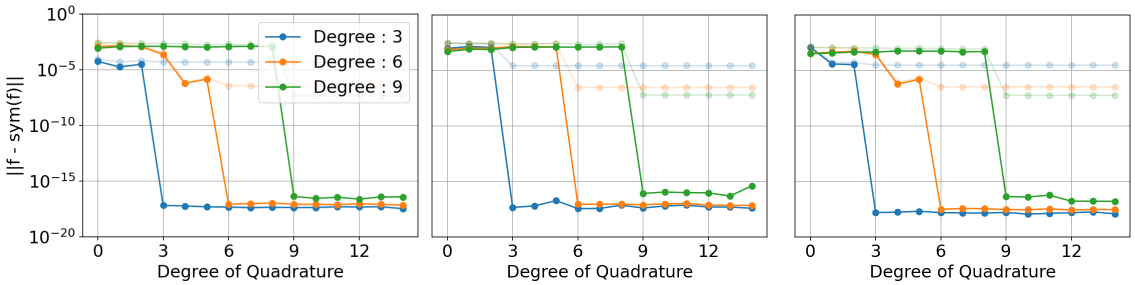
(a) UUU (b) δUU (c) $\delta^* UU$

图 4：ε_{sym}使用 UUU、δUU、$\delta^* UU$进行求积增强。未增强的训练集大小为 800，验证集大小为 200。所有图的 y轴相同。阴影线表示近似误差。



(a) $d = 1$

图 5： ε_{sym} 使用 $\delta^{*}UU$ 进行求积增强且不应用正则化时的结果。训练集大小为 800，验证集大小为 200。

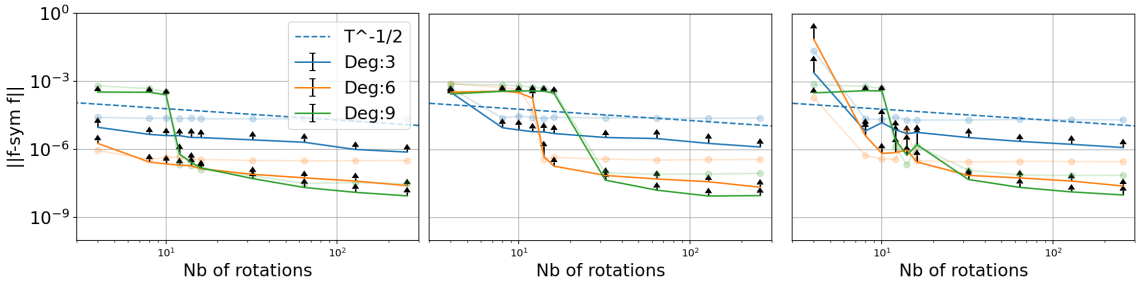


(a) UUU (b) δUU (c) $\delta^{*}UU$

图 6：对 UUU、δUU 和 $\delta^{*}UU$ 进行求积增强的结果。此时（未经增强的）训练数据集大小为 100。验证集大小为 100。所有图共用同一 y轴。阴影线表示近似误差。

这表明，即使数据集规模不足，求积增强最终也会实现完美的对称化。

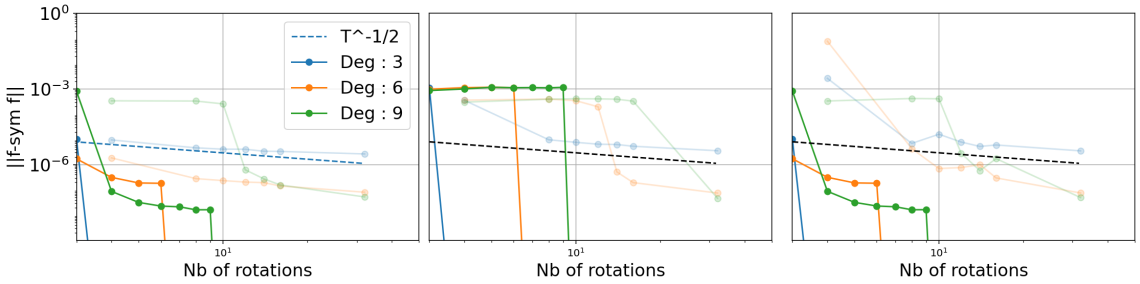
图7向我们展示了随机增强（30）的结果。在渐近区间中，我们清楚地观察到由2所预测的速率 $T^{-1/2}$ 。然而，在前渐近区间中，我们观察到对称化误差急剧下降，我们推测这对应于所有参数被完全确定的时刻。这一点在集中分布中尤其得到了很好的说明。事实上，由于粒子 1始终被约束在 $\theta = 0$，某些基元素在未增强的数据集中是未确定的。这也可以从 $\delta^{*}UU$ 相对较早的下降中观察到，如图7c所示；在该情形下，粒子 1并未被完全约束，因此这些基元素具有相对更多的信息。此外，在该下降附近，标准差相当高，这可以解释为最小二乘系统正处于拥有足够数据以完全确定这少数基元素的临界点。然而，存在一个先验上反直觉的观察。对于固定的



(a) UUU (b) δUU (c) $\delta^* UU$

图7： ε_{sym}对 UUU、 δUU和 $\delta^* UU$进行随机增强的结果。（未经增强的）训练数据集大小为100，验证集大小为100。旋转次数为4到256之间的2的幂。对于每个分布，我们模拟了10个独立场景。图中曲线表示平均结果，误差箭头表示1个标准差。为便于观察，没有绘制向下的箭头，因为在曲线下降附近模拟结果不稳定，误差区间会延伸至负值。蓝色虚线表示 $T^{-1/2}$ 衰减。阴影线对应第一次试验的近似误差。

增强 ε_{sym} 会随着次数的增加而减小。人们会预期较低次数的逼近更容易对称化，因为参数更少。（49）中的上界或许可以解释这一现象。由于我们的目标函数具有很高的正则性，对于拟合良好的不变最小二乘（29），项$\|Y - A_{\text{I}}\beta_{\text{I}}\|$ 随 K呈指数衰减。因此，即便 $\|D_T^* A_{\text{N}}^*\|$ 会随着K的增大而增长，总体而言，对于足够光滑的目标，$\|D_T^* A_{\text{N}}^*\|\|Y - A_{\text{I}}\beta_{\text{I}}\|$ 仍会随K 减小。我们在附录（第6节）中说明了这一行为。



(a) UUU (b) δUU (c) $\delta^* UU$

图8：图4和图7的叠加。阴影图表示随机增强，实线图表示求积增强。所有图共用同一个y轴。

图8比较了随机增强和求积增-强（30）的对称化误差 ε_{sym} 。虽然我们的证明无法保证在精确对称性学习之前，求积增强会

显著优于随机增强（参见注 2），但我们在实验中观察到，求积增强的效果比随机增强好一个数量级。需要注意的是，由于系统尚未进入渐近区间，因此无法观察到平方根行为。

4.2 三维中的结果

	粒子 1	粒子2	粒子3
UUU	U	U	U
δUU	δ	U	U
$\delta^* UU$	δ^*	U	U
$\delta H_1 U$	δ	H_1	U
$\delta^* H_1^* U$	δ^*	H_1^*	U

表3：所用测度为每一行中所有测度的乘积。 U 是 S^2上的均匀测度， δ 是北极处的狄拉克测度， H_1 是测地线（经过北极）上的均匀测度。符号 $*$ 表示对测度进行平滑化，即添加额外的高斯噪声。在实验中，我们选择用标准差为 0.1 的高斯噪声对球面角进行独立扰动。

三维实验中使用的数据分布汇总于表3，并展示在图9中。

由于我们处理的是 3 个粒子，可以使用 (28) 来构建近似方案，因为它枚举了空间 $B^{3,}_{K}{}^{2}$ 的一组标准正交基。同样，我们选择目标多项式，使其系数以速率 $\alpha = 2$ 指数衰减，从而模拟解析性。也就是说，我们可以将目标函数 f 写成以下形式：

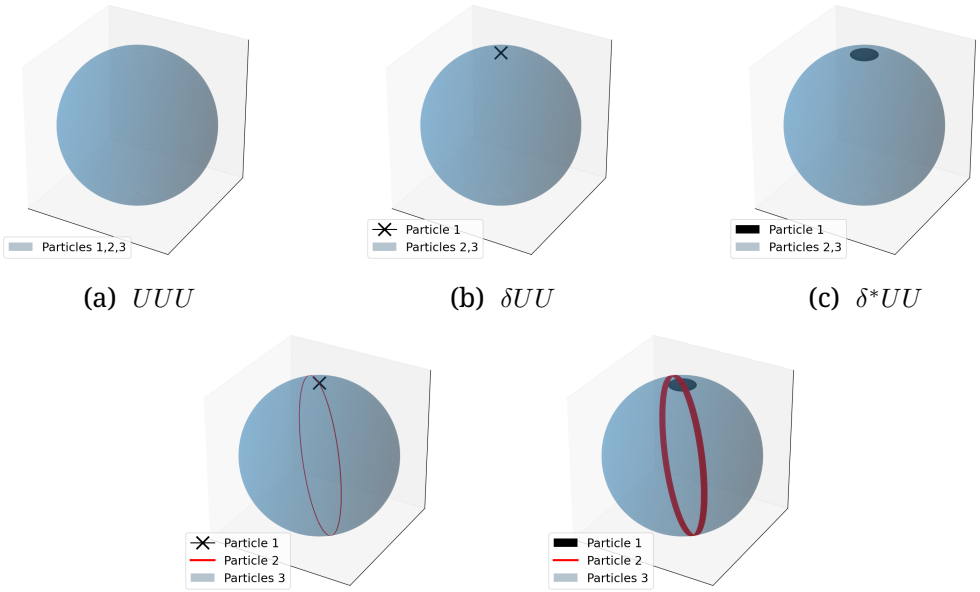
$$f = \sum_{\mathbf{l}} c_{\mathbf{l}} e^{-\alpha \|\mathbf{l}\|_1} \tilde{\varphi}_{\mathbf{l}}, \tag{66}$$

其中， $c_{\mathbf{l}}$ 在 $[-1, 1]$上均匀选取。在整个测试过程中，我们观察到的结果在定性上与二维测试中的结果非常相似。

在图10中，我们恢复了解析函数最优多项式逼近的谱收敛率。同样，我们必须对 $\delta^* UU$ 和$\delta^* H_1^* U$进行一些正则化。奇异值的最优截断值分别为 $10^{-3.4}$和10^{-5}。

接下来，为了测试求积增强，我们需要获得 SO(3) 上群平均的求积公式。一般而言，在流形（例如李群）上构造求积公式是一个非平凡的开放问题。文献 [35] 研究了 SO(3) 的特定情形，我们据此获得了次数为 1 到 11 的求积公式。

在图 11 中，我们观察到（与二维测试结果一致），当求积的阶数超过.....的阶数时，求积增强 (30) 可以实现完美的对称性学习

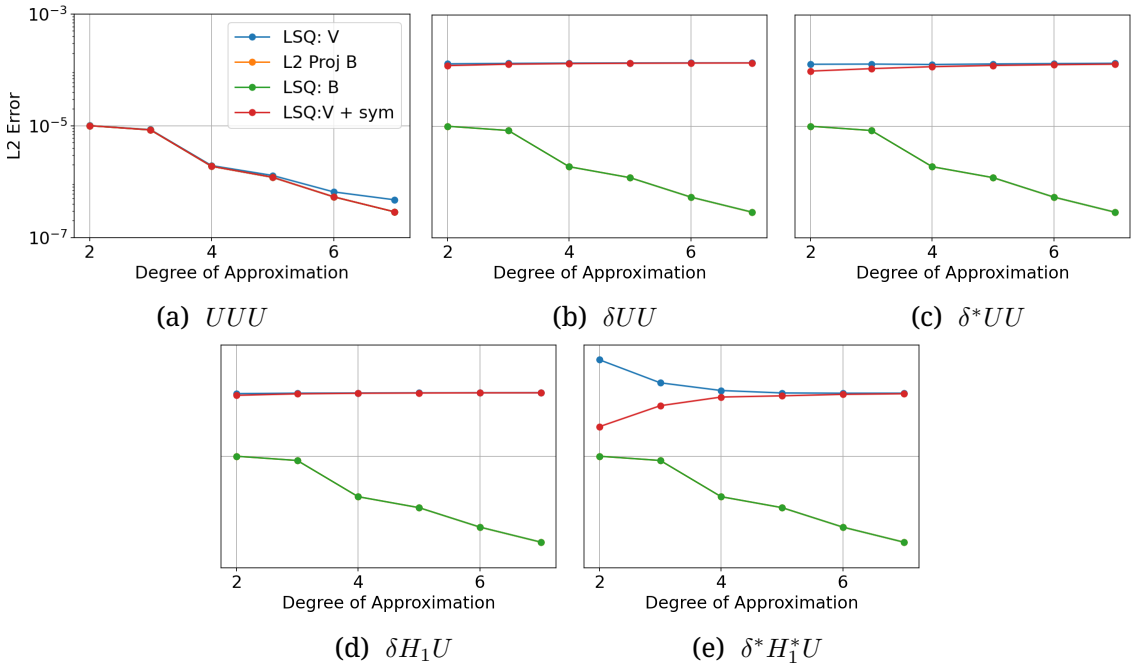


(a) UUU (b) δUU (c) $\delta^* UU$

(d) $\delta H^1 U$ (e) $\delta^* H_1^* U$

图9：表3中提及的分布示意图

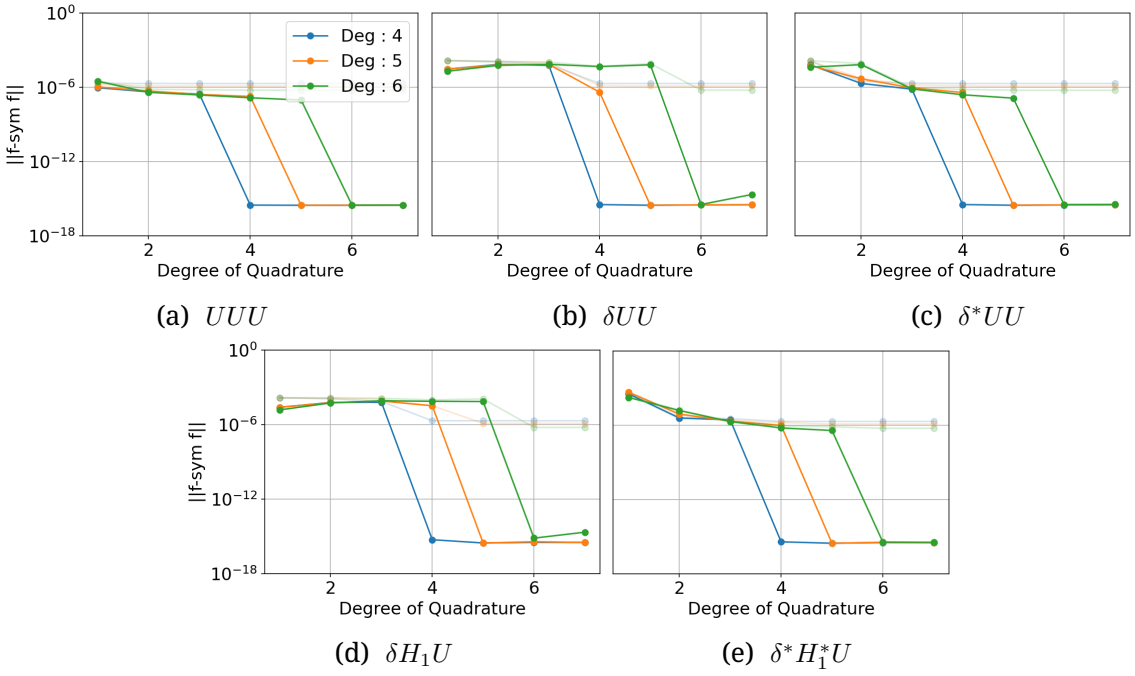
.



(a) UUU (b) δUU (c) $\delta^* UU$

(d) $\delta H_1 U$ (e) $\delta^* H_1^* U$

图10：表3中各分布的逼近率。我们使用训练集中的 10000 个点和验证集中的 2500 个点，对一个 11 次旋转不变函数进行逼近。与之前一样，扰动分布 $\delta^* UU$和 $\delta^* H_1^* U$需要进行正则化。所有图共用同一 y轴。

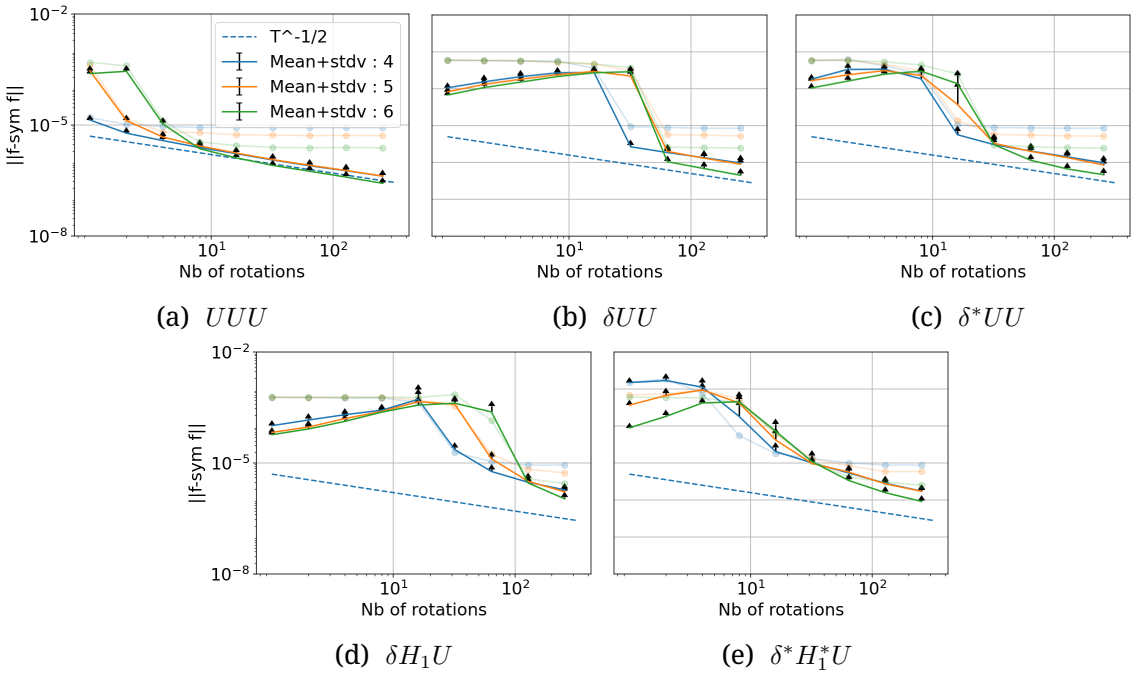


(a) UUU (b) δUU (c) $\delta^* UU$

(d) $\delta H_1 U$ (e) $\delta^* H_1^* U$

图 11：求积增强的对称化误差变化。我们近似一个 11 次目标，且（未增强的）训练集大小为 800，验证集大小为 200。所有图共用同一 y轴。阴影线表示近似误差。

该模型。值得特别注意的是，虽然 $\delta H_1 U$ 在未增强的最小二乘中无需事先进行正则化，但当数据增强不足时，最小二乘会变得极不稳定，与二维情形类似。我们推测，受限的数据会使某些基元素无法确定，直到旋转为数据集带来足够的多样性。



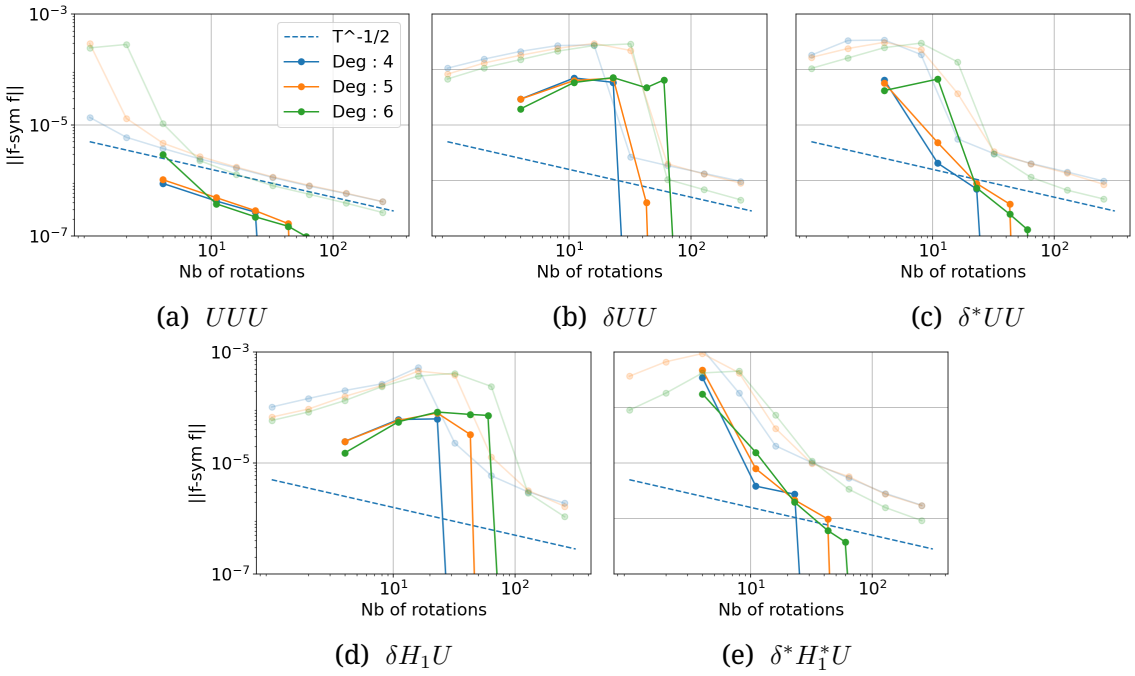
(a) UUU (b) δUU (c) $\delta^* UU$

(d) $\delta H_1 U$ (e) $\delta^* H_1^* U$

图12：随机增强下对称化误差的演变。我们逼近一个次数为 11 的目标，训练集大小为 800，验证集大小为 200。除 UUU 外，所有方案都经过正则化。所有图共用同一 y轴。我们先运行 10 次独立模拟，再绘制对称化误差的均值和标准差。阴影线对应第一次试验的近似误差。

作为比较，图12展示了随机增强（30）下对称化误差的演变。我们观察到与 $d = 1$ 相同的行为。除了出现预期的平方根衰减外，在施加足够多次旋转后， ε_{sym} 也会下降。

图13比较了随机增强与求积增强（30）的对称化误差 ε_{sym} 。正如在图8中一样，我们观察到，经验上求积增强比随机增强好一个数量级。此外，平方根行为可在图13a中看到。



(a) UUU (b) δUU (c) $\delta^* UU$

(d) $\delta H_1 U$ (e) $\delta^* H_1^* U$

图13：图12和图11的叠加。阴影曲线对应随机增强，实线曲线对应求积增强。所有曲线共用同一y轴。

5 结论

我们考虑了数据驱动设定下对群作用下不变函数的逼近。我们分析了两种用于增强数据集和最小二乘系统的方法，以改善所得逼近元的对称性：基于确定性求积的方法与概率方法。对于这两种情形，我们都给出了严格的收敛速率。关键结论是，当原始数据分布相对于对称群高度集中时，求积方法可以提供精确对称化，但代价是最小二乘系统显著扩大。在前渐近区间，它相对于概率方法并无可辨别的改进；后者不出所料地以通常的 $T^{-1/2}$ 速率收敛，其中 T 为增强次数（尽管对该速率的解释并非立即显而易见）。我们的两种方法都无法同时使数据增强既准确又计算成本低廉。

根据我们在第 2 节提出的论证，当守恒性质对下游模拟任务很重要时，尽可能严格地保持对称性

可能至关重要。在这种情况下，建议采用精确对称化，无论是通过在架构中强制实现，还是通过所提出的求积方法实现。

最后需要指出的是，本文仅限于多项式模型，而将我们的观察扩展到非线性方法和深度学习方法并非显而易见：这需要进一步的理论研究和实证测试，但实证证据表明，求积增强在这一场景中同样具有潜力，因为在我们的数值测试中，其表现比随机增强好一个数量级。

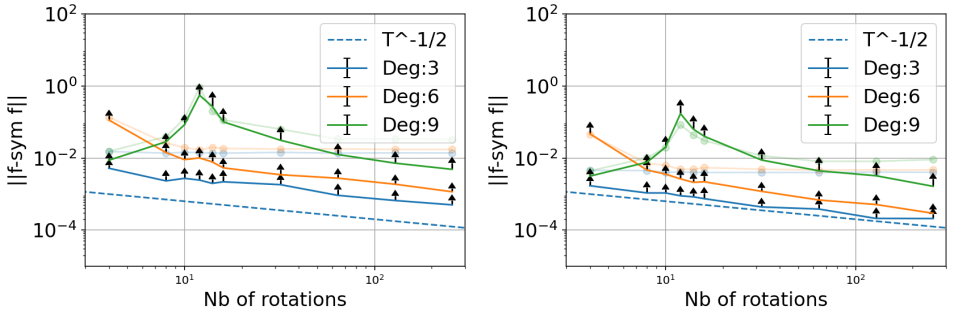
致谢

我们感谢 Timon Gutleb 和 Isaac Holt 在本研究项目早期阶段进行的富有启发性的讨论。作者得到加拿大自然科学与工程研究理事会的支持，资助来自 NSERC探索基金以及一项 NSERC-NSF 关于量子科学和人工智能的合作基金。研究还获得了加拿大数字研究联盟（alliancecan.ca）提供的计算资源和服务，以及不列颠哥伦比亚大学高级研究计算提供的支持。Christoph Ortner 是 Symmetric Group LLP 的合伙人，该公司对力场进行商业许可。

6 附录

在本附录中，我们讨论图7中观察到的异常现象。我们进行了模拟，从均匀分布中采样，针对具有不同正则性的目标函数开展研究。我们模拟了系数渐近表现为 $K^{-1/2}, K^{-1}, K^{-2}, K^{-3}$ 的目标，这对应于 $_1{}_2$-H¨older 、 C^1 和 C^2 和 C^3 函数。

从图14中，我们可以观察到上界（49）的影响。确实，对于正则性较低的目标函数，我们此前的直觉是正确的：对低次逼近元进行对称化更容易。但随着正则性提高，高次逼近元逐渐赶上并最终超过低次逼近元。这表明正则性越高，对称化就越容易。



(a) $1/2$-持有者　(b) C^1

(c) C^2　(d) C^3

图14：$\varepsilon_{\text{psym}}$ 目标函数正则性变化时的情况。未增强的训练集包含100个点，验证集大小为100，目标多项式的次数固定为30。

参考文献

[1]Jörg Behler 和 Michele Parrinello。高维势能面的广义神经网络表示。《物理评论快报》，98:146401，2007年4月。doi：10.1103/PhysRevLett.98.146401。URL https://link.aps.org/doi/10.1103/ PhysRevLett.98.146401.[2] Jiang Wang、Simon Olsson、Christoph Wehmeyer、Adrià Pérez、Nicholas E. Charron、Gianni de Fabritiis、Frank Noé 和 Cecilia Clementi。粗粒化分子动力学力场的机器学习。《*ACS*中央科学》，5(5):755–767，2019年05月。doi：10. 1021/acscentsci.8b00913。URL https://doi.org/10.1021/acscentsci.8b00913.[3] Rose Yu 和 Rui Wang。从数据中学习动力系统：物理引导的深度学习简介。《美国国家科学院院刊》，121(27): e2311808121，2026/01/26 2024。doi：10.1073/pnas.2311808121。URL https://doi.org/ 10.1073/pnas.2311808121.[4] Johannes Brandstetter、Max Welling 和 Daniel E Worrall。面向神经偏微分方程求解器的李点对称性数据增强。载于 Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvari、Gang Niu 和 Sivan Sabato（编），第*39*届国际机器学习会议论文集，第162卷，《机器学习研究论文集》，第2241–2256页。PMLR，2022年7月17–23日。URL https://proceedings.mlr.press/v162/brandstetter22a.html.[5] George Casella Christian P. Robert。《蒙特卡罗统计方法》。2004年。[6] Ilyes Batatia、David P Kovacs、Gregor Simm、Christoph Ortner 和 Gabor Csanyi。Mace：用于快速、准确力场的高阶等变消息传递神经网络。载于 S. Koyejo、S. Mohamed、A. Agarwal、D. Belgrave、K. Cho 和 A. Oh（编），《神经信息处理系统进展》，第35卷，第11423–11436页。Curran Associates, Inc.，2022年。URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf.[7] Akiyoshi Sannai、Makoto Kawano 和 Wataru Kumagai。不变与等变雷诺网络。《机器学习研究杂志》，25(42):1–36，2024年。URL http://jmlr.org/papers/v25/22-0891.html.[8] YuQing Xie 和 Tess Smidt。两个对称性的故事：探索等变模型的损失景观，2025年。URL https://arxiv.org/abs/2506.02269.[9] Andrei Manolache、Luiz F. O. Chamon 和 Mathias Niepert。通过约束优化学习（近似）等变网络，2025年。URL https://arxiv.org/ abs/2505.13631。

[10] Filippo Bigi、Paolo Pegolo、Arslan Mazitov 和 Michele Ceriotti。突破无约束机器学习原子间势的极限，2026年。URL https://arxiv.org/ abs/2601.16195.[11] Arslan Mazitov、Filippo Bigi、Matthias Kellner、Paolo Pegolo、Davide Tisi、Guillaume Fraux、Sergey Pozdnyakov、Philip Loche 和 Michele Ceriotti。Pet-mad：用于先进材料建模的轻量通用原子间势，2025年。URL https: //arxiv.org/abs/2503.14118.[12] E. Hairer、C. Lubich 和 G. Wanner。几何数值积分：常微分方程的保结构算法。《计算数学施普林格丛书》。施普林格柏林海德堡，2013年。ISBN 9783662050187。URL https://books.google.ca/books?id=cPTxCAAAQBAJ.[13] Atsushi Horikoshi。南部力学中保结构积分器的影子哈密顿量。理论与实验物理学进展，2024(5):053A05，2024年05月。ISSN 2050-3911。doi：10.1093/ptep/ptae067。URL https://doi.org/10.1093/ptep/ptae067.[14] V.I. Arnold。经典力学的数学方法，第60卷。施普林格，1989年。[15] William C. Swope、Hans C. Andersen、Peter H. Berens 和 Kent R. Wilson。一种计算分子物理团簇形成平衡常数的计算机模拟方法：在小水团簇中的应用。化学物理杂志，76(1):637–649，1982年01月。ISSN 0021-9606。doi：10.1063/1.442716。URL https://doi.org/10.1063/1.442716.[16] Salman Habib 和 Robert D. Ryne。李雅普诺夫指数的辛计算。《物理评论快报》，74:70–73，1995年1月。doi：10.1103/PhysRevLett.74.70。URL https://link. aps.org/doi/10.1103/PhysRevLett.74.70.[17] Giancarlo Benettin、Luigi Galgani、Antonio Giorgilli 和 Jean-Marie Strelcyn。光滑动力系统和哈密顿系统的李雅普诺夫特征指数；一种计算所有指数的方法。第一部分：理论。*Meccanica*，15(1):9–20，1980年。doi：10.1007/BF02128236。URL https://doi.org/10.1007/BF02128236.[18] Jairo Bochi 和 Marcelo Viana。一般体积保持映射和辛映射的李雅普诺夫指数。数学年刊，161(3):1423–1485，2005年。ISSN 0003486X。URL http://www.jstor.org/stable/3597359.[19] Mario Bessa 和 Paulo Varandas。哈密顿线性微分系统的正李雅普诺夫指数，2014年。URL https://arxiv.org/abs/1304.3794.[20] Marcelo Viana。李雅普诺夫指数讲义。《剑桥高等数学研究丛书》。剑桥大学出版社，2014年。

[21] V.I. Arnold、E. Khukhro、V.V. Kozlov 和 A.I. Neishtadt。《经典力学与天体力学的数学方面》。数学科学百科全书。施普林格柏林海德堡，2007年。ISBN 9783540489269。URL https://books.google.ca/books? id=25iQQvHe9awC。[22] Pauli Virtanen、Ralf Gommers、Travis E. Oliphant、Matt Haberland、Tyler Reddy、David Cournapeau、Evgeni Burovski、Pearu Peterson、Warren Weckesser、Jonathan Bright、St´efan J. van der Walt、Matthew Brett、Joshua Wilson、K. Jarrod Millman、Nikolay Mayorov、Andrew R. J. Nelson、Eric Jones、Robert Kern、Eric Larson、C J Carey、Ilhan Polat、Yu Feng、Eric W. Moore、Jake VanderPlas、Denis Laxalde、Josef Perktold、Robert Cimr- man、Ian Henriksen、E. A. Quintero、Charles R. Harris、Anne M. Archibald、Antˆonio H. Ribeiro、Fabian Pedregosa、Paul van Mulbregt 和 SciPy 1.0 贡献者。SciPy 1.0：Python科学计算基础算法。《自然·方法》，17:261–272，2020年。doi: 10.1038/s41592-019-0686-2。[23] Adolfas P. Yutsis、I. B. Levinson 和 V. V. Vanagas。《角动量理论的数学工具》。1962年。[24]D. Bump。《李群》。数学研究生教材。施普林格，2004年。ISBN 9780387211541。URL https://books.google.ca/books?id=wCE8dInDw5gC。[25] L. Gy¨orfi。《非参数回归的无分布理论》。施普林格统计学丛书。施普林格，2002年。ISBN 9780387954417。URL https://books.google.ca/ books?id=Q9NMAp-cKy0C。[26] Genevi`eve Dusson、Markus Bachmayr、G´abor Cs´anyi、Ralf Drautz、Simon Etter、Cas van der Oord 和 Christoph Ortner。原子簇展开：完备性、效率与稳定性。《计算物理学杂志》，454:110946，2022年。ISSN 0021-9991。doi: https://doi.org/10.1016/j.jcp.2022.110946。URL https://www.sciencedirect. com/science/article/pii/S0021999122000080。[27] Sifan Wang、Shyam Sankaran、Hanwen Wang 和 Paris Perdikaris。物理信息神经网络训练专家指南，2023年。URL https://arxiv.org/abs/2308. 08468。[28] Lek-Heng Lim 和 Bradley J. Nelson。什么是等变神经网络？，2022年。URL https://arxiv.org/abs/2205.07362。[29] Matteo Favoni、Andreas Ipp 和 David I. M¨uller。晶格规范等变神经网络的应用。《EPJ会议网络》，274:09001，2022年。ISSN 2100-014X。doi: 10.1051/epjconf/202227409001。URL http://dx.doi.org/10.1051/epjconf/ 202227409001。

[30] Karl Bengtsson Bernander、Ida-Maria Sintorn、Robin Strand 和 Ingela Nystr¨om。使用等变神经网络对旋转不变生物医学图像进行分类。《科学报告》，14(1):14995，2024年。doi: 10.1038/s41598-024-65597-x。URL https: //doi.org/10.1038/s41598-024-65597-x。[31] Georg B¨okman 和 Fredrik Kahl。研究 ReLU 网络如何编码对称性，2023年。URL https://arxiv.org/abs/2305.17017。[32] F. Zhang。《舒尔补及其应用》。数值方法与算法。Springer US，2006年。ISBN 9780387242736。URL https://books.google.ca/ books?id=EMEyg8NcuskC。[33] R. Goodman 和 N.R. Wallach。《对称性、表示与不变量》。数学研究生教材。Springer New York，2010年。ISBN 9781441927293。URL https: //books.google.ca/books?id=5CTQtgAACAAJ。[34] Lloyd N. Trefethen 和 J. A. C. Weideman。指数收敛梯形法则。《SIAM评论》，56(3):385–458，2014年。doi: 10.1137/130932132。URL https://doi.org/10.1137/130932132。[35] M. Gr¨af。《在黎曼流形上计算最优求积点的高效算法》。开姆尼茨工业大学，Universit¨atsverlag Chemnitz，2013年。ISBN 978-3-941003-89-7。[36] Manuel Gr¨af 和 Daniel Potts。旋转群上的采样集与求积公式。《数值泛函分析与优化 - NUMER FUNC ANAL OPTIMIZ》，30，2009年08月。doi: 10.1080/01630560903163508。[37] Gregory Benton、Marc Finzi、Pavel Izmailov 和 Andrew Gordon Wilson。学习神经网络中的不变性，2020年。URL https://arxiv.org/abs/2010.11882。[38]Tri Dao、Albert Gu、Alexander J. Ratner、Virginia Smith、Christopher De Sa 和 Christopher R´e。现代数据增强的核理论，2019年。URL https: //arxiv.org/abs/1803.06084。[39] Rick Durrett。《概率：理论与例子》。剑桥统计与概率数学丛书。剑桥大学出版社，第5版，2019年。