arXiv:2301.09308v3 [cs.LG] 3 Mar 2024

论几何图神经网络的表达能力 S

Chaitanya K. Joshi *[1] **Cristian Bodnar** *[1] **Simon V. Mathis** [1] **Taco Cohen** [2] **Pietro Li`o** [1]

摘要

图神经网络（GNN）的表达能力已通过 Weisfeiler-Leman（WL）图同构检验得到了广泛研究。然而，标准图神经网络和 WL 框架并不适用于嵌入欧几里得空间的几何图，例如生物分子、材料以及其他物理系统。本文提出几何版 WL 测试（GWL），用于区分几何图，同时遵循其底层物理对称性：置换、旋转、反射和平移。我们利用 GWL，根据区分几何图的能力，刻画对物理对称性不变或等变的几何图神经网络的表达能力。GWL 揭示了关键设计选择如何影响几何图神经网络的表达能力：（1）不变层的表达能力有限，因为它们无法区分一跳相同几何图；（2）等变层通过将几何信息传播到局部邻域之外，能够区分更大类别的图；（3）高阶张量和标量化使几何图神经网络具有最大表达能力；（4）基于区分能力的 GWL 视角等价于通用逼近。补充我们研究结果的合成实验可在 https://github.com/chaitjo/geometric-gnn-dojo 获取。

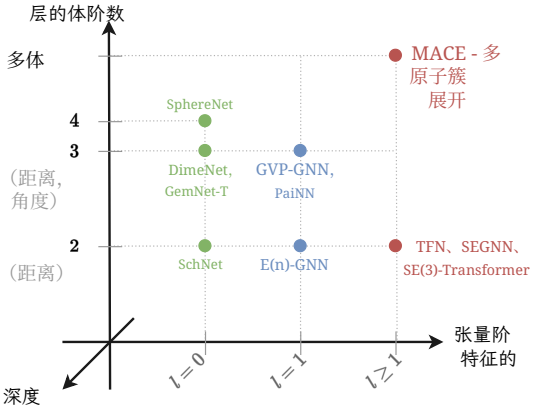


图 1：**几何 GNN 表达能力的维度：** (1) 体阶：提高标量化体序可构建更具表达能力的局部描述符；(2)张量阶：高阶张量决定邻域的相对方向；以及 (3)深度：深层等变层可将几何信息传播到局部邻域之外。

1. 引言

生物化学系统（Jamasb 等人，2022年）、材料科学（Chanussot 等人，2021）、物理模拟（Sanchez-Gonzalezetal.，2020年）以及多智能体机器人技术（Li 等人，2020年）都同时包含几何结构和关系结构。

此类系统可以通过嵌入欧几里得空间中的几何图来建模。例如，分子可表示为一组节点，其中包含关于每个原子及其三维空间坐标的信息，以及速度或加速度等其他几何量。值得注意的是，几何属性会随系统的欧几里得变换而变换，也就是说，它们对于旋转、反射和平移构成的对称群是等变的。未考虑空间对称性的标准图神经网络（GNN）不适用于几何图，因为几何属性将无法继续保留其物理意义和变换行为（Bogatskiy 等人，2022年；Bronstein 等人，2021）。

面向几何图的图神经网络遵循消息传递范式（Gilmer 等人，2017年），通过聚合局部邻域中的特征，以置换等变的方式更新节点特征。至关重要的是，除置换外，节点更新后的几何特征还保留了初始属性的变换行为，即它们对于旋转李群（$SO(d)$）或旋转与反射李群（$O(d)$）也具有等变性。我们用 $\mathfrak{G}$ 作为这些李群的通用符号。我们考虑两类几何图神经网络：（1） **$\mathfrak{G}$-等变模型**，

*共同第一作者 [1]英国剑桥大学 [2]荷兰高通人工智能研究院。高通人工智能研究院是高通技术公司的一个项目。通信作者：Chaitanya K. Joshi <chaitanya.joshi@cl.cam.ac.uk>。

论文集： 40th 国际机器学习大会，美国夏威夷州檀香山。PMLR 202，2023。版权所有 © 2023，归作者所有。

其中，中间特征和传播消息是等变几何量，例如向量（Satorras 等人，2021）或张量（Thomas 等人，2018）；以及（2）$\mathfrak{G}$-**不变模型**，这类模型只传播局部不变标量特征，例如距离和角度（Schütt 等人，2018；Gasteiger 等人，2020）。

这两类架构在诸多应用中都展现出良好的前景，包括蛋白质结构预测（Baek 等人，2021）与设计（Dauparas 等人，2022），以及分子动力学（Batzner 等人，2022）和催化（Gasteiger 等人，2021）。与此同时，一些关键理论问题仍未得到解答：（1）如何刻画几何图神经网络的表达能力？（2）$\mathfrak{G}$等变与 $\mathfrak{G}$不变图神经网络之间存在怎样的权衡?

图同构问题（Read 与 Corneil，1977年）以及用于区分非同构图的 Weisfeiler-Leman（WL）（Weisfeiler 与 Leman，1968）测试，已成为分析非几何图神经网络表达能力的有力工具（Xu 等人，2019；Morris 等人，2019）。WL 框架一直是推动图表示学习发展的重要动力（Chen 等人，2019；Maron 等人，2019；Dwivedi 等人，2023；Bodnar 等人，2021b；a）。然而，WL 并不直接适用于几何图，因为几何图具有更强的几何同构概念，能够考虑空间对称性。

贡献。 本文从区分非同构几何图的角度，研究几何图神经网络的表达能力。

- 在第3节中，我们提出了 Weisfeiler-Leman 图同构检验的几何版本，称为 GWL，它构成了几何图神经网络表达能力的理论上界。
- 在第4节中，我们使用 GWL 框架形式化地阐述关键设计选择如何影响几何图神经网络的表达能力，概述见图1。值得注意的是，不变图神经网络无法区分一跳邻域相同的图，也无法计算体积、质心等非局部几何性质。等变图神经网络能够区分更大范围的图，因为堆叠等变层可以将几何信息传播到局部邻域之外。
- 第5节的合成实验展示了构建表达能力最强的几何图神经网络的实际意义，例如随着深度增加而出现的几何过度挤压，以及凸显高阶球张量作用的反例。例如几何过度挤压会随深度增加而加剧，而反例则凸显了高阶球张量的作用。
- 最后，在第 6 节中，我们证明了模型区分几何图的能力与其普适逼近能力之间的等价性。尽管普适性是二元的，但 GWL 基于区分能力的视角能为理解几何图神经网络表达能力提供更细致且更具实践启发性的观察视角。

2. 背景

图同构。 属性图 $\mathcal{G} = (\boldsymbol{A},\ \boldsymbol{S})$ 是由边连接的一组 $\mathcal{V}$ 节点 n。 $\boldsymbol{A}$ 表示一个 $n \times n$邻接矩阵，其中每个条目 $a_{ij} \in \{0,1\}$ 表示节点 i 与 j 之间是否存在边。 $\boldsymbol{S} \in \mathbb{R}^{n \times f}$ 标量特征矩阵 存储与每个节点 i关联的属性$\boldsymbol{s}_i \in \mathbb{R}^f$，其中下标对应于相应矩阵中的行和列。

图同构问题询问两个图是否相同，只是绘制方式不同。两个带属性的图$\mathcal{G}$，$\mathcal{H}$ 若存在保边双射 $b\colon\ \mathcal{V}(\mathcal{G}) \to \mathcal{V}(\mathcal{H})$ 满足$\boldsymbol{s}_i^{(\mathcal{G})} = \boldsymbol{s}_{b\,i}^{(\mathcal{H})}$ （），则称它们同构（记作 $\mathcal{G} \simeq \mathcal{H}$）。

魏斯费勒-莱曼测试（WL）。WL 测试用于判断两个（带属性的）图是否同构。在第零次迭代中，WL 为每个节点 i分配一个来自可数颜色空间 C的颜色 $c_i^{(0)} \in C$。若节点的特征相同，则它们被赋予相同颜色；否则，赋予不同颜色。在后续迭代 t中，WL 会迭代更新着色$\boldsymbol{c}_i^{(t)} \in C$:

$$\boldsymbol{c}_i^{(t)} := \textsc{Hash}\left(\boldsymbol{c}_i^{(t-1)}, \{\!\{\boldsymbol{c}_j^{(t-1)} \mid j \in \mathcal{N}_i\}\!\}\right), \qquad (1)$$

其中，HASH 是一个单射映射（即完美哈希映射），为每个输入分配一个唯一的颜色。当颜色所诱导的节点划分趋于稳定时，测试终止。给定两个图 $\mathcal{G}$ 和 $\mathcal{H}$，若存在某次迭代 t 使得 $\{\!\{c_i^{(t)} \mid i \in \mathcal{V}(\mathcal{G})\}\!\} \neq \{\!\{c_i^{(t)} \mid i \in \mathcal{V}(\mathcal{H})\}\!\}$，则这两个图不同构。否则，WL 测试无法得出结论，我们也无法区分这两个图。

群论。我们假设读者对群论有基本了解，概述请参见 Zee（2016年）。我们用 $\mathfrak{g} \cdot x$ 表示群 $\mathfrak{G}$ 在空间 X 上的群作用。如果 $\mathfrak{G}$ 作用于空间 X 和 Y，我们称: (1) 函数 $f\colon\ X \to Y$ 是 $\mathfrak{G}$-不变的，如果 $f(\mathfrak{g} \cdot x) = f(x)$，即输出在输入的变换下保持不变；以及 (2) 函数 $f\colon\ X \to Y$ 是 $\mathfrak{G}$-等变的，如果 $f(\mathfrak{g} \cdot x) = \mathfrak{g} \cdot f(x)$，即输入的变换必须导致输出以等价方式变换。

我们还考虑 $\mathfrak{G}$-轨道单射 函数: $\mathfrak{G}$-轨道是 $x \in X$ 的 $\mathcal{O}_{\mathfrak{G}}(x) = \{\mathfrak{g} \cdot x \mid \mathfrak{g} \in \mathfrak{G}\} \subseteq X$。当 x 和 x' 属于同一轨道时，我们记作 $x \simeq x'$。如果对于任意 $x_1,\ x_2 \in X$，函数 $f\colon\ X \to Y$ 满足: $f(x_1) = f(x_2)$ 当且仅当 $x_1 \simeq x_2$，则称其为 $\mathfrak{G}$-轨道单射函数。由于 $f(\mathfrak{g} \cdot x) = f(x)$，此类函数是 $\mathfrak{G}$- 不变的。

我们研究作用于 n 个元素上的置换群 S_n，以及李群 $\mathfrak{G} = SO(d)$ 或 $\mathfrak{G} = O(d)$。通常通过相对位置来处理对平移 $T(d)$ 的不变性。对于上述任一群，其元素$\mathfrak{g}$ 的标准矩阵表示记作 $\boldsymbol{M}_{\mathfrak{g}}$（或其他大写字母）。

几何图。几何图 $\mathcal{G} = (\boldsymbol{A}, \boldsymbol{S}, \vec{\boldsymbol{V}}, \vec{\boldsymbol{X}})$ 是一种带有几何属性的属性图：节点坐标 $\vec{\boldsymbol{X}} \in \mathbb{R}^{n\times d}$ 和（可选的）向量特征 $\vec{\boldsymbol{V}} \in \mathbb{R}^{n\times d}$（例如 速度、加速度）。不失一般性，我们假设每个节点只有一个向量特征。我们的设定可推广到多个向量特征或高阶张量。几何属性在相关群的群作用下按如下方式变换：(1) S_n 通过 $\boldsymbol{P}_\sigma\mathcal{G}$ 作用于图：$= (\boldsymbol{P}_\sigma \boldsymbol{A} \boldsymbol{P}_\sigma^\top, \boldsymbol{P}_\sigma \boldsymbol{S}, \boldsymbol{P}_\sigma \vec{\boldsymbol{V}}, \boldsymbol{P}_\sigma \vec{\boldsymbol{X}})$；(2) 正交变换 $\vec{} \quad \vec{} \quad \vec{} \quad \vec{}$作用于 $\boldsymbol{V}, \boldsymbol{X}$，通过 $\boldsymbol{V}\boldsymbol{Q}_\mathfrak{g}$, $\boldsymbol{X}\boldsymbol{Q}_\mathfrak{g}$实现；(3) 平移 $\vec{\boldsymbol{t}} \in T(d)$对所有节点 i的坐标 $\vec{\boldsymbol{X}}$ 施加作用，变换为 $\vec{\boldsymbol{x}}_i + \vec{\boldsymbol{t}}$。

若存在属性图同构映射 b，使几何属性在全局群作用 $\boldsymbol{Q}_\mathfrak{g} \in \mathfrak{G}$ 和 $\vec{\boldsymbol{t}} \in T(d)$ 下等价，则称两个几何图 $\mathcal{G}$ 和 $\mathcal{H}$ 为几何同构的：

$$\left(\boldsymbol{s}_{b(i)}^{(\mathcal{H})},\ \boldsymbol{Q}_\mathfrak{g}\vec{\boldsymbol{v}}_{b(i)}^{(\mathcal{H})},\ \boldsymbol{Q}_\mathfrak{g}(\vec{\boldsymbol{x}}_{b(i)}^{(\mathcal{H})} + \vec{\boldsymbol{t}})\right) = \left(\boldsymbol{s}_i^{(\mathcal{G})}, \vec{\boldsymbol{v}}_i^{(\mathcal{G})}, \vec{\boldsymbol{x}}_i^{(\mathcal{G})}\right) \quad \text{for all } i \in \mathcal{V}(\mathcal{G}). \quad (2)$$

几何图同构以及区分（子）图的几何结构，对表示学习具有重要的实际影响。例如，例如，对于分子系统，理想的架构应将原子周围不同的局部结构环境映射为不同的表示（Bart´ok 等人，2013；Pozdnyakov 等人，2020年），随后可将这些表示用于下游预测任务。

几何图神经网络。我们考虑两大类几何图神经网络架构。(1) $\mathfrak{G}$- 等变图神经网络层（Thomas 等人，2018；Satorras 等人，2021）通过可学习的聚合与更新函数 AGG 和 UPD，分别将标量特征和几何向量特征从第 t 次迭代传播至第 $t+1$ 次迭代：

$$\boldsymbol{s}_i^{(t+1)}, \vec{\boldsymbol{v}}_i^{(t+1)} := \textsc{Upd}\Big((\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}),\ \textsc{Agg}\left(\{\!\{(\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}), (\boldsymbol{s}_j^{(t)}, \vec{\boldsymbol{v}}_j^{(t)}), \vec{\boldsymbol{x}}_{ij} \mid j \in \mathcal{N}_i\}\!\}\right)\Big), \quad (3)$$

其中 $\vec{\boldsymbol{x}}_{ij} = \vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j$ 表示相对位置向量。或者，(2) $\mathfrak{G}$-不变的 *GNN 层* (Sch¨utt 等人，2020年；Gasteiger 等人，2020年不更新向量特征，而仅传播利用局部邻域中的几何信息计算得到的不变标量：

$$\boldsymbol{s}_i^{(t+1)} := \textsc{Upd}\Big(\boldsymbol{s}_i^{(t)},\ \textsc{Agg}\left(\{\!\{(\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}), (\boldsymbol{s}_j^{(t)}, \vec{\boldsymbol{v}}_j^{(t)}), \vec{\boldsymbol{x}}_{ij} \mid j \in \mathcal{N}_i\}\!\}\right)\Big). \quad (4)$$

对于这两种模型，最终迭代中的标量特征通过置换不变的读出函数映射为图级预测 $f\colon \mathbb{R}^{n\times f} \to \mathbb{R}^{f'}$。**在附录B中，我们给出了本文框架所涵盖的几何图神经网络的统一方程和示例。**

3. 几何 Weisfeiler-Leman 测试

假设。与 WL 测试类似，本节余下部分假设我们要区分的所有几何图都是有限大小的，并且来自一个可数数据集。换言之，这些节点所具有的几何特征和标量特征分别来自可数子集$C \subset \mathbb{R}^d$ 和 $C' \subset \mathbb{R}$。因此，当我们要求函数为单射时，要求它们在通过 $\mathfrak{G}$作用于数据集而得到的 $\mathfrak{G}$-轨道这一可数集合上为单射。

直观理解。为了直观理解如何将 WL 推广到几何图，我们注意到，WL 使用一种局部的、以节点为中心的过程，利用每个节点 i 的 1 跳 邻域 $\mathcal{N}_i$的颜色来更新其颜色 $\mathcal{N}_i$，其中 i是中心节点周围的带属性点云 i。因此，每个邻域包含两类信息：(1) 邻域类型（对 $\mathfrak{G}$ 不变）和 (2) 邻域几何朝向（对 $\mathfrak{G}$等变）。这对公式 1 中邻域聚合过程的推广提出了约束。从公理化视角来看，我们对 WL 聚合过程的推广必须满足以下两个性质：

性质 1：颜色的轨道单射性。如果两个邻域在 $\mathfrak{G}$ 的某种群作用（例如旋转）下相同，那么相应中心节点的颜色应当相同。因此，在数据集中所有邻域轨道组成的可数集合上，着色必须是 $\mathfrak{G}$-轨道单射的，同时也是 $\mathfrak{G}$- 不变的。

性质 2：局部几何保持。WL 的一个关键性质是聚合具有单射性。仅满足性质 1 的 $\mathfrak{G}$<style id='6}-不变着色过程并不充分，因为根据定义，它会丢失每个邻域的空间属性，例如相对位姿或方向（Hinton 等人，2011年）。因此，我们还必须以一种$\mathfrak{G}$-等变且具有单射性的方式，更新辅助几何信息变量。

几何 Weisfeiler-Leman（GWL）。这些直觉促成了下面对 GWL 测试的定义。在初始化时，我们为每个节点 $i \in \mathcal{V}$ 分配一个标量节点颜色$c_i \in C'$ 以及一个辅助对象 $\boldsymbol{g}_i$，其中包含与该节点相关联的几何信息：

$$c_i^{(0)} := \textsc{Hash}(\boldsymbol{s}_i), \quad \boldsymbol{g}_i^{(0)} := \left(c_i^{(0)}, \vec{\boldsymbol{v}}_i\right), \quad (5)$$

其中，HASH 表示作用于节点 i的标量属性 $\boldsymbol{s}_i$ 上的单射映射，例如，标准 WL 中的 HASH。为定义归纳步骤，假设在迭代 $t-1$时，我们已获得节点的颜色及其对应的几何对象。然后，我们可以将节点 i 周围的几何信息聚合为一个新对象，如下所示：

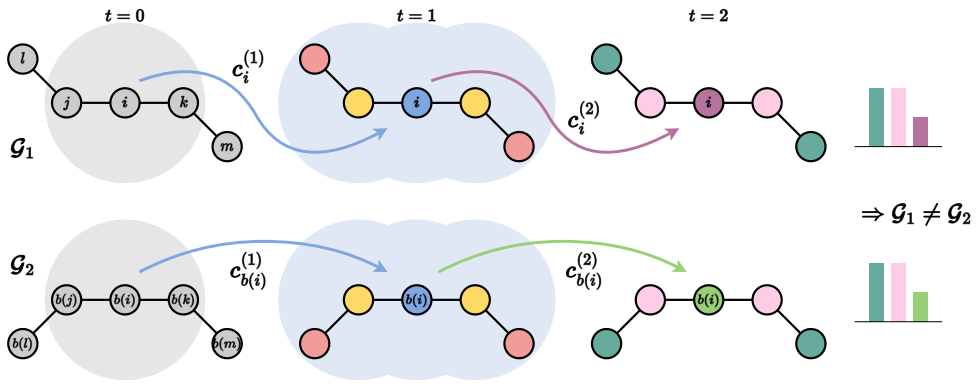


图 2：**几何 Weisfeiler-Leman 测试**。GWL 通过对不同的邻域模式进行单射颜色分配，在全局对称性意义下区分非同构几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。每次迭代都会扩展可收集几何信息的邻域范围（节点 i处以阴影标示）。示例受 Schütt 等人启发。(2021)， $\mathfrak{G} = O(d)$。

$$\boldsymbol{g}_i^{(t)} := \Big((c_i^{(t-1)}, \boldsymbol{g}_i^{(t-1)}),\ \{\!\{(c_j^{(t-1)}, \boldsymbol{g}_j^{(t-1)}, \vec{\boldsymbol{x}}_{ij}) \mid j \in \mathcal{N}_i\}\!\}\Big), \quad (6)$$

这里，$\{\!\{\cdot\}\!\}$ 表示一个多重集——一种元素可以重复出现的集合。重要的是，群 $\mathfrak{G}$ 可以通过作用于几何对象内部的几何信息，归纳地作用于上述几何对象。这相当于旋转（或反射）其中包含的整个 t跳邻域：

$$\mathfrak{g} \cdot \boldsymbol{g}_i^{(0)} := \Big(c_i^{(0)},\ \boldsymbol{Q}_{\mathfrak{g}} \vec{\boldsymbol{v}}_i\Big), \quad (7)$$

$$\mathfrak{g} \cdot \boldsymbol{g}_i^{(t)} := \Big((c_i^{(t-1)}, \mathfrak{g} \cdot \boldsymbol{g}_i^{(t-1)}),\ \{\!\{(c_j^{(t-1)},\ \mathfrak{g} \cdot \boldsymbol{g}_j^{(t-1)},\ \boldsymbol{Q}_{\mathfrak{g}} \vec{\boldsymbol{x}}_{ij}) \mid j \in \mathcal{N}_i\}\!\}\Big) \quad (8)$$

显然，对于任意时间步 t，构建聚合 $\boldsymbol{g}_i$ 都是单射且具有 $\mathfrak{G}$-等变性。最后，我们可以通过聚合节点 i 周围邻域中的几何信息，计算所有 $i \in \mathcal{V}$ 在迭代 t 时的节点颜色：

$$c_i^{(t)} := \text{I-HASH}^{(t)}\left(\boldsymbol{g}_i^{(t)}\right), \quad (9)$$

通过使用一个 $\mathfrak{G}$-轨道单射且 $\mathfrak{G}$-不变的函数，我们将其记为 I-HASH。也就是说，对于任意几何对象 $\boldsymbol{g}$，$\boldsymbol{g}'$，当且仅当存在 $\mathfrak{g} \in \mathfrak{G}$ 使得 $\boldsymbol{g} = \mathfrak{g} \cdot \boldsymbol{g}'$ 时，I-HASH($\boldsymbol{g}$) = I-HASH($\boldsymbol{g}'$)。注意，I-HASH 是一种理想化的 $\mathfrak{G}$-轨道单射函数，类似于 WL 中使用的 HASH 函数，而后者不一定是连续的。

概述。每次迭代中，$\boldsymbol{g}_i^{(t)}$ 都会聚合节点 i周围逐渐扩大的 t-跳子图邻域 $\mathcal{N}_i^{(t)}$ 中的几何信息。节点颜色概括了这些 t-跳邻域的结构，这通过 I-HASH 执行的 $\mathfrak{G}$-不变聚合 $\mathfrak{G}$来实现。当由颜色诱导的节点划分与上一次迭代相比不再变化时，该过程终止。最后，给定两个几何图 $\mathcal{G}$和 $\mathcal{H}$，如果存在某次迭代 t，使得 $\{\!\{c_i^{(t)} \mid i \in \mathcal{V}(\mathcal{G})\}\!\} \neq \{\!\{c_i^{(t)} \mid i \in \mathcal{V}(\mathcal{H})\}\!\}$，那么 GWL 会判定这两个图在几何上非同构。否则，GWL 无法区分它们。

不变 GWL。由于我们关注的是理解 $\mathfrak{G}$-等变性的作用，因此我们还考虑一种限制更严格的不变 GWL（IGWL）：它仅使用 $\mathfrak{G}$-轨道单射 I-HASH 函数更新节点颜色，且不传播几何信息：

$$c_i^{(t)} := \text{I-HASH}\Big((c_i^{(t-1)}, \vec{\boldsymbol{v}}_i),\ \{\!\{(c_j^{(t-1)}, \vec{\boldsymbol{v}}_j,\ \vec{\boldsymbol{x}}_{ij}) \mid j \in \mathcal{N}_i\}\!\}\Big). \quad (10)$$

具有 k-体阶标量的 IGWL。为进一步分析节点着色函数 I-HASH 的构造，我们考虑基于计算 $^{(k)}$ -不变标量时涉及的节点最大数量的 IGWL（也称为“体阶”（Batatia 等人，2022a））：

$$c_i^{(t)} := \text{I-HASH}_{(k)}\Big((c_i^{(t-1)}, \vec{\boldsymbol{v}}_i),\ \{\!\{(c_j^{(t-1)}, \vec{\boldsymbol{v}}_j,\ \vec{\boldsymbol{x}}_{ij}) \mid j \in \mathcal{N}_i\}\!\}\Big), \quad (11)$$

I-HASH$^{(k+1)}$ 的定义如下：

$$\text{HASH}\Big(\{\!\{\text{I-HASH}\Big((c_i^{(t-1)}, \vec{\boldsymbol{v}}_i), \{\!\{(c_{j_1}^{(t-1)}, \vec{\boldsymbol{v}}_{j_1},\ \vec{\boldsymbol{x}}_{ij_1}), \ldots, (c_{j_k}^{(t-1)}, \vec{\boldsymbol{v}}_{j_k},\ \vec{\boldsymbol{x}}_{ij_k})\}\!\}\Big) \mid \boldsymbol{j} \in (\mathcal{N}_i)^k\}\!\}\Big), \quad (12)$$

其中 $\boldsymbol{j} = [j_1, \ldots, j_k]$ 是由 $\mathcal{N}_i$ 中的元素组成的所有可能的 k 元组。因此，IGWL$^{(k)}$ 现在仅限于从邻域中所有可能的、大小为 k 的节点元组（包括中心节点）中提取信息。例如，I-HASH$^{(2)}$ 只能根据中心节点与其任一邻居之间的两两距离来识别邻域（即，一个二体标量），而 I-HASH$^{(3)}$ 则能识别到

任意两条边之间的距离和角度（即一个3体标量）。值得注意的是，仅凭距离和角度不足以完整描述局部几何，因为存在若干反例，需要借助高阶体序不变量才能加以区分（Bart´ok 等人，2013；Pozdnyakov 等人，2020年）。因此，标量化体序较低的 I-HASH$_{(k)}$ k 会使着色能力减弱。

3.1. 刻画几何图神经网络的表达能力

遵循 Xu 等人 (2019);Morris 等人 (2019), 我们通过 GWL 测试给出几何图神经网络表达能力的上界。我们证明，任何 $\mathfrak{G}$-等变 GNN 在区分非同构几何图方面的能力至多与 GWL 测试相当。证明见附录 E。

定理 1. 任何一对可由 $\mathfrak{G}$-等变 *GNN* 区分的几何图，也可由 *GWL* 区分。

经过足够多次迭代后，若聚合函数、更新函数和读出函数满足某些条件，$\mathfrak{G}$-等变图神经网络的输出可与 GWL 等价。

命题 2. $\mathfrak{G}$-等变图神经网络在满足以下条件时具有与 *GWL* 相同的表达能力：*(1)* 聚合 AGG 是一个单射的、 $\mathfrak{G}$-等变的多重集函数。*(2)* 更新函数 UPD$_S$ 的标量部分是一个 $\mathfrak{G}$-轨道单射的、 $\mathfrak{G}$-不变的多重集函数。*(3)* 更新函数 UPD$_v$ 的向量部分是一个单射的、 $\mathfrak{G}$-等变的多重集函数。*(4)* 图级读出 f 是一个单射的多重集函数。

对于 $\mathfrak{G}$-不变图神经网络和 IGWL，也可以得出类似的结论。因此，我们可以直接在 GWL/IGWL（抽象数学工具）与相应测试所限定的几何图神经网络类别之间迁移理论结果。这种联系具有若干有趣的实际影响，后文将对此进行讨论。

4. 通过 GWL 理解几何图神经网络的设计空间

概述。 我们展示了 GWL 框架对于理解几何图神经网络设计选择（Batatia 等人，2022a）如何影响表达能力的作用：（1）深度或层数；（2）不变消息传递与等变消息传递；以及（3）标量化体序。为此，我们形式化了当前架构的理论局限，并提出了设计最大表达能力模型的实际建议。证明见附录 D。

4.1. 深度的作用：传播几何信息

考虑两个几何图 $\mathcal{G}_1 = (\boldsymbol{A}_1, \boldsymbol{S}_1, \vec{\boldsymbol{V}}_1, \vec{\boldsymbol{X}}_1)$ 和 $\mathcal{G}_2 = (\boldsymbol{A}_2, \boldsymbol{S}_2, \vec{\boldsymbol{V}}_2, \vec{\boldsymbol{X}}_2)$ 的简化情形，其中底层属性图 $(\boldsymbol{A}_1, \boldsymbol{S}_1)$ 和 $(\boldsymbol{A}_2, \boldsymbol{S}_2)$ 同构。这种情况在化学中十分常见：分子会呈现不同的构象，但具有相同的图拓扑，这由共价键合结构决定。

GWL 的每次迭代都会从节点 i周围逐步扩大的邻域 $\mathcal{N}_i^{(k)}$ 中聚合几何信息$\boldsymbol{g}_i^{(k)}$，并通过比较 $\boldsymbol{g}_i^{(k)}$ 的$\mathfrak{G}$-轨道单射着色来区分（子）图。如果对于所有图同构 b, 都存在某个节点 $i \in \mathcal{V}_1$，$b\,(i) \in \mathcal{V}_2$，使得相应的 k-跳子图 $\mathcal{N}_i^{(k)}$ 和 $\mathcal{N}_{b\,i}^{(k)}{}_{()}$ 彼此不同，我们就称 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是相差 跳 的。否则，如果所有相应的 k-跳子图 $\mathcal{N}_i^{(k)}$ 和 $\mathcal{N}_{(k)\,b(i)}$ 在群作用下均相同，我们就称$\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是 k 相同的 跳 。现在，我们可以根据迭代次数，形式化说明 GWL 和表达能力最强的几何图神经网络能够区分和不能区分哪些几何图。

命题3. *GWL* 可以区分任意 k跳不同的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，其中底层属性图同构，且 k 次迭代就足够。

命题4. 经过最多 k 次迭代，*GWL* 无法区分任意 k跳相同的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，其中底层属性图同构。

此外，我们还可以给出关于约束更严格的 IGWL 的以下结果。

命题5. 对于任意1跳不同的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，若其底层属性图同构，则 *IGWL* 可以将其区分，且 *1* 次迭代就足够。

命题6. 无论进行多少次迭代，*IGWL* 都无法区分任意1跳相同的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，其中底层属性图同构。

命题3 和6 的示例如图2 和3所示。

现在我们可以考虑更一般的情形：$\mathcal{G}_1 = (\boldsymbol{A}_1, \boldsymbol{S}_1, \vec{\boldsymbol{V}}_1, \vec{\boldsymbol{X}}_1)$ 和 $\mathcal{G}_2 = (\boldsymbol{A}_2, \boldsymbol{S}_2, \vec{\boldsymbol{V}}_2, \vec{\boldsymbol{X}}_2)$ 的底层属性图并不同构，并且是从点云通过径向截断构造而成的，这种做法在生物化学和材料科学应用中很常见。

命题7。 假设几何图是从点云通过径向截断构造而成的，那么 *GWL* 可以区分任意底层属性图不同构的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。至多需要 k_{Max} 次迭代，其中 k_{Max} 是 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 中较大的图直径。

4.2. 不变消息传递的局限性：无法捕捉全局几何

命题 3 和6 使我们能够比较 GWL 和 IGWL 的表达能力。

定理 8。 *GWL* 的表达能力严格强于 *IGWL*。

这一论述形式化地阐明了 $\mathfrak{G}$-等变中间层在几何图中的优势，这正是几何深度学习蓝图（Bronstein 等人，2021）所倡导的，同时也呼应了计算机视觉领域的类似直觉。正如Hinton 等人（2011年）所指出的，平移不变模型无法理解图像各个部分之间的关系（这一现象被称为“毕加索问题”）。类似地，我们的结果表明，IGWL 和 $\mathfrak{G}$-不变图神经网络无法理解图中各个 1 跳邻域彼此之间的朝向。

因此，即使是最强大的 $\mathfrak{G}$-不变图神经网络，其计算全局和非局部几何性质的能力也受到限制。

命题 9. *IGWL* 和 $\mathfrak{G}$-不变图神经网络无法判定若干几何图性质：*(1)* 包围几何图的包围盒/球体的周长、表面积和体积；*(2)* 到质心或质心的距离；以及 *(3)* 二面角。

最后，我们找到了一个两种方法之间的这种区别不复存在的情形。

命题10. 对于全连接几何图，*IGWL* 与 *GWL* 具有相同的表达能力。

实际影响。 命题9 揭示了 $\mathfrak{G}$-不变图神经网络的关键理论局限。这表明，在处理大型几何图（例如包含数千个节点的大分子）时，应优先采用 $\mathfrak{G}$-等变图神经网络，因为此类图中的消息传递仅限于每个节点周围的局部径向邻域。堆叠多个 $\mathfrak{G}$-等变层可以计算组合几何特征。

克服 $\mathfrak{G}$-不变 GNN 局限性的两种直接方法可能是：(1) 预先计算非局部几何性质，作为输入特征，例如，GemNet（Gasteiger 等人，2021）和 ComENet（Wang 等人，2022）等模型使用两跳二面角。以及 (2) 使用全连接几何图，正如命题10 所示，在执行全对全消息传递时， $\mathfrak{G}$-等变和 $\mathfrak{G}$-不变 GNN 可以具有同等的表达能力。近期 $\mathfrak{G}$-不变图变换器（Joshi，2020；Shi 等人，2022）在含有数十个节点的小分子上的经验性成功，为此提供了支持；对于这类分子，使用完整图是可行的。

4.3. 标量化体序的作用：识别邻域 $\mathfrak{G}$-轨道

在 GWL 和 IGWL 的每次迭代中，I-HASH 函数会为不同的几何邻域模式分配一个 $\mathfrak{G}$-不变的颜色。I-HASH 是一个理想化的 $\mathfrak{G}$-轨道单射函数，但不一定是连续的。在几何图神经网络中，这对应于在更新标量特征时将局部几何信息进行标量化；示例为如方程 14 和方程 15所示。我们可以通过ASH 来分析 I-H k-主体变体 IGWL$^{(k)}$中 I-HASH 函数的构造和几何图神经网络中的标量化步骤，该变体在第 3节中进行了描述。在此过程中，我们将建立 IGWL 与非几何图上的 WL 之间的联系。

首先，我们形式化 I-HASH$^{(k)}$ 的单射性与给定数据集中局部邻域最大基数之间的关系。

命题 11. I-HASH$^{(m)}$ 对于 $m = max(\{|\mathcal{N}_i| \mid i \in \mathcal{V}\})$是 $\mathfrak{G}$-轨道单射的，其中 $m = max(\{|\mathcal{N}_i| \mid i \in \mathcal{V}\})$表示给定数据集中所有局部邻域的最大基数。

实际影响。尽管构建可证明为单射的 I-HASH$^{(k)}$ 函数可能需要高得难以处理的 k，但 IGWL$^{(k)}$ 测试的层级使我们能够研究当前几何图神经网络层中所用实用 $\mathfrak{G}$-不变聚合器的表达能力。例如，SchNet（Sch¨utt 等人，2018）、E-GNN（Satorras 等，2021）和 TFN（Thomas 等，2018）使用距离，而 DimeNet（Gasteiger 等人，2020年）使用距离和角度。值得注意的是，MACE（Batatia 等人，2022年b）通过原子簇展开（Dusson 等人，2019）构造了任意体阶 k 下的完整标量基；若满足命题 11 中的条件，该基可以是 $\mathfrak{G}$-轨道单射的。关于 IGWL$^{(k)}$ 层级及相应的 GNN，我们可以得出以下结论。

命题 12。 *IGWL*$^{(k)}$ 至少与 *IGWL*$^{(k-1)}$一样强。对于 $k \leq 5$，*IGWL*$^{(k)}$ 严格强于 *IGWL*$^{(k-1)}$。

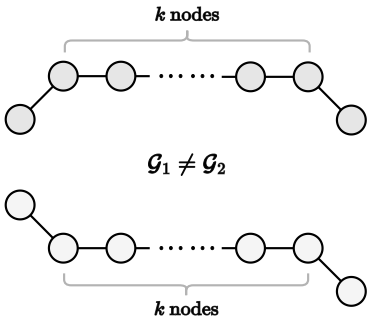
最后，我们证明当所有两两距离都相同时，IGWL$^{(2)}$ 与 Weisfeiler–Leman 等价。最近，Pozdnyakov 和 Ceriotti (2022）也提出了类似的观察。

命题 13。 设 $\mathcal{G}_1 = (\boldsymbol{A}_1, \boldsymbol{S}_1, \vec{\boldsymbol{X}}_1)$ 和 $\mathcal{G}_2 = (\boldsymbol{A}_2, \boldsymbol{S}_2, \vec{\boldsymbol{X}}_2)$ 是两个几何图，且所有边等长。那么，*IGWL*$^{(2)}$ 能够区分这两个图，当且仅当 *Weisfeiler–Leman* 能够区分属性图 $(\boldsymbol{A}_1, \boldsymbol{S}_1)$ 和 $(\boldsymbol{A}_1, \boldsymbol{S}_1)$。

这种等价性揭示了基于距离的 $\mathfrak{G}$不变模型（如 SchNet（Sch¨utt 等人，2018）的局限性。这些模型会遇到 Weisfeiler–Leman 的所有已知失效情形，例如，它们无法区分两个等边三角形和一个正六边形（Gasteiger 等人，2020年）。

5. 表达能力的合成实验

概述。 我们提供合成实验来补充理论结果，并凸显构建能力最强的几何图神经网络时面临的实际挑战，例如 随着深度增加，几何信息会过度压缩（表 1），以及在以下网络中使用高阶张量的必要性$\mathfrak{G}$-等变 GNN（表 2）。我们希望随附的



	（$k=4$-链）图神经网络层	层数 $\lfloor \frac{k}{2} \rfloor$	$\lfloor \frac{k}{2} \rfloor + 1 = 3$	$\lfloor \frac{k}{2} \rfloor + 2$	$\lfloor \frac{k}{2} \rfloor + 3$	$\lfloor \frac{k}{2} \rfloor + 4$
Equiv.	GWL	50%	**100%**	**100%**	**100%**	**100%**
	E-GNN	50。0 ± 0.0	50.0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0	**100.0 ± 0.0**
	GVP-GNN	50.0 ± 0.0	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**
	TFN	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	**80.0 ± 24.5**	**85.0 ± 22.9**
	MACE	50。0 ± 0.0	**90.0 ± 20.0**	**90.0 ± 20.0**	**95.0 ± 15.0**	**95.0 ± 15.0**
Inv.	IGWL	50%	50%	50%	50%	50%
	SchNet	50。0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0
	DimeNet	50。0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	SphereNet	50。0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50。0 ± 0.0	50。0 ± 0.0
	SchNet 完整图	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**
	SchNet 全局特征	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**

表 1： k 链几何图。 k 链是 $(\lfloor \frac{k}{2} \rfloor + 1)$ 跳可区分的，而 $(\lfloor \frac{k}{2} \rfloor + 1)$ 次 GWL 迭代理论上足以区分它们。我们训练具有递增层数的几何图神经网络来区分$k = 4$ 链。 **$\mathfrak{G}$ 等变图神经网络可能需要比 GWL 指定的更多迭代，这初步表明，当几何信息通过固定维度的特征空间跨多层传播时会出现过度压缩。** IGWL 和 $\mathfrak{G}$ 不变图神经网络无法区分 k 链，无论 $k \geq 2$ 和 $\mathfrak{G} = O(3)$ 如何。 $\mathfrak{G}$-不变图神经网络使用预先计算的非局部特征（包围盒体积），或在全连接图上进行消息传递，都可以轻易完成此任务。异常结果以红色标示，预期结果以绿色标示。

代码库 `geometric-gnn-dojo`[1]，能成为一种教学资源，以有原则的方式推动几何 GNN 架构的发展。实验设置和超参数的详情请参见附录 C 。

5.1. 表 1 中的实验1：深度、非局部几何性质与过度压缩

GWL 假设每次迭代都能完美传播 $\mathfrak{G}$-等变几何信息，这意味着测试可以在任意迭代次数下运行而不丢失信息。在几何图神经网络中， $\mathfrak{G}$-等变信息通过在固定维度的空间中对多个层的特征求和来传播，这可能导致来自远处节点的信息发生失真或丢失。

实验。 为了研究深度在局部邻域之外传播几何信息的实际影响，我们考虑 k-链几何图，它们推广了以下文献中的示例 Sch¨utt 等人 (2021)。每对 k-链由 $k+2$ 个节点组成，其中 k 个节点排成一条直线，并根据 2 个端点的方向加以区分。因此， k-链图可在 $(\lfloor \frac{k}{2} \rfloor + 1)$ 跳内区分，而理论上，$(\lfloor \frac{k}{2} \rfloor + 1)$ 次 GWL 迭代足以将它们区分开来。在表 1中，我们训练$\mathfrak{G}$-等变和 $\mathfrak{G}$-不变图神经网络，并逐步增加层数，以区分 k-链。

结果。 尽管这项任务看似简单，我们发现，常见的 $\mathfrak{G}$-等变图神经网络，例如 E-GNN (Satorras 等人，2021) 和 TFN (Thomas 等人，2018) 可能需要比 GWL 规定的更多迭代次数。值得注意的是，对于长度大于 $k = 4$ 的链，所有 $\mathfrak{G}$-等变图神经网络往往都需要超过 $(\lfloor \frac{k}{2} \rfloor + 1)$ 次迭代才能解决该任务。此外，IGWL 和 $\mathfrak{G}$-不变图神经网络无法区分 k-链。

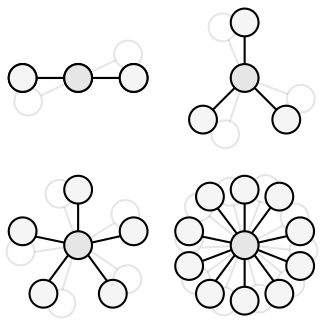
表 1 为几何图神经网络中的等变特征存在过度压缩现象提供了初步证据（Alon& Yahav，2021；Topping et al.，2022）。对于 E-GNN，这个问题尤为明显，因为它使用单个向量特征来聚合和传播几何信息。这可能会影响大分子的建模，因为长程相互作用往往发挥着重要作用。

5.2. 表 2中的实验：高阶张量与旋转对称结构

除了完美传播外，GWL 还能够利用辅助的嵌套几何对象 $\boldsymbol{g}_i$，对 $\mathfrak{G}$等变信息进行单射聚合。另一方面，$\mathfrak{G}$ 等变图神经网络通过对以笛卡尔向量（张量阶为 1）或更高阶球张量表示的邻域特征求和，来聚合几何信息。这种基的选择往往需要在计算可处理性与经验性能之间进行权衡。

实验。为了展示高阶张量在 $\mathfrak{G}$-等变 GNN 中的作用，我们研究旋转对称性与张量阶之间的相互作用。在表 2中，我们评估当前 $\mathfrak{G}$-等变层区分具有旋转对称性的结构方向的能力。一个

[1]GitHub 上提供的代码：
https://github.com/chaitjo/geometric-gnn-dojo



	图神经网络层	旋转对称性 二重	三重	五重	10 折
Cart.	E-GNN $L=1$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	GVP-GNN $L=1$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
Spherical	TFN/MACE $L=1$	50.0 ± 0.0	50。0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	TFN/MACE $L=2$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	TFN/MACE $L=3$	**100.0 ± 0.0**	**100.0 ± 0.0**	50。0 ± 0.0	50.0 ± 0.0
	TFN/MACE $L=5$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	50.0 ± 0.0
	TFN/MACE $L=10$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**

表2：旋转对称结构。我们训练单层 $\mathfrak{G}$ 等变 GNN，以区分每个 L 重旋转对称结构的两个不同旋转版本。**我们发现，使用阶数为 L 的张量的层，无法识别旋转对称性高于 L 重的结构的方向**。对于使用笛卡尔向量（张量阶为1）的层，这一问题尤为普遍。

L 重旋转对称结构绕某一点（二维）或某条轴（三维）旋转 $\frac{2\pi}{L}$ 角度后保持不变。我们选取每种对称结构的两个不同的旋转版本，并训练单层 $\mathfrak{G}$等变图神经网络，利用更新后的等变特征对这两个方向进行分类。

结果。我们发现，使用阶数为 L 的球张量的层，无法识别旋转对称性高于 L 重的结构的方向，即输入的两个不同旋转版本具有相同的等变特征。我们认为，这一现象源于球谐函数：它们作为球张量特征的正交归一基，本身也具有旋转对称性。

与信号的傅里叶展开类似，$\mathfrak{G}$ 等变图神经网络使用球谐展开将笛卡尔向量转换为球面信号。球谐基的张量阶决定了球体上近似函数的振荡速率。我们认为，这一振荡速率与一组对称向量的旋转对称阶数密切相关。

在傅里叶展开中，仅使用低频正弦波来准确逼近高频函数并不可行。同样，当球谐展开截断到低于旋转对称性的阶数时，旋转对称的向量会表现为高频函数。因此，低频基无法保留这些向量的方向。

诸如 E-GNN（Satorras 等人，2021）和 GVP-GNN（Jing 等人，2020年）等使用笛卡尔向量的网络颇受欢迎，因为在许多应用中，高阶张量的计算可能难以处理。然而，E-GNN 和 GVP-GNN 在区分旋转对称结构的方向方面表现尤其不佳。这可能会影响周期性材料的建模，因为这类材料天然具有此类对称性（Levine 和 Steinhardt，1984年）。

6. 判别性与普适性

概述。 按照 Chen 等人 (2019) 的研究，我们探讨几何图神经网络（Dym 和 Maron，2020年；Villar 等人，2021）的通用逼近能力与 GWL 提出的几何图判别视角之间的等价性，并将其结果推广到由紧李群诱导的任意同构 $\mathfrak{G}$。我们进一步研究在连续情形下，区分任意两个邻域所需的不变聚合器的数量。证明见附录 F。

为保持一般性，我们使用一个由紧群 $\mathfrak{G}$ 作用于其上的一般空间 X，并关注 $\mathfrak{G}$-不变函数在 X 上区分 Y 中各点的能力。将范围限制在较小的子集 Y 中很有用，因为我们希望单独考虑 Y 可数的情况，这是由于使用了可数特征。因此，一般而言，$\mathfrak{G}$ 在 Y 上的群作用可能没有严格定义，因为它可能产生 Y 之外的元素。对于我们的设定，读者可以将 $X = (\mathbb{R}^d \times \mathbb{R}^f)^{n\times n}$ 视为几何图的空间，并将 $Y = \mathcal{X}^{n\times n}$ 视为该空间，其中 $\mathcal{X} \subseteq \mathbb{R}^d \times \mathbb{R}^f$。

定义 14. 设 $\mathfrak{G}$ 为一个紧群，$\mathcal{C}$ 是从集合 X 到 $\mathbb{R}$的一组 $\mathfrak{G}$-不变函数。对于一个子集$Y \subseteq X$，如果对于任意满足 $y_1 \not\simeq y_2$的 $y_1, y_2 \in Y$，都存在一个函数$h \in \mathcal{C}$，使得 $h(y_1) \neq h(y_2)$，则称 $\mathcal{C}$ 是两两$\mathbf{Y}_{\mathfrak{G}}$ 判别的。

这里需要指出，h 不一定是单射，即可能存在 y_1', y_2'，使得 $h(y_1') = h(y_1')$。因此，两两判别是比 GWL 所采用的判别概念更弱的概念。

定义 15. 设 $\mathfrak{G}$ 是一个紧群，$\mathcal{C}$ 是一族从 X 到 $\mathbb{R}$的 $\mathfrak{G}$-不变函数。对于 $Y \subseteq X$，我们称该 $\mathcal{C}$ 在 Y 上是普适逼近的，如果对于所有从 X 到 $\mathbb{R}$ 的 $\mathfrak{G}$-不变函数 f，以及所有 $\epsilon > 0$，都存在 $h_{\varepsilon,f} \in \mathcal{C}$，使得 $\|f - h_{\varepsilon,f}\|_Y$:
$=$上确界$_{y\in Y} |f(y) - h(y)| < \varepsilon$。

可数特征。我们首先关注 GWL 测试所处的可数特征设定。因此，我们假设 Y 是 X 的一个可数子集。

定理 16. 如果 $\mathcal{C}$ 在 Y 上具有普适逼近能力，那么 $\mathcal{C}$ 也能成对地区分 $Y_{\mathfrak{G}}$。

这一结果进一步说明了区分几何图的重要性，因为无法区分两个非同构几何图的模型并非普适模型。进一步假设 Y 是有限的，我们可以得到一个方向相反的结果。给定一组函数 $\mathcal{C}$，我们仿照 Chen 等人(2019) 的做法，定义函数类 $\mathcal{C}^{+L}$，该类由所有形如 MLP($[f_1(x)\dots, f_k(x)]$) 的函数组成，其中 $f_i \in \mathcal{C}$，且 k 有限；MLP 具有 L 层，并采用 ReLU 隐藏层激活函数。

定理 17. 如果 $\mathcal{C}$ 能成对地区分 $Y_{\mathfrak{G}}$，那么 $\mathcal{C}^{+2}$ 在 Y 上具有普适逼近能力。

连续特征。几何图所具有的对称性自然是连续的（例如，旋转）。因此，自然会问上述结果如何推广到连续子空间 Y 上的连续 $\mathfrak{G}$-不变函数。因此，在本节余下部分，我们假设 (X, d) 是一个度量空间，Y 是 X 的紧子集，且 $\mathfrak{G}$ 在 X 上连续作用。

定理 18. 如果 $\mathcal{C}$ 能够对 Y 上任意连续的 $\mathfrak{G}$-不变函数进行通用逼近，那么 $\mathcal{C}$ 也具有成对 $Y_{\mathfrak{G}}$判别能力。

在附加假设 $X = Y$ 的条件下，我们也可以证明逆命题。

定理 19. 如果 $\mathcal{C}$（一类定义在 Y 上的函数）具有成对$Y_{\mathfrak{G}}$ 判别能力，那么 $\mathcal{C}^{+2}$ 也能够对定义在 Y 上的任意连续函数进行通用逼近。

现在，我们基于微分几何的结果（Lee，2013），给出学习流形 X 上的连续轨道空间单射函数所需聚合器数量的估计。若群 $\mathfrak{G}$ 在 X 上自由作用，则 $\mathfrak{g}x = x$ 意味着$\mathfrak{g} = e_{\mathfrak{G}}$，其中 $e_{\mathfrak{G}}$ 是 $\mathfrak{G}$中的单位元。

定理 20. 设 X 是一个光滑的 n 维流形，$\mathfrak{G}$ 是一个 m 维紧李群，并连续作用于 X。假设存在一个与 X 维数相同的光滑子流形 Y，使得 $\mathfrak{G}$ 在其上自由作用。那么，任意一个$\mathfrak{G}$-轨道单射函数 $f\colon X \to \mathbb{R}^d$ 都要求 $d \geq n - m$。

现在，我们将此定理应用于几何图神经网络执行的局部聚合操作。令 $X = \mathbb{R}^{n\times d}$ and$\mathfrak{G} = S_n \times O(d)$ 或 $S_n \times SO(d)$。令 $\boldsymbol{P}_{\mathfrak{g}}$ and $\boldsymbol{Q}_{\mathfrak{g}}$ 分别为与群元素 $\mathfrak{g} \in \mathfrak{G}$关联的置换矩阵和正交矩阵。则 $\mathfrak{g}$ 通过 $\boldsymbol{P}_{\mathfrak{g}}\boldsymbol{X}\boldsymbol{Q}_{\mathfrak{g}}^{\top}$连续地作用于矩阵$\boldsymbol{X} \in X$。因此，$\mathfrak{G}$ 在 X 上的轨道空间单射函数，是定义在大小为 n 的点云上的函数，能够区分任意两个不同的点云。

定理 21。对于 $n \geq d - 1 > 0$ 或 $n = d = 1$，任何连续的 $S_n \times SO(d)$ 轨道空间单射函数 f:$\mathbb{R}^{n\times d} \to \mathbb{R}^q$ 都要求 $q \geq nd - d(d-1)/2$。

我们也可以将其推广到 $O(d)$，但需要稍强的假设，即 $n \geq d$。

定理 22. 对于 $n \geq d > 0$，任何连续的 $S_n \times O(d)$轨道空间单射函数 $f\colon \mathbb{R}^{n\times d} \to \mathbb{R}^q$ 都要求$q \geq nd - d(d-1)/2$。

总体而言，这些结果表明，在 $\mathbb{R}^3$ 中处理点云时（这在分子或物理应用中很常见），至少需要 $q = 3(n-1)$ 个聚合器。这一结果说明了为何更大的表示宽度有助于区分邻域。最后，在零维子群 $S_n \times \{e_{\mathrm{SO(d)}}\} \simeq S_n$这一特例下，我们得到了一个对所有 n 都成立的结论，并推广了 Corso 等人 (2020年) 关于非几何图神经网络聚合器的结果。原始的 PNA 结果考虑的是 $d = 1$这一情形，而我们在此将其推广到任意 d。

命题 23. 任何 S_n-不变的单射函数 $f\colon \mathbb{R}^{n\times d} \to \mathbb{R}^q$ 都需要 $q \geq nd$。

7. 结论

我们提出了几何 Weisfeiler-Leman（GWL）框架，用于刻画几何图神经网络的表达能力。我们的工作填补了一项重要的研究空白，因为标准图神经网络及其相关理论工具并不适用于几何图。(1) 我们展示了 GWL 在理解关键设计选择如何影响几何图神经网络表达能力方面的作用；(2) 通过合成实验揭示构建高表达能力模型所面临的实际挑战；(3) 将 GWL 基于判别的视角与通用逼近联系起来。

未来工作。GWL 为研究几何图神经网络的理论极限提供了一种抽象框架。在实践中，构建能够满足命题 2条件且可证明具有强表达能力的模型颇具挑战，因为 GWL 依赖于完美的邻域聚合和着色函数。基于从 GWL 获得的认识，未来工作将致力于开发能力最强且实用的 几何图神经网络。

致谢。我们感谢 Andrew Blake、Challenger Mishra、Charles Harris、D´avid Kov´acs、Erik Thiede、Gabor Csanyi、Gregor Simm、Hannes St¨ark、Ilyes Batatia、Iulia Duta、Justin Tan、Limei Wang、Mario Geiger、Petar Veli˘ckovi´c、Ramon Vin˜as、Rob Hesselink、Soledad Villar、Weihua Hu 以及匿名审稿人提供的宝贵讨论。CKJ 获得了 A*STAR 新加坡国家科学奖学金（博士）的资助。SVM 获得了英国研究与创新署（UKRI）环境风险研究中的人工智能应用博士培训中心（EP/S022961/1）的资助。

参考文献

Alon、U. 和 Yahav、E. 图神经网络的瓶颈及其实际影响。发表于*ICLR*，2021。

Anderson, B.、Hy, T.S. 和 Kondor, R. Cormorant：协变分子神经网络。*NeurIPS*，2019。

Baek, M.、DiMaio, F.、Anishchenko, I.、Dauparas, J. 等。使用三轨神经网络准确预测蛋白质结构及其相互作用。*Science*，2021。

Bart´ok, A. P.、Kondor, R. 和 Cs´anyi, G. 论化学环境的表示。*PhysicalReviewB*，2013。

Batatia, I.、Batzner, S.、Kov´acs, D. P.、Musaelian, A.、Simm, G. N.、Drautz, R.、Ortner, C.、Kozinsky, B. 和 Cs´anyi, G. e (3) 等变原子中心原子间势的设计空间。*arXiv* 预印本，2022a。

Batatia, I., Kov´acs, D. P., Simm, G. N., Ortner, C. 和 Cs´anyi, G. Mace：用于快速、准确力场的高阶等变消息传递神经网络。载于*NeurIPS*，2022b。

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E. 和 Kozinsky, B. E (3)-等变图神经网络：用于高数据效率和准确的原子间势。《自然·通讯》，2022。

Bodnar, C., Frasca, F., Otter, N., Wang, Y., Lio, P., Montufar, G. F. 和 Bronstein, M. Weisfeiler和Lehman走向细胞复形：CW网络。*NeurIPS*，2021a。

Bodnar, C., Frasca, F., Wang, Y., Otter, N., Montufar, G. F., Lio, P., 和 Bronstein, M. Weisfeiler和Lehman走向拓扑：单纯形消息传递网络。发表于*ICML*，2021b。

Bogatskiy, A., Ganguly, S., Kipf, T., Kondor, R., Miller, D. W., Murnane, D., Offermann, J. T., Pettee, M., Shana- han, P., Shimmin, C., 等。面向物理学的对称群等变架构。*arXiv*预印本，2022。

Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. J., 和 Welling, M. 几何量和物理量可改进 E(3) 等变消息传递。发表于*ICLR*，2022。

Bredon, G. E. 拓扑与几何。Springer Science & Business Media，2013。

Bronstein, M. M., Bruna, J., Cohen, T. 和 Veli˘ckovi´c, P. 几何深度学习：网格、群、图、测地线与规范。*arXiv* 预印本，2021。

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., 等。开放催化剂 2020（oc20）数据集和社区挑战。*ACSCatalysis*，2021年。

Chen, Z., Villar, S., Chen, L., 和 Bruna, J. 关于图同构测试与基于 GNN 的函数逼近之间的等价性。*NeurIPS*，2019。Corso, G., Cavalleri, L., Beaini, D., Li`o, P., 和 Veli˘ckovi´c, P. 用于图网络的主邻域聚合。*NeurIPS*，2020年。Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I., Courbet, A., de Haas, R. J., Bethel, N., 等。使用 ProteinMPNN 进行基于稳健深度学习的蛋白质序列设计。*Science*，2022年。

Drautz, R. 用于精确且可迁移原子间势的原子簇展开。*Physical ReviewB*，2019。

Du, W., Zhang, H., Du, Y., Meng, Q., Chen, W., Zheng, N., Shao, B. 和 Liu, T.-Y. 基于完备局部标架的 Se (3) 等变图神经网络。载于*ICML*，2022。

Dusson, G., Bachmayr, M., Csanyi, G., Drautz, R., Etter, S., van der Oord, C. 和 Ortner, C. 原子簇展开：完备性、效率与稳定性。*arXiv* 预印本，2019。

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Li`o, P., Bengio, Y., and Bronstein, M.《几何图神经网络指南：面向三维原子系统》。*arXiv* 预印本，2023。

Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., andBresson, X.《图神经网络基准测试》。*JMLR*，2023年。

Dym, N. and Maron, H.《旋转等变点云网络的普适性》。发表于*ICLR*，2020年。

Fey, M. and Lenssen, J. E.《使用 PyTorch Geometric 进行快速图表示学习》。发表于*ICLR Workshop*，2019。

Fuchs, F., Worrall, D., Fischer, V., and Welling, M.《SE(3)-Transformer：三维旋转-平移等变注意力网络》。*NeurIPS*，2020年。

Garg, V., Jegelka, S., and Jaakkola, T. 图神经网络的泛化与表征能力极限。发表于*ICML*，2020年。

Gasteiger, J., Groß, J., and G¨unnemann, S. 面向分子图的方向性消息传递。发表于*ICLR*，2020年。

Gasteiger, J., Becker, F., and G¨unnemann, S. Gemnet：面向分子的通用方向图神经网络。发表于*NeurIPS*，2021。

Geiger, M. 和 Smidt, T. e3nn：欧几里得神经网络。*arXiv* 预印本，2022。

Gilmer, J.、Schoenholz, S. S.、Riley, P. F.、Vinyals, O. 和 Dahl, G. E. 用于量子化学的神经消息传递。在*ICML*，2017。

Hinton, G., Krizhevsky, A., 和 Wang, S. D. 变换自编码器。载于*ICANN*，2011年。

Jamasb, A. R., Torn´e, R. V., Ma, E. J., Du, Y., Harris, C., Huang, K., Hall, D., Lio, P., 和 Blundell, T. L. Graphein——用于生物分子结构与相互作用网络分析的几何深度学习 Python 库。载于*NeurIPS*，2022。

Jing, B., Eismann, S., Suriana, P., Townshend, R. J. L., 和 Dror, R. 利用几何向量感知器从蛋白质结构中学习。载于*ICLR*，2020年。

Joshi, C. Transformer 是图神经网络。*TheGradient*，2020年。Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. 等。借助 AlphaFold 实现高精度蛋白质结构预测。《自然》，2021年。Lee, J. M. 光滑流形。载于《光滑流形导论》，第540–555页。Springer，2013。Levine, D. 和 Steinhardt, P. J. 准晶体：一类新的有序结构。《物理评论快报》，1984年。Li, Q., Gama, F., Ribeiro, A., 和 Prorok, A. 用于去中心化多机器人路径规划的图神经网络。载于*IROS*，2020年。Liu, Y., Wang, L., Liu, M., Lin, Y., Zhang, X., Oztekin, B., 和 Ji, S. 面向三维分子图的球面消息传递。载于*ICLR*，2022。Maron, H., Ben-Hamu, H., Serviansky, H., 和 Lipman, Y. 可证明具有强大表达能力的图网络。*NeurIPS*，2019。Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., 和 Grohe, M. Weisfeiler–Leman 走向神经网络：高阶图神经网络。载于*AAAI*，2019。Pozdnyakov, S. N. 和 Ceriotti, M. 三维点云的图卷积神经网络并不完备。*arXiv*预印本，2022。Pozdnyakov, S. N., Willatt, M. J., Bart´ok, A. P., Ortner, C., Cs´anyi, G., 和 Ceriotti, M. 原子结构表示的不完备性。《物理评论快报》，2020年。Read, R. C. 和 Corneil, D. G. 图同构难题。《图论杂志》，1977年。

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., 和 Battaglia, P. 利用图网络学习模拟复杂物理过程。载于*ICML*，2020年。

Satorras, V. G., Hoogeboom, E. 和 Welling, M. E (n) 等变图神经网络。载于*ICML*，2021年。Sch¨utt, K., Unke, O. 和 Gastegger, M. 用于预测张量性质和分子光谱的等变消息传递。载于*ICML*，2021年。Sch¨utt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. 和 M¨uller, K.-R. SchNet——一种用于分子和材料的深度学习架构。《化学物理学杂志》，2018年。Shapeev, A. V. 矩张量势：一类可系统改进的原子间势。《多尺度建模与仿真》，2016年。S hi, Y., Zheng, S., Ke, G., Shen, Y., You, J., He, J., Luo, S., Liu, C., He, D. 和 Liu, T.-Y. 在大规模分子建模数据集上对 Graphormer 进行基准测试。*arXiv* 预印本，2022年。Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K. 和 Riley, P. 张量场网络：面向三维点云的旋转和平移等变神经网络。*arXiv* 预印本，2018年。Topping, J., Giovanni, F. D., Chamberlain, B. P., Dong, X. 和 Bronstein, M. M. 通过曲率理解图上的过度挤压和瓶颈。载于*ICLR*，2022年。Villar, S., Hogg, D. W., Storey-Fisher, K., Yao, W. 和 Blum-Smith, B. 标量具有普适性：结构类似经典物理学的等变机器学习。*NeurIPS*，2021年。Wang, L., Liu, Y., Lin, Y., Liu, H. 和 Ji, S. Co mENet：实现三维分子图完整且高效的消息传递。载于*NeurIPS*，2022年。Wang, L., Liu, H., Liu, Y., Kurtin, J. 和 Ji, S. 通过完备三维图网络学习分层蛋白质表征。载于*ICLR*，2023年。Weiler, M., Geiger, M., Welling, M., Boomsma, W. 和 Cohen, T. S. 三维可转向卷积神经网络：在体数据中学习旋转等变特征。*NeurIPS*，2018年。Weisfeiler, B. 和 Leman, A. 将图约化为典范形式及其中出现的代数。*NTI, Series*，1968年。Xie, T. 和 Grossman, J. C. 用于准确且可解释地预测材料性质的晶体图卷积神经网络。*Phys.Rev.Lett.*，2018年。Xu, K., Hu, W., Leskovec, J. 和 Jegelka, S. 图神经网络的表达能力有多强？载于*ICLR*，2019年。Zee, A. 物理学家速成群论。普林斯顿大学出版社，2016年。

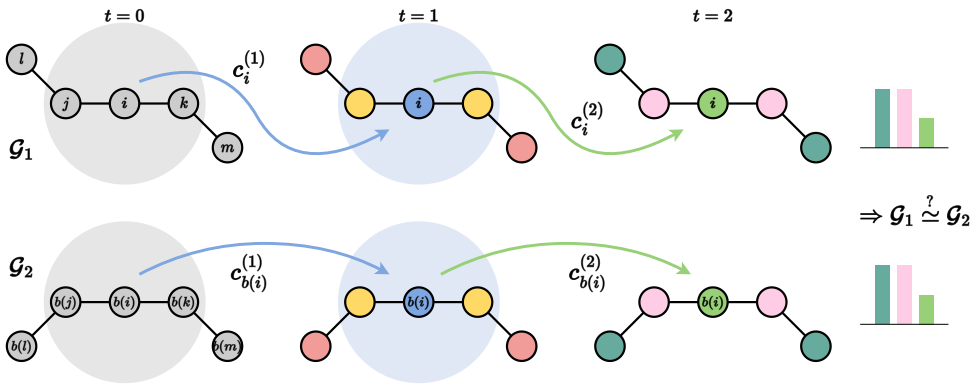


图3：**不变 GWL 测试**。IGWL 无法区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，因为它们是 1跳同一：每个节点周围的 1 跳邻域的 $\mathfrak{G}$-轨道都相同，而且 IGWL 无法将几何方向信息传播到 1 跳以外。

A. 讨论

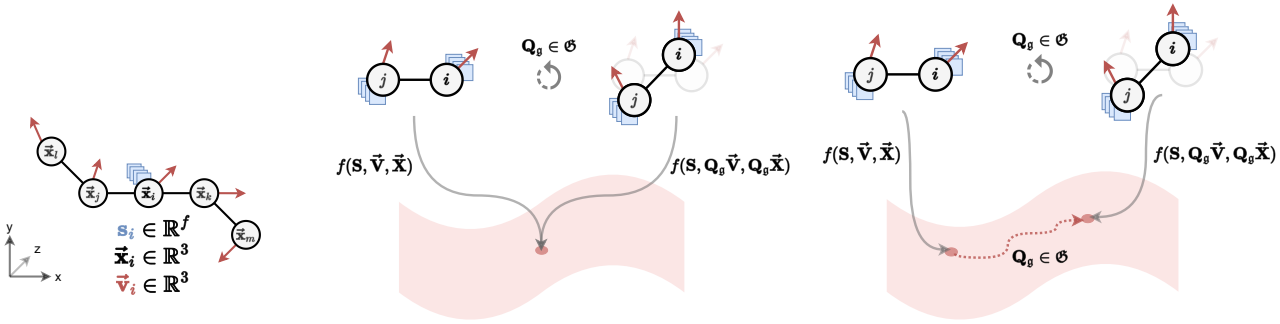
相关工作。 在分子模拟领域，关于 完备性 的研究着重于区分原子中心原子间势中的 1跳局部邻域（点云），方法是为连续的 $\mathfrak{G}$等变多重集函数构建生成集（Shapeev，2016年；Drautz，2019；Dusson 等人，2019；Pozdnyakov 等人，2020年）。GWL 将这一分析推广并扩展到一般几何图同构问题，不再局限于局部原子中心邻域。

近期关于几何图神经网络及其普适性的理论研究（Dym 和 Maron，2020年；Villar 等人，2021；Gasteiger 等人，2021；Jing 等人，2020年）表明，TFN、GemNet 和 GVP-GNN 等架构可以成为点云（ $\mathfrak{G}$-等变或 $\mathfrak{G}$-不变多重集函数的通用逼近器，即全连接图。相比之下，GWL 框架从区分几何图的角度研究在稀疏图上运行的几何图神经网络的表达能力。与普适性相比，区分能力这一视角可能更细致，也更具实践启发性。模型可能具有普适性，也可能不具有普适性。另一方面，根据能够区分和无法区分的几何图类别，区分能力可能存在多个层次，而我们的工作旨在将其形式化。

局限性。 理论上，*GWL* 能区分所有几何图吗？ 命题 7 表明，在实际应用中，若图结构是通过径向截断计算得到的，GWL 可以区分任意非同构几何图。然而，对于几何图的拓扑独立于坐标的一般情况，GWL 在理论上可能并不完备，因为可能存在一些病态的 边界情况 导致测试失败。例如，当所有坐标和向量特征都设为零时，GWL 与标准 1-WL 等同。在这种边界情况下，GWL 的表达能力与 1-WL 相同，并会继承 1-WL 所有众所周知的失效情形。

GWL 能否刻画所有类别的几何图神经网络？ GWL 是一种以节点为中心的框架，用于分析不使用规范参考系、执行局部消息传递的几何图神经网络的表达能力，而 GemNet-Q（Gasteiger 等人，2021）等非局部架构会聚合两跳邻域中的二面角，因此无法直接用 GWL 分析。类似地，基于规范参考系的几何图神经网络分析（Du 等人，2022；Wang 等人，2022）留待未来工作。这类架构使用局部或全局参考系，将等变特征标量化为不变特征；当规范参考系容易定义时，这提供了一种替代建模技术（例如 蛋白质主链结构（Jumper 等人，2021；Baek 等人，2021；Wang 等人，2023））。

值得注意的是，Wang 等人(2022) 提出了另一种完备性概念，这与 GWL 所依据的原子结构化表示完备性不同（Pozdnyakov 等人，2020年）。据我们所知，对于所有可能的几何图，依据任一定义构建可证明完备的几何图神经网络仍是一个尚未解决的问题，而我们的框架可帮助学界对此进行探索。例如，Wang 等人 (2022) 假设 SphereNet 等四体聚合器在局部上是完备的。尽管对于许多实际出现的几何图而言，这很可能成立，但表 3 中针对 Pozdnyakov 等人 (2020年) 的反例所给出的结果表明，SphereNet 确实并非 $\mathfrak{G}$-轨道单射。



(a) 几何图　　(b) $\mathfrak{G}$-不变函数　　(c) $\mathfrak{G}$-等变函数

图 4：**几何图上的不变函数和等变函数**。(a) 几何图是嵌入欧几里得空间中的属性图。几何属性会随系统的欧几里得变换而变化，例如全局旋转 $\mathfrak{G}$。(b) $\mathfrak{G}$-不变函数的输出在输入经过变换后保持不变。(c) 对于$\mathfrak{G}$-等变函数，输入的变换必须导致输出以相应的方式变换。

B. 关于几何图神经网络的补充背景

面向几何图的图神经网络可根据中间表示的类型分为两类：(1)等变模型，其中特征与传播消息均为几何量；(2) 不变模型，仅传播距离、角度等局部不变标量特征。如图 1所示，GWL 框架可用于刻画以下几类几何图神经网络的表达能力[2]。

B.1. $\mathfrak{G}$-不变 GNN

$\mathfrak{G}$-不变图神经网络层通过将几何信息标量化，从局部邻域中聚合标量。标量特征通过可学习的聚合函数 AGG 和更新函数 UPD，分别从迭代 t 更新至 $t+1$：

$$\boldsymbol{s}_i^{(t+1)} := \textsc{Upd}\left(\boldsymbol{s}_i^{(t)},\ \textsc{Agg}\left(\{\!\{(\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}), (\boldsymbol{s}_j^{(t)}, \vec{\boldsymbol{v}}_j^{(t)}), \vec{\boldsymbol{x}}_{ij} \mid j \in \mathcal{N}_i\}\!\}\right)\right). \tag{13}$$

SchNet (Sch¨utt 等人，2018年）和**CGCNN** （Xie 和 Grossman，2018年使用相对距离 $\|\vec{\boldsymbol{x}}_{ij}\|$这一二体不变量，将局部几何信息标量化：

$$\boldsymbol{s}_i^{(t+1)} := \boldsymbol{s}_i^{(t)} + \sum_{j \in \mathcal{N}_i} f_1\left(\boldsymbol{s}_j^{(t)},\ \|\vec{\boldsymbol{x}}_{ij}\|\right) \tag{14}$$

DimeNet （Gasteiger 等人，2020年）和**GemNet-T** （Gasteiger 等人，2021）使用三元组中的距离和角度 $\vec{\boldsymbol{x}}_{ij} \cdot \vec{\boldsymbol{x}}_{ik}$（三体不变量），如下所示：

$$\boldsymbol{s}_i^{(t+1)} := \sum_{j \in \mathcal{N}_i} f_1\Big(\boldsymbol{s}_i^{(t)},\ \boldsymbol{s}_j^{(t)},\ \sum_{k \in \mathcal{N}_i \setminus \{j\}} f_2\left(\boldsymbol{s}_j^{(t)},\ \boldsymbol{s}_k^{(t)},\ \|\vec{\boldsymbol{x}}_{ij}\|,\ \vec{\boldsymbol{x}}_{ij} \cdot \vec{\boldsymbol{x}}_{ik}\right)\Big) \tag{15}$$

更新后的标量特征既是 $\mathfrak{G}$-不变的，也是 $T(d)$-不变的，因为所用的唯一几何信息是相对距离和角度，而它们在 $\mathfrak{G}$ 群作用或平移下均保持不变。

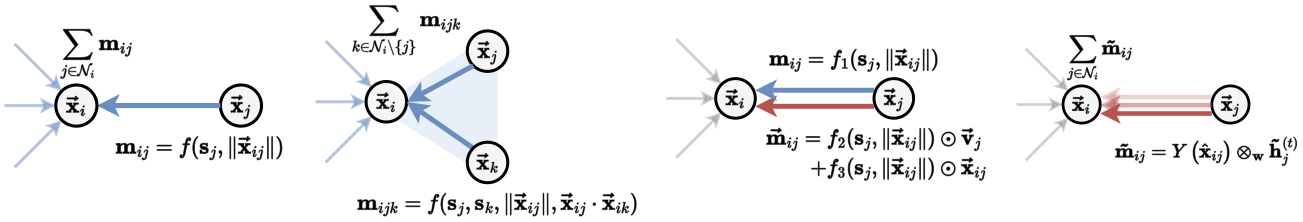
B.2. 使用笛卡尔向量的 $\mathfrak{G}$-等变 GNN

$\mathfrak{G}$-等变 GNN 层通过传播标量消息和向量消息，分别更新标量特征和向量特征，$\boldsymbol{m}_i^{(t)}$ 和 $\vec{\boldsymbol{m}}_i^{(t)}$ ：

$$\boldsymbol{m}_i^{(t)}, \vec{\boldsymbol{m}}_i^{(t)} := \textsc{Agg}\left(\{\!\{(\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}), (\boldsymbol{s}_j^{(t)}, \vec{\boldsymbol{v}}_j^{(t)}), \vec{\boldsymbol{x}}_{ij} \mid j \in \mathcal{N}_i\}\!\}\right) \qquad \text{(Aggregate)} \tag{16}$$

$$\boldsymbol{s}_i^{(t+1)}, \vec{\boldsymbol{v}}_i^{(t+1)} := \textsc{Upd}\left((\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}),\ (\boldsymbol{m}_i^{(t)}, \vec{\boldsymbol{m}}_i^{(t)})\right) \qquad \text{(Update)} \tag{17}$$

[2]请参阅 Duval 等人 (2023，了解几何图神经网络架构的全面介绍。



(a) SchNet (b) DimeNet (c) PaiNN (d) 张量场网络

图 5： **几何图神经网络消息传递**。 $\mathfrak{G}$-不变层仅传递距离等局部标量（SchNet，公式 14）或距离和角度（DimeNet，公式 15）。相比之下， $\mathfrak{G}$-等变层传递向量和相对位置等几何量（PaiNN，公式 19）或高阶张量（张量场网络，公式 21）。

PaiNN (Sch¨utt 等人，2021) 交互层通过根据相对距离（2-body 不变量）调节的学习滤波器，聚合标量和向量消息：

$$\boldsymbol{m}_i^{(t)} := \boldsymbol{s}_i^{(t)} + \sum_{j\in\mathcal{N}_i} f_1\left(\boldsymbol{s}_j^{(t)},\ \|\vec{\boldsymbol{x}}_{ij}\|\right) \tag{18}$$

$$\vec{\boldsymbol{m}}_i^{(t)} := \vec{\boldsymbol{v}}_i^{(t)} + \sum_{j\in\mathcal{N}_i} f_2\left(\boldsymbol{s}_j^{(t)},\ \|\vec{\boldsymbol{x}}_{ij}\|\right)\odot\vec{\boldsymbol{v}}_j^{(t)} + \sum_{j\in\mathcal{N}_i} f_3\left(\boldsymbol{s}_j^{(t)},\ \|\vec{\boldsymbol{x}}_{ij}\|\right)\odot\vec{\boldsymbol{x}}_{ij} \tag{19}$$

E-GNN (Satorras 等人，2021）和**GVP-GNN** (Jing 等人，2020年）采用了类似的操作。更新步骤对向量特征应用门控非线性（Weiler 等人，2018），通过将其范数与标量特征拼接，学习缩放向量特征的模长：

$$\boldsymbol{s}_i^{(t+1)} := \boldsymbol{m}_i^{(t)} + f_4\left(\boldsymbol{m}_i^{(t)}, \|\vec{\boldsymbol{m}}_i^{(t)}\|\right), \qquad \vec{\boldsymbol{v}}_i^{(t+1)} := \vec{\boldsymbol{m}}_i^{(t)} + f_5\left(\boldsymbol{m}_i^{(t)}, \|\vec{\boldsymbol{m}}_i^{(t)}\|\right)\odot\vec{\boldsymbol{m}}_i^{(t)}. \tag{20}$$

更新步骤会隐式构建三体不变量，因为计算范数 $\|\vec{\boldsymbol{m}}_i^{(t)}\| = \langle\vec{\boldsymbol{m}}_i^{(t)},\ \vec{\boldsymbol{m}}_i^{(t)}\rangle$ 相当于对所有成对的内积 $\langle\vec{\boldsymbol{x}}_{ij},\ \vec{\boldsymbol{x}}_{ik}\rangle$ 进行加权求和 $j,\ k\in\mathcal{N}_i$。

更新后的标量特征既是 $\mathfrak{G}$-不变的，也是 $T(d)$-不变的，因为所用的唯一几何信息是相对距离；而更新后的向量特征是 $\mathfrak{G}$-等变的，也是 $T(d)$-不变的，因为它们聚合了来自邻居的 $\mathfrak{G}$-等变、 $T(d)$-不变的向量量。

B.3. 使用球张量的 $\mathfrak{G}$-等变图神经网络

另一个 $\mathfrak{G}$-等变图神经网络的例子是 e3nn 框架（Geiger 和 Smidt，2022），它可用于实例化张量场网络（Thomas 等，2018）、Cormorant（Anderson 等人，2019）、SE(3)-Transformers（Fuchs 等人，2020）、SEGNN（Brandstetter 等人，2022）[3]，以及 MACE（Batatia 等人，2022b）。

这些模型使用高阶球张量 $\tilde{\boldsymbol{h}}_{i,\ l}\in\mathbb{R}^{2l+1\times f}$ 作为节点特征，阶数从 $l=0$ 一直到任意阶数$l=L$。前两阶分别对应标量特征 $\boldsymbol{s}_i$ 和向量特征 $\vec{\boldsymbol{v}}_i$。高阶张量 $\tilde{\boldsymbol{h}_i}$通过将邻域特征 $\boldsymbol{h}_j$ 与相对位移 $\frac{\boldsymbol{x}_{ij}}{\|\vec{\boldsymbol{x}}_{ij}\|}=\hat{\boldsymbol{x}}_{ij}$的高阶球谐函数$\vec{\ }$表示 Y 进行张量积 $\otimes$ 来更新，适用于所有 $j\in\mathcal{N}_i$：

$$\tilde{\boldsymbol{h}}_i^{(t+1)} := \tilde{\boldsymbol{h}}_i^{(t)} + \sum_{j\in\mathcal{N}_i} Y\left(\hat{\boldsymbol{x}}_{ij}\right)\otimes_{\boldsymbol{w}}\tilde{\boldsymbol{h}}_j^{(t)}, \tag{21}$$

其中，张量积的权重 $\boldsymbol{w}$ 通过相对距离的可学习径向基函数计算得出，即$\boldsymbol{w}=f(\|\vec{\boldsymbol{x}}_{ij}\|)$。为了获得更新后的高阶张量 $\tilde{\boldsymbol{h}}_i^{(t+1)}$ 中阶-l_3 部分的条目 $m_3\in\{-l_3,\ldots,+l_3\}$，

[3]我们建议 请参阅Brandstetter 等人（2022）的附录，以清晰简洁地了解这一模型类别 .

我们可以将方程21中的张量积展开为：

$$\tilde{\boldsymbol{h}}_{i,l_3m_3}^{(t+1)} := \tilde{\boldsymbol{h}}_{i,l_3m_3}^{(t)} + \sum_{l_1m_1,l_2m_2}^{l_3m_3} C_{l_1m_1,l_2m_2}^{l_3m_3} \sum_{j\in\mathcal{N}_i} f_{l_1l_2l_3}\left(\|\vec{\boldsymbol{x}}_{ij}\|\right) Y_{l_1}^{m_1}\left(\hat{\boldsymbol{x}}_{ij}\right) \tilde{\boldsymbol{h}}_{j,l_2m_2}^{(t)}, \tag{22}$$

其中 $C_{l_1m_1,l_2m_2}^{l_3m_3}$ 是克莱布施-戈丹系数，可确保更新后的特征具有 $\mathfrak{G}$-等变性。值得注意的是，当张量积仅限于标量（最高至 $l=0$）时，我们得到的更新形式与方程 14类似。同样，当仅使用标量和向量（最高至 $l=1$）时，我们得到的更新形式与方程 18 和方程 19类似。

MACE (Batatia 等人，2022b) 通过原子簇展开（高 k体阶特征的高效计算方法，在 `e3nn` 框架中实现（Dusson 等人，2019）：首先，他们按类似于方程 21的方式聚合邻域特征（ A 函数见 Batatia 等人 (2022b)（式 9）），然后对这些邻域特征进行 $k-1$ 多次自张量积。在我们的形式体系中，这对应于：

$$\tilde{\boldsymbol{h}}_i^{(t+1)} := \underbrace{\tilde{\boldsymbol{h}}_i^{(t+1)} \otimes_{\boldsymbol{w}} \cdots \otimes_{\boldsymbol{w}} \tilde{\boldsymbol{h}}_i^{(t+1)}}_{k-1 \text{ times}}, \tag{23}$$

这种方法省去了在更常见的多体展开中对所有 k元组进行对称化或生成它们的工作，例如DimeNet（Gasteiger 等人，2020年）。类比来说，方程23 相当于计算乘积$(a+b+\ldots)^k$，其中隐式地包含了诸如 $a^l b^{k-l}$这样的项，而不是逐一计算每个 $a^l b^{k-l}$ 项。

C. 合成实验详情

架构。geometric-gnn-dojo[4] 提供了若干种基于 GWL 的流行几何图神经网络架构的统一实现：

- $\mathfrak{G}$-不变图神经网络：SchNet（Sch¨utt 等人，2018）、DimeNet（Gasteiger 等人，2020年）和 SphereNet（Liu 等人，2022）；
- $\mathfrak{G}$-使用笛卡尔向量的等变图神经网络：E-GNN（Satorras 等人，2021）和 GVP-GNN（Jing 等人，2020年）；
- $\mathfrak{G}$使用球张量的等变图神经网络：TFN（Thomas 等人，2018）和 MACE（Batatia 等人，2022b）。

任务。我们设计合成实验，以凸显构建表达能力强的几何图神经网络时面临的实际挑战：

- 区分 k链，用于测试模型在非局部范围内传播几何信息的能力，并展示随着深度增加而出现的几何过度挤压；见表1。
- 旋转对称结构，用于测试层识别邻域方向的能力，并凸显高阶张量在 $\mathfrak{G}$等变图神经网络中的作用；见表 2。
- 来自*Pozdnyakov 等人（2020年）*的反例检验了某一层为局部邻域创建可区分指纹的能力，并凸显了采用更高体阶标量化的必要性；见表3。

超参数。对于 SchNet 和 DimeNet，我们采用 PyTorch Geometric 的实现（Fey 和 Lenssen，2019）。对于 SphereNet、E-GNN、GVP-GNN 和 MACE，我们改编了各自作者提供的实现。我们的 TFN 实现基于 `e3nn` （Geiger 和 Smidt，2022），此外，我们还将作者提供的`EquivariantProductBasisBlock` 整合到 TFN 层中，重新实现了 MACE。我们将 SchNet、DimeNet、SphereNet 和 E-GNN 的标量特征通道设为 128。将 GVP-GNN、TFN 和 MACE 的标量/向量/张量特征通道设为 64。TFN 和 MACE 默认使用阶数为 $L=2$ 的张量。MACE 默认使用局部体阶 4。我们使用 Adam 优化器训练所有模型 100 个 epoch，初始学习率为 $1e-4$；当性能停滞时，将学习率乘以 0.9，并设置 25 个 epoch 的耐心值。所有结果均为 10 个随机种子的平均值。

[4]代码可在 GitHub 上获取：`https://github.com/chaitjo/geometric-gnn-dojo`

	图神经网络层	来自Pozdnyakov 等人（2020年） 二体	三体 (图1(b))	四体 (图2(f))
Inv.	$\text{SchNet}_{\text{2-body}}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	$\text{DimeNet}_{\text{3-body}}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0
	$\text{SphereNet}_{\text{4-body}}$	**100.0 ± 0.0**	**100.0 ± 0.0**	50.0 ± 0.0
$O(3)$-Equiv.	$\text{E-GNN}_{\text{2-body}}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	$\text{GVP-GNN}_{\text{3-body}}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0
	$\text{TFN}_{\text{2-body}}$	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
	$\text{MACE}_{\text{3-body}}$	**100.0 ± 0.0**	50.0 ± 0.0	50.0 ± 0.0
	$\text{MACE}_{\text{4-body}}$	**100.0 ± 0.0**	**100.0 ± 0.0**	50.0 ± 0.0
	$\text{MACE}_{\text{5-body}}$	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**

表 3： 来自 *Pozdnyakov 等人 (2020年)*。 此任务评估单层几何图神经网络区分反例结构的能力，这些结构无法通过 k体标量化加以区分。 **体阶为k 的几何图神经网络层无法区分相应的反例**。 三体反例来自图1(b)（Pozdnyakov 等人，2020年），四体反例来自图2(f)（Pozdnyakov 等人，2020年），二体反例则基于我们贯穿全文的示例中的两个局部邻域。

D. 几何图神经网络设计空间证明

D.1. 深度的作用（第 4.1节）

以下结果是 GWL 的构造以及 k 跳不同几何图和 k 跳相同几何图的定义所导致的。注意， k 跳不同几何图也是 $(k+1)$ 跳不同几何图。类似地， k 跳相同几何图也是 $(k-1)$ 跳相同几何图，但不一定是 $(k+1)$ 跳不同几何图。

给定两个不同的邻域 $\mathcal{N}_1$ 和 $\mathcal{N}_2$，相应几何多重集 $\boldsymbol{g}_1$ 和 $\boldsymbol{g}_2$ 的 $\mathfrak{G}$-轨道彼此互斥，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_1) \cap \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_2) \equiv \emptyset$。根据 I-HASH 的性质，这意味着 $c_1 \neq c_2$。反过来，如果$\mathcal{N}_1$ 和 $\mathcal{N}_2$ 在群作用下相同，那么它们的 $\mathfrak{G}$-轨道就会重叠，即 对于某些 $\mathfrak{g} \in \mathfrak{G}$ 和$\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_1) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_2) \Rightarrow c_1 = c_2$，有 $\boldsymbol{g}_1 = \mathfrak{g}\,\boldsymbol{g}_2$。

命题3。 *GWL* 可以区分任意 k-跳不同的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，其中底层属性图同构，且 k 次迭代便足够。

命题 *3*的证明。 GWL 的第 k 次迭代通过几何多重集 $\boldsymbol{g}_i^{(k)}$，识别每个节点 i 处的 k-跳子图 $\mathcal{N}_i^{(k)}$ 的 $\mathfrak{G}$-轨道。$\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是 k-跳不同的，这意味着存在某个双射 b 和某个节点$i \in \mathcal{V}_1$，$b(i) \in \mathcal{V}_2$，使得对应的 k-跳子图 $\mathcal{N}_i^{(k)}$和 $\mathcal{N}_{b\,i\,()}^{(k)}$ 不同。因此，对应几何多重集 $\boldsymbol{g}_i^{(k)}$ 和 $\boldsymbol{g}_{b\,i\,()}^{(k)}$ 的 $\mathfrak{G}$-轨道互不相交，也就是说， $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k)}) \cap \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\,i\,()}^{(k)}) \equiv \emptyset \Rightarrow c_i^{(k)} \neq c_{b\,i\,()}^{(k)}$。因此，GWL 的 k 次迭代足以区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

□

命题4. 经过至多 k 次迭代，*GWL* 无法区分任何底层属性图同构的 k-跳相同几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

命题*4*的证明。 GWL 的第 k 次迭代通过几何多重集 $\boldsymbol{g}_i^{(k)}$，识别每个节点 i 处的 k-跳子图 $\mathcal{N}_i^{(k)}$ 的 $\mathfrak{G}$ 轨道。$\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是 k-跳相同的，这意味着对于所有双射 b 和所有节点 $i \in \mathcal{V}_1, b(i) \in \mathcal{V}_2$，相应的 k-跳子图 $\mathcal{N}_i^{(k)}$ 和 $\mathcal{N}_{b\,i\,()}^{(k)}$ 在群作用下相同。因此，相应几何多重集 $\boldsymbol{g}_i^{(k)}$ 和 $\boldsymbol{g}_{b\,i\,()}^{(k)}$ 的 $\mathfrak{G}$ 轨道有交集，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\,i\,()}^{(k)}) \Rightarrow c_i^{(k)} = c_{b\,i\,()}^{(k)}$。因此，经过至多 k 次 GWL 迭代，无法区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

□

命题5。 *IGWL* 可以区分任意一对底层属性图同构的 1跳不同几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，且只需迭代*1*次。

命题5的证明。 IGWL 的每次迭代都会识别每个节点 i处 1跳局部邻域 $\mathcal{N}_i^{(k=1)}$ 的 $\mathfrak{G}$-轨道。 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是1跳不同的，这意味着存在某个双射 b 以及某个节点 $i \in \mathcal{V}_1$， $b(i) \in \mathcal{V}_2$ ，使得对应的 1跳局部邻域 $\mathcal{N}_i^{(1)}$ 和 $\mathcal{N}_{b\,i\,()}^{(1)}$ 彼此不同。因此，对应的几何多重集 $\boldsymbol{g}_i^{(1)}$ 和 $\boldsymbol{g}_{b\,i\,()}^{(1)}$ 的 $\mathfrak{G}$-轨道互不相交，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(1)}) \cap \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\,i\,()}^{(1)}) \equiv \emptyset \Rightarrow c_i^{(1)} \neq c_{b\,i\,()}^{(1)}$。因此，IGWL 迭代1次即可区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

□

命题6. *IGWL* 的任意次迭代都无法区分任意1跳同一的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$，且它们的底层属性图同构。

命题6的证明. IGWL 的每次迭代都会识别每个节点 i处 1 跳局部邻域 $\mathcal{N}_i^{(k=1)}$ 的 $\mathfrak{G}$轨道，但根据 IGWL 的构造，它无法识别超出 1 跳范围的 $\mathfrak{G}$轨道，因为几何信息不会传播。$\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是 1 跳同一的，这意味着对于所有双射 b 以及所有节点 $i \in \mathcal{V}_1, b(i) \in \mathcal{V}_2$，相应的 1 跳局部邻域 $\mathcal{N}_i^{(k)}$ 和 $\mathcal{N}_{b\,i\,()}^{(k)}$ 在群作用下相同。因此，相应几何多重集 $\boldsymbol{g}_i^{(1)}$ 和 $\boldsymbol{g}_{b\,i\,()}^{(1)}$ 的 $\mathfrak{G}$轨道存在交叠，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(1)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\,i\,()}^{(1)}) \Rightarrow c_i^{(k)} = c_{b\,i\,()}^{(k)}$。因此，IGWL 的任意次迭代都无法区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

□

命题7。 假设几何图由点云通过径向截断构造而成，那么对于底层属性图非同构的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ ，*GWL* 可以将它们区分开来。至多需要 k_{Max} 次迭代，其中 k_{Max} 是 $\mathcal{G}_1$ 和 $\mathcal{G}_2$中较大的图直径。

命题 7的证明。 我们假设几何图 $\mathcal{G} = (\boldsymbol{A}, \boldsymbol{S}, \vec{\boldsymbol{V}}, \vec{\boldsymbol{X}})$ 由点云 $(\boldsymbol{S}, \vec{\boldsymbol{V}}, \vec{\boldsymbol{X}})$使用预先设定的径向截断 r构造而成。因此，对于所有 $a_{ij} \in \boldsymbol{A}$，邻接矩阵定义为：当 $a_{ij} = 1$ 时为 $\|\vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_j\|_2 \leq r$，否则为 0 。在生物化学和材料科学中，几何图通常按照此类构造流程构建。

给定底层属性图非同构的几何图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ ，找出 $\mathcal{G}_1$ 和 $\mathcal{G}_2$的图直径中的最大值 k_{Max} ，并任意选取节点 $i \in \mathcal{V}_1$和 $j \in \mathcal{V}_2$。我们可以分别定义以 i和 j为中心的 k_{Max}跳子图$\mathcal{N}(k_{\text{Max}})\,i$和 $\mathcal{N}^{(k_{\text{Max}})}\,j$。因此，对于所有 $i \in \mathcal{V}_1$， $\mathcal{N}(k_{\text{Max}})\,i = \mathcal{V}_1$；对于所有 $j \in \mathcal{V}_2$， $\mathcal{N}^{(k_{\text{Max}})}\,j = \mathcal{V}_2$。根据几何图的既定构造流程， $\mathcal{N}^{(k_{\text{Max}})}\,i$ 和 $\mathcal{N}^{(k_{\text{Max}})}\,j$ 必定不同。否则，若 $\mathcal{N}^{(k_{\text{Max}})}\,i$ 和 $\mathcal{N}^{(k_{\text{Max}})}\,j$ 在群作用下相同，则集合$(\boldsymbol{S}_1, \vec{\boldsymbol{V}}_1, \vec{\boldsymbol{X}}_1)$和$(\boldsymbol{S}_2, \vec{\boldsymbol{V}}_2, \vec{\boldsymbol{X}}_2)$将产生同构图。

GWL 的第 k_{Max} 次迭代通过几何多重集 $\boldsymbol{g}_i^{(k_{\text{Max}})}$ 在每个节点 i 处识别 k_{Max} 跳子图 $\mathcal{N}^{(k_{\text{Max}})}$ 的 $\mathfrak{G}$ 轨道 i 。由于对于任意节点 $i \in \mathcal{V}_1$、 $j \in \mathcal{V}_2$， $\mathcal{N}^{(k_{\text{Max}})}\,i$ 和 $\mathcal{N}(k_{\text{Max}})\,j$ 均不相同，相应几何多重集 $\boldsymbol{g}_i^{(k_{\text{Max}})}$ 和 $\boldsymbol{g}^{(k_{\text{Max}})}$ 的 $\mathfrak{G}$ 轨道彼此互斥，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k_{\text{Max}})}) \cap \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}^{(k_{\text{Max}})}\,j) \equiv \emptyset \Rightarrow c^{(k_{\text{Max}})}\,i \neq c^{(k_{\text{Max}})}\,j$。因此，GWL 的 k_{Max} 次迭代足以区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

□

D.2. 不变消息传递的局限性（第4.2节）

定理 8。 *GWL* 严格强于 *IGWL*。

定理8的证明。 首先，我们可以证明，如果 GWL 在更新时能够学习恒等映射$\boldsymbol{g}_i$ 对于所有 $i \in \mathcal{V}$，即 $\boldsymbol{g}_i^{(t)} = \boldsymbol{g}_i^{(t-1)} = \boldsymbol{g}_i^{(0)} \equiv (\boldsymbol{s}_i, \vec{\boldsymbol{v}}_i)$，则 GWL 类包含 IGWL。因此，GWL 至少与 IGWL 一样强，而 IGWL 不更新$\boldsymbol{g}_i$。

其次，要证明 GWL 的表达能力严格强于 IGWL，只需证明存在一对几何图，GWL 能够区分它们，而 IGWL 不能。我们可以考虑任意一对 k跳不同的几何图，其中 $k > 1$底层属性图同构。命题3指出，GWL 能够区分任意此类图，而命题6 指出，IGWL 无法区分它们。例如，图2 和3中的那对图。

□

命题 9. *IGWL* 和 $\mathfrak{G}$-不变图神经网络无法判定若干几何图性质：（*1*）包围几何图的包围盒/球体的周长、表面积和体积；（*2*）到质心或质心的距离；以及（*3*）二面角。

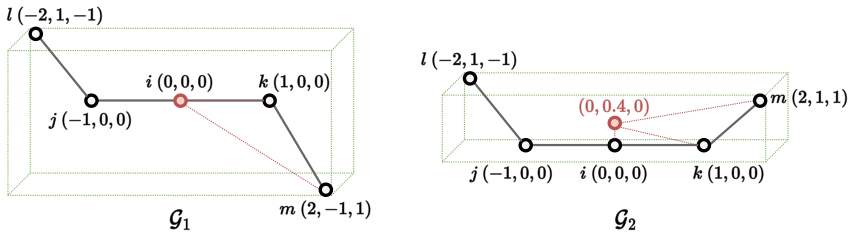


图6：IGWL 和 $\mathfrak{G}$-不变图神经网络无法区分两个几何图的周长、表面积、包围盒/球体体积、质心距离和二面角。质心以红点表示，与质心的距离以红色虚线表示。包围几何图的包围盒以绿色虚线表示。

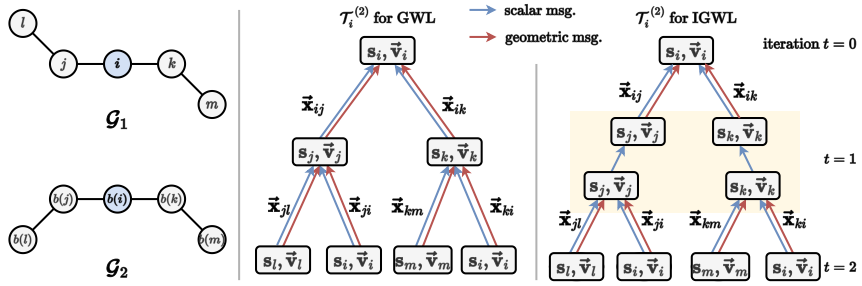


图7：**GWL 和 IGWL 的几何计算树**。与 GWL 不同，在 IGWL 中，几何方向信息无法从叶节点流向根节点，从而限制了其表达能力。由于所有 1 跳邻域在计算上都相同，IGWL 无法区分 $\mathcal{G}_1$ 和 $\mathcal{G}_2$。

命题 9 的证明。 参照 Garg等人 (2020年)，我们称某个模型类别 能够判定 某项几何图性质，是指该类别中存在一个模型，使得对于任意两个在该性质上不同的几何图，该模型都能够区分这两个几何图。

在图6中，我们给出了两个几何图的示例，以证明该命题。 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 在以下几何图属性方面存在差异：

- 包围几何图的包围盒的周长、表面积和体积[5]：（32 个单位、40 个单位$_2$、16 个单位^3）与（28 个单位、24 个单位$_2$、8 个单位^3）。
- 到质心或质心的距离多重集：$\{0.40, 1.08, 1.08, 2.32, 2.32\}$。 $\{0.00, 1.00, 1.00, 2.45, 2.45\}$ vs.
- 二面角：$\angle(ljkm) = \frac{(\vec{x}_{jk}\times\vec{x}_{lj})\cdot(\vec{x}_{jk}\times\vec{x}_{mk})}{|\vec{x}_{jk}\times\vec{x}_{lj}||\vec{x}_{jk}\times\vec{x}_{mk}|}$ 对于这两个图显然不同。

然而，根据命题 6 和定理 24，IGWL 和 $\mathfrak{G}$-不变图神经网络无法区分这两个几何图，因此也无法判定所有这些性质。

对于任意数量的 IGWL 或 $\mathfrak{G}$-不变图神经网络迭代，我们也可以通过计算树的几何版本证明这一点（Garg等人，2020年），如图 7所示。计算树 $\mathcal{T}_i^{(t)}$ 通过将消息传递过程“展开”，表示节点在 i 次迭代时，GWL/IGWL 颜色或图神经网络特征中包含的最大信息。我们可以直观地将 GWL、IGWL 及相应类别的图神经网络理解为对几何计算树进行着色。

几何计算树按递归方式构造：对于所有 $i \in \mathcal{V}$， $\mathcal{T}_{(0)\,i} = (\boldsymbol{s}_i, \vec{\boldsymbol{v}}_i)$。对于 $t > 0$，我们从一个根节点（$\boldsymbol{s}_i$， $\vec{\boldsymbol{v}}_i$）开始，并为所有 $j \in \mathcal{N}_i$ 添加一个子树 $\mathcal{T}_{(t-1)\,j}$，同时记录边上的相对位置 $\vec{\boldsymbol{x}}_{ij}$。为了获得根

[5]相同的结果也适用于图中未展示的有界球。

节点嵌入或颜色中，标量信息和几何信息都会从叶节点向上传播至根节点。因此，如果两个节点具有相同的几何计算树，它们将被映射为相同的节点嵌入或颜色。

关键在于，在 IGWL 和 $\mathfrak{G}$ 不变图神经网络的计算树中，几何方向信息无法从一个层级传递到另一个层级，因为它们只更新标量信息。在递归构造过程中，我们必须在添加子树 $\mathcal{T}_{(t-1)\ j}$ 之前插入一个连接节点（$\boldsymbol{s}_j,\ \vec{\boldsymbol{v}}_j$），并阻止它们之间的几何信息传播。

按照图 6中几何图的构造过程，我们发现，任意一对同构节点的 IGWL 计算树都相同，因为所有 1 跳邻域在计算上都相同。因此，节点颜色或节点标量特征也都相同，这意味着 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 无法区分。

□

Prop命题 10。 对于全连接几何图，*IGWL* 与 *GWL* 具有相同的表达能力 *hs.*

命题 *10*的证明。 我们将用反证法证明。假设存在一对全连接几何图$\mathcal{G}_1$ 和 $\mathcal{G}_2$，GWL 可以区分它们，但 IGWL 无法区分。

如果 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 的底层属性图同构，根据命题3和命题6，$\mathcal{G}_1$ 和 $\mathcal{G}_2$ 是 1跳同一，但对于某个 $k>1$，它们在 k跳时不同。对于所有双射 b 以及所有节点 $i\in\mathcal{V}_1$，$b(i)\in\mathcal{V}_2$，局部邻域$\mathcal{N}(1)_{\,i}$和 $\mathcal{N}^{(1)}_{b\ i}()$ 在群作用下相同，并且 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(1)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\ i}^{(1)}())\ \Rightarrow\ c_i^{(1)} = c_{b\ i}^{(1)}()$。此外，存在某个双射 b和某些节点 $i\in\mathcal{V}_1$，$b(i)\in\mathcal{V}_2$，使得 k 跳子图 $\mathcal{N}(k)_{\,i}$和 $\mathcal{N}^{(k)}_{b\ i}()$ 不同，并且$\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k)}) \cap \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\ i}^{(k)}()) \equiv \emptyset\ \Rightarrow\ c_i^{(k)} \neq c_{b\ i}^{(k)}()$。然而，由于 $\mathcal{G}_1$和 $\mathcal{G}_2$是完全连通的，对于任意 k，$\mathcal{N}(1_{\ i} = \mathcal{N}^{(k)}_{i})$ 和$\mathcal{N}(1)_{\,b(i)} = \mathcal{N}^{(k)}_{b\ i}()$ 在群作用下相同。因此，$\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(1)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\ i}^{(1)}()) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_{b\ i}^{(k)}())\ \Rightarrow\ c_i^{(1)} = c_i^{(k)} = c_{b\ i}^{(k)}() = c_{b\ i}^{(k)}()$。这是一个矛盾。

若 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 非同构且完全连通，对于任意的 $i\in\mathcal{V}_1$，$j\in\mathcal{V}_2$ 以及任意 k跳邻域，我们知道 $\mathcal{N}_i^{(1)} = \mathcal{N}(k)_{\,i}$和 $\mathcal{N}^{(1}_{\ j} = \mathcal{N}^{(k)}_{j})$。因此，GWL 和 IGWL 的一次迭代会识别出相同的 $\mathfrak{G}$-轨道，并赋予相同的节点颜色，即 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(1)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_i^{(k)})\ \Rightarrow\ c_i^{(1)} = c_i^{(k)}$ 和 $\mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_j^{(1)}) = \mathcal{O}_{\mathfrak{G}}(\boldsymbol{g}_j^{(k)})\ \Rightarrow\ c_j^{(1)} = c_j^{(k)}$。这与假设矛盾。

□

D.3. 标量化体阶的作用（第4.3节）

命题 11。 I-HASH(m) 对于 $m = max(\{|\mathcal{N}_i|\mid i\in\mathcal{V}\})$是 $\mathfrak{G}$-轨道单射，其中 *max* 是给定数据集中所有局部邻域 $\mathcal{N}_i$ 的最大基数。

命题 *11* 的证明。由于 m 是所考虑的所有局部邻域 $\mathcal{N}_i$ 中的最大基数，任意两个不同的邻域 $\mathcal{N}_1$ 和 $\mathcal{N}_2$ 必定具有不同的 m 阶标量多重集。由于 I-HASH(m) 计算涉及最多 m 个节点的标量，因此能够区分任意这类 $\mathcal{N}_1$ 和 $\mathcal{N}_2$。因此，I-HASH(m) 是 $\mathfrak{G}$- 轨道单射的。 □

命题 12。*IGWL*$^{(k)}$ 至少与 *IGWL*$^{(k-1)}$ 一样强。对于 $k\leq 5$，*IGWL*$^{(k)}$ 严格强于 *IGWL*$^{(k-1)}$。

命题 *12* 的证明。根据构造，I-HASH(k) 会根据邻域中的元素与中心节点所组成的、长度最多为 k 的所有可能元组，计算 $\mathfrak{G}$-不变标量。因此，I-HASH(k) 类包含 I-HASH$(k-1)$，且 I-HASH(k) 至少与 I-HASH$(k-1)$一样强。因此，相应的检验 IGWL(k) 至少与 IGWL$(k-1)$ 一样强。

其次，要证明对于 $k\leq 5$，IGWL(k) 的表达能力严格强于 IGWL$(k-1)$，只需证明存在一对几何邻域，能够被 IGWL(k) 区分，却无法被 IGWL$(k-1)$ 区分：

- 对于 $k=3$ 和 $\mathfrak{G}=O(3)$ 或 $SO(3)$，在（Sch¨utt 等人，2021）图 1 中的局部邻域里，两个相邻节点之间角度不同的配置可以被 IGWL(3) 区分，但无法被 IGWL(2)区分。

- 对于 $k=4$ 和 $\mathfrak{G}=O(3)$ 或 $SO(3)$，图1中来自（Pozdnyakov 等人，2020年）的这对局部邻域可由 IGWL(4) 区分，但无法由 IGWL(3)区分。

- 对于 $k=5$和 $\mathfrak{G}=O(3)$，图 2(e) 中来自（Pozdnyakov 等人，2020年）的这对局部邻域可由 IGWL$_{(5)}$ 区分，但无法由 IGWL$_{(4)}$区分。
- 对于 $k=5$ 和 $\mathfrak{G}=SO(3)$，(Pozdnyakov 等人，2020年) 中图 2(f) 所示的局部邻域对可由 IGWL$_{(5)}$ 区分，但无法由 IGWL$_{(4)}$区分。

□

命题 13. 设 $\mathcal{G}_1=(\boldsymbol{A}_1,\boldsymbol{S}_1,\vec{\boldsymbol{X}}_1)$ 和 $\mathcal{G}_2=(\boldsymbol{A}_2,\boldsymbol{S}_2,\vec{\boldsymbol{X}}_2)$ 是两个几何图，且所有边均等长。那么，*IGWL*$_{(2)}$ 能够区分这两个图，当且仅当 *Weisfeiler–Leman* 能够区分属性图$(\boldsymbol{A}_1,\boldsymbol{S}_1)$ 和 $(\boldsymbol{A}_1,\boldsymbol{S}_1)$。

命题 *13*的证明. 设 c 和 k 分别是 IGWL 产生的颜色$_{(2)}$ 和 Weisfeiler–Leman 产生的颜色，并设 i 和 j 是结果陈述中任意两个此类图中的两个节点。我们对该陈述进行归纳证明。
显然，$c_i^{(0)}=k_i^{(0)}$ 对所有节点 i 和 $c_i^{(0)}=c_j^{(0)}$ 成立，当且仅当 $k_i^{(0)}=k_j^{(0)}$。现在，假设该陈述对第 t 次迭代成立。也就是说，对所有 i，$c_i^{(t)}=c_j^{(t)}$ 当且仅当 $k_i^{(t)}=k_j^{(t)}$ 成立。注意，$c_i^{(t+1)}=c_j^{(t+1)}$ 当且仅当 $c_i^{(t)}=c_j^{(t)}$ 且$\{\!\{(c_p^{(t)},\ \|\vec{\boldsymbol{x}}_{ip}\|)\mid p\in\mathcal{N}_i\}\!\}=\{\!\{(c_p^{(t)},\ \|\vec{\boldsymbol{x}}_{jp}\|)\mid p\in\mathcal{N}_j\}\!\}$，因为相对向量的范数是 IGWL$_{(2)}$ 能够计算的唯一单射不变量（除缩放外）。由于所有范数都相等，根据归纳假设，这等价于$k_i^{(t)}=k_j^{(t)}$ 且
$\{\!\{k_p^{(t)}\mid p\in\mathcal{N}_i\}\!\}=\{\!\{k^{(t)}\mid p\in\mathcal{N}_j\}\!\}$。因此，这等价于 $k_i^{(t+1)}=k_j^{(t+1)}$ □

E. GWL 与几何图神经网络等价性的证明（第 3.1 节）

我们的证明改编自 Xu 等人 (2019)；Morris 等人 (2019将 Weisfeiler–Leman 与图神经网络联系起来时所用的技巧。请注意，为简洁起见，我们省略了 GWL 和几何图神经网络更新中的相对位置向量 $\vec{\boldsymbol{x}}_{ij}$，因为相对位置向量可以合并到向量特征中。

定理 1. 任意一对几何图，只要它们可由 等变 *GNN* 区分，也可由 *GWL* 区分 *L.*

定理 *1* 的证明。考虑两个几何图 $\mathcal{G}$ 和 $\mathcal{H}$。定理表明，如果图神经网络图级读出输出 $f(\mathcal{G})\neq f(\mathcal{H})$，那么 GWL 测试总能判定 $\mathcal{G}$ 和 $\mathcal{H}$ 是非同构的，即 $\mathcal{G}\not\simeq\mathcal{H}$。
我们将通过反证法证明。假设经过 T 次迭代后，图神经网络图级读出输出 $f(\mathcal{G})\neq f(\mathcal{H})$，但 GWL 测试无法判定 $\mathcal{G}$ 和 $\mathcal{H}$ 是非同构的，即 $\mathcal{G}$ 和 $\mathcal{H}$ 在第 0 至 T 次迭代中始终具有相同的节点颜色集合。因此，对于任意 $t=0\ldots T-1$，在第 t 次迭代和第 $t+1$ 次迭代时，$\mathcal{G}$和 $\mathcal{H}$具有相同的节点颜色集合 $\{c_i^{(t)}\}$，以及相同的邻域几何多重集 $\left\{(c_i^{(t)},\ \boldsymbol{g}_i^{(t)})\ \ \{\!\{(c_j^{(t)},\ \boldsymbol{g}_j^{(t)})\mid j\in\mathcal{N}_i\}\!\}\right\}$（差异仅在群作用下）。否则，由于不同的几何多重集会获得各自独特的新颜色，GWL 测试便会在第 $t+1$ 次迭代时为 $\mathcal{G}$和 $\mathcal{H}$ 赋予不同的节点颜色。

我们将证明，对于同一图上的节点 i和 k，如果 $(c_i^{(t)},\boldsymbol{g}_i^{(t)})=(c_k^{(t)},\mathfrak{g}\cdot\boldsymbol{g}_k^{(t)})$，那么对于任意迭代 t，我们总有 GNN 特征 $(\boldsymbol{s}_i^{(t)},\vec{\boldsymbol{v}}_i^{(t)})=(\boldsymbol{s}_k^{(t)},\boldsymbol{Q}_{\mathfrak{g}}\vec{\boldsymbol{v}}_k^{(t)})$。这对 $t=0$成立，因为 GWL 和 GNN 的初始化相同。假设这一结论在迭代 t时成立。在迭代 $t+1$时，如果对于任意 i和 k，$(c_i^{(t+1)},\boldsymbol{g}_i^{(t+1)})=(c_k^{(t+1)},\mathfrak{g}\cdot\boldsymbol{g}_k^{(t+1)})$，那么：

$$\left\{(c_i^{(t)},\boldsymbol{g}_i^{(t)}),\ \{\!\{(c_j^{(t)},\boldsymbol{g}_j^{(t)})\mid j\in\mathcal{N}_i\}\!\}\right\}=\left\{(c_k^{(t)},\mathfrak{g}\cdot\boldsymbol{g}_k^{(t)}),\ \{\!\{(c_j^{(t)},\mathfrak{g}\cdot\boldsymbol{g}_j^{(t)})\mid j\in\mathcal{N}_k\}\!\}\right\}\tag{24}$$

根据我们对迭代 t的假设，

$$\left\{(\boldsymbol{s}_i^{(t)},\vec{\boldsymbol{v}}_i^{(t)}),\ \{\!\{(\boldsymbol{s}_j^{(t)},\vec{\boldsymbol{v}}_j^{(t)})\mid j\in\mathcal{N}_i\}\!\}\right\}=\left\{(\boldsymbol{s}_k^{(t)},\boldsymbol{Q}_{\mathfrak{g}}\vec{\boldsymbol{v}}_k^{(t)}),\ \{\!\{(\boldsymbol{s}_j^{(t)},\boldsymbol{Q}_{\mathfrak{g}}\vec{\boldsymbol{v}}_j^{(t)})\mid j\in\mathcal{N}_k\}\!\}\right\}\tag{25}$$

由于 GNN 中每个节点都应用相同的聚合与更新操作，因此相同的输入，即邻域特征，会映射到相同的输出。因此，$(\boldsymbol{s}_i^{(t+1)},\vec{\boldsymbol{v}}_i^{(t+1)})=(\boldsymbol{s}_k^{(t+1)},\boldsymbol{Q}_{\mathfrak{g}}\vec{\boldsymbol{v}}_k^{(t+1)})$。根据归纳法，如果 $(c_i^{(t)},\boldsymbol{g}_i^{(t)})=(c_k^{(t)},\mathfrak{g}\cdot\boldsymbol{g}_k^{(t)})$，那么对于任意迭代 t，我们始终有图神经网络节点特征 $(\boldsymbol{s}_i^{(t)},\vec{\boldsymbol{v}}_i^{(t)})=(\boldsymbol{s}_k^{(t)},\boldsymbol{Q}_{\mathfrak{g}}\vec{\boldsymbol{v}}_k^{(t)})$。这会产生有效映射 ϕ_s,ϕ_v，使得 $\boldsymbol{s}_i^{(t)}=\phi_s(c_i^{(t)})$ 且 $\vec{\boldsymbol{v}}_i^{(t)}=\phi_v(c_i^{(t)},\boldsymbol{g}_i^{(t)})$，对于任意 $i\in\mathcal{V}$ 均成立。

因此，如果 $\mathcal{G}$ 和 $\mathcal{H}$ 具有相同的节点颜色集合和几何多重集，那么 $\mathcal{G}$ 和 $\mathcal{H}$ 也具有相同的 GNN 邻域特征集合

$$\left\{(\boldsymbol{s}_i^{(t)}, \vec{\boldsymbol{v}}_i^{(t)}),\ \{\!\{(\boldsymbol{s}_j^{(t)}, \vec{\boldsymbol{v}}_j^{(t)}) \mid j \in \mathcal{N}_i\}\!\}\right\} = \left\{(\phi_s(c_i^{(t)}), \phi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)}))\ ,\ \{\!\{(\phi_s(c_j^{(t)}), \phi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)})) \mid j \in \mathcal{N}_i\}\!\}\right\}$$

因此，GNN 将为 $\mathcal{G}$ 和 $\mathcal{H}$ 输出相同的节点标量特征集合 $\{\boldsymbol{s}_i^{(T)}\}$，而置换不变的图级读出将输出 $f(\mathcal{G}) = f(\mathcal{H})$。这与假设矛盾。 □

类似地， $\mathfrak{G}$-不变图神经网络的表达能力至多与 IGWL 相当。

定理 24. 任何一对可由 $\mathfrak{G}$-不变图神经网络区分的几何图，也可由 *IGWL* 区分。

证明。 证明过程与定理1的证明类似。 □

命题 2. $\mathfrak{G}$-等变图神经网络在满足以下条件时与 *GWL* 具有相同的表达能力：*(1)* 聚合 AGG 是一个单射的、$\mathfrak{G}$-等变的多重集函数。*(2)* 更新 UPD_s 的标量部分是一个 $\mathfrak{G}$-轨道单射的、 $\mathfrak{G}$-不变的多重集函数。*(3)* 更新 UPD_v 的向量部分是一个单射的、 $\mathfrak{G}$-等变的多重集函数。*(4)* 图级读出 f 是一个单射的多重集函数。

定理 *2* 的证明。 考虑一个满足这些条件的 GNN。我们将证明，经过足够多次迭代t，该 GNN 的输出与 GWL 等价，即 $\boldsymbol{s}^{(t)} \equiv c^{(t)}$。

设 $\mathcal{G}$和 $\mathcal{H}$是任意两个几何图，GWL 测试在第 T次迭代时判定它们非同构。由于图级读出函数是单射，即 它会将不同的节点标量特征多重集映射为唯一的嵌入，因此只需证明，经过足够多次迭代后，GNN 的邻域聚合过程会将 $\mathcal{G}$ 和 $\mathcal{H}$ 嵌入到不同的节点特征多重集中。

在此证明中，我们用定义在由 $\mathfrak{G}$的群作用生成的等价类上的单射函数，替代 $\mathfrak{G}$-轨道单射函数。因此，属于同一 $\mathfrak{G}$-轨道的所有元素首先都会被映射到该等价类的同一个代表元，记作方括号 $[\ldots]$，随后再经过一个单射映射。所得函数是 $\mathfrak{G}$-轨道单射的。

假设 GNN 按如下方式更新节点标量和向量特征：

$$\boldsymbol{s}_i^{(t)} = \text{UPD}_s\left(\left[(\boldsymbol{s}_i^{(t-1)}, \vec{\boldsymbol{v}}_i^{(t-1)})\ ,\ \text{AGG}\left(\{\!\{(\boldsymbol{s}_i^{(t-1)}, \boldsymbol{s}_j^{(t-1)}, \vec{\boldsymbol{v}}_i^{(t-1)}, \vec{\boldsymbol{v}}_j^{(t-1)}) \mid j \in \mathcal{N}_i\}\!\}\right)\right]\right) \tag{26}$$

$$\vec{\boldsymbol{v}}_i^{(t)} = \text{UPD}_v\left((\boldsymbol{s}_i^{(t-1)}, \vec{\boldsymbol{v}}_i^{(t-1)})\ ,\ \text{AGG}\left(\{\!\{(\boldsymbol{s}_i^{(t-1)}, \boldsymbol{s}_j^{(t-1)}, \vec{\boldsymbol{v}}_i^{(t-1)}, \vec{\boldsymbol{v}}_j^{(t-1)}) \mid j \in \mathcal{N}_i\}\!\}\right)\right) \tag{27}$$

其中，聚合函数 AGG 是 $\mathfrak{G}$-等变且单射的，标量更新函数 UPD_s 是 $\mathfrak{G}$-不变且单射的，而向量更新函数 UPD_v 是 $\mathfrak{G}$-等变且单射的。

GWL 测试按如下方式更新节点颜色 $c_i^{(t)}$ 和几何多重集 $\boldsymbol{g}_i^{(t)}$：

$$c_i^{(t)} = h_s\left(\left[(c_i^{(t-1)}, \boldsymbol{g}_i^{(t-1)})\ ,\ \{\!\{(c_j^{(t-1)}, \boldsymbol{g}_j^{(t-1)}) \mid j \in \mathcal{N}_i\}\!\}\right]\right), \tag{28}$$

$$\boldsymbol{g}_i^{(t)} = h_v\left((c_i^{(t-1)}, \boldsymbol{g}_i^{(t-1)})\ ,\ \{\!\{(c_j^{(t-1)}, \boldsymbol{g}_j^{(t-1)}) \mid j \in \mathcal{N}_i\}\!\}\right), \tag{29}$$

其中， h_s 是一个 $\mathfrak{G}$-不变且单射的映射，而 h_v 是一个 $\mathfrak{G}$-等变且单射的运算（例如，在方程6中，通过复制来扩展几何多重集）。

我们将通过归纳法证明，在任意迭代 t 中，始终存在单射函数 φ_s 和 φ_v，使得 $\boldsymbol{s}_i^{(t)} = \varphi_s(c_i^{(t)})$ 且 $\vec{\boldsymbol{v}}_i^{(t)} = \varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)})$。对于 $t = 0$，由于 GWL 和 GNN 的初始节点特征相同，且对于所有 $i \in \mathcal{V}(\mathcal{G})\ \mathcal{V}(\mathcal{H})$， $c_i^{(0)} \equiv \boldsymbol{s}_i^{(0)}$ 和 $\boldsymbol{g}_i^{(0)} \equiv (\boldsymbol{s}_i^{(0)}, \vec{\boldsymbol{v}}_i^{(0)})$，因此该结论成立。假设该结论对迭代 t 成立。在迭代 $t+1$ 中，将 $\boldsymbol{s}_i^{(t)}$ 替换为$\varphi_s(c_i^{(t)})$，并将 $\vec{\boldsymbol{v}}_i^{(t)}$ 替换为 $\varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)})$，可得

$$\boldsymbol{s}_i^{(t+1)} = \text{UPD}_s\left(\left[(\varphi_s(c_i^{(t)}), \varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)}))\ ,\ \text{AGG}\left(\{\!\{(\varphi_s(c_i^{(t)}), \varphi_s(c_j^{(t)}), \varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)}), \varphi_v(c_j^{(t)}, \boldsymbol{g}_j^{(t)})) \mid j \in \mathcal{N}_i\}\!\}\right)\right]\right)$$

$$\vec{\boldsymbol{v}}_i^{(t+1)} = \text{UPD}_v\left((\varphi_s(c_i^{(t)}), \varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)}))\ ,\ \text{AGG}\left(\{\!\{(\varphi_s(c_i^{(t)}), \varphi_s(c_j^{(t)}), \varphi_v(c_i^{(t)}, \boldsymbol{g}_i^{(t)}), \varphi_v(c_j^{(t)}, \boldsymbol{g}_j^{(t)})) \mid j \in \mathcal{N}_i\}\!\}\right)\right)$$

多个单射函数的复合仍是单射。因此，存在一些单射函数 g_s 和 g_v，使得：

$$\boldsymbol{s}_i^{(t+1)} = g_s\left(\left[(c_i^{(t)}, \boldsymbol{g}_i^{(t)}),\ \{\!\{(c_j^{(t)}, \boldsymbol{g}_j^{(t)}) \mid j \in \mathcal{N}_i\}\!\}\right]\right), \tag{30}$$

$$\vec{\boldsymbol{v}}_i^{(t+1)} = g_v\left((c_i^{(t)}, \boldsymbol{g}_i^{(t)}),\ \{\!\{(c_j^{(t)}, \boldsymbol{g}_j^{(t)}) \mid j \in \mathcal{N}_i\}\!\}\right), \tag{31}$$

接下来，我们可以考虑：

$$\boldsymbol{s}_i^{(t+1)} = g_s \circ h_s^{-1}\ h_s\left(\left[(c_i^{(t)}, \boldsymbol{g}_i^{(t)}),\ \{\!\{(c_j^{(t)}, \boldsymbol{g}_j^{(t)}) \mid j \in \mathcal{N}_i\}\!\}\right]\right), \tag{32}$$

$$\vec{\boldsymbol{v}}_i^{(t+1)} = g_v \circ h_v^{-1}\ h_v\left((c_i^{(t)}, \boldsymbol{g}_i^{(t)}),\ \{\!\{(c_j^{(t)}, \boldsymbol{g}_j^{(t)}) \mid j \in \mathcal{N}_i\}\!\}\right), \tag{33}$$

于是，我们可以将 $\varphi_s = g_s \circ h_s^{-1}$和 $\varphi_v = g_v \circ h_v^{-1}$视为单射函数，因为单射函数的复合仍是单射。因此，对于任意迭代 $t+1$，都存在单射函数 φ_s 和 φ_v，使得 $\boldsymbol{s}_i^{(t+1)} = \varphi_s\left(c_i^{(t+1)}\right)$ 并且
$\vec{\boldsymbol{v}}_i^{(t+1)} = \varphi_v\left(c_i^{(t+1)}, \boldsymbol{g}_i^{(t+1)}\right)$.

在第 T次迭代时，GWL 测试判定 $\mathcal{G}$ 和 $\mathcal{H}$ 非同构，这意味着对于 $\mathcal{G}$ 和 $\mathcal{H}$，节点颜色的多重集$\{c_i^{(T)}\}$ 不同。由于 φ_s的单射性，GNN 的节点标量特征 $\{\boldsymbol{s}_i^{(T)}\} = \{\varphi_s(c_i^{(T)})\}$ 对于 $\mathcal{G}$ 和 $\mathcal{H}$也必然不同。

□

较弱的一组条件就足以使 $\mathfrak{G}$-不变的 GNN 至少具有与 IGWL 同等的表达能力。

命题 25. $\mathfrak{G}$-不变图神经网络在满足以下条件时，与 *IGWL* 具有相同的表达能力：*(1)* 聚合函数 ψ 和更新函数 ϕ 是 $\mathfrak{G}$-轨道单射的 $\mathfrak{G}$-不变多重集函数。*(2)* 图级读出函数 f 是单射多重集函数。

证明. 证明与定理2的证明类似。 □

F. 普适性与判别性证明（第 6节）

F.1. 普适性与判别性的等价性

本小节的结果采用了 Chen 等人（2019）的证明，并做了少量调整。

定理 16. 如果 $\mathcal{C}$ 在 Y上普适逼近，那么 $\mathcal{C}$ 也具有成对 $Y_{\mathfrak{G}}$ 判别能力。

证明。 给定任意 $y \in Y$，我们可以构造一个定义在 X 上的 $\mathfrak{G}$-不变函数 $\delta_y(x) = 0$，当 $y \simeq x$ 时取该值，否则取 1。因此，δ_y 可以由某个函数 $h \in \mathcal{C}$ 通过 Y 上的某个 $\epsilon < 0.5$ 近似得到。因此，对于任意 y，$y' \in Y$，都有 $h(y) \neq h(y')$，且 $\mathcal{C}$ 是成对 $Y_{\mathfrak{G}}$ 判别的。 □

以下两个引理是对 Chen 等人(2019) 的结果稍作改编后得到的。

引理 26。 如果 $\mathcal{C}$ 是成对 $Y_{\mathfrak{G}}$ 判别的，那么对于所有 $y \in Y$，都存在一个函数 $\delta_y \in \mathcal{C}^{+1}$，使得对于所有 y'，$\delta_y(y') = 0$ 当且仅当 $y \simeq y'$。

证明。 对于任意满足 $y_1 \not\simeq y_2$ 的 $y_1, y_2 \in Y$，令 δ_{y_1,y_2} 为区分 y_1, y_2 的函数。也就是说，$\delta_{y_1,y_2}(y_1) \neq \delta_{y_1,y_2}(y_2)$。然后，我们可以定义函数 $\delta_{y,y'} \in \mathcal{C}$:

$$\overline{\delta}_{y,y'}(x) = |\delta_{y,y'}(x) - \delta_{y,y'}(y)| \to \begin{cases} = 0 & \text{if } x \simeq y \\ > 0 & \text{if } x \simeq y' \\ \geq 0 & \text{otherwise} \end{cases} \tag{34}$$

这个函数已经与我们想要证明其存在性的 δ_y 函数相似。为了得到一个在所有满足 $x \not\simeq y$ 的 $x \in Y$ 上都严格为正的函数，我们可以将所有 $\delta_{y,y'}$ 求和，构造出 δ_y：

$$\delta_y(x) = \sum_{y' \in Y, y' \not\simeq y} \overline{\delta}_{y,y'}(x) \to \begin{cases} = 0 & \text{if } x \simeq y \\ > 0 & \text{if } x \not\simeq y \text{ and } x \in \mathcal{O}_{\mathfrak{G}}(Y) \supseteq Y \\ \geq 0 & \text{otherwise} \end{cases} \tag{35}$$

给定函数有限集 $\{\delta_{y,y'}\}$，注意到

$$\overline{\delta}_{y,y'}(x) = \mathrm{ReLU}\big(\delta_{y,y'}(x) - \delta_{y,y'}(y)\big) \; + \; \mathrm{ReLU}\big(\delta_{y,y'}(y) - \delta_{y,y'}(x)\big).$$

那么，将这些函数在 $y' \in Y$ 上求和，并结合 $y' \not\simeq y$，即可得到 δ_y，因此 $\delta_y \in \mathcal{C}^{+1}$。 □

引理27. 设 $\mathcal{C}$ 是从 $X \to \mathbb{R}$ 到 $\mathfrak{G}$-不变函数的一个类，且对于任意 y、$y' \in Y \subseteq X$，其中 Y 为有限集，都存在一个 $\delta_y \in \mathcal{C}$ 满足性质 $\delta_y(y') = 0$ 当且仅当 $y \simeq y'$。那么，$\mathcal{C}^{+1}$ 在 Y 上是普适逼近的。

证明. 对于 $y \in Y$，定义 $r_y := \frac{1}{2}\min_{y' \in Y, y' \not\simeq y} \delta_y(y')$。定义半径为 $r > 0$ 的隆起函数 $b_r \colon \mathbb{R} \to \mathbb{R}$ 为 $b_r(s) = \psi(\frac{s}{r})$，其中

$$\psi(z) = \mathrm{ReLU}(z+1) + \mathrm{ReLU}(1-z) - 2\mathrm{ReLU}(z).$$

定义 $k_y := |Y \cap \mathcal{O}_{\mathfrak{G}}(y)|^{-1}$。由于 Y 是有限集，且与 y 的轨道的交集包含 y，因此 k_y 是有限且定义良好的。我们可以将从 X 到 $\mathbb{R}$ 的 $\mathfrak{G}$-不变函数 h 定义为：

$$h(x) = \sum_{y \in Y} k_y f(y) b_{r_y}(\delta_y(x)) \tag{36}$$

注意到 $h|_Y = f|_Y$ 和 $h \in \mathcal{C}^{+1}$。因此，$\mathcal{C}^{+1}$ 是普适逼近的。 □

定理 17. 如果 $\mathcal{C}$ 是成对 $Y_{\mathfrak{G}}$ 可区分的，那么 $\mathcal{C}^{+2}$ 在 Y 上是普适逼近的。

证明. 结论可直接由上述两个引理得出。 □

引理 28. 设 X、Y 为拓扑空间，且 $h \colon X \times Y \to \mathbb{R}$ 是连续函数。那么，若 Y 是紧的，$f(x) = \text{下确界}_{y \in Y} h(x,y)$ 是连续的。

证明. 开集 $(-\infty, a)$ 和 (b, ∞) 构成 $\mathbb{R}$ 的拓扑的一组基。因此，我们证明它们在 f 下的原像是开的。首先，注意到 $x \in f^{-1}((-\infty,a))$ 当且仅当对于某个 $y \in Y$，$(x,y) \in h^{-1}((-\infty,a))$。因此，$f^{-1}((-\infty,a)) = p_X(h^{-1}((-\infty,a)))$，其中 $p_X \colon X \times Y \to X$ 是投影到第一个变量的函数。由于 p_X 是连续且开的，因此 $p_X(h^{-1}((-\infty,a)))$ 是开的。

当 $x \in f^{-1}((b,\infty))$ 时，这意味着对于所有 $y \in Y, h(x,y) > b$。这意味着对于所有 $x \in f^{-1}((b,\infty))$ 和 $y \in Y$，都有 $(x,y) \in h^{-1}((b,\infty))$。由于 $h^{-1}((b,\infty)$ 是开集，因此存在一个开盒 $U_{x,y} \times V_{x,y} \subseteq h^{-1}((b,\infty))$，包含 (x,y)。于是，并集 $\cup_{y \in Y}\big(U_{x,y} \times V_{x,y}\big)$ 覆盖了 $\{x\} \times Y$。由于 Y 是紧的，因此存在一个大小为 K_x 的有限子覆盖 $\cup_{y_k}^{K_x}\big(U_{x,y_k} \times V_{x,y_k}\big)$。注意，开集 $A_x := \cap_{y_k}^{K_x} U_{x,y_k}$ 是 x 的一个邻域，且 $A_x \times Y \subseteq h^{-1}((b,\infty))$。因此，$A_x \subseteq f^{-1}((b,\infty))$；又因为 $x \in f^{-1}((b,\infty))$ 是任意选取的，所以 $f^{-1}((b,\infty))$ 是开集。 □

定理 18. 如果 $\mathcal{C}$ 能够对 Y 上任意连续的 $\mathfrak{G}$-不变函数进行通用逼近，那么 $\mathcal{C}$ 也具有成对 $Y_{\mathfrak{G}}$ 判别能力。

证明。取 $y, y' \in Y$ 使得 $y \not\simeq y'$。那么，函数 $\delta_y(x) = \inf_{\mathfrak{g} \in \mathfrak{G}} d(y, \mathfrak{g}x) = \min_{\mathfrak{g} \in \mathfrak{G}} d(y, \mathfrak{g}x) > 0$，其中第二个等式由 $\mathfrak{G}$ 的紧致性得出。该函数是 $\mathfrak{G}$-不变的。为证明其连续性，注意函数 $h(x, \mathfrak{g}) = d(y, \mathfrak{g}x)$ 可表示为复合函数 $d_y \circ a$，其中 $a \colon X \times \mathfrak{G} \to X$ 是连续的群作用，而 $d_y \colon X \to \mathbb{R}$ 由 $d_y(x) = d(y,x)$ 给出，且该函数也连续。由于连续函数的复合仍连续，且 $\delta_y(x) = \inf_{\mathfrak{g} \in \mathfrak{G}} h(x, \mathfrak{G})$，由引理 28 可知，$\delta_y$ 是连续函数。

给定一个普适逼近函数类 $\mathcal{C}$，我们可以找到一个函数 f，以精度 $\epsilon < {}^{\delta_y(y')}/_2$ 逼近 δ_y，因此，$f(y) \neq f(y')$。

□

定义 29. 设 $\mathcal{C}$ 是一个函数类 $X \to \mathbb{R}$ 和 $Y \subseteq X$。我们称 $\mathcal{C}$ 能够定位 Y 上的每个轨道，如果对于任意$y \in Y$和任意 $\varepsilon > 0$，都存在 $\delta_y \in \mathcal{C}$，使得：

1. 对所有 $y' \in Y, \delta_y(y') \geq 0$。

2. 对所有 $y' \in Y$，若 $y \simeq y'$，则 $\delta_y(y') = 0$。

3. 存在 $r_y > 0$，使得对于任意 $y' \in Y$，若 $\delta_y(y') < r_y$，则存在一个 $\mathfrak{g} \in \mathfrak{G}$，使得 $d(y', \mathfrak{g} \cdot y) < \varepsilon$。

注意，由于 $\delta_y \in \mathcal{C}$，它是 $\mathfrak{G}$-不变的，因此对于任意 $y^* \in \mathcal{O}_{\mathfrak{G}}(y')$，$\delta_y(y') = \delta_y(y^*)$，并且存在 $\mathfrak{g} \in \mathfrak{G}$，使得 $d(y^*, \mathfrak{g} \cdot y) < \varepsilon$。因此，直观而言，我们应将 δ_y 视为某种“距离函数”，用于衡量所有$y^* \in \mathcal{O}_{\mathfrak{G}}(y')$与 y的轨道之间的距离。换言之，当 $\delta_y(y^*)$ 较小时，意味着 y^* 的整个轨道都接近 y的轨道。

引理 30。 如果 $\mathcal{C}$ 是一组来自 $X \to \mathbb{R}$ 的连续 $\mathfrak{G}$-不变函数，且这些函数两两具有 $Y_{\mathfrak{G}}$-判别能力，那么 $\mathcal{C}^{+1}$ 能够定位 Y 上的每个轨道。

证明。 任取一个 $y \in Y$。对于 $y' \not\simeq y$，令 $\delta_{y,\,y'}$ 为 $\mathcal{C}$ 中用于区分 y 与 y' 的函数。考虑半径为 $r_{y,\,y'}$ 的情形：$= \frac{1}{2}|\delta_{y,\,y'}(y) - \delta_{y,\,y'}(y')| > 0$，并定义集合

$$A_{y'} := \delta_{y,y'}^{-1}\left(\delta_{y,y'}(y') - r_y, \delta_{y,y'}(y') + r_y\right)$$

由于 $\delta_{y,\,y'}$ 连续，因此 $A_{y'}$ 是开集。若 $y' \simeq y$，则定义 $A_y = B(y,\ \varepsilon)$，即 X 中以 y 为中心、半径为 ε 的开球。显然，$\bigcup_{y' \in Y} A_{y'}$ 构成了 Y 的一个覆盖。由于 Y 是紧集，因此存在一个由有限子集 $Y_0 \subseteq Y$ 给出的有限子覆盖，使得 $\bigcup_{y' \in Y_0} A_{y'}$。

我们在 X 上构造函数 $\delta_y(y')$，令其为 $\delta_y(y') := \sum_{y' \in Y_0 \setminus \mathcal{O}_{\mathfrak{G}}(y^*)} \bar{\delta}_{y,\,y'}(y^*)$，其中 $\bar{\delta}_{y,\,y'}(y^*) := \max(\frac{4}{3} r_{y,\,y'} - |\delta_{y,\,y'}(y^*) - \delta_{y,\,y'}(y')|,\ 0)$。由于 $\delta_{y,\,y'}$ 是连续且 $\mathfrak{G}$-不变的，δ_y 也如此。最后，可以证明 δ_y 确实能够在 Y 上定位 y 的轨道。

1. 显然，对于任意 $x \in X$，$\delta_y(x) \geq 0$。

2. 对于任意 $y^* \in Y$，若 $y \simeq y^*$，则由于 $\delta_{y,\,y'}$ 是 $\mathfrak{G}$-不变的，我们有 $\delta_{y,\,y'}(y^*) = \delta_{y,\,y'}(y)$，因此 $\delta_y(y^*) = 0$。

3. 首先，考虑 $y^* \in Y$使得 $\forall \mathfrak{g} \in \mathfrak{G}, d(\mathfrak{g} \cdot y^*, y) \geq \varepsilon$。然后，$y^* \in Y \setminus \bigcup_{y' \in \mathcal{O}_{\mathfrak{G}}(y)} A_{y'}$。接着，必定存在一个 $y' \in Y_0 \setminus \mathcal{O}_{\mathfrak{G}}(y)$，使得 $y^* \in A_{y'}$。因此，$|\delta_{y,y'}(y^*) - \delta_{y,y'}(y')| < r_{y,y'} < \frac{4}{3} r_{y,y'}$。然后，我们有 $\frac{4}{3} r_{y,y'} - |\delta_{y,y'}(y^*) - \delta_{y,y'}(y')| > \frac{4}{3} r_{y,y'} - r_{y,y'} = \frac{1}{3} r_{y,y'} > 0 \Rightarrow \bar{\delta}_{y,y'}(y^*) > \frac{1}{3} r_{y,y'}$。因此，我们可以令 $r_y := \frac{1}{3} \min_{y' \in Y_0 \setminus \mathcal{O}_{\mathfrak{G}}(y)} r_{y,y'} > 0$。如果 $\delta_y(y^*) < r_y$，则可推出对所有 $y' \in Y_0 \setminus \mathcal{O}_{\mathfrak{G}}(y^*)$，$\bar{\delta}_{y,y'}(y^*) < \frac{1}{3} r_{y,y'}$，这意味着 $y^* \in \bigcup_{\mathfrak{g} \in \mathfrak{G}} B(\mathfrak{g} \cdot y, \varepsilon)$。最后，这证明存在一个 $\mathfrak{g} \in \mathfrak{G}$ 使得 $d(y^*, \mathfrak{g} \cdot y) < \varepsilon$。

由于 绝对值函数可以使用 ReLU 激活实现，因此不难看出 $\delta_y \in \mathcal{C}^+$ [1]. □

引理 31. 如果 $\mathcal{C}$ 是紧集 Y 上的一个函数类，并且能够定位每个同构类，那么 $\mathcal{C}^{+2}$ 在 Y 上是通用逼近的。

证明。 考虑 X 上任意连续且具有 $\mathfrak{G}$ 不变性的函数。由于 $Y \subseteq X$ 是紧的，因此限制在 Y 上时，f 是一致连续的。换言之，对于所有 $\varepsilon > 0$，存在 $r > 0$，使得对于所有 $y_1, y_2 \in Y$，若 $d(y_1, y_2) < r$，则 $|f(y_1) - f(y_2)| < \varepsilon$。

令 $y \in Y$，并定义 $B_{\mathfrak{G}}(y, r) := \bigcup_{y' \in \mathcal{O}_{\mathfrak{G}}(y)} B(y', r)$ 为 y轨道上所有半径为 r 的开球的并集。利用定义 29 中的函数 δ_y，存在 r_y，使得对于任意 $y \in Y$，$\delta_y^{-1}([0, r_y)) \subseteq B_{\mathfrak{G}}(y, r)$。由于 δ_y 是连续的，$\delta_y^{-1}([0, r_y))$ 是开集。因此，$\{\delta_y^{-1}([0, r_y))\}_{y \in Y}$ 是 Y 的一个开覆盖。由于 Y 是紧的，我们可以找到一个有限子覆盖 $\{\delta_y^{-1}([0, r_y))\}_{y \in Y_0}$，其中 Y_0 是 Y 的有限子集。

现在，我们可以利用函数 δ_y 构造一组连续的 $\mathfrak{G}$-不变函数，为这个有限覆盖构成一个单位分解。对于 $y_0 \in Y_0$，我们构造函数 $\phi_{y_0}(y') = \max(r_{y_0} - \delta_{y_0}(y'),\ 0)$ 以及函数 $\phi(y') =$

$\sum_{y^* \in Y_0} \phi_{y^*}(y')$，这两个函数都是连续的。注意到 $\text{supp}(\phi_{y_0}) = \delta_{y_0}^{-1}([0, r_{y_0}))$，且对于任意$y' \in Y$，$\phi_{y^*}(y') > 0$，函数组 $\psi_{y_0}(y') = {}^{\phi_{y_0}(y')}/_{\phi(y')}$ 构成一个单位分解，其中 $\sum_{y_0 \in Y_0} \psi_{y_0}(y') = 1$ 对所有 $y' \in Y$ 都成立。

注意，我们可以将任意 $\mathfrak{G}$-不变函数 f 写成：

$$f(y') = f(y') \sum_{y_0 \in Y_0} \psi_{y_0}(y') = \sum_{y_0 \in Y_0 \,:\, y' \in \delta_{y_0}^{-1}([0, r_{y_0}))} f(y') \psi_{y_0}(y') \tag{37}$$

直观来看，由于 f 是连续的，我们可以用 $f(y_0)$的值来近似上述表达式中的 $f(y')$，因为y' 位于某个 y_0的邻域中。因此，近似 f 的函数为 $h(y') = \sum_{y_0 \in Y_0} f(y_0)\psi_{y_0}(y')$。现在我们证明，$h$ 可以以任意精度近似 f。

若 $y' \in \delta_{y_0}^{-1}([0, r_{y_0}))$，则存在 $\mathfrak{g} \in \mathfrak{G}$，使得 $d(y', \mathfrak{g} \cdot y_0) < r$。利用 f是连续的这一事实，可得$|f(y') - f(\mathfrak{g} \cdot y_0)| < \varepsilon$。由于 f是不变的，$f(y_0) = f(\mathfrak{g} \cdot y_0)$，这意味着 $|f(y') - f(y_0)| < \varepsilon$。于是我们有：

$$\left| f(y') - \sum_{y_0 \in Y_0} f(y_0)\psi_{y_0}(y') \right| = \left| f(y') - \sum_{y_0 \in Y_0 \,:\, y' \in \delta_{y_0}^{-1}([0, r_{y_0}))} f(y_0)\psi_{y_0}(y') \right| \tag{38}$$

$$= \sum_{y_0 \in Y_0 \,:\, y' \in \delta_{y_0}^{-1}([0, r_{y_0}))} |f(y') - f(y_0)|\, \psi_{y_0}(y') < \varepsilon \tag{39}$$

最后，为了证明 h 属于 $\mathcal{C} + 2$，我们可以使用一个具有一层隐藏层的多层感知机来近似 ψ_{y_0}，然后再通过一个最终层计算 ψ_{y_0}的线性组合。 □

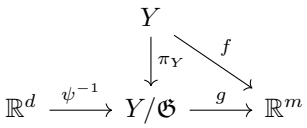
定理 19. 如果 $\mathcal{C}$（一个定义在 Y上的函数类）具有两两 $Y_{\mathfrak{G}}$ 判别性，那么 $\mathcal{C}^{+2}$ 也能在 Y上普遍逼近任意连续函数。

证明. 证明可由上述两个引理得出。 □

F.2. 连续情形下聚合器的数量

定理 20. 设 X 是一个光滑的 n维流形，且 $\mathfrak{G}$ 是一个 m维紧李群，它连续作用于 X。假设存在 X 中的一个光滑子流形 Y，其维数相同，并且 $\mathfrak{G}$ 在其上自由作用。那么，任何 $\mathfrak{G}$-轨道单射函数 $f\colon X \to \mathbb{R}^d$ 都要求 $d \geq n - m$。

证明. 为了导出矛盾，假设存在一个轨道空间单射且连续的函数 $f\colon X \to \mathbb{R}^m$，其中 $m < d$。由于 Y是一个与 X维数相同的子流形，因此 f 在 Y上也必须是单射。根据商流形定理（Lee，2013），$Y/\mathfrak{G}$ 是一个维数为 $d = \dim X - \dim \mathfrak{G}$的拓扑流形。映射 f诱导出一个单射函数 $g\colon Y/\mathfrak{G} \to \mathbb{R}^m$。该映射也是连续的，因为对于开集 $V \in \mathbb{R}^m$，$g^{-1}(V) = \pi_Y(f^{-1}(V))$。由于 f 连续且 π_Y是开映射，该集合是开集。



由于 $Y/\mathfrak{G}$ 是一个流形，因此存在一个开集 $U \subseteq Y/\mathfrak{G}$ 和一个同胚 $\psi\colon U \to \mathbb{R}^d$。那么复合映射$h = g \circ \psi^{-1}$ 是从 $\psi(U) \subseteq \mathbb{R}^m$ 到 $\mathbb{R}^m$的连续单射映射。根据区域不变性定理（Bredon，2013）（推论 19.9），h 是开映射，且它是到其像 $h(\psi(U)) \subseteq \mathbb{R}^m$的同胚。根据维数不变性定理（Bredon，2013）（推论 19.10），$d = m$。 □

定理 21. 对于 $n \geq d - 1 > 0$ 或 $n = d = 1$，任意连续的 $S_n \times SO(d)$ 轨道空间单射函数 $f\colon \mathbb{R}^{n \times d} \to \mathbb{R}^q$都要求 $q \geq nd - d(d-1)/2$。

证明。 现在考虑 $\mathfrak{G} = S_n \times SO(d)$ 的情形。首先请注意，上述证明对这个群同样成立，因为它是 $S_n \times O(d)$ 的子群。不过，我们可以得到一个更强的结果，并证明该结果对 $n \geq d-1$ 也成立。下文沿用上述证明中的记号。

我们定义集合

$$Z' = \{\boldsymbol{X} \in X \mid \exists\, 1 \leq i_1 < \ldots < i_{d-1} \leq n \text{ s.t } \boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_{d-1}} \text{ are linearly independent } \},$$

其中包含 $d-1$ 个线性无关的行向量。定义 $M_{\boldsymbol{X}}$ 为矩阵 $\boldsymbol{X}$的所有 $(d-1)$ 阶 $\times(d-1)$ 子式的集合。然后，我们可以构造一个连续函数 $h(\boldsymbol{X}) = \max_{m \in M_{\boldsymbol{X}}} |m|$，并注意到 Z' 与开集 $h^{-1}((0,\infty))$重合。于是，集合 $V = Y \cap Z'$ 也是开且非空的。因此， V是 X 的一个子流形，且维数相同； $\mathfrak{G}$ 在 W上的群作用定义良好且连续。我们还可以证明，该群作用是自由的。

如上面的证明所示，我们有 $\boldsymbol{P}_{\mathfrak{g}} = \boldsymbol{I}_n$，因此只需考察 $\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q}_{\mathfrak{g}}$的情形。任何非平凡旋转 $\boldsymbol{Q}_{\mathfrak{g}}$都必须旋转 $\mathbb{R}^d$中的至少一个二维子空间。由于矩阵 $\boldsymbol{X}$ 的各行张成$(d-1)$维子空间 $\mathbb{R}^d$，因此 $\boldsymbol{Q}_{\mathfrak{g}}$ 不可能在不满足 $\boldsymbol{X}$ 不变的情况下保持 $\boldsymbol{Q}_{\mathfrak{g}} = \boldsymbol{I}_d$。再次应用定理20 即可得到结论。

对于 $n = d = 1$, $\|\cdot\|$： $\mathbb{R}^{1\times d} \to \mathbb{R}$与之前一样，是 $\mathfrak{G}$ 轨道空间单射。 □

定理 22. 对于 $n \geq d > 0$，任意连续的 $S_n \times O(d)$ 轨道空间单射函数 $f\colon \mathbb{R}^{n\times d} \to \mathbb{R}^q$ 都要求$q \geq nd - d(d-1)/2$。

证明。 首先，假设 $n \geq d > 1$。考虑子空间 $Y = \{\boldsymbol{X} \in X \mid \|\boldsymbol{x}_i\| \neq \|\boldsymbol{x}_j\|, \forall i < j\}$，其中范数就是标准欧几里得范数。考虑函数 $g\colon X \to \mathbb{R}$，其定义为 $g(\boldsymbol{X}) = \min_{i<j} |\|\boldsymbol{x}_i\| - \|\boldsymbol{x}_j\||$。根据标准分析，该函数是连续的，并且注意到 $Y = g^{-1}((0,\infty))$，这意味着 Y在 X中是开的。我们还定义集合

$$Z = \{\boldsymbol{X} \in X \mid \exists\, i_1 < \ldots < i_d < n, |\det(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_d})| > 0\},$$

包含张成 $\mathbb{R}^d$ 的行向量。如上所述，该集合是绝对行列式在 $(0,\ \infty)$ 上的原像，因此 Z 在 X 中是开集。于是，集合 $W = Y \cap Z$ 也是开集且非空。因此， W 是 X 的一个同维子流形，并且 $\mathfrak{G}$ 在 W 上的群作用定义良好且连续。我们可以证明该群作用是自由的。

我们考察方程 $\boldsymbol{P}_{\mathfrak{g}}\boldsymbol{X}\boldsymbol{Q}_{\mathfrak{g}}^{\top} = \boldsymbol{X} \iff \boldsymbol{P}_{\mathfrak{g}}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q}_{\mathfrak{g}}$ 在 $\boldsymbol{X} \in W$上的解。由于正交变换保持范数，且 $\boldsymbol{X}$ 的各行范数互不相同，因此 $\boldsymbol{P}_{\mathfrak{g}} = \boldsymbol{I}_n \Rightarrow \boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q}_{\mathfrak{g}}$。我们知道， $\boldsymbol{X}$ 的某个行子集张成了整个 $\mathbb{R}^d$。将包含这些行的 $\boldsymbol{X}$ 子矩阵定义为 $\boldsymbol{X}^* \in \mathbb{R}^{d\times d}$。于是有 $\boldsymbol{X}^*\boldsymbol{Q}_{\mathfrak{g}} = \boldsymbol{X}^* \Rightarrow \boldsymbol{Q}_{\mathfrak{g}} = (\boldsymbol{X}^*)^{-1}\boldsymbol{X}^* = \boldsymbol{I}_d$。这证明了该群作用是自由的，并应用定理 20即可得到结论。

对于 $n = 1$ 的平凡情形，注意到 $\|\cdot\|\colon \mathbb{R}^{1\times d} \to \mathbb{R}$ 是 $\mathfrak{G}$ 轨道空间单射。 □

命题 23. 任意 S_n-不变单射函数 $f\colon \mathbb{R}^{n\times d} \to \mathbb{R}^q$ 都要求 $q \geq nd$。

证明。 沿用上文的记号，注意对于所有 $n \geq 1$， S_n 都在子流形 Y 上自由作用，如上所示。将 S_n 视为一个零维李群，并应用定理 20 即可得到结果。 □