利用分划代数快速计算置换等变层

Charles Godfrey，[1] Michael G. Rawson，[1] Davis Brown，[1] 和 Henry Kvinge[1,2]1太平洋西北国家实验室，[2] 数学系，华盛顿大学 {first}.{last}@pnnl.gov

arXiv:2303.06208v1 [cs.LG] 10 Mar 2023

摘要

对输入置换具有等变性或不变性的线性神经网络层，是现代深度学习架构的核心构建模块。例子包括 DeepSets 的层，以及 Transformer 注意力模块和某些图神经网络中的线性层。置换等变线性层的空间可被视为某个对称群表示的不变子空间。近期研究通过构造一组基来参数化该空间，其中的基向量是标准基元素在对称群作用下各轨道上的求和。这样的参数化使得通过梯度下降学习置换等变线性层的权重成为可能。置换等变线性层的空间是分划代数的一种推广。分划代数最初是在统计物理学中发现的，与对称群的表示论有着深刻联系；上述基则推广了分划代数中所谓的轨道基。我们提出另一种基，它推广了分划代数的图示基，并具有计算上的优势：构成该基的张量是低秩的，因而可以自然地分解为克罗内克积。正如使用秩一矩阵进行乘法的开销远低于使用任意矩阵，使用这些低秩张量进行乘法的开销也远低于使用轨道基中的元素。最后，我们描述一种实现与这些基元素相乘的算法。

1 引言

二十多年来，应用驱动的对称群的不变性或等变性一直是设计神经网络架构的指引，其历史至少可以追溯到卷积网络的提出（Fukushima，1980）。当底层对称性是置换时，过去五年间出现了若干类架构：DeepSets（Zaheer 等人，2018）（及其后续发展）、在图上运行的神经网络（其中对节点置换不变自然是一个期望性质）（Maron 等人，2018），以及 Transformer（Vaswani 等人，2017）。

置换等变线性层是线性映射 $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$ 相对于对称群 Σ_n 在定义域和目标空间的张量上以“对角”方式作用时保持等变的线性映射（第 3 节提供背景知识和更详细的定义）。这类线性层包括 DeepSets 中使用的线性层以及 Transformer 的一些层（关于后一种情况，参见（Kim 等人，2021 年）、（Bronstein 等人，2021年，§5.4）），它们是置换等变/不变模型的基本构建模块。（Maron 等人，2018，定理 1）的结果计算出了参数化置换等变线性层的向量空间的一组基，由此开启了 学习这类层的可能性，即将参数视为权重，并通过某种形式的梯度下降对其进行优化。

出于计算效率方面的充分考虑，我们不采用（Maron 等人，2018）中引入的基：在一种自然的对应关系下，置换等变线性层可以视为生活在 $(\mathbb{R}^n)^{\otimes(m+m')}$ 中的张量。正如对于形如 uv^T 的秩一矩阵（其中 u 和 v 是向量），使用顺序点积执行乘法，比与任意矩阵相乘快得多一样，与可分解为克罗内克积的 $(\mathbb{R}^n)^{\otimes(m+m')}$ 中的张量进行缩并（例如 $u \otimes v$，其中 $u \in (\mathbb{R}^n)^{\otimes p}$、$v \in (\mathbb{R}^n)^{\otimes p'}$ 和 $p + p' = m + m'$），通常会

比与 $(\mathbb{R}^n)^{\otimes(m+m')}$ 中的任意张量进行缩并的计算开销更低。遗憾的是，（Maron 等人，2018，定理 1）中找到的许多基向量都无法进行这种分解（附录 D 中给出了一个示例）。

我们的主要结果——下文的定理 5.4——给出了置换等变线性层的一种替代基，其中除一个基向量外，其余基向量都由克罗内克积显式构造。在这一构造中，我们利用了这样一个事实：从零阶到无穷阶的置换等变线性层的向量空间直和构成一个代数，其中的乘法就是克罗内克积。
在 $m = m'$ 的情况下，这个基是已知的：置换等变层 $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m}$恰好构成分划代数。这一对象最早于 20 世纪 90 年代中期在统计力学领域被发现，此后一直受到数学家的广泛研究（Halverson 与 Ram，2005；Benkart 与 Halverson，2017），而定理 5.4 中的基正是分划代数的“图示基”（Maron 等人，2018，定理 1）所用的基称为“轨道和”基）。我们将这种联系视为一个绝妙的例子：一个源自物理学的思想进入了深度学习架构，如今这些架构正被用于对物理现象建模，例如利用 ZINC 数据集预测分子几何（Sterling 与 Irwin，2015年）。

为便于具体说明，我们全文均在 $\mathbb{R}$ 上讨论，但本文的所有数学结果都适用于任意域，包括复数域 $\mathbb{C}$ 和有限域 $\mathbb{Z}/2$。这并非为了追求一般性而追求一般性，因为后者应用于量化神经网络。

本文的贡献包括：(i) 为置换等变线性层空间显式构造一种替代基；(ii) 验证当置换等变线性层构成分划代数时，我们的基与图示基一致；(iii) 概述一种利用这些基元素实现乘法的算法，并在此过程中说明该算法能够还原（Pan 与 Kondor，2022）中的高效运算。

2 相关工作

在附录 C 中，我们证明定理 5.4 的基能够还原（Pan 与 Kondor，2022，第 4–5 节）中描述的运算。不过请注意，虽然他们的工作确实证明了其定义的“求和/传递/广播”张量集合与（Maron 等人，2018，定理 1）中的基具有相同基数，但并未证明这些“求和/传递/广播”张量是线性无关的，也未证明它们张成置换等变线性层的空间。
根据（Pan 与 Kondor，2022，附录），定理 5.4 的基也与（Maron 等人，2018，附录）中使用的基一致。然而，后者也没有给出线性无关性或张成性的证明。置换等变线性层空间的维数会随着 $m + m'$ 极其迅速地增长，并且（当 n 足够大时）等于贝尔数 $B(m + m')$，见（Maron 等人，2018，定理 1）。因此，尽管他们的工作中考察的 $m = m' = 2$ 情形可能作为一个单独的特例尚可处理，但更高的 $m + m'$ 情形则无法处理。
分划代数最初由（Martin，1991；Jones，1994）独立发现。其结构（以生成元与关系的形式表示）以及与对称群表示论的联系，见（Halverson 与 Ram，2005）（近期进展另见（Benkart 与 Halverson，2017））。在 $m = m'$ 的情形下，我们的定理 5.4 只是用对角张量的克罗内克积，明确描述了上述图示基。
完成这项工作后，我们注意到近期发表的文章（Pearce-Crump，2023）也指出了置换等变线性层与分划代数之间的联系。该文章在技术上与我们的第 3 节有重叠（事实上，这部分也与多篇参考文献有所重叠），但没有包含我们的主要结果（即第 4 节和第 5 节的结果），也没有涉及用于与表示置换等变映射的张量相乘的算法（例如附录 C 中的算法），也没有讨论与（Pan 与 Kondor，2022，第 4–5 节）的运算之间的联系。

3 背景

我们固定一个自然数 $n \in \mathbb{N}$。在实际应用中，它表示置换等变/不变模型所作用的集合的基数。向量空间 $\mathbb{R}^n$ 带有一个自然作用，该作用来自

作用于 n 个元素上的对称群，记作 Σ_n，它置换基向量 $\{e_1,\dots,e_n\}$:

$$\sigma\cdot e_i = e_{\sigma(i)} \text{ for } \sigma\in\Sigma_n, i=1,\dots,n. \tag{3.1}$$

对于每个自然数 $m\in\mathbb{N}$， m阶张量幂

$$(\mathbb{R}^n)^{\otimes m} = \underbrace{\mathbb{R}^n\otimes\mathbb{R}^n\otimes\cdots\otimes\mathbb{R}^n}_{m \text{ times}} \tag{3.2}$$

也是 Σ_n 的一个表示，配备“对角”作用，其定义如下：对于一个 $\sigma\in\Sigma_n$ 和基张量 sor $v_1\otimes\cdots\otimes v_m\in(\mathbb{R}^n)^{\otimes m}$ 作为 $\sigma\cdot(v_1\otimes\cdots\otimes v_m)$: $=(\sigma\cdot v_1)\otimes\cdots\otimes(\sigma\cdot v_m)$。注释 3.3。当 $m=0$ 时，我们采用约定 $(\mathbb{R}^n)^{\otimes m}=\mathbb{R}$，并赋予平凡的 Σ_n 作用。

我们将所有（不一定是等变的）线性映射组成的向量空间 $(\mathbb{R}^n)^{\otimes m}\to(\mathbb{R}^n)^{\otimes m'}$记为 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})$。线性映射 $\varphi\in\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})$ 是 Σ_n- 等变的，当且仅当对于所有 $\sigma\in\Sigma_n$、 $v\in(\mathbb{R}^n)^{\otimes m}$，都有 $\varphi(\sigma\cdot v)=\sigma\cdot\varphi(v)$；我们将这类$\Sigma_n$-等变线性映射组成的向量空间记为 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$。

问题 3.4。参数化 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$，例如，将其作为 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})$ 的子空间，给出一个基。

接下来的两个初等结果将在后文中发挥作用。证明留待附录给出。引理 3.5 (参见 (Halverson 与 Ram, 2005, §3))。$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ 同构于张量幂 $(\mathbb{R}^n)^{\otimes(m+m'}$ 中的 Σ_n-不变子空间)。

推论 3.6。$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ 的维数是和 $m+m'$ 的函数（固定 n 时）。此外，给定某个$l\in\mathbb{N}$ 对应的 $(\mathbb{R}^n)^{\otimes l}$ 的 Σ_n-不变子空间的任意参数化，都有一种简单的方法可据此为所有$m+m'=l$生成 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m},(\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ 的参数化。

引理 3.5 和推论 3.6 将问题 3.4 简化为计算不变子空间 $((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$。现在我们介绍关于划分的记号。给定一个元组 $(i_1,\dots,i_l)\in\{1,\dots,n\}^l$，定义子集 $S(i_1,\dots,i_l)_j\subseteq\{1,\dots,l\}$，其中 $j=1,\dots,n$ 满足 $S(i_1,\dots,i_l)_j=\{k\in\{1,\dots,l\}\,|\,i_k=j\}$。根据构造，这些子集两两不相交，且其并集为 $\{1,\dots,l\}$；因此，它们构成 $\{1,\dots,l\}$ 的一个集合划分，我们将其记作

$$\Pi(i_1,\dots,i_l)=\{S(i_1,\dots,i_l)_j\,|\,j=1,\dots,n\}. \tag{3.7}$$

由此，我们得到一个从元组 $(i_1,\dots,i_l)\in\{1,\dots,n\}^l$ 到 $\{1,\dots,l\}$ 的集合划分 $\Pi(i_1,\dots,i_l)$ 的映射。下一个引理表明，划分 $\Pi(i_1,\dots,i_l)$ 唯一确定了 $(i_1,\dots,i_l)$ 的 Σ_n-轨道。关于该引理的具体示例说明，请参见附录 D。

引理 3.8（参见（Jones，1994，第 1 节）、（Benkart 与 Halverson，2017，第 5.2 节））。元组 $(i_1,\dots,i_l)$ 和 $(i'_1,\dots,i'_l)$ 属于同一个 Σ_n 轨道，当且仅当它们产生 $\{1,\dots,l\}$ 的同一个划分，即 $\Pi(i_1,\dots,i_l)=\Pi(i'_1,\dots,i'_l)$。

Maron 等人Maron 等人(2018）通过展示如下所述的一组基，解决了问题 3.4：设 $\mathcal{P}$ 是 $\{1,\dots,l\}$的一个固定集合划分，并定义

$$e_{\mathcal{P}}=\sum_{(i_1,\dots,i_l):\Pi(i_1,\dots,i_l)=\mathcal{P}} e_{i_1}\otimes e_{i_2}\otimes\cdots\otimes e_{i_l}\in\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{3.9}$$

定理 3.10 ((Maron 等人，2018，定理 1))。向量 $\{e_{\mathcal{P}}\,|\,\mathcal{P}\text{是 }\{1,\dots,l\}\}$的一组基 $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$。

我们在附录 A 中说明这一构造背后的直观思路。

4 分划代数 $T(\mathbb{R}^n)$ 及其 Σ_n-不变子代数

不变子空间 $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ 组合起来构成一个代数。该代数中的乘法可以解释为一种通过取张量积，将两个置换等变线性层 $(\mathbb{R}^n)^{\otimes m}\to(\mathbb{R}^n)^{\otimes m'}$ 和 $(\mathbb{R}^n)^{\otimes p}\to(\mathbb{R}^n)^{\otimes p'}$ 组合成第三个置换等变层 $(\mathbb{R}^n)^{\otimes(m+p)}\to(\mathbb{R}^n)^{\otimes(m'+p'}$)，从而“由旧层生成新的置换等变层”的运算。

引理 4.1. 对任意 $p, p' \in \mathbb{N}$，只要 $p + p' = l$，就存在一个自然双线性映射

$$\left((\mathbb{R}^n)^{\otimes p}\right)^{\Sigma_n} \times \left((\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \to \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \textit{ taking } (v, w) \mapsto v \otimes w. \tag{4.2}$$

直和 $\bigoplus_{l\in\mathbb{N}} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ 构成一个 $\mathbb{R}$-代数，因为引理 4.1 中的双线性映射在适当意义下满足结合律和单位律。

5 置换等变线性层的新旧基

对于 $\{1, \dots, l\}$ 的每个划分 $\mathcal{P}$，我们令 $e_{\mathcal{P}} \in \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ 表示由引理 3.8 和 A.2 关联的基元素。我们认为，还可以用另一种方式从划分生成基元素，这种方式可能具有计算优势：给定 $\mathcal{P} = \{S_1, \dots, S_n\}$，存在一个多重线性映射

$$\Phi_{\mathcal{P}} : \prod_{i=1}^{n} \left((\bigotimes_{j=1}^{|S_i|} \mathbb{R}^n)^{\Sigma_n} \right) \xrightarrow{\mu} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \xrightarrow{\tau} \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \tag{5.1}$$

其中，μ 是通过反复使用引理 4.1 中的双线性映射得到的，而 τ 是 $\{1, \dots, l\}$ 的一个置换（作用于张量因子），其逆置换会将集合 S_i 排列成连续相邻的块。 [1] 在每个 $(\bigotimes_{j=1}^{|S_i|} \mathbb{R}^n)^{\Sigma_n}$ 中，都有一个基元素 $e_{\{\{1,\dots,|S_i|\}\}}$ 对应于只有一个集合的划分，也就是一个对角张量。我们定义：

$$d_{\mathcal{P}} := \Phi_{\mathcal{P}}(e_{\{\{1,\dots,|S_1|\}\}}, e_{\{\{1,\dots,|S_2|\}\}}, \cdots, e_{\{\{1,\dots,|S_n|\}\}}). \tag{5.2}$$

注 5.3。用于定义 $\Phi_{\mathcal{P}}$ 的 $\{1, \dots, l\}$（标签）的置换选择无关紧要，因为任意两种选择都相差一系列对各个 S_1、.....、S_n 的置换，而 $e_{\{S_i\}}$ 对 S_i 的置换保持不变。

定理 5.4。 $\{1, \dots, l\}\Big\}$ 的集合划分 $\left\{ d_{\mathcal{P}} \,|\, \mathcal{P} \right.$ 向量构成 $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$ 的一组基

.

如引言中所述，置换等变线性层的空间 $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m}$ 是一个分划代数。

命题 5.5。 在 $m = m'$ 情形下，定理 5.4 中构造的 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m})^{\Sigma_n}$ 的基与分划代数的图示基一致，后者在（*Jones*，*1994*）中记作“$L_\sim$”。

根据构造，$d_{\mathcal{P}}$ 中的张量都是分解后的克罗内克积。在附录 C 中，我们介绍一种算法，用于计算 Σ_n等变映射$(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$，该映射对应于与张量 $d_{\mathcal{P}} \in ((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$相乘。尽管计算成本的分析（例如以浮点运算次数衡量）留待未来工作，但该算法表明，$d_{\mathcal{P}}$ 的分解使得与这些张量相乘的计算效率很高（尤其是比与元素 $e_{\mathcal{P}}$相乘更高效，见附录 D）。附录 C 中的说明还表明，$d_{\mathcal{P}}$ 能够实现（Pan 与 Kondor，2022）中的“求和／传递／广播”运算。

6 结论与开放问题

定理 5.4 为置换等变线性层提供了一组基，这些线性层经过专门设计，具有较高的计算效率，因为构成该基的张量是以克罗内克积构造的。未来一项可行的研究方向是利用该定理，为用户指定的任意 m实现置换等变线性层，m'（据我们所知，参考文献中的实现都针对特定的$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$对$m$、$m'$的参数化进行了硬编码）。另一个方向是利用分划代数的一种表示方法，以图示基的一个子集作为生成元（见（Jones，1994，§3）、（Halverson 与 Ram，2005，定理 1.11）），从而获得置换等变线性层中一个规模较小但仍具有表达能力的子空间，正如（Kim 等人，2021年，§6）所建议的那样。

[1] 具体而言，如果 $S_i = \{s_{i1}, \dots, s_{i|S_i|}\}$，那么在式 (5.1) 中，我们可以使用将 $(1, \dots, l) \mapsto (s_{11}, \dots, s_{1|S_1|}, \dots, s_{n1}, \dots, s_{n|S_n|})$ 映射过去的置换

参考文献

Georgia Benkart 和 Tom Halverson。《分划代数与对称群的不变量理论》，2017年9月。

A Bogatskiy、Sanmay Ganguly、Thomas Kipf、Risi Kondor、David W. Miller、Daniel Murnane、Jan T. Offermann、Mariel Pettee、Phiala E. Shanahan、Chase Owen Shimmin 和 Savannah T hais。用于物理学的对称群等变架构。ArXiv，abs/2203.06153，2022。

Michael M. Bronstein、Joan Bruna、Taco Cohen 和 Petar Veli˘ckovi´c。《几何深度学习：网格、群、图、测地线与规范》，2021年5月。

Kunihiko Fukushima。新认知机：一种用于模式识别的自组织神经网络模型，不受位置偏移的影响。《生物控制论》，36(4):193–202，1980年4月。ISSN 1432-0770。DOI：10.1007/BF00344251。

Tom Halverson 和 Arun Ram。《分割代数》。《欧洲组合数学杂志》，26(6): 869–921，2005年8月。ISSN 0195-6698。DOI：10.1016/j.ejc.2004.06.005。

Jason Hartford、Devon Graham、Kevin Leyton-Brown 和 Siamak Ravanbakhsh。《集合间交互的深度模型》。载于第35届国际机器学习会议论文集，第1909–1918页。PMLR，2018年7月。

V.F.R. Jones。《Potts模型与对称群》。《子因子：谷口算子代数研讨会论文集》（Kyuzeso，1993），1994。

Jinwoo Kim、Saeyoon Oh 和 Seunghoon Hong。《Transformer泛化了DeepSets，并可扩展至图和超图》。载于神经信息处理系统，2021年。

Y. LeCun、B. Boser、J. S. Denker、D. Henderson、R. E. Howard、W. Hubbard 和 L. D. Jackel。反向传播应用于手写邮政编码识别。《神经计算》，1(4):541– 551,1989年。DOI：10.1162/neco.1989.1.4.541。

Derek Lim、Joshua Robinson、Lingxiao Zhao、Tess Smidt、Suvrit Sra、Haggai Maron 和 Stefanie Jegelka。《用于谱图表示学习的符号与基不变网络》，2022年9月。

Haggai Maron、Heli Ben-Hamu、Nadav Shamir 和 Yaron Lipman。图的不变与等变 ph ne
tworks。arXiv 预印本 arXiv:1812.09902，2018。

P.P. Martin。《统计力学中的 Potts模型及相关问题》。统计力学进展丛书。世界科技出版公司，1991。ISBN 978-981-02-0075-6。

P. P. Nikitin、N. V. Tsilevich 和 A. M. Vershik。《对称群张量表示的分解》。《代数与表示论》，22(4):895–908，2019年8月。ISSN 1572-9079。DOI：10.1007/s10468-018-9804-6。

Horace Pan 和 Risi Kondor。《用于求解置换谜题的傅里叶基》。载于 第 24 届人工智能与统计国际会议论文集，第 172–180 页。PMLR，2021年3月。Horace Pan 和 Risi Kondor。《用于高阶交互的置换等变层》。载于第 25 届人工智能与统计国际会议论文集，第 5987–6001 页。PMLR，2022年5月。Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、An- dreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。《PyTorch：一种命令式风格、高性能的深度学习库》。载于 H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alché-Buc、E. Fox 和 R. Garnett（编），神经信息处理系统进展第 32 卷，第 8024–8035 页。Curran Associates, Inc.，2019年。网址 http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-高性能-d

Edward Pearce-Crump。《Brauer 群等变神经网络》，2022年12月。

Edward Pearce-Crump。《连接置换等变神经网络与划分图》，2023年1月。

J.P. Serre。《有限群的线性表示》。数学研究生教材。施普林格纽约，1977年。ISBN 978-1-4684-9458-7。

Teague Sterling 和 John J. Irwin。《Zinc 15——人人皆可进行配体发现》。《化学信息与建模杂志》，55(11):2324–2337，2015年。DOI：10.1021/acs.jcim.5b00559。网址 https://doi.org/10.1021/acs.jcim.5b00559。PMID：26479676。

Erik Henning Thiede、Truong Son Hy 和 Risi Kondor。《置换等变神经网络与高阶图变分编码器的一般理论》，2020年4月。

Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N. Gomez、Lukasz Kaiser 和 Illia Polosukhin。《注意力机制就是你所需要的一切》，2017年12月。

Soledad Villar、David W. Hogg、Kate Storey-Fisher、Weichi Yao 和 Ben Blum-Smith。《标量具有普适性：等变机器学习，结构类似于经典物理学》。载于《神经信息处理系统》，2021年。

Edward Wagstaff、Fabian B. Fuchs、Martin Engelcke、Michael A. Osborne 和 Ingmar Posner。《关于集合上函数的普适逼近。ArXiv，abs/2107.01959，2021年。

Manzil Zaheer、Satwik Kottur、Siamak Ravanbakhsh、Barnabas Poczos、Ruslan Salakhutdinov 和 Alexander Smola。《Deep Sets》，2018年4月。

A 置换表示的不变量

本节中，我们运用表示论中的几个基本结论以及一个组合引理，给出定理 3.10 的简明证明。

定义 A.1. 设 G 是一个有限群，X 是一个带有 G-作用的有限集。与之关联的置换表示，记作 $\mathbb{R}X$，是一个向量空间，其中对每个 $x \in X$ 都有一个基元素 e_x，且 G-作用定义为 $g \cdot e_x = e_{gx}$。

Σ_n 表示 $\mathbb{R}^n$ 是一个置换表示：$\mathbb{R}^n = \mathbb{R}\{1, \dots, n\}$。一般而言，若 X_1 和 X_2 是带有有限群 G 作用的有限集，则存在一个 G-表示的同构 $\mathbb{R}(X_1 \times X_2) = \mathbb{R}X_1 \otimes \mathbb{R}X_2$（理由基本相同：基本张量 $e_i \otimes e_j$ 对于 $i = 1$、.....、m、$j = 1$、$\dots n$ 构成了 $\mathbb{R}^m \otimes \mathbb{R}^n$ 的一组基）。因此，$(\mathbb{R}^n)^{\otimes l} = \mathbb{R}(\{1, \dots, n\}^l)$ 是与集合 $\{1, \dots, n\}^l$ 相关联的置换表示，其上带有“对角” Σ_n 作用。

引理 A.2 (参见（Jones，1994，§1），（Benkart 与 Halverson，2017，§5.2）。设 G 是一个有限群，且 X 是一个带有 G-作用的有限集。则存在一个自然向量空间同构 $\mathbb{R}(X/G) \xrightarrow{\simeq} (\mathbb{R}X)^G$，其中 X/G 是 X的轨道，定义如下：对每个轨道 $Gx \in X/G$，将基向量 e_{Gx} 映射到 $\sum_{x' \in Gx} e_{x'}$。

特别地，要计算$((\mathbb{R}^n)^{\otimes l})^{\Sigma_n}$，从而根据推论 3.6，计算所有 $m + m' = l$ 对应的置换等变线性层空间 $(\mathbb{R}^n)^{\otimes m} \to (\mathbb{R}^n)^{\otimes m'}$，我们只需计算 $\{1, \dots, n\}^l$ 上对角 Σ_n 作用的轨道——这正是引理 3.8 的内容。

B 证明

引理 3.5 的证明。一般而言，对于任意两个实向量空间 V, W，存在一个同构 $V^\vee \otimes W \simeq \mathrm{Hom}(V, W)$，它将基本张量 $\lambda \otimes w$ 映射为线性映射 $\varphi\colon V \to W$，其定义为 $\varphi(v) = \lambda(v) \cdot w$。在我们的情形中，这表明

$$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'}) \simeq \left((\mathbb{R}^n)^{\otimes m}\right)^\vee \otimes (\mathbb{R}^n)^{\otimes m'}. \tag{B.1}$$

由于 $(\mathbb{R}^n)^{\otimes m}$ 是有限群 Σ_n 的实表示，它具有等变欧几里得内积：一种自然且明确的选取方式是在标准基张量上定义为

$$\langle e_{i_1} \otimes \cdots \otimes e_{i_m}, e_{j_1} \otimes \cdots \otimes e_{j_m} \rangle = \prod_{k=1}^{m} \delta_{i_k j_k}. \tag{B.2}$$

这样的内积等价于一个同构 $(\mathbb{R}^n)^{\otimes m} \simeq \left((\mathbb{R}^n)^{\otimes m}\right)^{\vee}$，由映射 $v \mapsto \langle v, - \rangle$ 给出。因此

$$\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'}) \simeq (\mathbb{R}^n)^{\otimes m} \otimes (\mathbb{R}^n)^{\otimes m'} = (\mathbb{R}^n)^{\otimes (m+m')} \tag{B.3}$$

并对两边取不变量，即完成证明。 □

推论 3.6 的证明。引理 3.5 给出了一个从 $\left((\mathbb{R}^n)^{\otimes (m+m')}\right)^{\Sigma_n}$ 到 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ 的可逆映射。于是，$(\mathbb{R}^n)^{\otimes l}$ 的参数化基被映射为 $\mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})^{\Sigma_n}$ 的参数化基。 □

引理 A.2 的证明。注意，该引理本质上就是（Serre，1977年，习题 2.6 (a)）。一个 $\sum_{x \in X} c_x x \in \mathbb{R}X$ 属于不变子空间 $(\mathbb{R}X)^G$ 当且仅当对每个 $g \in G$

$$\sum_{x \in X} c_x x = g \cdot \sum_{x \in X} c_x x = \sum_{x \in X} c_x g x = \sum_{x \in X} c_{g^{-1}x} x \tag{B.4}$$

其中在最后一步中，我们对求和重新进行了指标重排。条件 $c_x = c_{g^{-1}x}$ 对于所有 $g \in G, x \in X$ 成立，恰好说明系数 c_x 在 G-轨道上为常数。换言之，若 c_{Gx} 是当 x' 遍历轨道 Gx 时 $c_{x'}$ 的共同值，则

$$\sum_{x \in X} c_x x = \sum_{Gx \in X/G} c_{Gx} \sum_{x' \in Gx} e_{x'} \tag{B.5}$$

这表明引理 A.2 中定义的映射是满射。它显然是单射，因为例如轨道和 $\sum_{x' \in Gx} e_{x'}$ 两两正交。 □

引理 3.8 的证明。由于对任意 l，引理 3.8 的证明与先前关于分划代数的研究中所讨论的 l-偶数情形的证明完全相同，因此我们省略证明，转而提供讨论 l-偶数情形的参考文献：见（Jones，1994，§1）、（Halverson 与 Ram，2005，§3）以及（Benkart 与 Halverson，2017，§5.1）。 □

引理 4.1 的证明。我们注意到，这个证明的本质与证明对称多项式的乘积仍为对称多项式相同。

张量（即克罗内克）积定义了一个双线性映射

$$(\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'} \to (\mathbb{R}^n)^{\otimes l} \tag{B.6}$$

将 (v, w) 映射为 $v \otimes w$。取 Σ_n-不变量后，得到一个映射

$$\left((\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \to \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{B.7}$$

最后，有一个包含关系 $\left((\mathbb{R}^n)^{\otimes p}\right)^{\Sigma_n} \times \left((\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n} \subseteq \left((\mathbb{R}^n)^{\otimes p} \times (\mathbb{R}^n)^{\otimes p'}\right)^{\Sigma_n}$，因为左边由一对张量 (v, w) 组成，它们在独立置换 $\sigma, \tau \in \Sigma_n$ 如 $(\sigma \cdot v, \tau \cdot w)$ 的作用下保持不变；而这一条件比在单个置换 σ 如 $(\sigma \cdot v, \sigma \cdot w)$ 的对角作用下保持不变更强。 □

引理 B.8. 设 V 是一个实向量空间，并设

$$\{v_\alpha \in V \mid \alpha \in \mathcal{A}\} \tag{B.9}$$

是由偏序集 $\mathcal{A}$ 索引的 V 中的一组向量。假设对于每个 $\alpha \in A$，都存在一个线性泛函 $\lambda_\alpha \colon V \to \mathbb{R}$，满足

$$\lambda_\alpha(v_\alpha) \neq 0 \textit{ and } \lambda_\alpha(v_\beta) = 0 \textit{ unless } \beta \preceq \alpha, \tag{B.10}$$

其中 $\preceq$ 表示 $\mathcal{A}$上的偏序关系。于是，$\{v_\alpha \in V \mid \alpha \in \mathcal{A}\}$ 线性无关。

证明。考虑如下形式的等式

$$\sum_{\alpha\in\mathcal{A}} c_\alpha v_\alpha = 0 \tag{B.11}$$

其中只有有限多个 c_α 不为 0。反设某个 $c_\alpha \neq 0$，并令

$$\mathcal{B} = \{\alpha \in \mathcal{A} \,|\, c_\alpha \neq 0\} \subseteq \mathcal{A}. \tag{B.12}$$

根据假设，$\mathcal{B}$ 是一个非空有限偏序集，因此至少存在一个极小元，即满足以下条件的 $\alpha^* \in \mathcal{B}$：若 $\alpha \in \mathcal{B}$ 且 $\alpha \preceq \alpha^*$，则 $\alpha = \alpha^*$。由此可得，对于 $\alpha \in \mathcal{B} \setminus \{\alpha^*\}$，$\lambda_{\alpha^*}(v_\alpha) = 0$；因此将 λ_{α^*} 应用于等式 (B.11)，得到

$$c_{\alpha^*}\lambda_{\alpha^*}(v_{\alpha^*}) = 0 \tag{B.13}$$

因此 $c_{\alpha^*} = 0$（因为 $\lambda_{\alpha^*}(v_{\alpha^*}) \neq 0$），矛盾。 □

引理 B.14. 对于 $\{1,\ldots,l\}$ 的任意集合划分 $\mathcal{P}$，向量 $d_\mathcal{P}$ 等于所有指标 $J = (j_1,\ldots,j_l)$ 上的 $e_J = e_{j_1}\otimes\cdots\otimes e_{j_l}$ 之和，其中这些指标在划分 $\mathcal{P}$ 中的每个集合上均为常数——等价地，$d_\mathcal{P} = \sum_{\mathcal{P}\preceq\mathcal{Q}} e_\mathcal{Q}$，其中 $\mathcal{Q} \preceq \mathcal{P}$ 当且仅当 $\mathcal{Q}$ 细化划分 $\mathcal{P}$。

证明。对 l 张量因子进行置换后，我们可以归结为 $\mathcal{P}$ 具有如下形式的情形

$$\mathcal{P} = \coprod_{i=1}^{n} \{\sum_{j<i} p_j + 1, \ldots, \sum_{j<i} p_j + p_i\} \tag{B.15}$$

其中有 p_1、……、$p_n \in \mathbb{N}$ 和 $\sum_{i=1}^n p_i = l$。于是，通过直接计算

$$d_\mathcal{P} = \sum_{j_1,\ldots,j_n=1}^{n} e_{\underbrace{j_1,\ldots,j_1}_{p_1\text{ times}}} \otimes e_{\underbrace{j_2,\ldots,j_2}_{p_2\text{ times}}} \otimes \cdots \otimes e_{\underbrace{j_n,\ldots,j_n}_{p_n\text{ times}}}, \tag{B.16}$$

并且显然，（j_1，……，j_1，……，j_n，……，j_n）求和中出现的指标恰好是在每个$\underbrace{\quad}_{p_1\text{次}}$ $\underbrace{\quad}_{p_n\text{次}}$恒定的指标

在每个集合 $\{\sum_{j<i} p_j + 1, \ldots, \sum_{j<i} p_j + p_i\}$上。 □

定理 5.4 的证明。由于 $e_\mathcal{P}$ 和 $d_\mathcal{P}$ 都由将 $\{1,\ldots,l\}$ 划分为至多 n 个非空子集的集合划分作指标，因此它们的基数相同；并且我们已经知道 $e_\mathcal{P}$ 是 $\left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}$的一组基。因此，只需证明 $d_\mathcal{P}$ 线性无关。我们将通过构造一组线性泛函来证明这一点

$$\{\lambda_\mathcal{P} : \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n} \to \mathbb{R} \,|\, \mathcal{P} \text{ is a set partition of } \{1,\ldots,l\}\} \tag{B.17}$$

使得

$$\lambda_\mathcal{P}(d_\mathcal{P}) \neq 0 \text{ and } \lambda_\mathcal{P}(d_\mathcal{Q}) = 0 \text{ unless } \mathcal{Q} \preceq \mathcal{P}, \tag{B.18}$$

其中 $\mathcal{Q} \preceq \mathcal{P}$ 当且仅当 $\mathcal{Q}$ 细化划分 $\mathcal{P}$，并满足下文引理 B.8 的性质。具体而言，对于每个划分 $\mathcal{P}$，选取一个指标 $I_\mathcal{P} = (i_1,\ldots,i_l) \in \{1,\ldots,n\}^l$，使得 $\mathcal{P}$ 是与 $I_\mathcal{P}$ 相关联的划分，如引理 3.8 所述，并令 $e_{I_\mathcal{P}} = e_{i_1}\otimes\cdots\otimes e_{i_l} \in (\mathbb{R}^n)^{\otimes l}$。等价地，$e_{I_\mathcal{P}}$ 是 $(\mathbb{R}^n)^{\otimes l}$ 的标准基向量之一，且出现在定义 $e_\mathcal{P}$的轨道和中。然后将 $\lambda_\mathcal{P}$ 定义为与 $e_{I_\mathcal{P}}$的点积：

$$\lambda_\mathcal{P}(v) = \langle e_{I_\mathcal{P}}, v\rangle \text{ for } v \in \left((\mathbb{R}^n)^{\otimes l}\right)^{\Sigma_n}. \tag{B.19}$$

首先，根据定义，$I_\mathcal{P}$ 在划分 $\mathcal{P}$的每个集合上均为常数，因此由引理 B.14 中对 $d_\mathcal{P}$ 的刻画可得 $\langle e_{I_\mathcal{P}}, d_\mathcal{P}\rangle = 1$。其次，若 $\mathcal{Q} \not\preceq \mathcal{P}$，则由于 $\mathcal{Q}$不细化 $\mathcal{P}$，写作

$$\mathcal{P} = \{S_1,\ldots,S_n\} \text{ and } \mathcal{Q} = \{T_1,\ldots,T_n\} \tag{B.20}$$

必定存在非空的 T_i，使得对所有 j，都有 $T_i \not\subseteq S_j$。由于 $T_i \subseteq \bigcup_j S_j$，必定存在不同的$S_j$、$S_k \in \mathcal{P}$，满足 $S_j \cap T_i \neq \emptyset$、$S_k \cap T_i \neq \emptyset$。根据设计，指标 $I_\mathcal{P}$在 S_j和 S_k上取不同值；但对于 $d_\mathcal{Q}$ 中系数非零的每个 $e_J = e_{j_1}\otimes\cdots\otimes e_{j_l}$，指标 J 在 T_i上都是常数。因此 $\langle e_{I_\mathcal{P}}$、$d_\mathcal{Q}\rangle = 0$。 □

命题 5.5 的证明。这由引理 B.14 和（Jones，1994，第 263 页）（另见（Benkart 与 Halverson，2017，§4.2-3））可得。 □

一种使用张量 $d_{\mathcal{P}}$进行乘法的算法

本节将说明如何将定理 5.4 中出现的张量 $d_{\mathcal{P}} \in \left((\mathbb{R}^n)^{\otimes(m+m')}\right)^{\Sigma_n} \cong \mathrm{Hom}((\mathbb{R}^n)^{\otimes m}, (\mathbb{R}^n)^{\otimes m'})$ Σ_n 作用于张量 $v \in (\mathbb{R}^n)^{\otimes m}$。我们将在这些张量积的通常基下进行讨论，并使用如下形式的记号

$$v = \sum_{i_1,\ldots,i_m=1}^{n} v_{i_1\ldots i_m} e_{i_1} \otimes \cdots \otimes e_{i_m}. \tag{C.1}$$

给定另一个张量

$$w = \sum_{j_1,\ldots,j_l=1}^{n} w_{j_1\ldots j_l} e_{j_1} \otimes \cdots \otimes e_{j_l} \in (\mathbb{R}^n)^{\otimes l} \tag{C.2}$$

以及指标的有序元组 T和 S，满足 set(T) $\subseteq \{1,\ldots,m\}$、set(S) $\subseteq \{1,\ldots,l\}$，且元组长度为 $d \leq \min\{m,l\}$，我们将使用记号 dot($v, w,$ (S, T)) 表示在 PyTorch 中通过 tensordot($v, w,$ dims $= (S, T)$) 实现的张量缩并运算（ Paszke 等人，2019 ）。特殊情况 $T = (m, m-1, \ldots, m-d+1)$ $S = (1, \ldots, d)$ 将简记为 dot(v, w, d)。更多详情请参阅文档：https://pytorch.org/docs/stable/index.html。

展开引理 3.5 可知，我们的目标是计算 dot($v, d_{\mathcal{P}}, m$)。我们想说明的是，对于许多划分 $\mathcal{P}$，可以通过对少于 m 个指标进行多次缩并来完成这一计算；此外，出现的缩并还可以替换为求和与索引操作。

缩并 dot($v, d_{\mathcal{P}}, m$) 对 v 的张量因子以及 $d_{\mathcal{P}}$ 的前 m 个张量因子的置换保持不变。此外，只要对 dot($v, d_{\mathcal{P}}, m$) 的 m' 个因子应用逆置换，我们就可以任意置换 $d_{\mathcal{P}}$ 的最后 m' 个因子。这两个观察使我们可以将问题化简为 $\mathcal{P}$ 具有如下形式的情形：

$$(\coprod_{i=1}^{a} S_i) \coprod (\coprod_{i=1}^{b} T_i) \coprod (\coprod_{i=1}^{c} B_i) \tag{C.3}$$

其中

- $a + b + c = n$,
- 其中 S_i 是 $\{1,\ldots,m\}$ 中从 1 开始、到例如 p 结束的连续指标集,
- 其中 B_i 是 $\{m+1,\ldots,m+m'\}$ 中从例如 $p+p'$ 开始、到 $m+m'$ 结束的连续指标集，并且
- 划分 T_i，此外它们分解为 $T_i = T_i' + T_i''$，其中 T_1'、.....、T_b' 是 $\{p+1,\ldots,m\}$ 中连续且相邻的指标集合，而 T_b''、.....、T_1'' 是 $\{p+p'+1,\ldots,m+m'\}$ 中连续且相邻的指标集合。

由此可得

$$\begin{aligned} d_{\mathcal{P}} = {} & e_{\{\{1,\ldots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|S_a|\}\}} \\ & \otimes \Psi(e_{\{\{1,\ldots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|T_b|\}\}}) \\ & \otimes e_{\{\{1,\ldots,|B_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\ldots,|B_c|\}\}} \end{aligned} \tag{C.4}$$

其中 Ψ 按照 $\{p, p+1,\ldots,p+p'\}$ 的某个置换对张量因子进行置换（该置换将每个 T_i 分成子集 T_i'、T_i''）。

现在我们反复使用两个简单的计算：首先,

$$\mathrm{dot}(v, w, d) = \mathrm{dot}(v, w', d) \otimes w'' \text{whenever } w = w' \otimes w'' \text{ where } w' \text{ has} \geq d \text{ indices.} \tag{C.5}$$

也就是说， w'' 可以从点中提取。另一方面,

$$\text{点 } \mathrm{dot}(v, w, d) = \mathrm{dot}(\mathrm{dot}(v, w', d'), w'', d - d') \text{whenever } w = w' \otimes w'' \text{ where } w' \text{ has } d' \leq d \text{ indice} \tag{C.6}$$

特别地，式 (C.5) 适用于式 (C.4) 中的因子 e_{B_i}，得出

$$\begin{aligned}&\mathrm{dot}(v, d_{\mathcal{P}}, m)\\&= \mathrm{dot}(v, e_{\{\{1,\dots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|S_a|\}\}} \otimes \Psi(e_{\{\{1,\dots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|T_b|\}\}}), m) \\&\otimes e_{\{\{1,\dots,|B_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|B_c|\}\}}.\end{aligned} \tag{C.7}$$

显式计算表明，张量运算 $x \mapsto x \otimes e_{\{1,\dots,p\}}$ 只会创建一个比 x 多出 p 个指标的张量，并沿着形如 $\dots, i, i, \dots, i$（p 次）的指标放置 x 的副本，其中其他位置均为零。因此，一旦

计算出点$(v, e_{\{\{1,\dots,|S_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|S_a|\}\}} \otimes \Psi(e_{\{\{1,\dots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|T_b|\}\}})\ m)$，就可以通过索引操作高效地完成与$e_{\{\{1,\dots,|B_i|\}\}}$ 的乘法。[2]

方程 (C.6) 表明，与 $e_{\{\{1,\dots,|S_i|\}\}}$ 的缩并可以“每次处理一个 i”，从左到右进行。根据我们对点的约定，$\mathrm{dot}(v, e_{\{\{1,\dots,p\}\}}, p)$ 的缩并会得到一个具有 $m-p$ 个指标的张量，其第 $(i_1, \dots, i_{m-p})$ 个分量为

$$\sum_{j=1}^{n} v_{i_1 \dots i_{m-p} \underbrace{j \dots j}_{p \text{ times}}} \tag{C.8}$$

因此，这些缩并可以通过指标运算和求和来实现。[3]

最后，张量 $\Psi(e_{T_1} \otimes \cdots \otimes e_{T_b})$ 必须加以处理。我们声称，与该张量进行 点 运算属于 传递运算，如 (Pan 与 Kondor, 2022, §5) 中所述。事实上，他们将传递运算定义为对应于划分 $\mathcal{P}$ 中与 $\{1, \dots, m\}$ 和 $\{m+1, \dots, m+m'\}$均有非空交集的集合的运算，这恰好就是 T_i所起的作用。既然我们已经说明了如何与 $e_{\{\{1,\dots,|S_i|\}\}}$ 和 $e_{\{\{1,\dots,|B_i|\}\}}$ 相乘，为简单起见，我们不妨假设 $d_{\mathcal{P}} = \Psi(e_{\{\{1,\dots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|T_b|\}\}})$。于是，

$$\mathrm{dot}(v, \Psi(e_{\{\{1,\dots,|T_1|\}\}} \otimes \cdots \otimes e_{\{\{1,\dots,|T_b|\}\}}))_{\underbrace{i_1 \dots i_1}_{|T_1''| \text{ times}} \dots \underbrace{i_b \dots i_b}_{|T_b''| \text{ times}}} = v_{\underbrace{i_1 \dots i_1}_{|T_1'| \text{ times}} \dots \underbrace{i_b \dots i_b}_{|T_b'| \text{ times}}}. \tag{C.9}$$

显然，这本质上是一种索引操作。[4]

D 示例

我们考察 $m = m' = 1$ 且 $n \geq 2$的情形。此时，置换等变线性层的空间为 $\left((\mathbb{R}^n)^{\otimes 2}\right)^{\Sigma_n}$，也就是说，它由在行和列同时置换下保持不变的矩阵构成，而这些矩阵通过矩阵向量乘法对应于等变映射 $\mathbb{R}^n \to \mathbb{R}^n$。

D.1 与指标元组相关的划分

引理 3.8 最简单的情形是 $l = 2$。此时，给定 $(i_1, i_2) \in \{1, \dots, n\}^2$，与之关联的集合划分为

$$\Pi(i_1, i_2) = \begin{cases} \{\{1,2\}\} & \text{if } i_1 = i_2 \\ \{\{1\},\{2\}\} & \text{if } i_1 \neq i_2 \end{cases} \tag{D.1}$$

令 $\mathcal{P}_1 = \{\{1,2\}\}$ 和 $\mathcal{P}_2 = \{\{1\},\{2\}\}$。接着，将 $\mathbb{R}^n \otimes \mathbb{R}^n$ 中的张量视为 $n \times n$ 矩阵，

$$\begin{aligned} e_{\mathcal{P}_1} &= \sum_i e_i \otimes e_i = I_n \text{ and} \\ e_{\mathcal{P}_2} &= \sum_{i \neq j} e_i \otimes e_j = \mathbf{1}\mathbf{1}^T - I \end{aligned} \tag{D.2}$$

[2]注意，与一般张量 w'' 进行克罗内克乘法时，若该张量有 p 个条目，则对 x的每个条目都需要进行 n^p 次标量乘法。因此，例如，若 x 有 q 个指标，则总共需要进行 n^{p+q} 次标量乘法。[3]注意，公式 (C.6) 将张量缩并的计算成本从大约 n^d 次乘法和加法降低到 $n^{d'} + n^{d-d'}$ 次乘法和加法。在 $w' = e_{\{1,\dots,d'\}}$ 这一特殊情形下，项 $n^{d'}$ 实际上简化为 n。[4]换言之，我们无需进行 $n^{\sum |T_i'|}$ 次乘法和加法，只需复制数组。

其中 I 是 $n \times n$ 单位矩阵，而 $\mathbf{1} = [1, \ldots, 1]^T$ 是全1向量。

D.2 基向量的秩

注意，$\mathbf{1}\mathbf{1}^T - I$ 不是克罗内克积 uv^T。事实上，它的秩为 n，因为 $det(-I + \mathbf{1}\mathbf{1}^T) = (1 + \mathbf{1}^T(-I)\mathbf{1})det(I) = (1-n)n \neq 0$，而任何 uv^T 的秩都是 1。
另一方面，定理 5.4 的基为

$$d_{\mathcal{P}_1} = I \text{ and } d_{\mathcal{P}_2} = \mathbf{1}\mathbf{1}^T \tag{D.3}$$

且 $\mathbf{1}\mathbf{1}^T$ 当然秩为 1。这当然是 DeepSets (Zaheer 等人，2018) (I 是恒等映射，$\mathbf{1}\mathbf{1}^T$ 是乘以 $\mathbf{1}$ 的向量和)。