

基于齐性空间与李群表示论的旋转位置编码(RoPE)推导

Gemini Thought Partner

2025 年 11 月 22 日

摘要

本文旨在通过齐性空间 (Homogeneous Spaces) 的表示论视角, 严格推导旋转位置编码 (Rotary Positional Embeddings, RoPE)。我们首先定义平移群在空间上的作用, 利用内积不变性导出位置编码必须是正交群的线性表示。随后, 利用 $SO(2N)$ 李代数中嘉当子代数 (Cartan Subalgebra) 的结构, 处理高维交换群 ($\mathbb{Z}^2, \mathbb{Z}^3$) 的情形, 自然地导出了多维 RoPE 的一般公式。

1 一般框架: 齐性空间与群表示

1.1 问题设定

考虑齐性空间 $X = \mathbb{R}^{d_{\text{model}}} \times G$, 其中 G 是离散平移群 (如 \mathbb{Z}, \mathbb{Z}^2 等), $V = \mathbb{R}^{d_{\text{model}}}$ 是词嵌入空间。群 G 在 X 上的作用 τ_t 定义为:

$$\tau_t(\mathbf{x}, p) = (\mathbf{x}, p + t), \quad \forall t, p \in G, \mathbf{x} \in V \quad (1)$$

即平移仅作用于位置分量。我们寻找映射 $f: X \rightarrow V$, 形式为分离变量:

$$f(\mathbf{x}, p) = \Phi(p)\psi(\mathbf{x}) \quad (2)$$

其中 $\Phi: G \rightarrow GL(V)$ 是矩阵值函数, 且 $\Phi(0) = I$ 。

1.2 不变性与同态性质

RoPE 的核心要求是内积的相对位置不变性:

$$\langle f(\mathbf{x}, p), f(\mathbf{y}, q) \rangle = \langle f(\mathbf{x}, p + t), f(\mathbf{y}, q + t) \rangle \quad (3)$$

代入分离变量形式, 得:

$$\psi(\mathbf{x})^T \Phi(p)^T \Phi(q) \psi(\mathbf{y}) = \psi(\mathbf{x})^T \Phi(p + t)^T \Phi(q + t) \psi(\mathbf{y}) \quad (4)$$

由于 \mathbf{x}, \mathbf{y} 的任意性, 这蕴含矩阵恒等式:

$$\Phi(p)^T \Phi(q) = \Phi(p + t)^T \Phi(q + t) \quad (5)$$

由此我们可以导出两个关键性质:

- **正交性 (Orthogonality):** 令 $q = p, t = -p$, 得 $\Phi(p)^T \Phi(p) = I$ 。故 $\Phi(p) \in O(d_{\text{model}})$ 。
- **群同态 (Group Homomorphism):** 利用正交性 $\Phi^T = \Phi^{-1}$ 及相对位置依赖性, 可推导得:

$$\Phi(p + k) = \Phi(p)\Phi(k) \quad (6)$$

这表明 Φ 是群 G 到正交群 $O(d_{\text{model}})$ 的一个 (实数域上的) 正交表示。

2 李代数结构: $\mathfrak{so}(2N)$ 的嘉当子代数

为了处理多维位置 (如 $\mathbb{Z}^2, \mathbb{Z}^3$), 我们需要研究交换群的表示。这涉及到李代数中的交换子代数结构。

2.1 李代数与指数映射

对于李群 G , 其同态由李代数 \mathfrak{g} 中的生成元决定。通过指数映射 $\exp: \mathfrak{g} \rightarrow G$, 我们可以将群元素表示为 e^A 。对于 $SO(d)$, 其李代数 $\mathfrak{so}(d)$ 由反对称矩阵组成。

2.2 嘉当子代数 (Cartan Subalgebra)

对于高维平移群 $G = \mathbb{Z}^k$ ($k \geq 2$), 由于它是阿贝尔群 (交换群), 其在 $O(d)$ 中的像也必须相互交换。事实上我们可以确定 $O(d)$ 的极大交换子群, 这在Lie理论中称为紧李群 $O(d)$ 的极大环面。

注记 1. 事实上不用Lie理论, 只要线性代数即可知道: 一组相互交换的实反对称矩阵可以被同时块对角化。做简单的数学归纳法即可。

在李理论中, 这对应于寻找 $\mathfrak{so}(d)$ 的嘉当子代数 \mathfrak{h} 。设 $d_{\text{model}} = 2N$ 。 $\mathfrak{so}(2N)$ 的嘉当子代数 \mathfrak{h} 是所有形式如下的矩阵组成的集合:

$$H(\theta) = \bigoplus_{j=1}^N \begin{pmatrix} 0 & -\theta_j \\ \theta_j & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\theta_1 & & & \\ \theta_1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & -\theta_N \\ & & & \theta_N & 0 \end{pmatrix} \quad (7)$$

任何相互交换的生成元集合都可以 (通过全局共轭作用) 被包含在同一个嘉当子代数中。这不仅保证了它们可以同时对角化, 还确定了其几何形式为 N 个独立平面的旋转。

注记 2. 事实上绝大部分李群的 *maximal torus* (或其李代数的 *Cartan subalgebra*) 的结构是存在且在共轭意义下唯一的; 这一点也有相当多有趣的证明, 比如基于 *Morse theory* 的; 而后者正是关于梯度下降的最优美的数学理论。

3 具体推导：1D, 2D 与 3D RoPE

3.1 1D RoPE ($G = \mathbb{Z}$)

群由单个生成元 $1 \in \mathbb{Z}$ 生成。设其对应的李代数元素为 $A \in \mathfrak{so}(d)$ 。

$$\Phi(m) = \exp(A \cdot m) \quad (8)$$

根据谱分解，我们总可以找到基底使得 A 位于嘉当子代数中，即 A 为块对角矩阵，第 j 个块为 $\begin{pmatrix} 0 & -\theta_j \\ \theta_j & 0 \end{pmatrix}$ 。则作用在第 j 个子空间上的算子为：

$$\Phi(m)_j = \begin{pmatrix} \cos(m\theta_j) & -\sin(m\theta_j) \\ \sin(m\theta_j) & \cos(m\theta_j) \end{pmatrix} \quad (9)$$

3.2 2D RoPE ($G = \mathbb{Z}^2$)

群由两个生成元 e_1, e_2 生成。设对应的李代数元素为 A, B 。由于 \mathbb{Z}^2 交换，必须有 $[A, B] = AB - BA = 0$ 。这意味着 A, B 属于同一个嘉当子代数 \mathfrak{h} 。因此它们拥有相同的块对角结构：

$$A = \bigoplus_{j=1}^N \begin{pmatrix} 0 & -\theta_j^A \\ \theta_j^A & 0 \end{pmatrix}, \quad B = \bigoplus_{j=1}^N \begin{pmatrix} 0 & -\theta_j^B \\ \theta_j^B & 0 \end{pmatrix} \quad (10)$$

位置编码矩阵为：

$$\Phi(m, n) = e^{Am} e^{Bn} = e^{Am+Bn} \quad (11)$$

在第 j 个子空间上，旋转角度叠加：

$$\alpha_j(m, n) = m\theta_j^A + n\theta_j^B \quad (12)$$

最终公式为：

$$\Phi(m, n)_j = \begin{pmatrix} \cos(m\theta_j^A + n\theta_j^B) & -\sin(m\theta_j^A + n\theta_j^B) \\ \sin(m\theta_j^A + n\theta_j^B) & \cos(m\theta_j^A + n\theta_j^B) \end{pmatrix} \quad (13)$$

3.3 3D RoPE ($G = \mathbb{Z}^3$)

对于视频或体素数据，群为 \mathbb{Z}^3 ，生成元对应李代数元素 A, B, C 。根据最大环面 (Maximal Torus) 原理，这三个相互交换的矩阵均位于 $\mathfrak{so}(2N)$ 的嘉当子代数 \mathfrak{h} 中。

$$\Phi(m, n, k) = \exp(Am + Bn + Ck) \quad (14)$$

对于第 j 个特征对，总旋转角为三个维度的线性组合：

$$\Theta_j = m\theta_j^A + n\theta_j^B + k\theta_j^C \quad (15)$$

这就构成了 3D RoPE 的一般形式。实际应用中，通常通过设置某些 θ_j 为 0 来实现维度的正交切分（例如前 1/3 通道编码 m ，中间 1/3 编码 n 等），但这只是本框架下的一个特例。

注记 3. *qwen3-vl* 的技术报告提出的交错式 *3DRoPE* 也不过是该框架的一个小小例子。

4 结论

通过引入 $X = \mathbb{R}^d \times G$ 的齐性空间结构，RoPE 被自然地解释为平移群 G 在 L^2 空间上的酉表示。利用李代数 $\mathfrak{so}(d)$ 的嘉当子代数理论，我们严格证明了对于任意维度的阿贝尔平移群 \mathbb{Z}^k ，其位置编码必然表现为特征空间内的成对旋转，且旋转角度是位置坐标的线性函数。