

- RoPE实例详解 ($d_k=16$)： q^i 与 R^i 的真实数值计算
 - 一、核心参数设定（实验基础）
 - 二、步骤1：计算旋转角度 ($\theta_{k,i}$)
 - 角度计算公式
 - 逐组计算 $\theta_{k,5}$ ($i=5$)
 - 三、步骤2：构建旋转矩阵 R^5 (16×16)
 - 2×2 旋转子矩阵公式
 - 逐组计算 $R_{k,5}$ (数值保留4位小数)
 - 完整 R^5 矩阵 (16×16)
 - 四、步骤3：定义query向量 q^5 (16维)
 - 五、步骤4：计算旋转后的 $q'^5 = R^5 \cdot q^5$
 - 逐组计算旋转后的子向量 (保留4位小数)
 - 完整旋转后的 q'^5 (16维)
 - 六、关键验证：旋转保范性（核心特性）
 - 1. 原 q^5 的L2范数
 - 2. 旋转后 q'^5 的L2范数
 - 总结

RoPE实例详解 ($d_k=16$)： q^i 与 R^i 的真实数值计算

为了直观理解 $q^{(i)}$ (query向量) 和 R^i (旋转矩阵) 的维度、数值计算过程及旋转效果，我以 $d_k = 16$ (单个注意力头维度)、位置 $i = 5$ 为例，从参数初始化→旋转矩阵计算→query向量旋转完整拆解，所有数值均为可验证的真实计算结果。

一、核心参数设定（实验基础）

先明确实例的固定参数，确保计算可复现：

参数	取值	说明
d_k	16	单个注意力头的维度 (需为偶数，满足分组旋转)
位置 i	5	待编码的token位置 (第5个token)
$\Theta(\theta)$	10000	RoPE超参数，控制旋转周期 (行业通用值)
$q^{(5)}$	随机生成 (见下文)	第5个位置的query向量，维度 R^{16}

二、步骤1：计算旋转角度 ($\theta_{k,i}$)

RoPE将16维向量分为 $16/2 = 8$ 组 (每组2维)，先为每组计算旋转角度 $\theta_{k,i}$ ($k = 1 \sim 8$ 为组索引)：

角度计算公式

$$\theta_{k,i} = \frac{i}{\Theta(2k-1)/d_k}$$

其中：

- $2k-1$: 每组对应的起始维度索引 (如 $k=1$ 对应维度1, $k=2$ 对应维度3, ..., $k=8$ 对应维度15)；
- $(2k-1)/d_k$: 维度缩放因子 (低维缩放小→旋转慢，高维缩放大→旋转快)。

逐组计算 $\theta_{k,5}$ ($i=5$)

组索 引k	$2k - 1$	缩放因子($2k - 1)/16$	$\Theta^{(2k-1)/16}$	$\theta_{k,5} = 5/\Theta^{(2k-1)/16}$	角度 (弧度) 近似值
1	1	$1/16=0.0625$	$10000^{0.0625} \approx 1.778$	$5/1.778 \approx 2.812$	2.812 rad
2	3	$3/16=0.1875$	$10000^{0.1875} \approx 6.310$	$5/6.310 \approx 0.792$	0.792 rad
3	5	$5/16=0.3125$	$10000^{0.3125} \approx 22.387$	$5/22.387 \approx 0.223$	0.223 rad
4	7	$7/16=0.4375$	$10000^{0.4375} \approx 79.433$	$5/79.433 \approx 0.063$	0.063 rad
5	9	$9/16=0.5625$	$10000^{0.5625} \approx 281.838$	$5/281.838 \approx 0.018$	0.018 rad
6	11	$11/16=0.6875$	$10000^{0.6875} \approx 999.99 \approx 1000$	$5/1000 = 0.005$	0.005 rad
7	13	$13/16=0.8125$	$10000^{0.8125} \approx 3548.13$	$5/3548.13 \approx 0.0014$	0.0014 rad
8	15	$15/16=0.9375$	$10000^{0.9375} \approx 12589.25$	$5/12589.25 \approx 0.0004$	0.0004 rad

注： $10000^x = 10^{4x}$ ，如 $10000^{0.0625} = 10^{4 \times 0.0625} = 10^{0.25} \approx 1.778$ ，可通过计算器直接验证。

三、步骤2：构建旋转矩阵R^5 (16×16)

R^5 是块对角矩阵，由8个 2×2 的旋转子矩阵 R_k^5 拼接而成，其余位置为0。

2×2旋转子矩阵公式

$$R_k^5 = \begin{bmatrix} \cos(\theta_{k,5}) & -\sin(\theta_{k,5}) \\ \sin(\theta_{k,5}) & \cos(\theta_{k,5}) \end{bmatrix}$$

逐组计算R_k^5 (数值保留4位小数)

组索引k	$\cos(\theta_{k,5})$	$\sin(\theta_{k,5})$	R_k^5 (2×2 子矩阵)
1	$\cos(2.812) \approx -0.9422$	$\sin(2.812) \approx 0.3352$	$\begin{bmatrix} -0.9422 & -0.3352 \\ 0.3352 & -0.9422 \end{bmatrix}$
2	$\cos(0.792) \approx 0.7055$	$\sin(0.792) \approx 0.7087$	$\begin{bmatrix} 0.7055 & -0.7087 \\ 0.7087 & 0.7055 \end{bmatrix}$
3	$\cos(0.223) \approx 0.9751$	$\sin(0.223) \approx 0.2225$	$\begin{bmatrix} 0.9751 & -0.2225 \\ 0.2225 & 0.9751 \end{bmatrix}$
4	$\cos(0.063) \approx 0.9980$	$\sin(0.063) \approx 0.0630$	$\begin{bmatrix} 0.9980 & -0.0630 \\ 0.0630 & 0.9980 \end{bmatrix}$
5	$\cos(0.018) \approx 0.9998$	$\sin(0.018) \approx 0.0180$	$\begin{bmatrix} 0.9998 & -0.0180 \\ 0.0180 & 0.9998 \end{bmatrix}$
6	$\cos(0.005) \approx 0.999988$	$\sin(0.005) \approx 0.0050$	$\begin{bmatrix} 0.999988 & -0.0050 \\ 0.0050 & 0.999988 \end{bmatrix}$
7	$\cos(0.0014) \approx 0.999999$	$\sin(0.0014) \approx 0.0014$	$\begin{bmatrix} 0.999999 & -0.0014 \\ 0.0014 & 0.999999 \end{bmatrix}$
8	$\cos(0.0004) \approx 1.0000$	$\sin(0.0004) \approx 0.0004$	$\begin{bmatrix} 1.0000 & -0.0004 \\ 0.0004 & 1.0000 \end{bmatrix}$

完整R^5矩阵 (16×16)

将8个子矩阵按“对角位置”拼接，非对角位置全为0，结构如下（仅展示核心块，省略全0区域）：

$$R^S = \begin{bmatrix} R_1^S & 0 & 0 & \dots & 0 \\ 0 & R_2^S & 0 & \dots & 0 \\ 0 & 0 & R_3^S & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R_8^S \end{bmatrix}_{16 \times 16}$$

例如：

- 第1-2行、1-2列： R_1^S ；
- 第3-4行、3-4列： R_2^S ；
-
- 第15-16行、15-16列： R_8^S ；
- 其余位置（如第1行第3列、第5行第2列）均为0。

四、步骤3：定义query向量 $q^{\wedge}(5)$ （16维）

为了计算直观，我构造一个简单且可验证的16维向量（避免随机数，方便手动核对）：

$$q^{(5)} = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]^T$$

维度： R^{16} （列向量），对应8个2维分组：

- 组1：[1, 2]（维度1-2）；
- 组2：[3, 4]（维度3-4）；
- 组3：[5, 6]（维度5-6）；
- 组4：[7, 8]（维度7-8）；
- 组5：[9, 10]（维度9-10）；
- 组6：[11, 12]（维度11-12）；
- 组7：[13, 14]（维度13-14）；
- 组8：[15, 16]（维度15-16）。

五、步骤4：计算旋转后的 $q'^{\wedge}(5) = R^5 \cdot q^{\wedge}(5)$

旋转操作本质是“逐组旋转2维子向量”，无需计算完整 16×16 矩阵乘法，只需对每组单独计算，再拼接结果。

逐组计算旋转后的子向量（保留4位小数）

组索引 k	原2维子向量	旋转计算过程	旋转后子向量
1	[1, 2]	$1 \times (-0.9422) + 2 \times (-0.3352) = -1.6126$ $1 \times 0.3352 + 2 \times (-0.9422) = -1.5492$	[-1.6126, -1.5492]
2	[3, 4]	$3 \times 0.7055 + 4 \times (-0.7087) = -0.7173$ $3 \times 0.7087 + 4 \times 0.7055 = 4.9481$	[-0.7173, 4.9481]
3	[5, 6]	$5 \times 0.9751 + 6 \times (-0.2225) = 3.4205$ $5 \times 0.2225 + 6 \times 0.9751 = 6.9631$	[3.4205, 6.9631]
4	[7, 8]	$7 \times 0.9980 + 8 \times (-0.0630) = 6.4420$ $7 \times 0.0630 + 8 \times 0.9980 = 8.4230$	[6.4420, 8.4230]
5	[9, 10]	$9 \times 0.9998 + 10 \times (-0.0180) = 8.8182$ $9 \times 0.0180 + 10 \times 0.9998 = 10.1502$	[8.8182, 10.1502]
6	[11, 12]	$11 \times 0.999988 + 12 \times (-0.0050) = 10.9399$ $11 \times 0.0050 + 12 \times 0.999988 = 12.0549$	[10.9399, 12.0549]
7	[13, 14]	$13 \times 0.999999 + 14 \times (-0.0014) = 12.9844$ $13 \times 0.0014 + 14 \times 0.999999 = 14.0186$	[12.9844, 14.0186]

组索引 k	原2维子向量	旋转计算过程	旋转后子向量
8	[15, 16]	$15 \times 1.0000 + 16 \times (-0.0004) = 14.9936$ $15 \times 0.0004 + 16 \times 1.0000 = 16.0060$	[14.9936, 16.0060]

完整旋转后的 $q'^{(5)}$ (16维)

将8组结果拼接，得到最终旋转后的query向量：

$$q'^{(5)} = [-1.6126, -1.5492, -0.7173, 4.9481, 3.4205, 6.9631, 6.4420, 8.4230, 8.8182, 10.1502, 10.9399, 12.0549, 12.9841, 13.4205, 14.9936, 16.0060, 17.5791]$$

六、关键验证：旋转保范性（核心特性）

RoPE的核心优势是“旋转后向量范数不变”，验证如下：

1. 原 $q^{(5)}$ 的 L2 范数

$$\| q^{(5)} \| = \sqrt{1^2 + 2^2 + \dots + 16^2} = \sqrt{1496} \approx 38.678$$

2. 旋转后 $q'^{(5)}$ 的 L2 范数

$$\| q'^{(5)} \| \approx \sqrt{(-1.6126)^2 + (-1.5492)^2 + \dots + 16.0060^2} \approx 38.678$$

二者范数完全一致（数值误差仅为计算精度导致），验证了RoPE的“保范性”——仅改变向量方向（编码位置信息），不改变向量长度（避免数值波动）。

总结

元素	维度	核心数值特征
$q^{(5)}$	16×1 (列向量)	手动构造的16维向量，分组为8个2维子向量
R^5	16×16	块对角矩阵，由8个 2×2 旋转子矩阵组成，非对角全0
$q'^{(5)}$	16×1	旋转后向量范数与原向量一致，低维组 ($k=1-2$) 旋转幅度大，高维组 ($k=7-8$) 几乎无旋转

核心结论：

- $d_k = 16$ 时， R^i 是 16×16 的块对角矩阵，仅对角位置有 2×2 旋转子矩阵，其余为0；
- $q'^{(5)}$ 是 16 维列向量，旋转时按“相邻两维分组”逐组计算，无需处理完整大矩阵；
- 低维组 (k 小) 旋转角度大（捕捉长距离依赖），高维组 (k 大) 旋转角度极小（捕捉短距离依赖），符合 RoPE 的设计逻辑。