What Happens When We Stretch the RoPE Clocks?

Due: Wednesday, December 9 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

A transformer with even head dimension $d \geq 4$ was trained through context length L using RoPE frequencies

$$\theta_j = B^{-2j/d}, \qquad j = 0, \ldots, d/2 - 1, \qquad B > 1.$$

It must now operate through length cL for a symbolic extension factor $c > 1$, without retraining. One proposal changes the base from B to B' just enough to multiply the slowest clock's period by c. A second keeps B fixed but replaces every position m by m/c inside the rotations.

The original clocks below each advance one position, with $d = 4$ and $B = 16$.



θ_0 θ_1

1. The first proposal stretches the period of the slowest clock by a factor of c. What new base B' does this require?

2. At adjacent positions, the fastest clock advances by θ_0. Which proposal preserves this phase difference?

3. Two tokens originally Δ positions apart now sit $c\Delta$ positions apart. Which proposal preserves their original phase difference at every frequency for every Δ?

Hint: You may quote $(\mathbf{R}_m)^\top \mathbf{R}_n = \mathbf{R}_{n-m}$, norm preservation of rotations, and that a clock of angular frequency θ has period $2\pi/\theta$. Replacing position m by m/c is equivalent to using the effective frequency θ_j/c.

Takeaway: *Changing the RoPE base preserves the fastest clock. Dividing all positions by a constant instead preserves every old phase at proportionally stretched distances. Extending the context window involves choosing which positional relationships to preserve.*