CSCI 145 Problem 22: Stretch the RoPE Clock

November 4, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, December 9**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- A complete, good-faith attempt earns submission credit. The grader may select any recent assigned problem for a brief conversation, so keep your work and be prepared to explain each solution and answer a related question. Scheduling details will be announced in class.

Problem 22: Stretch the RoPE Clock

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.

1. What effective frequency ladder does each proposal produce, including the base B' required by the first proposal?

2. Which proposal multiplies the slowest period by c while preserving the fastest adjacent-position phase, and which makes every new phase at stretched offset $c\Delta$ equal the old phase at 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.*