CSCI 145 Problem 7: Can Units Cure Poor Conditioning?

October 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on Wednesday, September 30**.

- 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 7: Can Units Cure Poor Conditioning?

A calibration model has two nonzero feature columns $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ with lengths $a, b > 0$ and cosine similarity

$$\gamma = \frac{\mathbf{x}_1^\top \mathbf{x}_2}{ab}, \qquad |\gamma| < 1.$$

An engineer may divide each column by any positive scale before minimizing MSE with fixed-step gradient descent. For each scaling, the learning rate is chosen to minimize the worst-direction contraction factor.

The scaling must preserve the set of attainable predictions.

1. Which diagonal scaling gives the fastest such convergence? Prove that no other positive diagonal scaling gives a smaller worst-direction factor.

2. Determine that factor and use it to decide whether changing units can guarantee fast convergence even for nearly redundant sensors.

Hint: For a two-dimensional symmetric matrix, the eigenvalue sum is its trace and their product is its determinant.

Analyze q^2, where

$$q = \frac{\kappa - 1}{\kappa + 1}$$

is the best fixed-step contraction factor and κ is the Hessian condition number. Use the arithmetic–geometric mean (AM–GM) inequality.

Takeaway: *Rescaling the two features to have equal lengths gives the fastest convergence among these choices of units. It cannot fix features that are nearly parallel. Good scaling helps gradient descent, but redundant features can still make it slow.*