CSCI 145 Problem 8: The Slowest Direction Sets the Pace

August 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on DATE**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- Problems are grouped into three-week **cycles**. Once per cycle, you will meet with the grader and present the solution to *one* problem from that cycle, selected at random when you arrive — so bring your work and be ready to explain any of them.

Problem 8: The Slowest Direction Sets the Pace 2 points

Consider the quadratic loss

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}(\mathbf{w} - \mathbf{w}^\star)^\top \mathbf{A}(\mathbf{w} - \mathbf{w}^\star),$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$ is symmetric positive definite with eigenvalues in $[\lambda_{\min}, \lambda_{\max}]$. You must choose one fixed learning rate α before seeing the initial error or which eigendirections it contains.

Choose the learning rate that minimizes the worst possible fraction of the error norm surviving one gradient-descent step. Prove that your choice is optimal, derive its contraction factor in terms of $\kappa = \lambda_{\max}/\lambda_{\min}$, and determine the number of steps needed to reduce every initial error by a factor of 10.

Use the result to explain the conditioning experiment from lecture: before feature scaling, the least-squares bowl had $\kappa \approx 9.5 \times 10^5$. Estimate the worst-case error remaining after 20,000 optimally tuned steps and the number of steps needed for a tenfold reduction. Compare this with a perfectly scaled bowl, and assess the claim that a model is slow to optimize primarily because it has many parameters.

Results you may use. Along an eigenvector of curvature λ, one gradient-descent step multiplies the error component by $1 - \alpha\lambda$. Thus the worst one-step contraction is $\max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |1 - \alpha\lambda|$, whose maximum occurs at an endpoint. For small u, you may use $\log(1 - u) \approx -u$.