CME 307: Optimization CME 307 / MS&E 311 / OIT 676

Quiz 5: Practice questions

Fall 2025 *Prof. Udell*

Exercise. Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth. Show that for any x and any $t \in (0, 1/L]$, the gradient step $x^+ = x - t\nabla f(x)$ satisfies $f(x^+) \le f(x) - \frac{t}{2}|\nabla f(x)|^2$ by using the quadratic upper bound for L-smooth functions.

Exercise. Armijo backtracking terminates. Let f be L-smooth, fix $c \in (0,1)$ and $\beta \in (0,1)$. Starting from $t_0 = 1$, repeatedly set $t \leftarrow \beta t$ until the Armijo condition $f(x - t\nabla f(x)) \le f(x) - c, t|\nabla f(x)|^2$ holds. Show Armijo holds whenever $0 < t \le \frac{2(1-c)}{L}$, and deduce an upper bound on the number of backtracks when starting at $t_0 = 1$.

Exercise. Exact line search on a quadratic. Let $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ with $A \succeq 0$. If $g = \nabla f(x) = Ax - b$ and $p = -g$, compute $t^\star = \arg\min_{t\ge 0} f(x + tp)$ and show $t^\star = \dfrac{g^\top g}{g^\top Ag}$ (assuming $g \ne 0$ and $g^\top Ag > 0$).

Exercise. Upper-bound viewpoint. Assume f is L-smooth. Show that the minimizer of $m_y(z) = f(y) + \nabla f(y)^\top(z-y) + \frac{L}{2}|z-y|^2$ is $z^\star = y - \frac{1}{L}\nabla f(y)$. Conclude gradient descent with stepsize $1/L$ arises by minimizing this upper model at each iterate.

Exercise. Quadratic loss. For $f(x) = |Ax - b|^2$, compute ∇f and $\nabla^2 f$ and prove L-smoothness with $L = 2\lambda_{\max}(A^\top A)$. When is f strongly convex?

Exercise. Logistic loss. For $f(x) = \sum_{i=1}^m \log!\big(1 + \exp(b_i a_i^\top x)\big)$, show $\nabla^2 f(x) = \sum_i \sigma_i(1-\sigma_i)a_i a_i^\top$ with $\sigma_i = \frac{1}{1+\exp(-b_i a_i^\top x)}$, deduce L-smoothness with $L \le \frac{1}{4}, \lambda_{\max}(A^\top A)$, and state a condition under which f is strongly convex on a compact domain.

Exercise. Sanity check in $\mathbb{R}^n$. For $f(x) = \frac{1}{2}|x|^2$, gradient descent gives $x^{k+1} = (1-t)x^k$. Determine precisely for which $t > 0$ the iterates converge and give the linear rate as a function of t.

Exercise. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, explain geometrically why the negative gradient $-\nabla f(x)$ gives the direction of steepest decrease of f at x. You may reason using the first-order Taylor approximation or by considering the directional derivative.

Exercise. In gradient descent,

$$x^{(k+1)} = x^{(k)} - t\,\nabla f(x^{(k)}).$$

Suppose x represents a position vector with physical units of meters, and $f(x)$ represents energy (joules).

(a) What are the units of $\nabla f(x)$? Recall the definition of the gradient.

(b) What are the units of the step size t?

(c) How does this affect your intuition about why it is difficult to choose a good step size in practice?

Exercise. Consider $f(x) = \frac{1}{2}x^T Ax$ for $A \succeq 0$. Show that gradient descent updates each eigen-direction of A independently. If the eigenvalues of A lie in $[\mu, L]$, how does this explain the convergence rate bound in the previous exercise?

Exercise. For an L-smooth function f, derive the gradient descent update rule by minimizing the quadratic upper bound

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2.$$

Conceptually, what are we minimizing at each step, and why does this guarantee descent when $t \leq 1/L$?

Exercise. You apply gradient descent to minimize a smooth loss $f(w)$ in machine learning. During training you observe oscillations: the objective decreases for a few steps, then increases. What does this behavior suggest about your current step size? Explain how you could adjust it using ideas from smoothness or the Armijo rule.