Required reading: Wright and Recht, Ch. 3.

1 Let a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $x \in \mathbb{R}^n$ be given, such that $\nabla f(x) \neq 0$. Prove that the set of all descent directions for f at x is convex.

2 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable, m-strongly convex function. Prove that

$$\nabla f(x)^T x \geq \frac{m}{2}\|x\|^2 - \frac{1}{2m}\|\nabla f(0)\|^2, \qquad \text{for all } x \in \mathbb{R}^n.$$

Hint: Use the fact that f satisfies

$$[\nabla f(x) - \nabla f(y)]^T (x - y) \geq m\|x - y\|^2, \qquad \text{for all } x, y \in \mathbb{R}^n.$$

3 (Exercise 5 from Wright and Recht, Ch. 3) Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is m-strongly convex, L-smooth, and has (unique) minimizer x^*. Use the co-coercivity property proved in Homework 1 and the fact that $\nabla f(x^*) = 0$ to prove that the kth iterate of the steepest descent method applied to f with constant steplength $\frac{2}{m+L}$ satisfies

$$\|x^k - x^*\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|x^0 - x^*\|,$$

where $\kappa := \frac{L}{m}$ is the *condition number* of the problem.

4 (Exercise 7(a-f) from Wright and Recht, Ch. 3) Let A be an $N \times d$ matrix with $N < d$ and $\text{rank}(A) = N$ [that is, the rows of N are linearly independent]. Consider the least-squares optimization problem

$$\min_x f(x), \qquad \text{where } f(x) := \frac{1}{N}\|Ax - b\|^2$$

where $b \in \mathbb{R}^N$ is a given vector.

(a) Assume that there exists a z such that $Az = b$. Characterize the solution space of the linear system $Ax = b$.

Hint: Recall the definition of the *nullspace* (or the *kernel*) of A, $\ker(A) := \{x \in \mathbb{R}^d : Ax = 0\}$.

(b) Compute the Lipschitz constant of the gradient of f in terms of A.

(c) If you run the steepest descent method on f starting at $x^0 = 0$ with appropriate choice of steplength, how many iterations are required to find a solution with $\frac{1}{N}\|Ax - b\|^2 \leq \varepsilon$?

(d) Consider the *regularized* problem

$$\min_x f_\mu(x), \qquad \text{where } f_\mu(x) := \frac{1}{N}\|Ax - b\|^2 + \mu\|x\|^2$$

for some $\mu > 0$. Find the minimizer x_μ of f_μ in closed form.

(e) If you run the steepest descent method on f_μ starting at $x^0 = 0$ with appropriate choice of steplength, how many iterations are required to find a solution with $f_\mu(x) - f_\mu(x_\mu) \leq \varepsilon$?

(f) Suppose $\hat{x}$ satisfies $f_\mu(\hat{x}) - f_\mu(x_\mu) \leq \varepsilon$. Find a tight upper bound on $f(\hat{x})$.