Required reading: Wright and Recht, Ch. 9 and 10.

1 (Exercise 4 in Wright and Recht, Ch. 9) Let f be m-strongly convex and L-smooth. Define the function

$$f_m(x) := f(x) - \frac{m}{2}\|x\|_2^2.$$

(a) Prove that f_m is convex with $(L-m)$-Lipschitz gradients.

(b) Write down the proximal-gradient algorithm for the function

$$f_m(x) + \frac{m}{2}\|x\|_2^2$$

where we take f_m to be the "smooth" part and $\frac{m}{2}\|x\|_2^2$ to be the "convex but possibly nonsmooth" part.

(c) Does there exist a steplength α such that this proximal-gradient algorithm has the same iterates as gradient descent applied to f for some (possibly different) constant steplength? Explain.

(d) Find a steplength for the proximal-gradient method such that

$$\|x^k - x^*\|_2 \le \left(1 - \frac{m}{L}\right)\|x^{k-1} - x^*\|_2$$

where x^* is the unique minimizer of f.

2 (Exercise 1 in Wright and Recht, Ch. 10) Consider minimization of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ over the polyhedral set defined by a combination of linear equalities and inequalities as follows:

$$\{x \in \mathbb{R}^n : Ex = g, Cx \ge d\}$$

where $E \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}^m$, $C \in \mathbb{R}^{p\times n}$, and $d \in \mathbb{R}^p$, and where $Cx \ge d$ means that each coordinate of the vector $Cx - d \in \mathbb{R}^p$ is nonnegative.

Show that the first-order necessary condition for x^* to be a solution of this problem is that there exist vectors $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$, such that

$$\nabla f(x^*) - E^T\lambda - C^T\mu = 0, \quad Ex^* = g, \quad 0 \le \mu \perp Cx^* - d \ge 0$$

where $0 \le u \perp v \ge 0$ for two vectors $u, v \in \mathbb{R}^p$ indicates that, for all $i = 1, 2, \ldots, p$, we have $u_i \ge 0$, $v_i \ge 0$, and $u_i v_i = 0$.

Hint: Introduce *slack variables* $s \in \mathbb{R}^p$ and reformulate the problem to the equivalent problem

$$\min_{x\in\mathbb{R}^n, s\in\mathbb{R}^p} f(x)$$
$$\text{subject to } Ex = g,\ Cx - s = d,\ s \ge 0.$$

Now define $\mathcal{X}$, A, and b appropriately and use Theorem 10.5 to find the optimality conditions for this reformulation, then eliminate s.