Required reading: Wright and Recht, Ch. 6 and 7.

Throughout the problem set, $\|\cdot\|$ stands for the 2-norm $\|\cdot\|_2$.

1 (Exercise 1 in Wright and Recht, Ch. 6) In the ERM example of Section 6.1, assume that the objective function f is known at the current point x, along with the vector $g = Ax$. Show that the cost for computing $f(x + \gamma e_i)$ for some $i = 1, 2, \ldots, n$ is $O(|A_{\cdot i}|)$ [here, $|A_{\cdot i}|$ denotes the number of nonzero entries in the ith column of A] – the same order as the cost of updating the gradient ∇f. Show that a similar observation holds for the graph example in Section 6.1.

2 (Exercise 1 in Wright and Recht, Ch. 7) Show that the Euclidean projection $P_\Omega(x)$ onto the unit ball $\Omega = \{x \in \mathbb{R}^n : \|x\| \le 1\}$ is given by

$$P_\Omega(x) = \begin{cases} x, & \|x\| \le 1 \\ x/\|x\|, & \|x\| > 1 \end{cases}.$$

3 (Exercise 4 in Wright and Recht, Ch. 7) Consider the linear function $f(x) := c^T x$ of $x \in \mathbb{R}^n$, where $c \in \mathbb{R}^n$ is a fixed vector. Find the minimizer of f over each of the following sets:

(a) The unit ball $\{x \in \mathbb{R}^n : \|x\| \le 1\}$.

(b) The unit simplex $\{x \in \mathbb{R}^n : x_i \ge 0 \text{ for all } i, \sum_{i=1}^n x_i = 1\}$.

(c) The box $\{x \in \mathbb{R}^n : 0 \le x_i \le 1 \text{ for all } i\}$.

4 (Exercise 6 in Wright and Recht, Ch. 7) Let $\Omega \subseteq \mathbb{R}^n$ be a closed and convex set and let a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ be given. Consider one step of projected gradient descent

$$x^{k+1} = P_\Omega\Big(x^k - \alpha_k \nabla f(x^k)\Big), \qquad k = 0, 1, 2, \ldots$$

with an arbitrary initial point $x^0 \in \Omega$. Prove that, for any $\alpha_k > 0$,

$$x^{k+1} = \operatorname*{arg\,min}_{x \in \Omega} \left(f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right)$$

and

$$\|x^{k+1} - x^k\|^2 \le \alpha_k \nabla f(x^k)^T (x^k - x^{k+1}).$$