Optimization for Machine Learning (Homework #1)

Assignment date: Sep 19
Due date: Oct 3 (noon)

Read Chapter 2.5 (Convex Duality)

Theoretical Problems (10 points)

1. (1 points) Let $f(x) = \|x\|_1 + \|x\|_2^4/4$, where $x \in \mathbb{R}^d$. Find its conjugate $f^*(x)$.
2. (2 points) Let $x \in \mathbb{R}$ and $y \in \mathbb{R}_+$. Is $f(x, y) = x^2/y$ a convex function? Prove your claim.
3. (2 points) Consider the convex set $C = \{x \in \mathbb{R}^d : \|x\|_\infty \leq 1\}$. Given $y \in \mathbb{R}^d$, compute the projection $\mathrm{proj}_C(y)$.
4. (3 points) Compute $\partial f(x)$ for the following functions of $x \in \mathbb{R}^d$
   - $f(x) = \|x\|_2$
   - $f(x) = \mathbb{1}(\|x\|_\infty \leq 1)$
   - $f(x) = \|x\|_2 + \|x\|_\infty$
5. (3 points) Consider the square root Lasso method. Given $X \in \mathbb{R}^{n \times d}$ and $y \in \mathbb{R}^n$, we want to find $w \in \mathbb{R}^d$ to solve

$$[w_*, \xi_*] = \arg\min_{w,b,\xi} \left[\|Xw - y\|_2 + \lambda \sum_{j=1}^{d} \xi_j \right], \tag{1}$$

$$\text{subject to } \xi_j \geq w_j, \quad \xi_j \geq -w_j \quad (j = 1, \ldots, d). \tag{2}$$

   Lasso produces sparse solutions. Define the support of the solution as

$$S = \{j : w_{*,j} \neq 0\}.$$

   Write down the KKT conditions under the assumption that $Xw_* \neq y$. Simplify in terms of $S, X_S, X_{\bar{S}}, y, w_S$. Here X_S contains the columns of X in S, $X_{\bar{S}}$ contains the columns of X not in S, and w_S contains the nonzero components of w_*.

Programming Problem (4 points)

We consider ridge regression problem with randomly generated data. The goal is to implement gradient descent and experiment with different strong-convexity settings and different learning rates.

- Use the python template "prog-template.py", and implement functions marked with '# implement'.
- Submit your code and outputs. Compare to the theoretical convergence rates in class, and discuss your experimental results.