Homework 4 for Math 173A - Fall 2019

1. Find an expression for the orthogonal projection of a point $x \in \mathbb{R}^n$ onto each of the following convex sets. **Make sure to show your work**:

   (a) $B_2^n := \{z \in \mathbb{R}^n : \|z\| \leq 1\}$. **Remark:** This set is called the ℓ_2-ball in $\mathbb{R}^n$.

   (b) $B_\infty^n := \{z \in \mathbb{R}^n : |z_i| \leq 1 \text{ for each } i = 1, ..., n\}$. **Remark:** This set is called the ℓ_∞-ball in $\mathbb{R}^n$.

   (c) $S = \{z \in \mathbb{R}^3 : z_1 + z_2 + z_3 = 0\}$. **Hint:** The method of Lagrange multipliers may be helpful

   (d) $T = \{z \in \mathbb{R}^3 : z_1 + z_2 + z_3 = 0 \text{ and } z_1 + 2z_2 + 3z_3 = 0\}$. **Hint:** The method of Lagrange multipliers may be helpful

   (e) $U = \{z \in \mathbb{R}^n : Az = 0 \text{ where } A \text{ is an } m \times n \text{ matrix with } m \leq n\}$. You may assume that AA^T is invertible.
   Hint: The set T from part (d) can be written as $T = \{z \in \mathbb{R}^3 : Az = 0\}$ for an appropriate choice of the matrix A, so this part is just a generalization of what you did in (d).

2. Let
$$f(x) = a^T x$$
where $a \in \mathbb{R}^n$. Using the results of Problem 1, write a projected gradient descent algorithm to solve
$$\min_{x \in \Omega} f(x)$$
where

   (a) $\Omega = B_2^n$.

   (b) $\Omega = B_\infty^n$.

   (c) $\Omega = \{z \in \mathbb{R}^n : Az = 0 \text{ where } A \text{ is an } m \times n \text{ matrix with } m \leq n\}$.

   You do not need to specify the step-size for this problem.

3. Repeat Problem 2, with the function
$$f(x) = \sum_{i=1}^n a_i x_i^2$$
where $a_i \geq 0$.

4. Show that each of the following functions is L-smooth, and find an upper bound on the smoothness constant L.

   (a) $f(x) = \sum_{i=1}^n a_i x_i^2$, where $a_i \geq 0$.

   (b) $f(x) = x^T Q x + a^T x + b$ where $Q \in \mathbb{R}^{n \times n}, a \in \mathbb{R}^n, b \in \mathbb{R}$.

(c) $f(x) = f_1(x) + f_2(x)$ where f_1 is L_1-smooth and f_2 is L_2-smooth and both are convex.

(d) Suppose that we wish to use gradient descent to minimize each of the functions from parts (a)-(c) above, what choice of step-size μ should we employ to guarantee that at the t-th iteration, we have

$$f(x^{(t)}) - f(x^*) \leq \frac{\|x^{(t)} - x^{(0)}\|}{2t\mu}.$$

5. **Computer problem:** This question refers to Question 6 from HW3, so $\tilde{\rho}(x) := \log(1 + e^x)$.

(a) Select an ambient dimension n (preferably $n \geq 20$), and generate $T > n$ random points, $x_1, ..., x_T$, in $\mathbb{R}^n$ (generate them from your favorite distribution).

(b) Select (or randomly generate) a "ground truth" vector $w_0 \in \mathbb{R}^n$, and generate $c_i = \tilde{\rho}(w_0^T x_i)$, $i = 1, ..., T$.

(c) Pretend that you don't know the true w_0, and find it by running a gradient descent algorithm with backtracking line-search to minimize the function

$$g(w) = \sum_{i=1,...,T} |\tilde{\rho}(w^T x_i) - c_i|^2.$$

(d) For various values of γ and β plot (as a function of t) the evolution of $g(w^{(t)})$, where $w^{(t)}$ are the iterates from your algorithm.