IE 539 Convex Optimization Assignment 2

Fall 2022

Out: 27th September 2022

Due: 9th October 2022 at 11:59pm

Instructions

- Submit a PDF document with your solutions through the assignment portal on KLMS by the due date. Please ensure that your name and student ID are on the front page.
- Late assignments will **not** be accepted except in extenuating circumstances. Special consideration should be applied for in this case.
- It is **required** that you typeset your solutions in LaTeX. Handwritten solutions will not be accepted.
- Spend some time ensuring your arguments are **coherent** and your solutions **clearly** communicate your ideas.

Question:	1	2	3	4	5	6	Total
Points:	15	15	15	10	20	25	100

1. For a fixed z, derive closed form optimal solutions for
   (a) (5 points) $\min_{x\in\mathbb{R}^d}\left\{\frac{1}{2}\|x-z\|_2^2+\lambda\|x\|_2^2\right\}$
   (b) (10 points) $\min_{x\in\mathbb{R}^d}\left\{\frac{1}{2}\|x-z\|_2^2+\lambda\|x\|_1\right\}$.

2. (15 points) Let $X=\{x\in\mathbb{R}^d:\|x-x_0\|_2\le r\}$ be a ball of radius r centred at the point x_0. Find an expression for the projection onto X. In other words, for $z\in\mathbb{R}^d$, find $\arg\min_{x\in X}\left\{\frac{1}{2}\|x-z\|_2^2 : x\in X\right\}$. To get full marks, you must also *prove* that your expression is correct.

3. (15 points) Recall the ℓ_1- and ℓ_∞-norms are defined as

$$\|x\|_1 := \sum_{i\in[d]}|x_i|, \quad \|x\|_\infty := \max_{i\in[d]}|x_i|.$$

The dual norm is defined as

$$\|x\|_* := \max_{z:\|z\|\le 1} x^\top z.$$

Show that the ℓ_1-norm is the dual norm to the ℓ_∞-norm, and vice versa.

[Hint: given x, which vector would give $x^\top z=\sum_{i\in[d]}|x_i|=\|x\|_1$? Since $|x_i|=\text{sign}(x_i)x_i$, we can see that such a vector must satisfy $\sum_{i\in[d]}(\text{sign}(x_i)-z_i)x_i=0$. Derive a guess for z from this, and argue why it satisfies the optimality condition for $\max_{z:\|z\|_\infty} x^\top z$.

Modify this strategy to show that the dual of the ℓ_1-norm is the ℓ_∞-norm. For this part, it may help to write $\{z:\|z\|_1\le 1\}=\{z: z_i=s_iw_i, s_i\in\{-1,0,1\}, w_i\ge 0\ \forall i\in[d],\ \sum_{i\in[d]}w_i\le 1\}$.]

4. (10 points) Given a convex set C, suppose that x^* solves $\min_{x\in C}\frac{1}{2}\|x\|_2^2$. Show that for any $x\in C$,

$$\frac{1}{2}\|x-x^*\|_2^2\le\frac{1}{2}\|x\|_2^2-\frac{1}{2}\|x^*\|_2^2.$$

5. In this question we will work through the convergence analysis of the subgradient method for functions that are strongly convex and Lipschitz continuous. Let $f:\mathbb{R}^d\to\mathbb{R}$ be a function that is α-strongly convex with respect to the ℓ_2 norm and L-Lipschitz continuous in the ℓ_2 norm. Recall that the subgradient method proceeds as follows.
   - Choose $x_1\in\mathbb{R}^d$.
   - For $t=1,2,3,\ldots,T+1$:
     - Select any subgradient $g_t\in\partial f(x_t)$ and step size $\eta_t>0$.
     - Compute $x_{t+1}=x_t-\eta_t g_t$.

   (a) (10 points) Show that for each t, we have

$$f(x_t)-f(x^*)\le\left(\frac{1}{2\eta_t}-\frac{\alpha}{2}\right)\|x_t-x^*\|_2^2-\frac{1}{2\eta_t}\|x_{t+1}-x^*\|_2^2+\eta_t^2\|g_t\|_2^2$$

   where x^* is an optimal solution to $\min_{x\in\mathbb{R}^d}f(x)$.

   (b) (10 points) Set $\eta_t=\frac{2}{\alpha(t+1)}$. Then use part (a) to show that

$$f\left(\sum_{t=1}^{T}\frac{2t}{T(T+1)}x_t\right)-f(x^*)\le\frac{2L^2}{\alpha(T+1)}.$$

6. In this question we will work through the proof of projected subgradient descent and derive rates for Lipschitz continuous functions. The algorithm proceeds as follows:
   - Choose $x_1\in C$.
   - For $t=1,2,3,\ldots,T+1$:
     - Select any subgradient $g_t\in\partial f(x_t)$ and step size $\eta_t>0$.
     - Compute $x_{t+1}=\arg\min_{x\in C}\|x_t-\eta_t g_t-x\|_2^2$.

   (a) (10 points) Show that for each t, we have

$$\|x_{t+1}-x^*\|_2^2\le\|x_t-\eta_t g_t-x^*\|_2^2$$

   where x^* is an optimal solution to $\min_{x\in\mathbb{R}^d}f(x)$.

(b) (10 points) The rate of convergence of projected subgradient descent depends on the step sizes η_t. Derive the rates of convergence for the following three step size rules:

$$\eta_t = \frac{1}{\sqrt{t}}, \quad \eta_t = \frac{1}{t}, \quad \eta_t = \eta.$$

You may use the following facts without proof: there exists constants c_1, c_2, c_3, c_4 such that

$$c_1\sqrt{T} \leq \sum_{t \in [T]} \frac{1}{\sqrt{t}}, \quad c_2 \log(T) \leq \sum_{t \in [T]} \frac{1}{t} \leq c_3 \log(T), \quad \sum_{t \in [T]} \frac{1}{t^2} \leq c_4.$$

Show that if we know T ahead of time, we can choose η in such a way that the third rate is faster than the first two.

(c) (5 points) Set $\eta_t = \frac{\|x_1 - x^*\|_2}{L\sqrt{T}}$. Then use part (a) to show that

$$f\left(\frac{1}{T}\sum_{t=1}^{T} x_t\right) - f(x^*) \leq \frac{L\|x_1 - x^*\|_2}{\sqrt{T}}.$$