

IE 539 Convex Optimization Assignment 3

Fall 2022

Out: 18th October 2022

Due: 1st November 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 be subject to a penalty. 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	Total
Points:	20	20	20	30	10	100

1. In this question we prove the convergence of the projected subgradient method for functions that are strongly convex and Lipschitz continuous. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 projected subgradient method proceeds as follows.

- Choose $x_1 \in C$.
- For $t = 1, 2, 3, \dots, T - 1$:
 - Select any subgradient $g_t \in \partial f(x_t)$ and step size $\eta_t > 0$.
 - Compute $x_{t+1} = \text{Proj}_C\{x_t - \eta_t g_t\}$.

- (a) (10 points) Set $\eta_t = \frac{2}{\alpha(t+1)}$. Show that

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

where $x^* \in \arg \min_{x \in C} f(x)$.

- (b) (10 points) Set $\eta_t = \frac{1}{\alpha t}$. Show that

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

where $x^* \in \arg \min_{x \in C} f(x)$.

2. In this question we will work through the convergence analysis of the online (projected) subgradient method for online convex optimization where the loss functions are strongly convex and Lipschitz continuous. Let $f_1, \dots, f_T : \mathbb{R}^d \rightarrow \mathbb{R}$ be a loss functions that are α -strongly convex with respect to the ℓ_2 norm and L -Lipschitz continuous in the ℓ_2 norm. Recall that the online (projected) subgradient method proceeds as follows.

- Choose $x_1 \in C$.
- For $t = 1, 2, 3, \dots, T - 1$:
 - Observe f_t and Select any subgradient $g_t \in \partial f_t(x_t)$ and step size $\eta_t > 0$.
 - Compute $x_{t+1} = \text{Proj}_C\{x_t - \eta_t g_t\}$.

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

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

where x^* is an optimal solution to $\min_{x \in \mathbb{R}^d} \sum_{t=1}^T f_t(x^*)$.

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

$$\sum_{t=1}^T f_t(x_t) - \min_{x \in C} \sum_{t=1}^T f_t(x) \leq \frac{L^2}{\alpha} (1 + \ln T).$$

3. This question is about analyzing the following algorithm for online convex optimization.

- Choose $x_1 \in C$.
- For $t = 1, 2, 3, \dots, T$:
 - Observe f_t .
 - Compute $x_{t+1} \in \arg \min_{x \in C} \sum_{s=1}^t f_s(x)$.

- (a) (10 points) Use induction on T to show the following bound on the regret of the algorithm up to time T .

$$\text{Regret}(T) = \sum_{t=1}^T f_t(x_t) - \min_{x \in C} \sum_{t=1}^T f_t(x) \leq \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x_{t+1}).$$

- (b) (10 points) Suppose that $C = [-1, 1] \subseteq \mathbb{R}$ is the interval between -1 and 1 . At each $t \geq 2$, the adversary chooses a function f_t given by $f_t(x) = g_t x$ where $g_t \in [-1, 1]$ is chosen so that

$$g_t \in \arg \max_{g \in [-1, 1]} g_t x_t.$$

The adversary chooses $f_1(x) = 0.5x$. Then prove that the regret of the algorithm starting with $x_1 = 0$ up to time T is $T - 0.5$.

4. This question is about analyzing the following algorithm for online convex optimization where each f_t is given by $f_t(x) = \ell_t^\top x$ for some $\ell_t \in \mathbb{R}^d$.

- Choose $x_1 = 0$.
- For $t = 1, 2, 3, \dots, T$:
 - Observe ℓ_t .
 - Compute $x_{t+1} \in \arg \min_{x \in \mathbb{R}^d} \left\{ \frac{1}{2\eta} \|x\|_2^2 + \sum_{s=1}^t \ell_s^\top x \right\}$.

(a) (10 points) Prove that $x_{t+1} = x_t - \eta \ell_t$ for $t \geq 1$.

(b) (10 points) Show the following bound on the regret of the algorithm up to time T .

$$\text{Regret}(T) = \sum_{t=1}^T f_t(x_t) - \min_{x \in C} \sum_{t=1}^T f_t(x) \leq \frac{1}{2\eta} \|x^*\|_2^2 + \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x_{t+1})$$

where $x^* \in \arg \min_{x \in \mathbb{R}^d} \sum_{t=1}^T f_t(x)$.

(c) (5 points) Prove that

$$\text{Regret}(T) \leq \frac{1}{2\eta} \|x^*\|_2^2 + \eta \sum_{t=1}^T \|\ell_t\|_2^2.$$

(d) (5 points) Prove that by setting a proper value for η ,

$$\text{Regret}(T) \leq LR\sqrt{2T}$$

if $\|x^*\| \leq R$ and $\|\ell_t\|_2 \leq L$ for all t .

5. (10 points) In this question we prove the convergence of stochastic gradient descent for functions that are strongly convex and Lipschitz continuous. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 stochastic gradient descent proceeds as follows.

- Choose $x_1 \in \mathbb{R}^d$.
- For $t = 1, 2, 3, \dots, T-1$:
 - Obtain an estimator \hat{g}_{x_t} of some $g \in \partial f(x_t)$.
 - Update $x_{t+1} = x_t - \eta_t \hat{g}_{x_t}$ for a step size $\eta_t > 0$.

Set $\eta_t = \frac{1}{\alpha t}$. Assuming $\|\hat{g}_{x_t}\|_2 \leq L$ for all t , show that

$$\mathbb{E} \left[f \left(\frac{1}{T} \sum_{t=1}^T x_t \right) \right] - f(x^*) \leq \frac{L^2(1 + \log T)}{\alpha T}$$

where $x^* \in \arg \min_{x \in \mathbb{R}^d} f(x)$.