IE 539 Convex Optimization Final Exam

Fall 2022

Out: 13rd December 2022

Due: 14th December 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 submissions 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	Total
Points:	40	30	30	100

1. We consider the standard convex optimization problem

$$\min_{x\in C} f(x),$$

under the assumption that f is β-smooth. In this question we will analyse the following algorithm:

- Choose $v_1 := \arg\min_{x\in C} \frac{1}{2}\|x\|_2^2$.
- For $t = 1, \dots, T$:

$$\begin{aligned} x_t &:= \text{proj}_C\left(v_t - \eta\nabla f(v_t)\right) \\ v_{t+1} &:= \text{proj}_C\left(v_t - \eta\nabla f(x_t)\right). \end{aligned}$$

(a) (10 points) First show that for any $x \in C$,

$$\frac{1}{2}\|x - v_1\|_2^2 \le \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|v_1\|_2^2.$$

(b) (10 points) Consider two consecutive iterates $t, t+1$ and their associated points $v_t, x_t, v_{t+1}, x_{t+1}$. Show that for any $z \in C$,

$$\begin{aligned} \eta\nabla f(v_t)^\top(x_t - z) &\le \frac{1}{2}\|v_t - z\|_2^2 - \frac{1}{2}\|x_t - z\|_2^2 - \frac{1}{2}\|v_t - x_t\|_2^2, \\ \eta\nabla f(x_t)^\top(v_{t+1} - z) &\le \frac{1}{2}\|v_t - z\|_2^2 - \frac{1}{2}\|v_{t+1} - z\|_2^2 - \frac{1}{2}\|v_t - v_{t+1}\|_2^2. \end{aligned}$$

(c) (5 points) Consider two consecutive iterates $t, t+1$ and their associated points $v_t, x_t, v_{t+1}, x_{t+1}$. Show that for any $z \in C$,

$$\begin{aligned} \eta\nabla f(x_t)^\top(x_t - z) \le\ & \frac{1}{2}\|v_t - z\|_2^2 - \frac{1}{2}\|v_{t+1} - z\|_2^2 - \frac{1}{2}\|x_t - v_{t+1}\|_2^2 - \frac{1}{2}\|v_t - x_t\|_2^2 \\ & + \eta\left(\nabla f(v_t) - \nabla f(x_t)\right)^\top(v_{t+1} - x_t). \end{aligned}$$

(d) (5 points) Show that

$$-\frac{1}{2}\|x_t - v_{t+1}\|_2^2 - \frac{1}{2}\|v_t - x_t\|_2^2 + \eta\left(\nabla f(v_t) - \nabla f(x_t)\right)^\top(v_{t+1} - x_t) \le (\eta\beta - 1)\|v_t - x_t\|_2\|v_{t+1} - x_t\|_2,$$

(e) (5 points) Show that choosing $\eta = 1/\beta$ guarantees that for any $z \in C$

$$\eta\nabla f(x_t)^\top(x_t - z) \le \frac{1}{2}\|v_t - z\|_2^2 - \frac{1}{2}\|v_{t+1} - z\|_2^2.$$

(f) (5 points) Suppose we run the algorithm with $\eta = 1/\beta$. Define $\bar{x}_T = \frac{1}{T}\sum_{t\in[T]} x_t$ and let x^* be an optimal solution to the optimization problem. Show that

$$f(\bar{x}_T) - f(x^*) \le \frac{\beta(\|x^*\|_2^2 - \|v_1\|_2^2)}{2T}.$$

2. We consider the following constrained convex programming problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) \le 0 \end{aligned}$$

where $f, g : \mathbb{R}^d \to \mathbb{R}$. What follows is a penalized version of this convex program.

$$\text{minimize} \quad f(x) + \frac{\eta}{2}\left(g(x)\right)_+^2$$

where $(\delta)_+ = \max\{\delta, 0\}$.

(a) (15 points) Let

$$h(\delta) = \frac{\eta}{2}(\delta)_+^2, \quad \delta \in \mathbb{R}.$$

Show that

$$h^*(\lambda) = \begin{cases} \frac{1}{2\eta}\lambda^2, & \text{if } \lambda \ge 0, \\ +\infty, & \text{if } \lambda < 0. \end{cases}$$

(b) (15 points) Show that the penalized problem is equivalent to

$$\min_{x\in\mathbb{R}^d} \max_{\lambda\geq 0} \left\{ f(x) + \lambda g(x) - \frac{1}{2\eta}\lambda^2 \right\}.$$

3. Suppose we have binary-labelled data $\{(x_i, y_i)\}_{i\in[n]}$ where $x_i \in \mathbb{R}^d$ and $y_i \in \{\pm 1\}$. We wish to build a classifier, i.e., a function $h : \mathbb{R}^d \to \{\pm 1\}$ that can predict the label $y \in \{\pm 1\}$ from the feature $x \in \mathbb{R}^d$ of a new point (x, y). That is, given $x \in \mathbb{R}^d$, the prediction of y is $h(x) \in \{\pm 1\}$.

   In order to meaningfully solve this problem, we need to make assumptions on the form of h. We consider the following structural form:

$$h(x) = \operatorname{sign}(w^\top x + b),$$

   that is, h classifies x according to which side of the hyperplane $H = \{z : w^\top z + b = 0\}$ that x belongs to.

   Our job is to choose the best fitting w, b using the data $\{(x_i, y_i)\}_{i\in[n]}$. A well-known optimization model to do this is the *support vector machine* (SMV):

$$\min_{w,b} \frac{1}{n} \sum_{i\in[n]} \max\{0, 1 - y_i(w^\top x_i + b)\}.$$

   (a) (20 points) To improve out-of-sample predictions, a regularization term $\frac{\tau}{2}\|w\|_2^2$ is often added to the objective, where $\tau > 0$. Therefore the problem we solve is

$$\min_{w,b} \left\{ \frac{\tau}{2}\|w\|_2^2 + \frac{1}{n} \sum_{i\in[n]} \max\{0, 1 - y_i(w^\top x_i + b)\} \right\}.$$

   Formulate this as a convex optimization problem with quadratic objective and linear constraints, then derive the Lagrange dual.

   (b) (10 points) At optimality, the primal optimal vector w takes the form of a weighted sum of data points:

$$w = \sum_{i\in[n]} \gamma_i (y_i x_i).$$

   The indices i where $\gamma_i > 0$ are the so-called *support vectors*. Describe the relationship between s_i, the data point (x_i, y_i) and the function $w^\top z + b$ for the support vectors.