SK Hynix ML Courses - Convex Optimization in ML

Solutions to Exercise Problems 1

TA Gyeongjo Hwang*, Doyeon Kim†

July 1, 2019

1. *(1st order condition of convex functions)* Suppose $f : \mathbf{R}^d \to \mathbf{R}$ is differentiable (i.e., its gradient ∇f exists at each point in the domain of the function f, denoted by $\mathbf{dom} f$). Suppose that $\mathbf{dom} f$ is a convex set, i.e. for $x_1, x_2 \in \mathbf{dom} f$, $\lambda x_1 + (1-\lambda)x_2 \in \mathbf{dom} f$ for all $\lambda \in [0,1]$. This problem explores the proof of the following claim: f is convex if and only if

$$f(x) \geq \nabla f(x^*)^T (x - x^*) + f(x^*) \quad \forall x, x^* \in \mathbf{dom} f. \tag{1}$$

(a) Suppose that $d = 1$. Show that if $f(x)$ is convex, then (1) holds.

Solution)
If $x = x^*$, it is straightforward that (1) holds. So let's focus on $x \neq x^*$ case. From the convexity of $f(x)$, we get the following inequality:

$$f(\lambda x + (1-\lambda)x^*) \leq \lambda f(x) + (1-\lambda) f(x^*) \quad \forall x, x^* \in \mathbf{dom} f,\ \lambda \in [0,1].$$

We can rewrite the above inequality as:

$$f(x) - f(x^*) \geq \frac{f(\lambda x + (1-\lambda)x^*) - f(x^*)}{\lambda(x - x^*)}(x - x^*).$$

From the definition of derivative, as $\lambda \to 0$, we get

$$f(x) - f(x^*) \geq f'(x^*)(x - x^*).$$

(b) Suppose that $d = 1$. Show that if (1) holds, then $f(x)$ is convex.

Solution)
Let $z = \lambda x + (1-\lambda)y$ for all $x, y \in \mathbf{dom} f$. Then, we can write

$$\begin{aligned} f(x) &\geq f'(z)(x - z) + f(z) \quad \cdots \quad ① \\ f(y) &\geq f'(z)(y - z) + f(z) \quad \cdots \quad ② \end{aligned}$$

Then, $\lambda \cdot ① + (1-\lambda) \cdot ②$ can be written as

$$\begin{aligned} \lambda f(x) + (1-\lambda) f(y) &\geq f(z) + f'(z)(\lambda x + (1-\lambda)y - z) \\ &\overset{(1)}{=} f(z) \\ &= f(\lambda x + (1-\lambda)y) \end{aligned}$$

where (1) follows from the fact that $z = \lambda x + (1-\lambda)y$. By definition, $f(x)$ is convex.

*hkj4276@kaist.ac.kr
†dy.kim@kaist.ac.kr

(c) Prove the claim for arbitrary d.

Solution)
(i) (If $f(x)$ is convex, then (1) holds.)
From the convexity of $f(x)$, we get the following inequality:

$$f(\lambda x + (1-\lambda)x^*) \leq \lambda f(x) + (1-\lambda)f(x^*) \quad \forall x, x^* \in \mathbf{dom} f,\ \lambda \in [0,1].$$

We can rewrite the above inequality as:

$$f(x) - f(x^*) \geq \frac{f(\lambda x + (1-\lambda)x^*) - f(x^*)}{\lambda}.$$

Now let $g(\lambda) := f(x^* + \lambda(x - x^*))$. Then we get

$$f(x) - f(x^*) \geq \frac{g(\lambda) - g(0)}{\lambda}.$$

As $\lambda \to 0$, we get

$$f(x) - f(x^*) \geq g'(0).$$

Since $g'(\lambda) = \nabla f(x^* + \lambda(x - x^*))^T(x - x^*)$, $g'(0)$ is $\nabla f(x^*)^T(x - x^*)$. So if we substitute $\nabla f(x^*)^T(x - x^*)$ for $g'(0)$ in the above inequality, we get

$$f(x) \geq \nabla f(x^*)^T(x - x^*) + f(x^*).$$

Therefore, $f(x)$ is convex.

(ii) (If (1) holds, then $f(x)$ is convex.)
Let $z = \lambda x + (1-\lambda)y$ for all $x, y \in \mathbf{dom} f$. Then, we can write

$$\begin{aligned} f(x) &\geq \nabla f(z)^T(x - z) + f(z) \quad \cdots \quad ① \\ f(y) &\geq \nabla f(z)^T(y - z) + f(z) \quad \cdots \quad ② \end{aligned}$$

Then, $\lambda \cdot ① + (1-\lambda) \cdot ②$ can be written as

$$\begin{aligned} \lambda f(x) + (1-\lambda)f(y) &\geq \nabla f(z)^T(\lambda x + (1-\lambda)y - z) + f(z) \\ &\overset{(1)}{=} f(z) \\ &= f(\lambda x + (1-\lambda)y) \end{aligned}$$

where (1) follows from the fact that $z = \lambda x + (1-\lambda)y$. By definition of convexity, this implies that $f(x)$ is convex.

2. *(Optimality condition for a convex function)* Consider the same function f in Problem 1. Suppose that f is convex. In class, it was claimed that

$$\nabla f(x^*) = 0 \iff f(x) \geq f(x^*) \ \ \forall x. \tag{2}$$

(a) Using (1), prove the forward direction of the claim (2).

Solution)
By the inequality (1) above, we have

$$f(x_2) \geq \nabla f(x_1)^T(x_2 - x_1) + f(x_1) \quad \text{for } \forall x_1, x_2 \in \mathbf{dom} f.$$

Let $x_1 = x^*$ s.t. $\nabla f(x_1) = \nabla f(x^*) = 0$. This implies that $f(x_2) \geq \nabla f(x^*)^T(x_2 - x^*) + f(x^*) = f(x^*)$ for $\forall x_2 \in \mathbf{dom} f$. Therefore, $\nabla f(x^*) = 0 \implies f(x) \geq f(x^*) \ \ \forall x$.

(b) For $x, x^* \in \mathbf{dom} f$, consider a point that lies in between:

$$z(\lambda) = \lambda x + (1-\lambda)x^* \in \mathbf{dom} f \tag{3}$$

where $\lambda \in [0,1]$. Show that

$$\frac{d}{d\lambda} f(z(\lambda))\Big|_{\lambda=0} = \nabla f(x^*)^T (x - x^*). \tag{4}$$

Solution)
Note that $\frac{d}{d\lambda} f(z(\lambda)) \overset{(i)}{=} \nabla f(z(\lambda))^T \frac{d}{d\lambda} z(\lambda) \overset{(ii)}{=} \nabla f(z(\lambda))^T (x - x^*)$ where (i) follows from the chain rule and (ii) follows from (3). Hence, $\frac{d}{d\lambda} f(z(\lambda))\Big|_{\lambda=0} = \nabla f(x^*)^T (x - x^*)$ as $z(0) = x^*$.

(c) Using the result in part (b), prove:

$$f(x) \geq f(x^*) \ \ \forall x \implies \nabla f(x^*)^T (x - x^*) \geq 0 \ \ \forall x. \tag{5}$$

Solution)
(Proof by contradiction) Suppose that there exists some $x \in \mathbf{dom} f$ s.t. $\nabla f(x^*)^T (x - x^*) < 0$. Then using the $z(\lambda)$ defined above, we get $\frac{d}{d\lambda} f(z(\lambda))\Big|_{\lambda=0} = \nabla f(x^*)^T (x - x^*) < 0$, i.e. $f(z(\lambda))$ is a decreasing function in λ (in the regime where λ is close to 0). Therefore, $f(z(\lambda)) < f(z(0)) = f(x^*)$ for $\lambda \approx 0$. This contradicts our assumption that $f(x) \geq f(x^*)$ $\forall x \in \mathbf{dom} f$, so the claim is proved.

(d) Using the result in part (c), prove the backward direction of the claim (2).

Solution)
(Proof by contradiction) Suppose that $\nabla f(x^*) \neq 0$ when $f(x) \geq f(x^*)$ $\forall x \in \mathbf{dom} f$. Here the key thing to note is that there is no constraint on x, except that $x \in \mathbf{dom} f$. So one can choose x such that $x - x^*$ points to an *arbitrary* direction. This implies that we can easily choose x such that $\nabla f(x^*)^T (x - x^*) < 0$. This contradicts (c) above that $\nabla f(x^*)^T (x - x^*) \geq 0$ $\ \forall x$. Therefore we conclude that $\nabla f(x^*)^T (x - x^*) \geq 0$ $\ \forall x \implies \nabla f(x^*) = 0$.
Through (a) $\sim$ (d), we conclude that the statement (2) holds indeed.

3. *(Convex functions)* Let $f_i : \mathbf{R}^d \to \mathbf{R}$ be convex functions for $i = 1, 2$.

(a) Show that $f_1(x) + f_2(x)$ is convex.

Solution)
Let $h(x) = f(x) + g(x)$. For all $x, y \in \mathbf{dom} f \cap \mathbf{dom} g$ and $\lambda \in [0,1]$,

$$\begin{aligned} h(\lambda x + (1-\lambda)y) &= f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \\ &\overset{(1)}{\leq} \lambda f(x) + (1-\lambda)f(y) + \lambda g(x) + (1-\lambda)g(y) \\ &\leq \lambda(f(x) + g(x)) + (1-\lambda)(f(y) + g(y)) \\ &= \lambda h(x) + (1-\lambda)h(y) \end{aligned}$$

where (1) follows from the fact that $f(x)$ and $g(x)$ are convex.

(b) Show that $\max\{f_1(x), f_2(x)\}$ is convex.

Solution)

Let $f(x) = \max\{f_1(x), f_2(x)\}$. For all $x, y \in \mathbf{dom} f_1 \cap \mathbf{dom} f_2$ and $\lambda \in [0, 1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \max\{f_1(\lambda x + (1-\lambda)y), f_2(\lambda x + (1-\lambda)y)\} \\ &\le \max\{\lambda f_1(x) + (1-\lambda) f_1(y), \lambda f_2(x) + (1-\lambda) f_2(y)\} \\ &\overset{(1)}{\le} \lambda \max\{f_1(x), f_2(x)\} + (1-\lambda) \max\{f_1(y), f_2(y)\} \\ &= \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

where (1) follows from the fact that $f_i(x) \le \max\{f_1(x), f_2(x)\}$ for any $i = 1, 2$.

4. *(2nd-order condition of convex functions)* Suppose $f : \mathbf{R}^d \to \mathbf{R}$ is twice differentiable, i.e., its Hessian (the second derivative) $\nabla^2 f$ exists at each point in $\mathbf{dom} f$. Suppose that $\mathbf{dom} f$ is a convex set, i.e., for $x_1, x_2 \in \mathbf{dom} f$, $\lambda x_1 + (1-\lambda) x_2 \in \mathbf{dom} f$ for all $\lambda \in [0, 1]$. This problem explores the proof of the following claim: f is convex if and only if

$$f(x) \text{ is positive semi-definite (PSD), i.e., } \nabla^2 f(x) \succeq 0 \quad \forall x \in \mathbf{dom} f. \tag{6}$$

(a) State the definition of a *positive semi-definite matrix*.

Solution)
Suppose a matrix $A \in S^n$, i.e. A is a real symmetric $n \times n$ matrix. Then, A is called *positive semi-definite* if A satisfies $x^T A x \ge 0 \ \ \forall x \in \mathbf{R}^n$.

(b) Suppose that $d = 1$. Show that if $f(x)$ is convex, then (6) holds.

Solution)
Let $x, y \in \mathbf{dom} f, y > x$. From the 1st-order condition of convexity, the followings hold:

$$\begin{aligned} f(y) &\ge f(x) + f'(x)(y-x) \\ f(x) &\ge f(y) + f'(y)(x-y) \end{aligned}$$

Then, with some substractions, we have

$$f'(x)(y-x) \le f(y) - f(x) \le f'(y)(y-x).$$

Dividing both the LHS and the RHS by $(y-x)^2$ gives

$$\frac{f'(y) - f'(x)}{y - x} \ge 0, \quad \forall x, y, \ x \neq y.$$

By letting $y \to x$, we get

$$f''(x) \ge 0, \quad \forall x \in \mathbf{dom} f.$$

(c) Suppose that $d = 1$. Show that if (6) holds, then $f(x)$ is convex.

Solution)
We employ the **Taylor's theorem** without a proof which states the following:

$$f(y) = f(x) + f'(x)(y-x) + \frac{f''(z)}{2}(y-x)^2$$

for some point $z \in [x, y]$.
Since $f''(x) \ge 0$,

$$f(y) \ge f(x) + f'(x)(y-x)$$

Thus, f(x) in convex.

(d) Using the 1st-order condition of convex functions or otherwise, prove the 2nd-order condition for arbitrary d.

Solution)
Recall that convexity is equivalent to convexity along all lines; i.e., $f : \mathbf{R}^d \to \mathbf{R}$ is convex if $g(\alpha) = f(x_0 + \alpha v)$ is convex $\forall x_0 \in \mathbf{dom} f$ and $\forall v \in \mathbf{R}^d$. Then, it's enough to show that $g(\alpha)$ is convex iff

$$g''(\alpha) = v^T \nabla^2 f(x_0 + \alpha v) v \geq 0,$$

$\forall x_0 \in \mathbf{dom} f,\ \forall v \in \mathbf{R}^d$ and $\forall \alpha$ s.t. $x_0 + \alpha v \in \mathbf{dom} f$. Since it is the case of $d = 1$, we have shown it from (b) and (c). Hence, f is convex iff $\nabla^2 f(x) \succeq 0 \ \ \forall x \in \mathbf{dom} f$.