ENGG 5501: Foundations of Optimization	2018–19 First Term

Homework Set 2 Solution

Instructor: Anthony Man–Cho So — October 1, 2018

Problem 1.

(a) **(10pts).** By Theorem 12 of Handout 2, we have

$$\begin{aligned} f(y) &\geq f(x) + (\nabla f(x))^T(y-x), \\ f(x) &\geq f(y) + (\nabla f(y))^T(x-y). \end{aligned}$$

Adding the above two inequalities yields the desired result.

(b) **(10pts).** Let $x, y \in \mathbb{R}^n$ be arbitrary. Consider the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(t) = f(x + t(y-x))$. By the Fundamental Theorem of Calculus and the Chain Rule, we have

$$f(y) - f(x) = g(1) - g(0) = \int_0^1 g'(t)\, dt = \int_0^1 \nabla f(x + t(y-x))^T (y-x)\, dt.$$

Upon writing

$$\int_0^1 \nabla f(x+t(y-x))^T(y-x)\, dt = \int_0^1 \left[\nabla f(x + t(y-x)) - \nabla f(x)\right]^T (y-x)\, dt + \nabla f(x)^T(y-x)$$

and using the Cauchy–Schwarz inequality together with the Lipschitz continuity of ∇f, we obtain

$$\begin{aligned} \left| f(y) - f(x) - \nabla f(x)^T(y-x) \right| &\leq \int_0^1 \left| \left[\nabla f(x+t(y-x)) - \nabla f(x)\right]^T (y-x) \right| dt \\ &\leq L\|y-x\|_2^2 \int_0^1 t\, dt \\ &= \frac{L}{2}\|y-x\|_2^2, \end{aligned}$$

as desired.

Problem 2 (10pts). If the function $x \mapsto c^T x$ is constant on S, then its minimum is attained at any point on S. By Theorem 3 of Handout 2, we have $\operatorname{rel\,int}(S) \neq \emptyset$. It follows that the minimum is attained at a point $\bar{x} \in \operatorname{rel\,int}(S)$.

Conversely, suppose that the minimum of the function $x \mapsto c^T x$ is attained at a point $\bar{x} \in \operatorname{rel\,int}(S)$. If $S = \{\bar{x}\}$, then there is nothing to prove. Hence, suppose that $S \neq \{\bar{x}\}$ and let $y \in S \setminus \{\bar{x}\}$ be arbitrary. Let $\mathcal{L}$ be the line that passes through $\bar{x}$ and y. Since $\bar{x} \in \operatorname{rel\,int}(S)$, there exists an $\epsilon > 0$ such that $B(\bar{x}, \epsilon) \cap \operatorname{aff}(S) \subseteq S$. On the other hand, since $\mathcal{L} \subseteq \operatorname{aff}(S)$, we have $B(\bar{x}, \epsilon) \cap \mathcal{L} \subseteq S$. Thus, there exist $z \in S$ and $\theta \in (0,1)$ such that $\bar{x} = \theta y + (1-\theta)z$. Now, since

$$c^T\bar{x} = \theta c^T y + (1-\theta)c^T z$$

and $c^T\bar{x} \leq \min\{c^T y, c^T z\}$ by the minimality of $\bar{x}$ on S, we conclude that $c^T\bar{x} = c^T y = c^T z$. Since $y \in S \setminus \{\bar{x}\}$ is arbitrary, we see that $x \mapsto c^T x$ is constant on S.

Problem 3.

(a) Let $A = U\Lambda U^T$ be the spectral decomposition of A. Then, we have $\mathrm{tr}(AX) = \mathrm{tr}(U\Lambda U^T X) = \mathrm{tr}(\Lambda U^T XU)$. Since

$$\mathrm{tr}(X) = \mathrm{tr}(XUU^T) = \mathrm{tr}(U^T XU)$$

and

$$v^T Xv = (U^T v)^T (U^T XU)(U^T v) \quad \text{for any } v \in \mathbb{R}^n,$$

we see that $X \in \mathcal{U}_k$ iff $U^T XU \in \mathcal{U}_k$, where

$$\mathcal{U}_k \equiv \{Z \in \mathcal{S}^n : \mathrm{tr}(Z) = k,\ I \succeq Z \succeq \mathbf{0}\}.$$

In particular, the given optimization problem is equivalent to

$$\begin{array}{ll} \text{maximize} & \mathrm{tr}(\Lambda X) \\ \text{subject to} & \mathrm{tr}(X) = k, \\ & I \succeq X \succeq \mathbf{0}. \end{array} \tag{1}$$

Now, we claim that there exists an optimal solution to (1) that is diagonal. To see this, observe that $\mathrm{tr}(\Lambda X) = \sum_{i=1}^n \Lambda_{ii} X_{ii}$, and $I \succeq X \succeq \mathbf{0}$ implies that $X_{ii} \in [0,1]$ for $i = 1, 2, \ldots, n$. In particular, if X^* is an optimal solution to (1), then the diagonal matrix $\tilde{X}^* = \mathrm{diag}(X^*_{11}, X^*_{22}, \ldots, X^*_{nn})$ is feasible for (1) and has the same objective value as X^*. This establishes the claim. Consequently, Problem (1) is equivalent to the following linear program:

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^n \Lambda_{ii} x_i \\ \text{subject to} & \displaystyle\sum_{i=1}^n x_i = k, \\ & \mathbf{0} \le x \le e. \end{array} \tag{2}$$

It is easy to verify that the optimal value of (2) is the sum of the largest k quantites in the set $\{\Lambda_{11}, \ldots, \Lambda_{nn}\}$, which is precisely equal to $\lambda_1^k(A)$. This completes the proof.

(b) For each $X \in \mathcal{S}^n$, define the function $f_X : \mathcal{S}^n \to \mathbb{R}$ by $f_X(A) = \mathrm{tr}(AX)$. Clearly, f_X is linear for each $X \in \mathcal{S}^n$. Then, by the result in (a), we have

$$\lambda_1^k(A) = \max_{X \in \mathcal{U}_k} f_X(A);$$

i.e., λ_1^k is the pointwise supremum of a collection of linear functions. Thus, λ_1^k is convex.

Problem 4.

(a) By Theorem 10 of Handout 2 (which requires the closedness of epi(f)), we have

$$f(x) = \sup_{(y,c) \in S_f} \left\{ y^T x - c \right\}, \tag{3}$$

where $S_f = \left\{(y,c) \in \mathbb{R}^n \times \mathbb{R} : y^T x - c \le f(x) \text{ for all } x \in \mathbb{R}^n\right\}$. Moreover, the discussion below Theorem 10 of Handout 2 shows that $S_f = \mathrm{epi}(f^*)$. Hence, we have $(y,c) \in S_f$ iff $f^*(y) \le c$. This, together with (3), implies that

$$f(x) = \sup_{y \in \mathbb{R}^n} \left\{ y^T x - f^*(y) \right\}, \tag{4}$$

which in turn implies that $f = f^{**}$.

(b) Suppose that (i) holds; i.e., $y \in \partial f(x)$. Then, by definition, we have $f(z) \ge f(x) + y^T(z - x)$ for all $z \in \mathbb{R}^n$, or equivalently,

$$y^T x - f(x) \ge y^T z - f(z) \quad \text{for all } z \in \mathbb{R}^n.$$

In particular, we have $y^T x - f(x) \ge \sup_{z \in \mathbb{R}^n} \left\{ y^T z - f(z) \right\} = f^*(y)$. On the other hand, we have $f(x) \ge y^T x - f^*(y)$ from (4). Hence, we obtain $f(x) + f^*(y) = x^T y$; i.e., (ii) holds. By reversing the preceding argument, we see that the converse also holds.

Next, suppose that (ii) holds; i.e., $f(x) + f^*(y) = x^T y$. By the result in (a), we have $f^{**}(x) + f^*(y) = x^T y$. Since $f^{**}(x) \ge z^T x - f^*(z)$ for all $z \in \mathbb{R}^n$, we obtain $y^T x \ge f^*(y) + z^T x - f^*(z)$ for all $z \in \mathbb{R}^n$, or equivalently,

$$f^*(z) \ge f^*(y) + x^T(z - y) \quad \text{for all } z \in \mathbb{R}^n.$$

This shows that $x \in \partial f^*(y)$; i.e., (iii) holds. Again, the converse follows by reversing the preceding argument.