ENGG 5501: Foundations of Optimization **2021–22 First Term**

Homework Set 1 Solution

Instructor: Anthony Man–Cho So September 24, 2021

Problem 1 (15pts). By introducing an extra variable $t \in \mathbb{R}$, the given optimization problem is equivalent to

$$\begin{array}{rl} \text{minimize} & t \\ \text{subject to} & -t \le a_i^T x - b_i \le t, \quad \text{for } i = 1, \ldots, m, \\ & \displaystyle\sum_{i=1}^n |x_i| \le 1, \end{array} \tag{1}$$

where a_i^T denotes the i-th row of A. Now, observe that for any $x \in \mathbb{R}^n$, we can write $x = x^+ - x^-$, where $x^+, x^- \in \mathbb{R}_+^n$. Consider the following optimization problem:

$$\begin{array}{rl} \text{minimize} & t \\ \text{subject to} & -t \le a_i^T (x^+ - x^-) - b_i \le t, \quad \text{for } i = 1, \ldots, m, \\ & \displaystyle\sum_{i=1}^n (x_i^+ + x^-) \le 1, \\ & x^+, x^- \ge \mathbf{0}. \end{array} \tag{2}$$

Given a feasible solution $(\bar{x}, \bar{t})$ to Problem (1), by setting $\bar{x}_i^+ = \max\{\bar{x}_i, 0\}$ and $\bar{x}_i^- = \max\{-\bar{x}_i, 0\}$ for $i = 1, \ldots, n$, we see that $(\bar{x}^+, \bar{x}^-, \bar{t})$ is feasible for Problem (2). Conversely, given a feasible solution $(\bar{x}^+, \bar{x}^-, \bar{t})$ to Problem (2), by setting $\bar{x} = \bar{x}^+ - \bar{x}^-$, we see that $(\bar{x}, \bar{t})$ is feasible for Problem (1). This shows that Problem (1) and (2) are equivalent. Moreover, Problem (2) is a linear program. This completes the reformulation.

Problem 2 (30pts).

(a) **(10pts).** The claim is true. Indeed, let $x \in \text{conv}(S)$. Then, by Handout 2, Proposition 3(b), there exist vectors $x_1, \ldots, x_\ell \in S$ and scalars $\alpha_1, \ldots, \alpha_\ell \ge 0$ such that $x = \sum_{k=1}^\ell \alpha_k x_k$ and $\sum_{k=1}^\ell \alpha_k = 1$. Since A is an affine map, there exist $A_0 \in \mathbb{R}^{m \times n}$ and $y_0 \in \mathbb{R}^m$ such that $A(x) = A_0 x + y_0$ for all $x \in \mathbb{R}^n$. It follows that

$$A(x) = A\left(\sum_{k=1}^\ell \alpha_k x_k\right) = \sum_{k=1}^\ell \alpha_k A(x_k) \in \text{conv}(A(S)).$$

Conversely, let $y \in \text{conv}(A(S))$. Again, by Handout 2, Proposition 3(b), there exist vectors $y_1, \ldots, y_q \in A(S)$ and scalars $\beta_1, \ldots, \beta_q \ge 0$ such that $y = \sum_{k=1}^q \beta_k y_k$ and $\sum_{k=1}^q \beta_k = 1$. Moreover, we can find $x_1, \ldots, x_q \in S$ such that $y_k = A(x_k)$ for $k = 1, \ldots, q$. It follows that

$$y = \sum_{k=1}^q \beta_k A(x_k) = A\left(\sum_{k=1}^q \beta_k x_k\right) \in A(\text{conv}(S)),$$

as desired.

(b) **(10pts).** Recall from Handout 1, Proposition 1 that $\lambda_{\max}(X) \leq 1$ if and only if $I \succeq X$. Hence, we can write $S = \{X \in \mathcal{S}^n : I \succeq X \succeq \mathbf{0}\}$. Now, let $X_1, X_2 \in S$ and $\alpha \in [0,1]$. Set $\bar{X} = \alpha X_1 + (1-\alpha)X_2$. Clearly, we have $\bar{X} \succeq \mathbf{0}$. Moreover, observe that for any $u \in \mathbb{R}^n$,

$$u^T(I-\bar{X})u = u^T(\alpha(I-X_1)+(1-\alpha)(I-X_2))u = \alpha u^T(I-X_1)u + (1-\alpha)u^T(I-X_2)u \geq 0,$$

as $I - X_1 \succeq \mathbf{0}$ and $I - X_2 \succeq \mathbf{0}$. This shows that $\bar{X} \in S$, as desired.

(c) **(10pts).** The set $S = \{X \in \mathcal{S}^n : \text{rank(X)} \leq 1\}$ is not convex. Indeed, let $X_1 = e_1e_1^T$ and $X_2 = e_2e_2^T$, where $e_i \in \mathbb{R}^n$ is the i-th basis vector. Note that both X_1, X_2 are of rank 1 and hence $X_1, X_2 \in S$. However, the matrix $\frac{1}{2}(X_1+X_2)$ is of rank 2.

Problem 3 (25pts).

(a) **(10pts).** Let $x \in \mathbb{R}^n$ be fixed. For any $\alpha > 0$ and $u \in N(x)$, we have $\alpha u^T(y-x) \leq 0$ for all $y \in B(\mathbf{0},1)$. Hence, $N(x)$ is a cone. Moreover, for any $u, v \in N(x)$ and $\alpha \in (0,1)$, we have

$$(\alpha u + (1-\alpha)v)^T(y-x) = \alpha u^T(y-x) + (1-\alpha)v^T(y-x) \leq 0 \text{ for all } y \in B(\mathbf{0},1).$$

It follows that $N(x)$ is convex.

(b) **(15pts).** We claim that

$$N(x) = \begin{cases} \{\mathbf{0}\} & \text{if } \|x\|_2 < 1, \\ \{\alpha x : \alpha \geq 0\} & \text{if } \|x\|_2 = 1. \end{cases}$$

Indeed, if $\|x\|_2 = \ell < 1$, then $x + (1-\ell)v \in B(\mathbf{0},1)$ for all $v \in S^{n-1}$ (recall that $S^{n-1} = \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$). Hence, if $u \in N(x)$, then we must have $u^T(x+(1-\ell)v-x) = (1-\ell)u^Tv \leq 0$ for all $v \in S^{n-1}$. This implies that $u = \mathbf{0}$.

On the other hand, if $\|x\|_2 = 1$, then by the Cauchy-Schwarz inequality, for any $\alpha \geq 0$, we have

$$\alpha x^T(y-x) = \alpha(x^Ty - \|x\|_2) \leq \alpha(\|y\|_2 - 1) \leq 0 \text{ for all } y \in B(\mathbf{0},1).$$

This shows that $\{\alpha x : \alpha \geq 0\} \subseteq N(x)$. Conversely, let $u \in N(x)$. By the result in (a), we may assume without loss of generality that $\|u\|_2 = 1$. Then, since $u \in B(\mathbf{0},1)$, we have $u^T(u-x) \leq 0$. This implies that $u^Tu = 1 \leq u^Tx \leq \|u\|_2 \cdot \|x\|_2 = 1$, which, together with the Cauchy-Schwarz inequality, yields $u = x$. It follows that $N(x) \subseteq \{\alpha x : \alpha \geq 0\}$, as desired.

Problem 4 (30pts).

(a) **(15pts).** If $x \in H^-(s,c)$, then we clearly have $\Pi_{H^-(s,c)}(x) = x$. Suppose that $x \notin H^-(s,c)$. Intuitively, the point $\Pi_{H^-(s,c)}(x)$ should lie on the hyperplane $H(s,c)$ and the vector $x - \Pi_{H^-(s,c)}(x)$ should be normal to the hyperplane $H(s,c)$. In other words, we should have $\Pi_{H^-(s,c)}(x) \in H(s,c)$ and $x - \Pi_{H^-(s,c)}(x) = \alpha s$ for some $\alpha \in \mathbb{R}$. These imply that $s^T(x-\alpha s) = c$, or equivalently, $\alpha = (s^Tx - c)/s^Ts$. Based on the above discussion, we have the following candidate for $\Pi_{H^-(s,c)}(x)$:

$$\Pi_{H^-(s,c)}(x) = x - \frac{(s^Tx - c)_+}{s^Ts}s. \tag{3}$$

Here, $(x)_+ = \max\{x, 0\}$. To prove the correctness of the above formula, we consider two cases:

Case 1: $x \in H^-(s, c)$.

Then, we have $s^T x \leq c$, which implies that $(s^T x - c)_+ = 0$. Hence, the formula (3) yields $\Pi_{H^-(s,c)}(x) = x$, which is the desired result.

Case 2: $x \notin H^-(s, c)$.

Let $y \in H^-(s, c)$ be arbitrary. Since $s^T x > c$ and $s^T y \leq c$, we obtain

$$\begin{aligned}
\left(y - \Pi_{H^-(s,c)}(x)\right)^T \left(x - \Pi_{H^-(s,c)}(x)\right) &= \left(y - x + \frac{s^T x - c}{s^T s} s\right)^T \left(\frac{s^T x - c}{s^T s} s\right) \\
&= \frac{(s^T x - c)(s^T (y - x))}{s^T s} + \frac{(s^T x - c)^2}{s^T s} \\
&= \frac{(s^T x - c)(s^T y - c)}{s^T s} \\
&\leq 0.
\end{aligned}$$

The correctness of the formula (3) then follows from Theorem 5 of Handout 2.

(b) **(15pts).** Let $v \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$ be fixed. Observe that for any $x \in \Delta$, we have

$$\|v + \delta e - x\|_2^2 = \|v - x\|_2^2 + \delta^2 \|e\|_2^2 + 2\delta (v - x)^T e = \|v - x\|_2^2 + \delta^2 n + 2\delta e^T v - 2\delta.$$

Since $\delta^2 n + 2\delta e^T v - 2\delta$ is a constant, it follows that

$$\Pi_\Delta(v + \delta e) = \arg\min_{x \in \Delta} \|v + \delta e - x\|_2^2 = \arg\min_{x \in \Delta} \|v - x\|_2^2 = \Pi_\Delta(v),$$

as desired.