ENGG 5501: Foundations of Optimization	2017–18 First Term
Final Examination	
Time Limit: 2 Hours	December 7, 2017

SOLVE THE FOLLOWING PROBLEMS:

√ **Problem 1 (20pts).** Let $\Delta = \{x \in \mathbb{R}^n : e^T x = 1, x \geq 0\}$ be the standard simplex. Show that for any $v \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$, we have

$$\Pi_\Delta(v) = \Pi_\Delta(v + \delta e).$$

As usual, $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$ is the vector of all ones.

√ **Problem 2 (15pts).** Give an example of a primal–dual pair of LPs, both of which have multiple optimal solutions.

Problem 3 (20pts). Consider the following SDP:

$$\begin{array}{ll} \inf & X_{11} \\ \text{subject to} & \begin{bmatrix} X_{11} & 1 \\ 1 & X_{22} \end{bmatrix} \succeq 0. \end{array} \qquad (P)$$

(a) **(10pts).** Write down the dual (D) of (P) and show that the duality gap between (P) and (D) is zero.

(b) **(10pts).** Is the common optimal value of (P) and (D) attained by a primal feasible solution and a dual feasible solution? Justify your answer.

Problem 4 (15pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function. Consider the problem

$$\inf_{x \in \Delta} f(x), \qquad (S)$$

where Δ is the standard simplex defined in Problem 1. Show that x^* is an optimal solution to (S) if and only if $x^* \in \Delta$ and there exists a $w^* \in \mathbb{R}$ such that

$$\begin{aligned} \frac{\partial f(x^*)}{\partial x_i} &\geq w^* \quad \text{for } i = 1, \ldots, n, \\ \frac{\partial f(x^*)}{\partial x_i} &= w^* \quad \text{if } x_i^* > 0, \text{ for } i = 1, \ldots, n. \end{aligned}$$

√ **Problem 5 (30pts).** Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}$ be given convex functions and $A \in \mathbb{R}^{m \times n}$ be a given matrix. Consider the following problems:

$$\begin{aligned} v_p^* &= \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax)\}, \\ v_d^* &= \sup_{w \in \mathbb{R}^m} \{-f^*(A^T w) - g^*(-w)\}. \end{aligned}$$

Here, $f^*(z) = \sup_{x \in \mathbb{R}^n} \{z^T x - f(x)\}$ is the conjugate of f.

(a) **(15pts).** Show that $v_p^* \geq v_d^*$.

(b) **(15pts).** Suppose that we take $f(x) = \|x\|_1$. Derive an explicit expression for f^*. (*Hint: Consider the dual norm of* $\|\cdot\|_1$.)

ENGG 5501: Foundations of Optimization	2017–18 First Term
Solution to Final Examination	
Time Limit: 2 Hours	December 7, 2017

Simplex: nonempty convex. closed.

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $\Delta = \{x \in \mathbb{R}^n : e^T x = 1, x \geq 0\}$ be the standard simplex. Show that for any $v \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$, we have

$$\Pi_\Delta(v) = \Pi_\Delta(v + \delta e).$$

As usual, $e = (1, 1, \dots, 1) \in \mathbb{R}^n$ is the vector of all ones.

ANSWER: 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 (by the definition)

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

as desired.

Problem 2 (15pts). Give an example of a primal–dual pair of LPs, both of which have multiple optimal solutions.

ANSWER: For instance, consider the following LP in standard primal form:

$$\begin{array}{ll} \min & x_1 \\ \text{subject to} & x_1 = 0, \\ & (x_1, x_2) \geq 0. \end{array}$$

$(1,0)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

min $(0 \cdot 0)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$
s.t. $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
$(x_1, x_2) \geq 0$

Its optimal solution set is given by $\{(0, x_2) \in \mathbb{R}^2 : x_2 \geq 0\}$. Moreover, its dual is

$$\begin{array}{ll} \max & 0 \\ \text{subject to} & \begin{bmatrix} 1 \\ 0 \end{bmatrix} y \leq \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \end{array}$$

$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

max $(0 \cdot 0)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$
s.t. $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq$

whose optimal solution set is $\{y \in \mathbb{R} : y \leq 1\}$. $y_1 \leq 1$.

Problem 3 (20pts). Consider the following SDP:

$$\begin{array}{ll} \inf & X_{11} \\ \text{subject to} & \begin{bmatrix} X_{11} & 1 \\ 1 & X_{22} \end{bmatrix} \succeq 0. \end{array} \qquad (P)$$

(a) **(10pts).** Write down the dual (D) of (P) and show that the duality gap between (P) and (D) is zero.

$$\begin{cases} X_{11} \geq 0 \\ X_{22} \geq 0 \\ X_{11} X_{22} \geq 1. \end{cases}$$

ANSWER: The given SDP can be written as

$$\begin{array}{ll} \inf & C \bullet X \\ \text{subject to} & A \bullet X = 2, \quad \Leftrightarrow X_{12} + X_{12} = 2 \Leftrightarrow X_{12} = 1. \\ & X \succeq 0, \end{array}$$

where

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12} & X_{22} \end{bmatrix}.$$

Hence, the dual (D) is given by

$$\begin{array}{ll} \sup & 2y \\ \text{subject to} & S = \begin{bmatrix} 1 & -y \\ -y & 0 \end{bmatrix} \succeq 0. \qquad -y^2 \geq 0 \Leftrightarrow y \leq 0 \Leftarrow \end{array}$$

Now, observe that $X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is a ① strictly feasible solution to (P). Moreover, the optimal value of ② (P) is bounded below by 0, since the diagonal entries of a symmetric positive semidefinite matrix must be non-negative. Thus, by the CLP strong duality theorem, we conclude ~~that~~ the duality gap between ~~(P) and (D) is zero.~~

(b) **(10pts).** Is the common optimal value of (P) and (D) attained by a primal feasible solution and a dual feasible solution? Justify your answer.

ANSWER: A necessary condition for $S \succeq 0$ is $\det(S) \geq 0$; i.e. $y^2 \leq 0$. Thus, we see that $y = 0$ is the only feasible solution to (D), which implies that the optimal value of (D) (and hence of (P) by ~~the result~~ of (a)) is 0. The dual optimal value is attained by $y = 0$. On the other hand, the primal optimal value is not attained. Indeed, the feasible set of (P) is given by

$$\left\{ X \in \mathcal{S}^2 : X_{11} \geq 0,\ X_{22} \geq 0,\ X_{11} X_{22} \geq 1 \right\},$$

which implies that $X(\epsilon) = \begin{bmatrix} \epsilon & 1 \\ 1 & \epsilon^{-1} \end{bmatrix}$ is feasible for (P) for any $\epsilon > 0$. However, any point $X \in \mathcal{S}^2$ with $X_{11} = 0$ is not feasible for (P).

Problem 4 (15pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function. Consider the problem

$$\inf_{x \in \Delta} f(x), \tag{S}$$

where Δ is the standard simplex defined in Problem 1. Show that x^* is an optimal solution to (S) if and only if $x^* \in \Delta$ and there exists a $w^* \in \mathbb{R}$ such that

$$\begin{array}{rcll} \dfrac{\partial f(x^*)}{\partial x_i} & \geq & w^* & \text{for } i = 1, \ldots, n, \\ \dfrac{\partial f(x^*)}{\partial x_i} & = & w^* & \text{if } x_i^* > 0, \text{ for } i = 1, \ldots, n. \end{array}$$

ANSWER: Since (S) is a linearly constrained convex optimization problem, its associated first-order conditions, which are given by

KKT.

$$
\begin{aligned}
x^* &\in \Delta, \\
\nabla f(x^*) - v^* - w^* e &= 0, && (1) \\
v^* &\geq 0, && (2) \\
v_i^* x_i^* &= 0 \quad \text{for } i = 1, \ldots, n, && (3)
\end{aligned}
$$

are both necessary and sufficient for optimality. It is straightforward to verify that (1)–(3) together is equivalent to

$$
\begin{aligned}
\frac{\partial f(x^*)}{\partial x_i} &\geq w^* \quad \text{for } i = 1, \ldots, n, \\
\frac{\partial f(x^*)}{\partial x_i} &= w^* \quad \text{if } x_i^* > 0, \text{ for } i = 1, \ldots, n.
\end{aligned}
$$

This completes the proof.

Problem 5 (30pts). Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}$ be given convex functions and $A \in \mathbb{R}^{m \times n}$ be a given matrix. Consider the following problems:

$$
\begin{aligned}
v_p^* &= \inf_{x \in \Re^n} \{f(x) + g(Ax)\}, \\
v_d^* &= \sup_{w \in \Re^m} \{-f^*(A^T w) - g^*(-w)\}.
\end{aligned}
$$

Here, $f^*(z) = \sup_{x \in \Re^n} \{z^T x - f(x)\}$ is the conjugate of f.

$f^*(A^T w) = \sup_x \{w^T A x - f(x)\}$

$g^*(-w) = \sup_x \{-w^T x - g(x)\}$.

(a) **(15pts).** Show that $v_p^* \geq v_d^*$.

ANSWER: Observe that

$$
\begin{aligned}
v_p^* = \quad &\inf \quad f(x) + g(y) \\
&\text{subject to} \quad Ax = y, \\
&\qquad x \in \mathbb{R}^n,\ y \in \mathbb{R}^m.
\end{aligned}
$$

$v_d^* = \sup_w \{-\sup_x \{w^T A x - f(x)\} - \sup_x \{-w^T x - g(x)\}\}$

We can take the Lagrangian dual of the above problem as follows:

$= \sup_w \{\inf_x \{f(x) - w^T A x\} + \inf_x \{g(x) + w^T x\}\}$

$$v_d^* = \sup_{w \in \Re^m} \theta(w),$$

where

$\leq \sup_w \{\inf_x \{f(x) + g(x) + w^T x - w^T A x$

$$
\begin{aligned}
\theta(w) &= \inf_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \{f(x) + g(y) + w^T(y - Ax)\} \\
&= -\sup_{x \in \mathbb{R}^n} \{(A^T w)^T x - f(x)\} - \sup_{y \in \mathbb{R}^m} \{(-w)^T y - g(y)\} \\
&= -f^*(A^T w) - g^*(-w).
\end{aligned}
$$

$\leq \inf_x \sup_w \{f(x) + g(x) + w^T x - w^T A x\}$

It then follows from the Lagrangian weak duality theorem that $v_p^* \geq v_d^*$.

$\sup_w \{g(x) + w^T(x - Ax)\} = \begin{cases} +\infty, & x \neq Ax \\ g(Ax), & x = Ax \end{cases} \Rightarrow \inf_x \{f(x) + g(Ax)\}$

$= v_p^*$.

~~$\nabla_w \{g(x) + w^T(x - Ax)\} = x - Ax = 0 \Rightarrow x = Ax$~~

~~$\therefore \sup_w \{g(x) + w^T(x - Ax)\} =$~~

(b) **(15pts)**. Suppose that we take $f(x) = \|x\|_1$. Derive an explicit expression for f^*. (*Hint: Consider the dual norm of* $\|\cdot\|_1$.)

ANSWER: By definition of the conjugate function and the dual norm, we have

$$(\|\cdot\|_1)^*(z) = \sup_{x\in\mathbb{R}^n} \left\{ z^T x - \|x\|_1 \right\},$$

$$\|z\|_\infty = \max_{x\in\mathbb{R}^n : \|x\|_1 \le 1} z^T x.$$

Consider the following cases:

Case 1: $\|z\|_\infty \le 1$. $\cancel{\max_{\|x\|_1 \le 1} z^T x = \max_{\|x\|_1 = 1} z^T x = 1}$

Then, we have $z^T x \le \|x\|_1$ for all $x \in \mathbb{R}^n$ and equality holds when $x = 0$. It follows that $\sup_{x\in\mathbb{R}^n} \left\{ z^T x - \|x\|_1 \right\} = 0$. $z^T x \le \|z\|_\infty \|x\|_1 \le \|x\|_1$

Case 2: $\|z\|_\infty > 1$.

Then, there exists an $\bar{x} \in \mathbb{R}^n$ such that $\|\bar{x}\|_1 \le 1$ and $z^T \bar{x} > 1$. Observe that

$$\sup_{x\in\mathbb{R}^n} \left\{ z^T x - \|x\|_1 \right\} \ge z^T (t\bar{x}) - \|t\bar{x}\|_1 = t(\underbrace{z^T \bar{x}}_{>1} - \underbrace{\|\bar{x}\|_1}_{\le 1}) > 0$$

for any $t > 0$. It follows that $\sup_{x\in\mathbb{R}^n} \left\{ z^T x - \|x\|_1 \right\} = +\infty$.

To summarize, we obtain

$$(\|\cdot\|_1)^*(z) = \begin{cases} 0 & \text{if } \|z\|_\infty \le 1, \\ +\infty & \text{otherwise.} \end{cases}$$

$-2(y-x)+w\cdot 2x=0$ compare $\|y - y/\|y\|_2\|_2$ vs. $\|y + \frac{y}{\|y\|_2}\|_2$

$(\|x\|_2^2-1)=2x\neq 0$ for $\forall x\in\{x:\|x\|_2^2=1\}$. $\Rightarrow x=y/\|y\|_2$ is optimal

1: $y=0$. Any x is optimal.

2: $y\neq 0$ $w\neq -1$. $y=(1+w)x \Rightarrow \|y\|_2=|1+w| \Rightarrow \pm\|y\|_2=1+w$ $\begin{cases} w=\|y\|_2-1 \Rightarrow x=\frac{y}{\|y\|_2} \\ w=-\|y\|_2-1 \Rightarrow x=-\frac{y}{\|y\|_2}\end{cases}$

ENGG 5501: Foundations of Optimization **2016–17 First Term**

Final Examination

Time Limit: 2 Hours December 12, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (15pts). Let $y \in \mathbb{R}^n$ be given. Consider the following problem:

$\begin{bmatrix} \end{bmatrix}\succeq 0 \Leftrightarrow A, C\succeq 0, C-B^TA^+B\succeq 0$

$$\begin{array}{ll} \text{minimize} & \|y-x\|_2^2 \\ \text{subject to} & \|x\|_2^2 = 1. \end{array} \tag{S}$$

By considering the first-order optimality condition of (S), determine its optimal solution.

Schur complement

Problem 2 (20pts). Let $m \geq 1$ be given. Consider the following SDP:

$y_1 \geq 4$

0, $y_1\geq 2^2$, $y_k\geq y_{k-1}^2$

$y_m \geq 2^{2^m}$

$$\begin{array}{rcll} v_m^* & = & \inf & y_m \\ & & \text{subject to} & \begin{bmatrix} 1 & 2 \\ 2 & y_1 \end{bmatrix} \succeq 0, \\ & & & \begin{bmatrix} 1 & y_{k-1} \\ y_{k-1} & y_k \end{bmatrix} \succeq 0 \quad \text{for } k=2,\ldots,m, \\ & & & y \in \mathbb{R}^m. \end{array} \tag{SDP}$$

(a) **(10pts).** Determine the optimal value v_m^* of (SDP) as a function of m.

(b) **(10pts).** Is (SDP) strictly feasible? Explain.

Problem 3 (30pts). Let $a_1,\ldots,a_m \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$ be given, with $m \geq n$. Consider the cone

$$C = \left\{ x \in \mathbb{R}^n : a_i^T x \geq 0 \text{ for } i=1,\ldots,m \right\}.$$

(a) **(10pts).** Show that C is pointed if and only if there exist n vectors in the collection $\{a_1,\ldots,a_m\}$ that are linearly independent.

(b) **(20pts).** We say that $d \in C\backslash\{0\}$ is an *extreme ray* of C if there are $n-1$ linearly independent constraints that are active at d. Now, suppose that C is pointed. Consider the LP

$$v^* = \min_{x\in C} c^T x.$$

Show that $v^* = -\infty$ if and only if there exists an extreme ray d of C satisfying $c^T d < 0$. (*Hint: If $v^* = -\infty$, then there exists an $\bar{x} \in C$ such that $c^T\bar{x} < 0$.*)

Problem 4 (35pts). Let $f, h_1, \ldots, h_m : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable functions. Consider the following problem:

$$\begin{array}{ll} \inf & f(x) \\ \text{subject to} & h_i(x) = 0 \quad \text{for } i=1,\ldots,m, \\ & x \in \mathbb{R}^n. \end{array} \tag{NP}$$

Let $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be the Lagrangian function associated with (NP); i.e.,

$$L(x, w) = f(x) + \sum_{i=1}^{m} w_i h_i(x).$$

For any given $c \geq 0$, define the so-called *augmented Lagrangian function* $L_c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ associated with (NP) as

$$L_c(x, w) = f(x) + \sum_{i=1}^{m} w_i h_i(x) + \frac{c}{2} \sum_{i=1}^{m} h_i(x)^2.$$

Suppose that $(\bar{x}, \bar{w}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfies the following conditions:

$\nabla_x L_c(\bar{x}, \bar{w}) = \nabla f(\bar{x}) + \sum_i \bar{w}_i \nabla h_i(\bar{x}) + c \sum_i h_i(\bar{x}) \nabla h_i(\bar{x})$

(I) $\nabla_x L(\bar{x}, \bar{w}) = 0$.

$\nabla^2_{xx} L_c(\bar{x}, \bar{w}) = \nabla^2 f(\bar{x}) + \sum_i (\bar{w}_i + c h_i(\bar{x})) \nabla^2 h_i(\bar{x}) + \sum_i c \nabla h_i(\bar{x})$

(II) $h_i(\bar{x}) = 0$ for $i = 1, \ldots, m$.

(III) $y^T \nabla^2_{xx} L(\bar{x}, \bar{w}) y > 0$ for all $y \neq 0$ satisfying $\nabla h_i(\bar{x})^T y = 0$, where $i = 1, \ldots, m$.

(a) **(10pts).** Compute $\nabla_x L_c(\bar{x}, \bar{w})$ and $\nabla^2_{xx} L_c(\bar{x}, \bar{w})$.

(b) **(15pts).** Show that there exists a constant $\bar{c} \geq 0$ such that $\nabla^2_{xx} L_c(\bar{x}, \bar{w}) \succ 0$ for all $c > \bar{c}$. Hence, show that $\bar{x}$ is an unconstrained local minimum of $x \mapsto L_c(x, \bar{w})$ for all $c > \bar{c}$. You may use without proof the following fact:

Fact 1 *Let $P \in S^n$ and $Q \in S^n_+$ be such that $x^T P x > 0$ for all $x \neq 0$ satisfying $Qx = 0$. Then, there exists a constant $\bar{c} \geq 0$ such that $x^T(P + cQ)x > 0$ for all $x \neq 0$ and $c > \bar{c}$.*

(c) **(10pts).** Using the results in (a) and (b), or otherwise, show that $\bar{x}$ is a strict local minimum of (NP); i.e., there exists an $\epsilon > 0$ such that for all $x \in B(\bar{x}, \epsilon) \setminus \{\bar{x}\}$ satisfying $h_i(x) = 0$ for $i = 1, \ldots, m$, we have $f(\bar{x}) < f(x)$.

$\underbrace{y^T \nabla^2_{xx} L(\bar{x}, \bar{w}) y}_{P} > 0 \quad \forall y \neq 0 \quad \underbrace{\sum_i \nabla h_i(\bar{x}) \nabla h_i(\bar{x})^T}_{Q} y = 0 \qquad x^T(P + cQ)x > 0$

$\exists \varepsilon > 0 \quad L_c(\bar{x}, \bar{w}) \leq L_c(x, \bar{w})$ and $\underline{\nabla^2_{xx} L_c(x, \bar{w}) > 0} \quad \forall x \in B(\bar{x}, \varepsilon)$

Define $g_d(\alpha) = L_c(\bar{x} + \alpha d, \bar{w}) \quad \|d\|_2 = 1$

$\underline{g_d(\alpha) = g_d(0) + \alpha g_d'(0) + \frac{\alpha^2}{2} g_d''(\alpha_0)} \quad \alpha_0 \in (0, \alpha)$

$= L_c(\bar{x}, \bar{w}) + \alpha \nabla_x L_c(\bar{x}, \bar{w})^T d + \frac{\alpha^2}{2} d^T \nabla^2_{xx} L_c(\bar{x} + \alpha_0 d$

$> f(\bar{x})$

$\therefore L_c(\bar{x} + \alpha d, \bar{w}) > f(\bar{x}) \quad \forall \alpha \in (0, \varepsilon)$. $\therefore$ If $x = \bar{x} + \alpha d$, $h_i(x) = 0$

then

$f(x) = L_c(x, \bar{w}) > f(\bar{x})$

ENGG 5501: Foundations of Optimization	2015–16 First Term
Final Examination	
Time Limit: 2 Hours	December 17, 2015

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (25pts). Let $\mathcal{E} \subset \{(i,j) : 1 \le i < j \le n\}$ with $\mathcal{E} \neq \emptyset$ and $d_{ij} > 0$ be given. Define $e = (1,1,\dots,1) \in \mathbb{R}^n$ and $E_{ij} = (e_i - e_j)(e_i - e_j)^T \in \mathcal{S}^n_+$, where $e_i \in \mathbb{R}^n$ is the i–th basis vector. Consider the SDP

$$\begin{array}{ll} \inf & 0 \\ \text{subject to} & E_{ij} \bullet X = d_{ij} \quad \text{for } (i,j) \in \mathcal{E}, \\ & ee^T \bullet X = 0, \\ & X \succeq \mathbf{0}. \end{array} \tag{L}$$

(a) **(10pts).** Let y_{ij} and θ be the dual variables corresponding to the constraints $E_{ij} \bullet X = d_{ij}$ and $ee^T \bullet X = 0$, respectively. Write down the dual of (L). ✓

(b) **(15pts).** Suppose that the dual of (L) is feasible. Let $(\bar{y}, \bar{\theta})$ be any dual feasible solution, where $\bar{y} = (\bar{y}_{ij})_{(i,j)\in\mathcal{E}}$. Show that the matrix $\sum_{(i,j)\in\mathcal{E}} \bar{y}_{ij} E_{ij}$ is semidefinite (i.e., either positive semidefinite or negative semidefinite). (*Hint: Evaluate $E_{ij} \bullet ee^T$.*).

Problem 2 (40pts). Let $n \ge 1$ be given. Consider the set

$$\mathcal{Q}_r^{n+2} = \left\{(u,v,x) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n : 2uv \ge \|x\|_2^2,\ u,v \ge 0\right\}.$$

(a) **(15pts).** Show, by first principles, that $\mathcal{Q}_r^{n+2}$ is a pointed cone.

(b) **(10pts).** Show that $(u,v,x) \in \mathcal{Q}_r^{n+2}$ if and only if $(\bar{u},\bar{v},x) \in \mathcal{Q}^{n+2}$, where $\mathcal{Q}^{n+2} = \left\{(t,z) \in \mathbb{R} \times \mathbb{R}^{n+1} : t \ge \|z\|_2\right\}$ and

$$\bar{u} = \frac{1}{\sqrt{2}}(u+v), \quad \bar{v} = \frac{1}{\sqrt{2}}(u-v).$$

(c) **(15pts).** Using the result in (b) and the fact that $(\mathcal{Q}^{n+2})^* = \mathcal{Q}^{n+2}$, or otherwise, show that $(\mathcal{Q}_r^{n+2})^* = \mathcal{Q}_r^{n+2}$.

Problem 3 (20pts). Let $a_1, \dots, a_n > 0$ be given. Consider the problem

$$\begin{array}{ll} \text{maximize} & x_1 x_2 \cdots x_n \\ \text{subject to} & \displaystyle\sum_{i=1}^n \frac{x_i}{a_i} = 1. \end{array} \tag{P}$$

(a) **(10pts).** Write down the first–order optimality conditions of (P) and explain why they are necessary for optimality.

nponent (b) **(10pts).** Using the result in (a), or otherwise, determine the optimal solution to (P).

$\prod_{j\neq i} x_j - \frac{w}{a_i} = 0 \ (i = 1, \cdots, n)$, constraint is linear $\Rightarrow$ FO necessary

…iply by x_i and sum:

$\prod x_j \sum_{i=1}^n \frac{w x_i}{a_i} = w \Rightarrow n \prod_{j=1}^n x_j = w \Rightarrow \prod_{j\neq i} x_j = \frac{n}{a_i}\prod_{j=1}^n x_j \Rightarrow x_i = \frac{a_i}{n}$

since $x_i \neq 0$ when do…

Problem 4 (15pts). Consider the standard form LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \qquad \text{(LP)}$$

$Ax \leq b$

Suppose that (LP) has a *unique* optimal solution. Does this necessarily imply that the dual of (LP) also have a *unique* optimal solution? If so, give a proof. If not, construct an example (any dimension will do) to explain why.

min $3x_1 + 8x_2$

s.t. $x_1 + 2x_2 = 1$

$x_1 + 3x_2 = 1$

$x \geq 0$

$(x_1, x_2) = (1, 0)$

max $y_1 + y_2$

s.t. $y_1 + y_2 \leq 3$

$2y_2 + 3y_2 \leq 8$

$(y_1, y_2) = (1,2)$, $(2,1)$

$A = I$, $b = \vec{0}$

min $x_1 + x_2$

s.t. $x_1 = 0$

$x_2 \geq 0$

$x_1, x_2 \geq 0$

(D)

max $0y_1 + 0y_2$

s.t. $y_1 \leq 1$

$y_2 \leq 1$

$Ax = b$

$A^T y \leq c$

min $x_1 + x_2$ $c = [1, 1]$

s.t. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$x \geq 0$

max $\begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = y_1$

s.t. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

max y_1

s.t. $y_1 \leq 1$ 芀ㄣ吗

Scanned with CamScanner

ENGG 5501: Foundations of Optimization	2014–15 First Term
Final Examination	
Time Limit: 2 Hours	December 16, 2014

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (50pts). Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $\mu > 0$ be given. Suppose that the set $S = \{x \in \mathbb{R}^n : Ax = b,\ x > 0\}$ is non–empty. Consider the following problem:

$$
\begin{array}{rrl}
v_p^* = & \inf & c^T x - \mu \sum_{j=1}^n \ln(x_j) \\
& \text{subject to} & Ax = b, \\
& & x \geq 0
\end{array} \qquad (P)
$$

$A^T w$ $\Rightarrow x > 0.$ so $s_i = 0.$ $\{s_i \nabla g_i(x)\}$ $x_i \geq 0$ $x \in X$

Let $y \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$ be the Lagrangian multipliers associated with the constraints $Ax = b$ and $x \geq 0$, respectively.

KKT. & b/c linear contraints $\Rightarrow$ necessary & b/c convex opt $\Rightarrow$ sufficient.

(a) **(15pts).** Write down the first–order optimality conditions of (P) and explain why they are necessary and sufficient for optimality.

$\} \Rightarrow x_j(c - A^T y)_j = \mu$

(b) **(15pts).** Consider the following Lagrangian dual of (P):

$$v_d^* = \sup_{y \in \mathbb{R}^m} \theta(y), \qquad (D)$$

$\theta(y) = b^T y + \inf_{x>0} \{c^T x - \mu \sum \ln(x_j) - y^T Ax\}$

$= b^T y + \inf_{x>0} [\sum_{j=1}^n (c_j - (A^T y)_j) x_j - \mu \ln(\ldots$

where

$$\theta(y) = \inf_{x \geq 0} \left\{ c^T x - \mu \sum_{j=1}^n \ln(x_j) + y^T(b - Ax) \right\}.$$

$= b^T y + \sum_{j=1}^n \inf [(c_j - (A^T y)_j) x_j - \mu \ln(\ldots$ $\quad x > 0.$

Show that (D) is equivalent to the following problem:

$$
\begin{array}{rl}
\sup & b^T y + \mu \sum_{j=1}^n \ln\left(c_j - (A^T y)_j\right) \\
\text{subject to} & A^T y \leq c.
\end{array} \qquad (D')
$$

(c) **(20pts).** Suppose that (P) has an optimal solution. By considering the first–order optimality conditions of (D') and comparing them with those in (a), show that (D) also has an optimal solution, and that $v_p^* = v_d^*$.

Problem 2 (35pts). Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$ be given. Consider the following LP:

$$
\begin{array}{rl}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b, \\
& x \geq 0.
\end{array} \qquad (Q)
$$

Suppose that (Q) has an optimal solution.

~~$V_1^* = V^*$~~ ~~① $Q' \subseteq Q$~~

② $BFS(Q') \subseteq BFS(Q)$

Q' Q

~~$V_1^* \le V^*$~~ ~~$BFS(Q') \supseteq BFS(Q)$~~

(a) **(20pts).** Show that there exists an $M > 0$ such that every optimal solution to the following LP is an optimal solution to (Q):

$M = \max\{e^T x : x \in BFS(Q)\}$

(Q) has a basic opt soln

Set $M = \max\{e^T x : x$ is a

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & e^T x \le M, \\ & x \ge 0. \end{aligned}$$

$(Q) \Rightarrow$ not change feasible region.

① opt val $(Q') \ge$ opt val (Q)

② Let x^* be the vertex opt.

Then x^* is feasible

③ Every feasible soln to (Q) feasible for

Here, $e = (1, \ldots, 1)$ is the vector of all ones of appropriate dimension. (*Hint: Consider the basic feasible solutions to (Q).*)

(b) **(15pts).** Using the result in (a), or otherwise, show that for some suitably chosen matrix $\bar{A}$ and vectors $\bar{b}$ and $\bar{c}$, every optimal solution to the LP

$$\begin{aligned} \min \quad & \frac{c^T}{M} x \\ \text{s.t.} \quad & \frac{A}{M} x = \frac{b}{M} \\ & e^T x = 1 \\ & x \ge 0 \end{aligned}$$

$$\begin{aligned} \text{minimize} \quad & \bar{c}^T \bar{x} \\ \text{subject to} \quad & \bar{A}\bar{x} = \bar{b}, \\ & e^T \bar{x} = 1, \\ & \bar{x} \ge 0 \end{aligned}$$

can be converted into an optimal solution to (Q).

Problem 3 (15pts). Let $Q \in \mathcal{S}_+^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$ be given. Give an equivalent SOCP formulation of the following problem:

$$\begin{aligned} \inf \quad & t \\ \text{subject to} \quad & Ax = b, \\ & x^T Q x \le t. \end{aligned}$$

Justify your answer.

Typical approach

$x^T Q x \le t$

$\Leftrightarrow \|Q^{1/2} x\|_2^2 \le t.$

$\Leftrightarrow$ ~~$\|Q^{1/2} x\|_2 \le y$~~

$\inf y$

s.t. $Ax = b$,

$\|Q^{1/2} x\|_2 \le y$.

Alternative approach.

$x^T Q x \le t$

$\Leftrightarrow (t - \frac{1}{4})^2 + \|Q^{1/2} x\|_2^2 \le (t + \frac{1}{4})^2$

$\Leftrightarrow \left\| \begin{bmatrix} 1 & 0 \\ 0 & Q^{1/2} \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} - \begin{bmatrix} \frac{1}{4} \\ 0 \end{bmatrix} \right\| \le (t + \frac{1}{4})$

$x^T Q x \le t$

$(t - \frac{1}{4})^2 + \|Q^{\frac{1}{2}} x\|_2^2 \le (t + \frac{1}{4})^2$

$(1; 0)\begin{pmatrix} t \\ x \end{pmatrix} + \frac{1}{4}$

$\left\| \begin{pmatrix} 1 & 0 \\ 0 & Q^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} - \begin{pmatrix} \frac{1}{4} \\ 0 \end{pmatrix} \right\|_2 \le t + \frac{1}{4}$

$\left(t + \frac{1}{4}, \begin{pmatrix} 1 & 0 \\ 0 & Q^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} - \begin{pmatrix} \frac{1}{4} \\ 0 \end{pmatrix} \right) \in Q^{n+2}$

1) $x_1^2+(x_2+1)^2 \ge x_1^2+(e^{x_1}+1)^2 = e^{2x_1}+2e^{x_1}+x_1^2+1 = h(x_1)$

$h'(x_1) = 4e^{x_1}+2x_1$

$h''(x_1) = 4e^{x_1}+2>0 \Rightarrow h$ cvx $\Rightarrow 4e^{\bar{x}_1}+2\bar{x}_1=0 \Leftrightarrow e^{\bar{x}_1} = -\frac{1}{2}\bar{x}_1$ (has a soln)

$x_1^2+(x_2+1)^2 \ge h(x_1) \ge h(\bar{x}_1)$. when $(x_1,x_2)=(\bar{x}_1, e^{\bar{x}_2})$, $h(\bar{x}_1)$ is attained.

SEG 5520: Optimization I — **2010–11 First Term**

Final Examination

Time Limit: 2 Hours — December 14, 2010

(Q1 soln 2) minimized continuous function over a compact set

min $x_1^2+(x_2+1)^2$
s.t. $e^{x_1}\le x_2$ → closed
$x_1^2+(x_2+1)^2\le 4$ → closed & bound
$x_1,x_2\in\mathbb{R}$

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & x_1^2+(x_2+1)^2 \\ \text{subject to} & x_2 \ge e^{x_1}, \\ & x_1, x_2 \in \mathbb{R}. \end{array} \qquad (1)$$

(0,1) feasible

(a) **(10pts).** Show that the optimal value of Problem (1) is finite and can be attained by some $x_1^*, x_2^* \in \mathbb{R}$.

(b) **(10pts).** Write down the KKT conditions associated with Problem (1), and explain why they are necessary for optimality.

Problem 2 (25pts). Consider the following SDP problem:

$$\begin{array}{ll} \inf & x_{11} \\ \text{subject to} & X = \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0. \end{array} \qquad (2)$$

$\Leftrightarrow$ $X\succeq 0$, $X_{12}=1$ constraint

$s\in\mathbb{R}^n$

for all i

第一种 $\theta(u) = \inf_{X\succeq 0}\{L(X,u) = X_{11}+u(1-X_{12})\}$

Let $L: \mathcal{S}^2 \times \mathcal{S}_+^2 \to \mathbb{R}$, where

第二种 $$L(X,Q) = x_{11} - Q \bullet X,$$

be the Lagrangian function associated with Problem (2).

(a) **(10pts).** Write down the saddle point optimality conditions associated with Problem (2).

(b) **(15pts).** Using the result in (a), find a saddle point of L, or show that none can exist.

Problem 3 (20pts). Let $A^1, \ldots, A^m : \mathbb{R}^n \to \mathcal{S}^d$ be affine functions that take values in the space of $d \times d$ real symmetric matrices. In other words, for any given $x \in \mathbb{R}^n$, the matrix $A^i(x)$ can be expressed as

$$A^i(x) = \begin{bmatrix} A^i_{11}(x) & A^i_{12}(x) & \cdots & A^i_{1d}(x) \\ A^i_{12}(x) & A^i_{22}(x) & \cdots & A^i_{2d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ A^i_{1d}(x) & A^i_{2d}(x) & \cdots & A^i_{dd}(x) \end{bmatrix}$$

$L(X,Q) = X_{11} - Q\cdot X$

$\theta(Q) = \inf_{X_{12}=1} L(X,Q)$

$\sup_{Q\succeq 0} \theta(Q)$

for some real–valued affine functions $A^i_{jk} : \mathbb{R}^n \to \mathbb{R}$, where $1 \le i \le m$ and $1 \le j \le k \le d$. Consider the following optimization problem:

$\inf\{(1-Q_{11})X_{11} - 2Q_{12} + X_{22}Q_{22}\}$

$$\begin{array}{ll} \inf & c^T x \\ \text{subject to} & \displaystyle\sum_{i=1}^m b_i^2 (A^i(x))^2 \preceq b_0^2 I_d, \quad (\dagger) \\ & x \in \mathbb{R}^n, \end{array} \qquad (3)$$

$= \begin{cases} -2Q_{12} & \text{if } Q_{11}=1,\ Q_{22}=0 \\ -\infty & \text{else.} \end{cases}$

$b_0^2 I_d \succeq \sum_{i=1}^m b_i^2 (A^i(x))^2$

Dual: $\sup_{Q\succeq 0,\ Q_{11}=1,\ Q_{22}=0} -2Q_{12} = 0$

$b_1^2(A^1(x))^2 \le b_0^2 I$

$A^2 = A\cdot A$

$\begin{pmatrix} b_0 I & b_1 A_1 \cdots b_0 A_d \\ b_1 A_1 & \\ \vdots & b_0 I \\ b_0 A_d & \end{pmatrix} \succeq 0$

$b_0 I - b_1 A^1 (b_0 I)^{-1} b_0 A^1 \succeq 0$

where $c \in \mathbb{R}^n$, $b_0, b_1, \ldots, b_m \in \mathbb{R}$ are given, and I_d is the $d \times d$ identity matrix. Reformulate Problem (3) as an SDP problem. Justify your answer.

Hint: It suffices to show that the constraint (†) can be written as $F(x) \succeq 0$, where $F : \mathbb{R}^n \to \mathcal{S}^d$ is some affine function that takes values in $\mathcal{S}^d$.

Problem 4 (35pts). Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ be given. Consider the following primal-dual pair of LP problems:

$$(P)\quad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq 0. \end{array} \qquad (D)\quad \begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \leq c. \end{array}$$

$a_i^T y \leqslant c_i$

Suppose that both (P) and (D) have an optimal solution.

(a) **(20pts).** Consider a fixed index $j \in \{1, \ldots, n\}$. Suppose that every optimal solution $\bar{x}$ to (P) satisfies $\bar{x}_j = 0$. Show that there exists an optimal solution $\bar{y}$ to (D) such that $a_j^T \bar{y} < c_j$, where a_j is the j-th column of A.

Hint: Let v_p^* be the optimal value of (P). Consider the dual of the following LP problem:

$$\begin{array}{ll} \text{minimize} & -x_j \\ \text{subject to} & Ax = b, \\ & c^T x \leq v_p^*, \\ & x \geq 0. \end{array}$$

$-e_j^T x$

$\begin{pmatrix} A & 0 \\ c^T & I \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} = \begin{pmatrix} b \\ v_p^* \end{pmatrix}$

$(x, s) \geqslant 0$

$c^T x \geqslant v_p^*$

$\bar{x}_j = 0.$

$c^T \bar{x} = v_p^*,$

(b) **(15pts).** Using the result in (a), show that there exist optimal solutions x^* and y^* to (P) and (D), respectively, such that for each $j \in \{1, \ldots, n\}$, we either have $x_j^* > 0$ or $a_j^T y^* < c_j$.

$\begin{pmatrix} A^T & -c \\ 0 & -I \end{pmatrix}\begin{pmatrix} z \\ \theta \end{pmatrix} \leq -e_j^T$

$\exists\, (z, \theta)$ dual feasible

$b^T z - v_p \theta = 0$

① $\theta = 0$, $b^T z = 0$, y^*

② $\theta > 0$ $b^T \frac{z}{\theta} = v_p$

$A^T\left(\frac{z}{\theta}\right) \leq \frac{-e_j}{\theta} + c$ <

min $(-e_j^T, 0)\begin{pmatrix} x \\ s \end{pmatrix}$

s.t. $Ax = b$

$c^T x + s = v_p^*$

$(x, s) \geqslant 0.$

$\begin{pmatrix} A & 0 \\ c^T & 1 \end{pmatrix}\begin{pmatrix} x \\ s \end{pmatrix} = \begin{pmatrix} b \\ v_p^* \end{pmatrix}$

$(x, s) \geqslant 0$

$b^T \bar{y} + v_p^* \bar{t} = 0$

max $b^T y + v_p^* t$

$\begin{pmatrix} A^T & c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y \\ t \end{pmatrix} \leqslant \begin{pmatrix} -e_j \\ 0 \end{pmatrix}$

$\boxed{A^T y + ct \leq -e_j}$

$t \leq 0$

$a_j^T y + c_j t \leqslant -1.$

$a_j^T y \leqslant -c_j t - 1$

$< c_j.$

WANG Xiaolu 1155099202

100

ENGG 5501: Foundations of Optimization **2017–18 First Term**

Homework Set 1

Instructor: Anthony Man–Cho So Due: September 29, 2017

INSTRUCTIONS: Problem 1 is compulsory. For Problems 2 to 4, two of them will be graded. Nevertheless, you are advised to solve all the problems.

Problem 1. Let $S = \{x \in \mathbb{R}^n : x^T Ax + b^T x + c \leq 0\}$, where $A \in \mathcal{S}^n$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$ are given.

(a) Show that S is convex if $A \succeq \mathbf{0}$. Is the converse true? Explain.

(b) Let $H = \{x \in \mathbb{R}^n : g^T x + h = 0\}$, where $g \in \mathbb{R}^n \backslash \{0\}$ and $h \in \mathbb{R}$. Show that $S \cap H$ is convex if $A + \lambda gg^T \succeq \mathbf{0}$ for some $\lambda \in \mathbb{R}$.

Problem 2. Let $A \in \mathbb{R}^{m \times n}$ be given. We are interested in finding a vector $x \in \mathbb{R}^n_+$ such that $Ax = \mathbf{0}$ and the number of positive components of x is maximized. Formulate this problem as a linear program. Justify your answer.

Problem 3. Let $M \in \mathbb{R}^{m \times n}$ be a given matrix and $r \geq 1$ be a given integer. Define the function $g : \mathbb{R}^{m \times r} \times \mathbb{R}^{n \times r} \to \mathbb{R}_+$ by

$$g(U, V) = \frac{1}{2}\|UV^T - M\|_F^2 + \frac{1}{4}\|U^T U - V^T V\|_F^2.$$

Determine $\nabla g(U, V)$. Show your work.

Remark: The function g above has featured in recent non-convex formulations of the problem of finding the best rank-r factorization of a given matrix M.

Problem 4.

(a) Consider the hyperplane $H(s, c) \doteq \{x \in \mathbb{R}^n : s^T x = c\}$, where $s \in \mathbb{R}^n \setminus \{0\}$ and $c \in \mathbb{R}$ are given. Let $x \in \mathbb{R}^n$ be arbitrary. Find a formula for $\Pi_{H(s,c)}(x)$ in terms of s and c and prove its correctness.

(b) Consider the space $\mathcal{S}^n$ of $n \times n$ real symmetric matrices equipped with the inner product $\bullet$, where

$$A \bullet B = \sum_{i,j=1}^{n} A_{ij} B_{ij} \quad \text{for any } A, B \in \mathcal{S}^n.$$

Let $A \in \mathcal{S}^n$ be arbitrary and $A = U \Lambda U^T$ be its spectral decomposition. Prove that $\Pi_{\mathcal{S}^n_+}(A) = U \Lambda^+ U^T$, where Λ^+ is the $n \times n$ diagonal matrix given by

$$\Lambda^+_{ii} = \max\{\Lambda_{ii}, 0\} \quad \text{for } i = 1, \ldots, n.$$

ENGG 5501: Foundations of Optimization	2017–18 First Term

Homework Set 1 Solution

Instructor: Anthony Man–Cho So — September 29, 2017

Problem 1.

(a) Let $x_1, x_2 \in S$ and $\alpha \in (0,1)$. Then, we have

$$\begin{aligned} x_1^T A x_1 + b^T x_1 + c &\le 0, \qquad (1)\\ x_2^T A x_2 + b^T x_2 + c &\le 0. \qquad (2) \end{aligned}$$

Now, we compute

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T(\alpha x_1 + (1-\alpha)x_2) + c \\ = {} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + \alpha\left(b^T x_1 + c\right) + (1-\alpha)\left(b^T x_2 + c\right) \\ \le {} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) - \alpha x_1^T A x_1 - (1-\alpha) x_2^T A x_2 \qquad (3) \\ = {} & -\alpha(1-\alpha)x_1^T A x_1 - (1-\alpha)(1-(1-\alpha))x_2^T A x_2 + 2\alpha(1-\alpha)x_1^T A x_2 \\ = {} & -\alpha(1-\alpha)\left(x_1^T A x_1 - 2x_1^T A x_2 + x_2^T A x_2\right) \\ = {} & -\alpha(1-\alpha)(x_1 - x_2)^T A (x_1 - x_2) \\ \le {} & 0, \qquad (4) \end{aligned}$$

where (3) follows from the fact that $b^T x_i + c \le -x_i^T A x_i$ for $i = 1,2$ (by (1) and (2)), and (4) follows from the assumption that $A \succeq \mathbf{0}$. This proves that S is convex if $A \succeq \mathbf{0}$.

Note that the converse of the claim need not be true. Indeed, let $n = 1$, and let $A = -1$, $b = c = 0$. Then, we have $S = \{x \in \mathbb{R} : -x^2 \le 0\} = \mathbb{R}$, which is trivially convex.

(b) Let $x_1, x_2 \in S \cap H$ and $\alpha \in (0,1)$. ① From the calculations in part (a), we have

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T(\alpha x_1 + (1-\alpha)x_2) + c \\ \le {} & -\alpha(1-\alpha)(x_1 - x_2)^T A (x_1 - x_2). \qquad (5) \end{aligned}$$

Since $A + \gamma g g^T \succeq \mathbf{0}$, we have

$$0 \le (x_1 - x_2)^T\left(A + \gamma g g^T\right)(x_1 - x_2) = (x_1 - x_2)^T A (x_1 - x_2) + \gamma\left(g^T(x_1 - x_2)\right)^2.$$

It follows from (5) that

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T(\alpha x_1 + (1-\alpha)x_2) + c \\ \le {} & -\alpha(1-\alpha)(x_1 - x_2)^T A (x_1 - x_2) \\ \le {} & \alpha(1-\alpha)\gamma\left(g^T(x_1 - x_2)\right)^2 \\ = {} & 0, \qquad \Rightarrow \alpha x_1 + (1-\alpha) x_2 \in S \end{aligned}$$

where the last equality follows from the fact that $g^T x_1 + h = g^T x_2 + h = 0$.

② $\cancel{\alpha g^T x_1 + (1-\alpha) g^T x_2 \Rightarrow g^T 1}$

② $g^T[\alpha x_1 + (1-\alpha)x_2] + h = \alpha g^T x_1 + \alpha h + (1-\alpha) g^T x_2 + (1-\alpha) h$

$= \alpha(g^T x_1 + h) + (1-\alpha)(g^T x_1 + h) = 0 \Rightarrow \alpha x_1 + (1-\alpha) x_2 \in H$

Problem 2. Consider the following system:

$$Ax = 0, \quad x \geq 0. \tag{6}$$

Suppose that $\bar{x} \in \mathbb{R}^n$ is feasible for (6). Then, $\alpha\bar{x}$ is also feasible for (6) for any $\alpha > 0$, and $\bar{x}$ and $\alpha\bar{x}$ have the same number of positive components. Thus, as far as the number of positive components is concerned, we may assume that all the non-zero entries of $\bar{x}$ have magnitude at least 1. Consequently, we can decompose $\bar{x}$ as $\bar{x} = \bar{y}+\bar{z}$, where $\bar{y}_i = \max\{\bar{x}_i - 1, 0\}$ and $\bar{z}_i = \bar{x}_i - \bar{y}_i \in \{0,1\}$ for $i = 1,\dots,n$. Note that $e^T\bar{z}$ is precisely the number of positive components in $\bar{x}$. The above observation motivates us to consider the following LP:

$$\begin{array}{lll} \text{maximize} & e^T z & \\ \text{subject to} & A(y+z) = 0, & \\ & z_i \leq 1 & \text{for } i = 1,\dots,n, \\ & y, z \geq 0. & \end{array} \tag{7}$$

It is not hard to verify that if (y^*, z^*) is an optimal solution to (7), then we must have $z^* \in \{0,1\}^n$. Moreover, we have $z_i^* = 0$ if and only if every solution $\bar{x}$ to (6) satisfies $\bar{x}_i = 0$. Thus, the above LP indeed finds a solution to (6) that has the largest number of positive components.

Problem 3. Observe that

$$\begin{aligned} \|UV^T - M\|_F^2 &= \operatorname{tr}\left((UV^T - M)^T(UV^T - M)\right) \\ &= \operatorname{tr}\left(VU^TUV^T - VU^TM - M^TUV^T + M^TM\right), \\ \|U^TU - V^TV\|_F^2 &= \operatorname{tr}\left(U^TUU^TU - U^TUV^TV - V^TVU^TU + V^TVV^TV\right). \end{aligned}$$

Following the developments in [1, Chapter 8], we compute the differentials of each term as follows:

$$\begin{aligned} d_U\|UV^T - M\|_F^2 &= \operatorname{tr}\left(Vd_U(U^TU)V^T - V(d_UU)^TM - M^T(d_UU)V^T\right) \\ &= \operatorname{tr}\left(V(d_UU)^TUV^T + VU^T(dU)V^T\right) - 2\operatorname{tr}\left(V^TM^Td_UU\right) \\ &= 2\operatorname{tr}\left(\left(V^TVU^T - V^TM^T\right)d_UU\right), \\ d_U\|U^TU - V^TV\|_F^2 &= 4\operatorname{tr}\left(\left(U^TUU^T - V^TVU^T\right)d_UU\right), \\ d_V\|UV^T - M\|_F^2 &= 2\operatorname{tr}\left(\left(U^TUV^T - U^TM\right)d_VV\right), \\ d_V\|U^TU - V^TV\|_F^2 &= 4\operatorname{tr}\left(\left(V^TVV^T - U^TUV^T\right)d_VV\right). \end{aligned}$$

It follows that

$$\begin{aligned} \nabla_U g(U,V) &= \left(V^TVU^T - V^TM^T\right)^T + \left(U^TUU^T - V^TVU^T\right)^T \\ &= (UV^T - M)V + U(U^TU - V^TV), \\ \nabla_V g(U,V) &= \left(U^TUV^T - U^TM\right)^T + \left(V^TVV^T - U^TUV^T\right)^T \\ &= (UV^T - M)^TU - V(U^TU - V^TV). \end{aligned}$$

Problem 4.

(a) Intuitively, the vector $x - \Pi_{H(s,c)}(x)$ should be normal to the hyperplane $H(s,c)$. Hence, we should have $x - \Pi_{H(s,c)}(x) = \alpha s$ for some $\alpha \in \mathbb{R}$. Since $\Pi_{H(s,c)}(x) \in H(s,c)$, this requires that $s^T(x - \alpha s) = c$, which implies that $\alpha = (s^Tx - c)/s^Ts$. This yields the following candidate for $\Pi_{H(s,c)}(x)$:

$$\Pi_{H(s,c)}(x) = x - \frac{s^Tx - c}{s^Ts}s. \tag{8}$$

To prove the correctness of the above formula, we use Theorem 5 of Handout 2. Let $y \in H(s,c)$ be arbitrary. Since $s^Ty = c$, we obtain

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

This establishes the correctness of the formula in (8).

(b) **(15pts).** Let $A \in \mathcal{S}^n$ be arbitrary and $A = U\Lambda U^T$ be its spectral decomposition. Observe that

$$(\Lambda - \Lambda^+)_{ii} = \min\{\Lambda_{ii}, 0\} \quad \text{for } i = 1, \ldots, n.$$

Hence, for any $Q \in \mathcal{S}^n_+$, we have

$$\begin{aligned}
(Q - \Pi_{\mathcal{S}^n_+}(A)) \bullet (A - \Pi_{\mathcal{S}^n_+}(A)) &= (Q - \Pi_{\mathcal{S}^n_+}(A)) \bullet U(\Lambda - \Lambda^+)U^T \\
&= (U^TQU - \Lambda^+) \bullet (\Lambda - \Lambda^+) \\
&= \sum_{i=1}^n \left[(U^TQU)_{ii} \cdot \min\{\Lambda_{ii}, 0\} - \Lambda^+_{ii} \cdot \min\{\Lambda_{ii}, 0\}\right] \\
&= \sum_{i:\Lambda_{ii} \leq 0} (U^TQU)_{ii} \cdot \Lambda_{ii} \\
&\leq 0,
\end{aligned}$$

where the last inequality follows from the fact that $U^TQU \in \mathcal{S}^n_+$ and the diagonal entries of a psd matrix are non-negative. This, together with Theorem 5 of Handout 2, completes the proof.

References

[1] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Chichester, England, revised edition, 1999.

ENGG 5501 Homework Set 1 (WANG Xiaolu 1155099202)

First we prove that: A set Υ is convex $\Leftrightarrow$ $\Upsilon \cap L$ is convex for all line $L = \{x_0 + tv \mid t \in R\}$

($\Rightarrow$) Υ is convex, obviously L is convex, thus $\Upsilon \cap L$ is convex

($\Leftarrow$) $\forall x_1, x_2 \in \Upsilon$, there exist a line L passing through the two points. $\Upsilon \cap L$ is convex, ~~thus~~ and $x_1, x_2 \in \Upsilon \cap L$. Thus $\mu x_1 + (1-\mu) x_2 \in \Upsilon \cap L \subset \Upsilon$ ($\forall \mu \in [0,1]$) $\Rightarrow$ Υ is convex.

(a) For any $L = \{x_0 + tv \mid t \in R\}$,

$$L \cap S = \{x_0 + tv \mid (v^T A v)t^2 + (2x_0^T A v + b^T v)t + (x_0^T A x_0 + b^T x_0 + c) \le 0\}.$$

Since $A \succeq 0$, $v^T A v \ge 0$ for all v.

Obviously $f(t) = (v^T A v)t^2 + (2x_0^T A v + b^T v)t + (x_0^T A x_0 + b^T x_0 + c)$ is a quadratic function of t.

$\{t \mid f(t) \le 0\}$ is a convex ~~function~~ set belonging to R

Hence $\{x_0 + tv \mid f(t) \le 0\}$ is a convex set of R^n.

Conversely, if $S = \{x \in R : -x^2 \le 0\}$, S is surely convex

However, $A = -1 < 0$ in this case.

(b) Choose $x_0 \in H$, thus $g^T x_0 + h = 0$.

Let $L = \{x_0 + tv \mid t \in R\}$.

$L \cap H = \{x_0 + tv \mid g^T(x_0 + tv) + h = 0\} = \{x_0 + tv \mid g^T v t = 0\}$

Hence $L \cap S \cap H = \{x_0 + tv \mid f(t) \le 0, (g^T v)t = 0\}$.

Since $g^T \ne 0$, $g^T v$ can be non-zero.

① If $g^T v \neq 0$, then $t=0 \Rightarrow f(t) = x_0^T A x_0 + b^T x_0 + c$ ..

Hence $L \cap S \cap H = \{x_0 \mid x_0^T A x_0 + b^T x_0 + c \leq 0\} = \{x_0\}$ or $\emptyset$, either set i

② If $g^T v = 0$, $L \cap S \cap H = \{x_0 + tv \mid f(t) \leq 0\}$.

(Hence $L \cap S \cap H$ can be convex if $A \succeq 0$)

If ~~...~~ $A + \lambda g^T g \succeq 0$ for some λ, we have

$$v^T A v = v^T A v + \lambda \|g^T v\|^2 = v^T (A + \lambda g^T g) v \geq 0,$$

thus $A \succeq 0 \Rightarrow L \cap S \cap H$ is convex.

Since L can be arbitrary, $S \cap H$ is convex

2. The primal Problem can be formulated as

$$(P) \begin{cases} \max \|x\|_0 \\ \text{s.t. } Ax = 0 \\ \quad x \geq 0 \end{cases}$$

Consider the $\underbrace{\text{transformed}}_{\text{Linear programming}}$ problem (T)

$$\begin{cases} \max \sum_{i=1}^{n} y_i \\ \text{s.t. } A(z+y) = \\ \quad y_i \leq 1 \\ \quad y, z \geq 0 \end{cases}$$

Now prove the equivalence of (P) and (T).

① (P) ⇒ (T). Suppose x^* is the optimal solution of (P). For all $\alpha > 0$,

still the solution of (P). Choose $\alpha > 0$ s.t. $\alpha x_i^* \geq 1$ for $x_i^* > 0$

for $x_i^* = 0$.

Then, αx^* can be decomposed as $\alpha x^* = y^* + z^*$, where $y_i^* = 1$ for a

and $y_i^* = 0$ for $\alpha x_i^* = 0$. Obviously, $\sum_{i=1}^{n} y_i^* = \|\alpha x^*\|_0 = \|y^* + z^*\|_0$.

Denote $C_P = \{(y, z) \mid A(y+z) = 0, y+z \geq 0\}$

$C_T = \{(y, z) \mid A(y+z) = 0, y_i \leq 1, y \geq 0, z \geq 0\}$.

Of course we have $C_T \subset C_P$.

Since αx^* is the optimal solution of (P), and $(y^*, z^*) \in C_P$ ~~is feasible to (P)~~ we have

$(y^*, z^*) = \arg\max_{(y,z) \in C_P} \|y+z\|_0$. Fortunately $(y^*, z^*) \in C_T$

Hence $(y^*, z^*) = \arg\max_{(y,z) \in C_T} \|y+z\|_0 = \arg\max_{(y,z) \in C_T} \sum_{i=1}^{n} y_i$. i.e. (y^*, z^*) is t

(T) $\Rightarrow$ (P). Suppose (y^*, z^*) is the optimal solution of (T).

Let $x^* = y^* + z^* \Rightarrow Ax^* = A(y^*+z^*) = 0$, $x^* = y^* + z^* \geq 0 \Rightarrow x^*$ is feasible to (P).

Assume x^* is not the optimal solution of (P), and the optimal solution of (P) is $\hat{x}^* \neq x^*$.

Scale $\hat{x}^*$ with $a > 0$ s.t. all the non-zero entry of $a\hat{x}^*$ is at least

Then decompose $c\hat{x}^*$ as before: $c\hat{x}^* = \hat{y} + \hat{z}$, thus $\hat{y}$ is the optimal solution of (T) according to ①.

Hence, $\sum_{i=1}^{n} \hat{y}_i = \|a\hat{x}^*\|_0 > \|x^*\|_0 = \sum_{i=1}^{n} y_i^* \Rightarrow y^*$ is not optimal to (T), which is a contradiction. Therefore, x^* is the optimal solution of (P)

Let $g_1(U,V) = \frac{1}{2}\|UV^T - M\|_F^2$, $g_2(U,V) = \frac{1}{4}\|U^TU - V^TV\|_F^2$

$$g_1(U,V) = \frac{1}{2}\operatorname{tr}(UV^T - M)^T(UV^T - M) = \frac{1}{2}\operatorname{tr}(VU^TUV^T - VU^TM - M^TUV^T + M^TM)$$

$$\frac{\partial g_1(U,V)}{\partial U} = \frac{1}{2}U(V^TV + V^TV) - \frac{1}{2}(MV + MV) = UV^TV - MV$$

$$\frac{\partial g_1(U,V)}{\partial V} = \frac{1}{2}(VU^TU + VU^TU) - M^TU = VU^TU - M^TU$$

$$g_2(U,V) = \frac{1}{4}\operatorname{tr}(U^TUU^TU - U^TUV^TV - V^TVU^TU + V^TVV^TV)$$
$$= \frac{1}{4}\operatorname{tr}(U^TUU^TU) - \frac{1}{2}\operatorname{tr}(U^TUV^TV) + \frac{1}{4}\operatorname{tr}(V^TVV^TV)$$

$$\frac{\partial g_2(U,V)}{\partial U} = \frac{1}{4}\cdot 4\,UU^TU - \frac{1}{2}(2V^TVU^T)^T = UU^TU - UV^TV$$

$$\frac{\partial g_2(U,V)}{\partial V} = \frac{1}{4}\,4\,VV^TV - \frac{1}{2}(2U^TUV^T)^T = VV^TV - VU^TU$$

$$\text{Hence, } \nabla g(U,V) = \begin{pmatrix} \frac{\partial g_1}{\partial U} + \frac{\partial g_2}{\partial U} \\ \frac{\partial g_1}{\partial V} + \frac{\partial g_2}{\partial V} \end{pmatrix} = \begin{pmatrix} UU^TU - MV \\ VV^TV - M^TU \end{pmatrix}$$

4. (a) Suppose x_0 is the projection of x on $H(s,c)$.

We have $x_0 + \lambda s = x \quad (\lambda \in \mathbb{R})$.

$\therefore$ $s^T x_0 = c$

Hence $x_0 = x - \lambda s \Rightarrow s^T(x - \lambda s) = c \Rightarrow \lambda = \frac{s^T x - c}{\|s\|^2} \Rightarrow x_0 = x - \lambda s = x - \frac{s^T x - c}{\|s\|^2} s$

i.e. $\Pi_{H(s,c)}(x) = x - \frac{s^T x - c}{\|s\|^2} s$

(b) Find X s.t. $\min\limits_{X \in S_+^n} \|A - X\|_F^2$.

$\|A - X\|_F^2 = \|U\Lambda U^T - X\|_F^2 = \|U\Lambda U^T\|_F^2 + \|X\|_F^2 - 2\operatorname{tr}(U\Lambda U^T X)$.

since $\|U\Lambda U^T\|_F^2 = \operatorname{tr}(U\Lambda U^T)(U\Lambda U^T) = \operatorname{tr}(U\Lambda^2 U^T) = \operatorname{tr}(U^T U \Lambda^2) = \operatorname{tr}\Lambda^2 = \|\Lambda\|_F^2$

similary $\|X\|_F^2 = \|U^T X U\|_F^2$, $\operatorname{tr}(U\Lambda U^T X) = \operatorname{tr}(U\Lambda U^T X) = \operatorname{tr}(\Lambda U^T X U)$

We have $\|U\Lambda U^T - X\|_F^2 = \|\Lambda\|_F^2 + \|U^T X U\|_F^2 - 2\operatorname{tr}\Lambda U^T X U = \|\Lambda - U^T X U\|_F^2$.

Let $Y = U^T X U$. For $\forall v \in \mathbb{R}^n$, $v^T Y v = v^T U^T X U v = (Uv)^T X (Uv) \geq 0$

(Because $X \succeq 0$) we know $Y \succeq 0$.

The object turns to be finding Y s.t. $\min\limits_{Y \in S_+^n} \|\Lambda - Y\|_F^2$

$\|\Lambda - Y\|_F^2 = \sum\limits_{i \neq j} y_{ij}^2 + \sum\limits_{i=1}^n (\Lambda_{ii} - y_{ii})^2$

For $i \neq j$, let $y_{ij} = 0$ s.t. $\sum\limits_{i \neq j} y_{ij}^2 = 0$

For $i = j$, if $\Lambda_{ii} \geq 0$, let $y_{ii} = \Lambda_{ii}$ s.t. $(\Lambda_{ii} - y_{ii})^2 = 0$

if $\Lambda_{ii} < 0$, since $y_{ii} \geq 0$ because $Y \succeq 0$

let $y_{ii} = 0$ s.t. $(\Lambda_{ii} - y_{ii})^2 = \Lambda_{ii}^2$

Hence $\min \|\Lambda - Y\|_F^2 = \sum\limits_{\Lambda_{ii} < 0}^{n} \Lambda_{ii}^2$,

and the minimizer $Y^* = \operatorname{diag}(\Lambda_{11}^+, \Lambda_{22}^+, \cdots, \Lambda_{nn}^+) = \Lambda^+$

Then $U X^* U^T = Y^* \Rightarrow X^* = U Y^* U^T = U \Lambda^+ U^T$

$\because$ Obviously $U \Lambda^+ U^T \in S_+^n$

Hence, $\Pi_{S_+^n}(A) = U\Lambda^+ U^T$.

WANG Xiaolu 1155099202

60

ENGG 5501: Foundations of Optimization	2017–18 First Term
Homework Set 2	
Instructor: Anthony Man–Cho So	Due: October 20, 2017

INSTRUCTIONS: Problem 1 is compulsory. For Problems 2 to 4, two of them will be graded. Nevertheless, you are advised to solve all the problems.

Problem 1. Let S be a non-empty convex subset of $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ be a convex function. Prove that if f attains its maximum at a point $\bar{x} \in \operatorname{int} S$, then f must be a constant function.

Problem 2. Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ be an arbitrary norm on $\mathbb{R}^n$. The *dual norm* of $\|\cdot\|$, which is denoted by $\|\cdot\|_*$, is defined as

$$\|x\|_* = \sup_{\|d\|=1} d^T x.$$

Show that

$$\partial\|x\| = \left\{ s \in \mathbb{R}^n : \|s\|_* \leq 1,\ s^T x = \|x\| \right\}.$$

Problem 3. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function such that $\mathrm{epi}(f)$ is closed and f is not identically $+\infty$.

(a) Show that $f = f^{**}$, where $f^{**} = (f^*)^*$ is the conjugate of f^*.

(b) Using the result in (a), show that for any $x, y \in \mathbb{R}^n$, the following statements are equivalent:

(i) $y \in \partial f(x)$.

(ii) $f(x) + f^*(y) = x^T y$

(iii) $x \in \partial f^*(y)$

REMARK: The subdifferential of f, ∂f, is a *set-valued mapping* in the sense that it assigns a set $\partial f(x) \subset \mathbb{R}^n$ to each $x \in \mathbb{R}^n$. The *inverse mapping* of ∂f, denoted by $(\partial f)^{-1}$, is simply defined as

$$(\partial f)^{-1}(y) = \{x \in \mathbb{R}^n : y \in \partial f(x)\}.$$

With these notations, the equivalence of (i) and (iii) above can be expressed as $(\partial f)^{-1} = \partial f^*$, which is another important relationship between f and f^*.

Problem 4.

(a) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and $n_1, n_2 \geq 1$ be integers such that $n = n_1 + n_2$. Consider the function $g : \mathbb{R}^{n_1} \to \mathbb{R} \cup \{+\infty\}$ defined by

$$g(x) = \inf_{y \in \mathbb{R}^{n_2}} f(x, y). \tag{1}$$

(Here, we implicitly assume that $g(x) > -\infty$ for all $x \in \mathbb{R}^{n_1}$.) Show that g is convex.

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $n_1, n_2 \geq 1$ be integers such that $n = n_1 + n_2$. Suppose that for any $y \in \mathbb{R}^{n_2}$, the function $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ defined by $f_1(x) = f(x, y)$ is convex, and that for any $x \in \mathbb{R}^{n_1}$, the function $f_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ defined by $f_2(y) = f(x, y)$ is convex. Is the function g defined in (1) convex (again, we assume that $g(x) > -\infty$ for all $x \in \mathbb{R}^{n_1}$)? Justify your answer.

ENGG 5501: Foundations of Optimization	2017–18 First Term
Homework Set 2 Solution	
Instructor: Anthony Man–Cho So	October 20, 2017

Problem 1. Let $x \in S$ be arbitrary. We show that $f(x) = f(\bar{x})$. Since $\bar{x} \in \text{int}(S)$, there exists an $\epsilon > 0$ such that $B(\bar{x}, \epsilon) \subset S$. Now, consider the line $\mathscr{L}$ that passes through x and $\bar{x}$. It is not hard to see that there exist $x' \in \mathscr{L} \cap B(\bar{x}, \epsilon)$ and $\alpha \in (0,1)$ such that $\bar{x} = \alpha x + (1-\alpha)x'$. Note that $x' \in S$, since $B(\bar{x}, \epsilon) \subset S$. Moreover, we have

$$f(\bar{x}) = f(\alpha x + (1-\alpha)x') \le \alpha f(x) + (1-\alpha) f(x') \le \alpha f(\bar{x}) + (1-\alpha) f(\bar{x}) = f(\bar{x}), \tag{1}$$

where the second inequality follows from the fact that f attains its maximum at $\bar{x}$. In particular, all the inequalities in (1) hold as equality, from which we conclude that $f(x) = f(\bar{x})$.

Problem 2. (a) Consider a fixed $x \in \mathbb{R}^n$. Let $s \in \mathbb{R}^n$ be such that $\|s\|_* \le 1$ and $s^T x = \|x\|$. By definition of the dual norm, for any $\bar{x} \in \mathbb{R}^n \backslash \{0\}$, we have

$$1 \ge \|s\|_* = \sup_{d \ne 0} \frac{d^T s}{\|d\|} \ge \frac{s^T \bar{x}}{\|\bar{x}\|}.$$

It follows that

$$\|\bar{x}\| \ge s^T \bar{x} = s^T x + s^T(\bar{x} - x) = \|x\| + s^T(\bar{x} - x).$$

Note that the above inequality is also valid at $\bar{x} = 0$, because $s^T x = \|x\|$ by assumption. Hence, we conclude that $s \in \partial\|x\|$.

(b) Conversely, suppose that $s \in \partial\|x\|$. Consider first the case where (1) $x \ne 0$. We have

$$\begin{aligned} y = 2x \Rightarrow \quad 2\|x\| &= \|x + x\| \ge \|x\| + s^T x, \quad s \in \partial\|x\| = \{s : \|y\| \ge \|x\| + s^T(y - x) \ \forall \\ y = 0 \Rightarrow \quad 0 &= \|x - x\| \ge \|x\| - s^T x, \end{aligned}$$

which together imply that $s^T x = \|x\|$. Since $x \ne 0$, it follows that

$$\|s\|_* \le 1 \qquad \|s\|_* = \sup_{\|d\|=1} d^T s \ge \frac{s^T x}{\|x\|} = 1.$$

We claim that $\|s\|_* = 1$. Suppose that this is not the case. Then, we have $\|s\|_* > 1$, which implies that $s^T d > 1$ for some $d \in \mathbb{R}^n$ with $\|d\| = 1$. We compute

$$\|x\| + 1 = \|x\| + \|d\| \ge \|x + d\| \ge \|x\| + s^T d > \|x\| + 1,$$

take $y = x + d$

which is a contradiction. This establishes the claim.

Now, consider the case where (2) $x = 0$. The condition $s^T x = \|x\|$ is automatically satisfied. On the other hand, let $f : \mathbb{R}^n \to \mathbb{R}_+$ be given by $f(x) = \|x\|$. The directional derivative of f at 0 in the direction $d \in \mathbb{R}^n \backslash \{0\}$ is, by definition,

$$f'(0, d) = \lim_{t \searrow 0} \frac{\|td\|}{t} = \|d\|.$$

$s \in \partial\|d\| \Leftrightarrow \|y\| \ge s^T y \quad \forall y$

$\Leftrightarrow \frac{s^T y}{\|y\|} \le 1 \quad \forall y \ne 0 \Rightarrow \|s\|_* = \sup_{y \ne 0} \frac{s^T y}{\|y\|} \le 1.$

$\cancel{\Rightarrow \|s\|_* = \sup \ldots s^T y}$

Hence, by Theorem 16(a) of Handout 2, we have $f'(0,d) = \|d\| \geq s^T d$ for all $s \in \partial\|0\|$. T
together with the definition of the dual norm, yields

$$\|s\|_* = \sup_{d \neq 0} \frac{s^T d}{\|d\|} \leq 1,$$

as desired.

Problem 3.

(a) By Theorem 12 of Handout 2 (which requires the closedness of epi(f)), we have

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

where $S_f = \left\{ (y,c) \in \mathbb{R}^n \times \mathbb{R} : y^T x - c \leq f(x) \text{ for all } x \in \mathbb{R}^n \right\}$. Moreover, the discussion low Theorem 12 of Handout 2 shows that $S_f = \text{epi}(f^*)$. Hence, we have $(y,c) \in S_f$ $f^*(y) \leq c$. This, together with (2), implies that

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

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

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

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

In particular, we have $y^T x - f(x) \geq \sup_{z \in \mathbb{R}^n} \left\{ y^T z - f(z) \right\} = f^*(y)$. On the other hand, have $f(x) \geq y^T x - f^*(y)$ from (3). Hence, we obtain $f(x) + f^*(y) = x^T y$; i.e., (ii) holds. B 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 $f^{**}(x) \geq z^T x - f^*(z)$ for all $z \in \mathbb{R}^n$, we obtain $y^T x \geq f^*(y) + z^T x - f^*($ for all $z \in \mathbb{R}^n$, or equivalently,

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

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

Problem 4.

(a) Let $x_1, x_2 \in \mathbb{R}^{n_1}$ and $\alpha \in (0,1)$ be arbitrary. Then, we have

$$\begin{aligned} g(\alpha x_1 + (1-\alpha)x_2) &= \inf_{y \in \mathbb{R}^{n_2}} f(\alpha x_1 + (1-\alpha)x_2, y) \\ &= \inf_{y_1, y_2 \in \mathbb{R}^{n_2}} f(\alpha(x_1, y_1) + (1-\alpha)(x_2, y_2)) \\ &\leq \alpha \inf_{y_1 \in \mathbb{R}^{n_2}} f(x_1, y_1) + (1-\alpha) \inf_{y_2 \in \mathbb{R}^{n_2}} f(x_2, y_2) \\ &= \alpha g(x_1) + (1-\alpha) g(x_2), \end{aligned}$$

as desired.

(b) Consider the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $f(x, y) = x^2 + y^2 - 4xy$. Then, for any given $\bar{y} \in \mathbb{R}$, the function $f_1 : \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = f(x, \bar{y}) = x^2 + \bar{y}^2 - 4x\bar{y}$ is convex in x. Similarly, for any given $\bar{x} \in \mathbb{R}$, the function $f_2 : \mathbb{R} \to \mathbb{R}$ given by $f_2(y) = f(\bar{x}, y) = \bar{x}^2 + y^2 - 4\bar{x}y$ is convex in y. However, the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) = \inf_{y \in \mathbb{R}} f(x, y) = \inf_{y \in \mathbb{R}} \{x^2 + y^2 - 4xy\} = -3x^2$$

is not convex.

$f^{**}(x) + f^{*}(y) = x^T y$

$f^{**}(x) \geq z^T x - f^{*}(z) \quad \forall z \in \mathbb{R}^n.$

$y^T x - f^{*}(y) \geq z^T x - f(z)$

ENGG 5501 Homework Set 2 (WANG Xiaolu, 1155099202)

Problem 1.

Suppose f is not a constant function. Since f attains its maximum at $\bar{x}$, there must exists $x_1 \in S$, s.t. $f(x_1) < f(\bar{x})$.

Since $\bar{x} \in \text{int}\, S$, there exists $\varepsilon > 0$, s.t. $B(\bar{x}, \varepsilon) \subseteq S$.

i.e. $\forall x \in S$, $\|x - \bar{x}\|_2 < \varepsilon$ implies $x \in S$.

Let $x_2 = \frac{1}{1-\lambda}(\bar{x} - \lambda x_1)$, where $1 > \lambda > \frac{\|x_1 - \bar{x}\|_2}{\|x_1 - \bar{x}\|_2 + \varepsilon}$. Then,

$$\|x_2 - \bar{x}\|_2 = \left\|\frac{1}{1-\lambda}(\bar{x} - \lambda x_1) - \bar{x}\right\|_2 = \frac{\lambda}{1-\lambda}\|\bar{x} - x_1\|_2 < \varepsilon.$$

Hence, $x_2 \in S$ and $\bar{x} = \lambda x_1 + (1-\lambda) x_2$

Since f is a convex function, we have

$$f(\bar{x}) = f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$
$$< \lambda f(\bar{x}) + (1-\lambda) f(\bar{x}) = f(\bar{x})$$

which is an obvious contradiction.

Therefore, f is a constant function.

15

25

Problem 2.

By definition, $\partial\|x\| = \{ s : \|x\| + s^T(y-x) \le \|y\| \text{ for all } y \}$

Denote $A = \{ s : \|s\|_* \le 1,\ s^T x = \|x\|. \}$

① $\forall s \in A$, we have $\|s\|_* \le 1$, $s^T x = \|x\|$.

Hence, $\|x\| + s^T(y-x) = \|x\| + s^T y - s^T x = s^T y = \langle s, y \rangle$

$\le \|s\|_* \|y\| \le \|y\|$, for all y it holds.

The first inequality is from Hölder's Inequality for a pair of dual norms, and the second inequality is because $\|s\|_* \le 1$

Now, we have $s \in \partial\|x\|$ for $\forall s \in A$

Hence, $A \subseteq \partial\|x\|$

② $\forall s \in \partial\|x\|$, we have for all $y \in \mathbb{R}^n$, $\|y\| \geq \|x\| + s^T(y-x)$

i.e. for $\forall y$, $s^T y - \|y\| \leq s^T x - \|x\|$

which is equivalent to $\sup_y \{s^T y - \|y\|\} \leq s^T x - \|x\|$ (*)

Let $f(y) = s^T y - \|y\| = \langle s, y\rangle - \|y\|$

(i) If $\|s\|_* \leq 1$, $\langle s, y\rangle \leq \|s\|_* \|y\| \leq \|y\|$.

Then, $f(y) = \langle s, y\rangle - \|y\| \leq 0$ for $\forall y$ (#)

Take $y = 0$, we get $f(0) = \langle s, 0\rangle - \|0\| = 0$

Hence we have $\sup_y \{f(y)\} = 0$

(ii) If $\|s\|_* > 1$, then $\|s\|_* = \sup_{\|d\|=1} \langle s, d\rangle > 1$

Thus, there must exist d_0 s.t. $\langle s, d_0\rangle > 1$ and $\|d_0\|$

Take $y = t d_0$ $(t > 0)$, then

$$f(y) = \langle s, t d_0\rangle - \|t d_0\| = t(\langle s, d_0\rangle - \|d_0\|)$$

$$= t(\langle s, d_0\rangle - 1) > t$$

Since t can be arbitrarily large, $\sup_y \{f(y)\} = +\infty$

Since (*) holds, and $s^T x - \|x\|$ is finite for fixed x and

we can know that $\|s\|_* \leq 1$. Thus (#) holds.

Take $y = x$, we get $0 \leq \langle s, x\rangle - \|x\| \leq \|s\|_* \|x\| - \|x\| \leq 0$

Hence, $s^T x = \langle s, x\rangle = \|x\|$.

Now, $s \in A$ for $\forall s \in \partial\|x\|$ is shown.

Then $A \subseteq \partial\|x\|$

Combining ① ②, we get $A = \partial\|x\|$

i.e. $\partial\|x\| = \{s \in \mathbb{R}^n : \|s\|_* \leq 1,\ s^T x = \|x\|\}$

blem 3.

) $f^*(x^*) = \sup_x \{ \langle x, x^* \rangle - f(x) \}$ ①

Thus $f^{**}(x) = \sup_{x^*} \{ \langle x, x^* \rangle - f^*(x^*) \} \triangleq \sup_{x^*} B$

By Theorem 12, we know $f(x) = \sup_{x^*, \mu^*} \{ \langle x, x^* \rangle - \mu^* \mid (\forall x)\ \langle x, x^* \rangle - \mu^* \le f(x) \} \triangleq \sup_{x^*, \mu^*} A$

For each $\alpha = \langle x, x^* \rangle - \mu^* \in A$, we have

$\langle x, x^* \rangle - \mu^* \le f(x)$ holds for $\forall x$.

$\Leftrightarrow$ $\langle x, x^* \rangle - f(x) \le \mu^*$ $\forall x$

$\Leftrightarrow$ $\sup_x \{ \langle x, x^* \rangle - f(x) \} \le \mu^*$

$\Leftrightarrow$ $f^*(x^*) \le \mu^*$

$\Leftrightarrow$ $\langle x, x^* \rangle - \mu^* = \alpha \le \langle x, x^* \rangle - f^*(x^*) \triangleq \beta \in B$

Hence, $\forall \alpha \in A$, $\exists \beta \in B$, s.t. $\alpha \le \beta$. ②

For each $\beta \in B$, $\beta = \langle x, x^* \rangle - f^*(x^*)$

Form ①, we know $\langle x, x^* \rangle - f(x) \le f^*(x^*)$ $(\forall x)$

$\Leftrightarrow$ $\langle x, x^* \rangle - f^*(x^*) \le f(x)$ $(\forall x)$

$\Rightarrow \beta \in A$

Thus, $B \subseteq A$ ③

Combining ② ③, we get $\sup_{x^*, \mu^*} A = \sup_{x^*} B$ i.e. $f^{**}(x) = f(x)$

(From ③ we know that $\sup B \le \sup A$. Suppose $\sup B \ne \sup A$,

then $\sup B < \sup A$. Thus $\exists\ \varepsilon > 0$, s.t. $\sup B < (\sup A) - \varepsilon \in A$. due

ity of
ction. i.e. $\exists\ (\sup A) - \varepsilon \notin B$, $\forall \beta \in B$, $(\sup A) - \varepsilon > \beta$, which contradicts ②

(b) $y \in \partial f(x)$ $\Leftrightarrow$ $f(z) \ge y^T(z - x) + f(x)$ $(\forall z)$

$\Leftrightarrow$ $y^T z - f(z) \le y^T x - f(x)$ $(\forall z)$

$\Leftrightarrow$ $\sup_z \{ y^T z - f(z) \} \le y^T x - f(x)$

$\Leftrightarrow$ $f^*(y) \le y^T x - f(x)$

$\Leftrightarrow$ $f(x) + f^*(y) = x^T y$

i.e. (i) $\Leftrightarrow$ (ii)

Similarly, we can get $x \in \partial f^*(y)$

$\Leftrightarrow f^*(y) + f^{**}(x) = x^T y$

$\Leftrightarrow f^*(y) + f(x) = x^T y$ (since $f^{**} = f$)

i.e. (iii) $\Leftrightarrow$ (ii)

Hence (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii)

Problem 4.

(a) $g(x) = \inf_{y \in R^{n_2}} f(x, y) \Rightarrow g(x_1) = \inf_{y \in R^{n_2}} f(x_1, y)$, $g(x_2) = \inf_{y \in R^{n_2}} f(x_2, y)$

$\forall \varepsilon > 0, \exists y_1 \in R^{n_2}$ s.t. $f(x_1, y_1) \le g(x_1) + \varepsilon$

$\exists y_2 \in R^{n_2}$ s.t. $f(x_2, y_2) \le g(x_2) + \varepsilon$

Hence, for all $0 \le \alpha \le 1$, we have

$g(\alpha x_1 + (1-\alpha) x_2) = \inf_y f(\alpha x_1 + (1-\alpha) x_2, y) \le f(\alpha x_1 + (1-\alpha) x_2, \alpha y_1 +$

$= f(\alpha(x_1, y_1) + (1-\alpha)(x_2, y_2)) \le \alpha f(x_1, y_1) + (1-\alpha) f(x_2, y_2)$

$\le \alpha(g(x_1) + \varepsilon) + (1-\alpha)(g(x_2) + \varepsilon) = \alpha g(x_1) + (1-\alpha) g(x_2) +$

$\therefore g(\alpha x_1 + (1-\alpha) x_2) \le \inf_\varepsilon \{\alpha g(x_1) + (1-\alpha) g(x_2) + \varepsilon\} = \alpha g(x_1) + (1-\alpha) g$

i.e. g is a convex function.

(b) Let $f(x, y) = xy + y^2$. $x, y \in R$

$g(x) = \inf_y f(x, y) = \inf_y \{xy + y^2\} = -\frac{x^2}{4}$.

Hence, $g(x)$ is a nonconvex function.

However $f_1(x) = yx + y^2$ is linear, which is convex.

$f_2(y) = y^2 + xy$ is also convex.

Hence, g may not be convex.

Scanned with CamScanner

blem 1 Solution 2''

to the subgradient inequality, there exists $s \in \mathbb{R}^n$, s.t.

$$f(x) \geq f(\bar{x}) + s^T(x - \bar{x}) \quad \text{for } \forall x \in C. \quad ①$$

sides, $f(\bar{x}) \geq f(x)$ for $\forall x \in C$ ②

can get $f(\bar{x}) \geq f(\bar{x}) + s^T(x - \bar{x})$ for $\forall x \in C$.

$$\Leftrightarrow \quad s^T x \leq s^T \bar{x} \quad \text{for } \forall x \in C$$

$s \neq 0$, we have $C \subseteq H^-(s, s^T\bar{x})$ ③

Then, we claim that $\bar{x} \notin \operatorname{int} C$.

Suppose not, $\bar{x} \in \operatorname{int} C \Rightarrow \exists \varepsilon > 0$, s.t. $B(\bar{x}, \varepsilon) \subseteq C$.

take $x_0 = \bar{x} + \frac{\varepsilon}{2} \frac{s}{\|s\|}$,

$$s^T x_0 = s^T\left(\bar{x} + \frac{\varepsilon}{2}\frac{s}{\|s\|}\right) = s^T\bar{x} + \frac{\varepsilon}{2}\cancel{\|s\|} > s^T\bar{x} \Rightarrow x_0 \notin H^-(s, s^T\bar{x})$$

$$\|x_0 - \bar{x}\| = \frac{\varepsilon}{2}\left\|\frac{s}{\|s\|}\right\| = \frac{\varepsilon}{2} < \varepsilon \Rightarrow x_0 \in B(\bar{x}, \varepsilon) \subseteq C$$

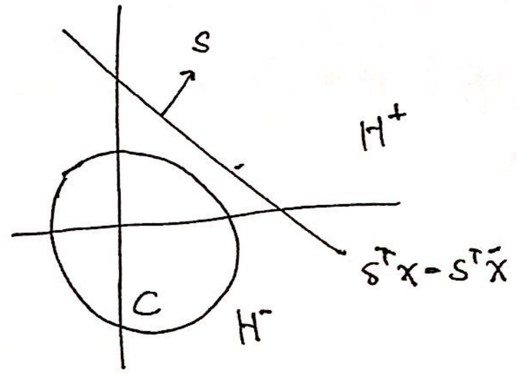
i.e. $\exists x_0 \in C$ while $x_0 \notin H^-(s, s^T\bar{x})$, which contradicts ③.

Hence, $\bar{x} \notin \operatorname{int} C$, However, it contradicts the given condition.

Therefore, $s = 0$.

From ① and ②, we get $f(x) \geq f(\bar{x})$ and $f(\bar{x}) \geq f(x)$

i.e. $f(x) = f(\bar{x})$ $\forall x \in C$. f is a constant function.



blem 3.

If $s \geq \|y\|_2$. Let $t = s$, $x = y$, then $(t, x) \in Q^{n+1}$.

$\|(t, x) - (s, y)\|^2 = 0$ is minimized. Thus $\Pi_{Q^{n+1}}(s, y) = (s, y)$.

) If $s \leq -\|y\|$, then

$\|(t, x) - (s, y)\|^2 = (t-s)^2 + \|x-y\|^2 = t^2 - 2st + s^2 + \|x\|^2 - 2x^T y + \|y\|^2$.

$\geq s^2 + \|y\|^2 + 2\|x\|^2 + 2t\|y\| - 2\|x\|\|y\|$

$\geq s^2 + \|y\|^2 + 2\|x\|^2 + 2\|x\|\|y\| - 2\|x\|\|y\|$

$\geq s^2 + \|y\|^2 = \|(0, 0) - (s, y)\|^2$

i.e. $\|(t, x) - (s, y)\|^2 \geq \|(0, 0) - (s, y)\|^2$ for all $(t, x) \in Q^{n+1}$.

when $(t, x) = (0, 0)$, the equality holds.

$\therefore \quad \Pi_{Q^{n+1}}(s, y) = (0, 0)$.

) If $-\|y\| < s < \|y\|$. Q^{n+1} is non-empty, closed and convex

$\therefore (t^*, x^*)$ is the projection of $(s, y) \Leftrightarrow \langle (t - t^*, x - x^*), (s - t^*, y - x^*) \rangle \leq 0 \; (\forall (t, x) \in Q^{n+1})$ (

et $(t, x) = (0, 0) \in Q^{n+1}$ in $(*)$, we get $\langle (t^*, x^*), (s - t^*, y - y^*) \rangle \leq 0$

et $(t, x) = (2t^*, 2x^*) \in Q^{n+1}$ in $(*)$ we get $\langle (t^*, x^*), (s - t^*, y - y^*) \rangle \leq 0$

$\therefore \langle (t^*, x^*), (s - t^*, y - x^*) \rangle = 0 \qquad (a)$

Further, we get $\langle (t, x), (s - t^*, y - x^*) \rangle \leq 0 \quad \forall (t, x) \in Q^{n+1} \qquad (b)$

Hence $(*) \Rightarrow (a)$ and (b).

On the contrary, if (a) and (b) hold, we can get

$-t^*, x - x^*), (s - t^*, y - x^*) \rangle = \langle (t, x), (s - t^*, y - y^*) \rangle - \langle (t^*, x^*), (s - t^*, y - x^*) \rangle$

$= \langle (t, x), (s - t^*, y - y^*) \rangle \leq 0 \qquad \forall (t, x) \in Q^{n+1}$

Hence (a) and $(b) \Rightarrow (*)$

i.e. $(*) \Leftrightarrow (a)$ and (b)

Thus (t^*, x^*) is the projection of (s, y) on Q^{n+1}

$\Leftrightarrow \langle (t^*, x^*), (s - t^*, y - x^*) \rangle = 0$ and $\langle (t, x), (s - t^*, y - x^*) \rangle \leq 0 \quad \forall (t, x) \in Q^{n+1}$.

Next, we show that $\frac{s+\|y\|_2}{2\|y\|_2}(\|y\|_2, y)$ satisfies (a)

1° $\langle(\|y\|_2, y), (s-\frac{s+\|y\|_2}{2}, y-\frac{s+\|y\|_2}{2\|y\|_2}y)\rangle$

$= s\|y\|_2 - \frac{s\|y\|_2+\|y\|_2^2}{2} + \|y\|_2^2 - \frac{s+\|y\|_2}{2}\|y\|_2 = 0$

(a) is satisfied

2° $\forall (t,x)\in Q^{n+1}$, $t\geq\|x\|_2$, then

$\langle(t,x), (s-\frac{s+\|y\|_2}{2}, y-\frac{s+\|y\|_2}{2\|y\|_2}y)\rangle$

$= t(s-\frac{s+\|y\|_2}{2}) + x^T y(1-\frac{s+\|y\|_2}{2\|y\|_2})$

$= \frac{s-\|y\|_2}{2}(t-\frac{x^Ty}{\|y\|_2}) \overset{\Delta}{=} \gamma$

Since $t-\frac{x^Ty}{\|y\|} \geq t-\frac{\|x\|\|y\|}{\|y\|} = t-\|x\|_2 \geq 0$ (Cau

$\frac{s-\|y\|_2}{2} < 0$ (since $s < \|y\|_2$)

$\therefore$ $\gamma < 0$

$\therefore$ (b) is satisfied

combine 1° and 2°, we obtain that

$\frac{s+\|y\|_2}{2\|y\|_2}(\|y\|_2, y)$ is the projection of (s,y) on Q^{n+1}

Finally, we have

$$\Pi_{Q^{n+1}}(s,y) = \begin{cases} (s,y), & \text{if } \|y\|_2\leq s \\ (0,0), & \text{if } \|y\|_2\leq -s \\ \frac{s+\|y\|_2}{2\|y\|_2}(\|y\|_2, y), & \text{otherwise} \end{cases}$$

WANG Xiaolu 1155099202

100

ENGG 5501: Foundations of Optimization	2017–18 First Term
Homework Set 3	
Instructor: Anthony Man–Cho So	Due: November 13, 2017

INSTRUCTIONS: Problem 1 is compulsory. For Problems 2 to 4, two of them will be graded. Nevertheless, you are advised to solve all the problems.

Problem 1. Let $A \in \mathbb{R}^{m \times n}$ be given. Use the Farkas lemma to show that exactly one of the following systems has a solution:

$$\begin{array}{ll} \text{(I)} & Ax \le 0,\ Ax \ne 0,\ x \ge 0. \\ \text{(II)} & A^T y \ge 0,\ y > 0. \end{array}$$

Problem 2. Let $a_1, \ldots, a_m \in \mathbb{R}^n$ and $c \in \mathbb{R}^n$ be given, with $m \ge n$. Consider the cone

$$C = \left\{ x \in \mathbb{R}^n : a_i^T x \ge 0 \text{ for } i = 1, \ldots, m \right\}.$$

(a) Show that C is pointed if and only if there exist n vectors in the collection $\{a_1, \ldots, a_m\}$ that are linearly independent.

(b) We say that $d \in C \setminus \{0\}$ is an *extreme ray* of C if there are $n-1$ linearly independent constraints that are active at d. Now, suppose that C is pointed. Consider the LP

$$v^* = \min_{x \in C} c^T x.$$

Show that $v^* = -\infty$ if and only if there exists an extreme ray d of C satisfying $c^T d < 0$. (*Hint: If $v^* = -\infty$, then there exists an $\bar{x} \in C$ such that $c^T \bar{x} < 0$.*)

Problem 3. Given $n \ge 1$, consider the second-order cone $\mathcal{Q}^{n+1} = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t \ge \|x\|_2\}$. Show that for any $(s, y) \in \mathbb{R} \times \mathbb{R}^n$, we have

$$\Pi_{\mathcal{Q}^{n+1}}(y) = \begin{cases} (s, y) & \text{if } \|y\|_2 \le s, \\ (0, 0) & \text{if } \|y\|_2 \le -s, \\ \dfrac{s + \|y\|_2}{2\|y\|_2}(\|y\|_2, y) & \text{otherwise.} \end{cases}$$

Problem 4. Consider the following SDP:

$$\begin{array}{ll} \sup & Z_{11} \\ \text{subject to} & Z_{11} - 2Z_{12} + Z_{22} = 1, \\ & Z_{11} - 2Z_{13} + Z_{33} = 1, \\ & Z_{22} - 2Z_{23} + Z_{33} = 4, \\ & Z_{22} = Z_{33} = 2, \\ & Z \in \mathcal{S}^3_+. \end{array} \qquad (T)$$

(a) Write down the dual of (T).

(b) Determine the feasible region of (T). Hence, or otherwise, determine the optimal solut
and optimal value of (T).

(c) Let $A, B \in S^n_+$. Show that $A \bullet B = 0$ if and only if $AB = 0$.

(d) Using the results in (a)–(c), show that there is no dual feasible solution that attains t
optimal value computed in (b).

$Ax \leq 0,\ Ax \neq 0,\ x \geq 0$

$A^T y \geq 0,\ y > 0$

ENGG 5501: Foundations of Optimization	2017–18 First Term
Homework Set 3 Solution	
Instructor: Anthony Man-Cho So	November 13, 2017

Problem 1. First, let us show that systems (I) and (II) cannot both have solutions. Suppose to the contrary that there exist vectors $\bar{x} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$ satisfying (I) and (II), respectively. Then, since $\bar{y} > 0$, $A\bar{x} \leq 0$ and $A\bar{x} \neq 0$, we have $\bar{y}^T A\bar{x} < 0$. On the other hand, since $\bar{x} \geq 0$ and $A^T\bar{y} \geq 0$, we have $\bar{y}^T A\bar{x} \geq 0$. This results in a contradiction.

Now, suppose that system (I) does not have a solution. Then, by a simple scaling argument, the system

$$Ax \leq 0,\ e^T Ax = -1,\ x \geq 0$$

$e^T Ax = \alpha \neq 0 \Leftrightarrow e^T A\left(\frac{x}{\alpha}\right) = -1$

does not have a solution either. By introducing slack variables, we see that system (I′) is equivalent to

$$\begin{bmatrix} A & I \\ e^T A & 0^T \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix},\ (x, s) \geq 0.$$

Hence, by Farkas' lemma, there exists a $\bar{z} = (\bar{u}, \bar{t}) \in \mathbb{R}^{m+1}$ such that

$$\begin{bmatrix} A^T & A^T e \\ I & 0 \end{bmatrix} \begin{bmatrix} \bar{u} \\ \bar{t} \end{bmatrix} \geq 0,\ \bar{t} > 0,$$

or equivalently,

$$A^T(\bar{u} + \bar{t}e) \geq 0,\ \bar{u} \geq 0,\ \bar{t} > 0.$$

Now, let $\bar{y} = \bar{u} + \bar{t}e \in \mathbb{R}^m$. Clearly, we have $A^T\bar{y} \geq 0$. Moreover, since $\bar{u} \geq 0$ and $\bar{t} > 0$, we have $\bar{y} \geq \bar{t}e > 0$. This completes the proof.

Problem 2.

(a) Let $A \in \mathbb{R}^{m\times n}$ be the $m \times n$ matrix whose i–th row is a_i^T, where $i = 1, \ldots, m$. Then,

$$\begin{aligned} & C \text{ is pointed} \\ \iff\ & \text{null}(A) = \{0\} \\ \iff\ & A \text{ has full column rank} \\ \iff\ & \text{there exist } n \text{ vectors in the collection } \{a_1, \ldots, a_m\} \text{ that are linearly independent.} \end{aligned}$$

(b) Suppose that there exists an extreme ray d of C satisfying $c^T d < 0$. Then, we have $\lambda d \in C$ for any $\lambda > 0$, which implies that $v^* = -\infty$.

Conversely, suppose that $v^* = -\infty$. Then, by scaling if necessary, there exists an $\bar{x} \in C$ such that $c^T\bar{x} = -1$. This implies that the polyhedron

$$P = \{x \in \mathbb{R}^n : a_i^T x \geq 0 \text{ for } i = 1, \ldots, m,\ c^T x = -1\}$$

is non-empty. Since C is pointed, by the result in (a), there exist n vectors in the collec $\{a_1,\dots,a_m\}$ that are linearly independent. Hence, P has at least one extreme point, $d \in \mathbb{R}^n$. As there are n linearly independent active constraints at d, we can find $n-1$ line independent constraints of the form $a_i^T x \geq 0$ that are active at d. It follows that d is extreme ray of C.

Problem 3. Let $(s,y) \in \mathbb{R}\times\mathbb{R}^n$ be fixed. If $\|y\|_2 \leq s$, then $(s,y) \in \mathcal{Q}^{n+1}$ and hence $\Pi_{\mathcal{Q}^{n+1}}((s,y)$ (s,y). Next, if $\|y\|_2 \leq -s$, then for any $(t,z) \in \mathcal{Q}^{n+1}$, we have

$$((s,y)-(0,0))^T((t,z)-(0,0)) = st + y^Tz \leq st + \|y\|_2 \cdot \|z\|_2 \leq st - st = 0.$$

Hence, by Theorem 5 of Handout 2, we have $\Pi_{\mathcal{Q}^{n+1}}((s,y)) = (0,0)$. Lastly, suppose that $\|y\|_2 >$ Let $s_0 = \frac{s+\|y\|_2}{2}$ and $y_0 = \frac{y}{\|y\|_2}$. Then, we have $s_0 > 0$ and $\|y\|_2 - s_0 = \frac{\|y\|_2 - s}{2} > 0$, which imp that $(s_0, s_0y_0) \in \mathcal{Q}^{n+1}$. Now, for any $(t,z) \in \mathcal{Q}^{n+1}$, we compute

$$\begin{aligned}
& ((s,y)-(s_0,s_0y_0))^T((t,z)-(s_0,s_0y_0)) = (s-s_0,(\|y\|_2-s_0)y_0)^T(t-s_0,z-s_0y_0) \\
= \;& (s-s_0)(t-s_0) + (\|y\|_2 - s_0)y_0^T(z-s_0y_0) \\
= \;& (s-s_0)(t-s_0) + (\|y\|_2 - s_0)y_0^Tz + s_0(s_0 - \|y\|_2) \\
\leq \;& (s-s_0)(t-s_0) + (\|y\|_2 - s_0)\|z\|_2 + s_0(s_0 - \|y\|_2) \\
\leq \;& (s-s_0)(t-s_0) + (\|y\|_2 - s_0)t + s_0(s_0 - \|y\|_2) \\
= \;& (s-s_0)(t-s_0) + (\|y\|_2 - s_0)(t-s_0) \\
= \;& (t-s_0)(s-s_0+\|y\|_2 - s_0) = 0.
\end{aligned}$$

Hence, by Theorem 5 of Handout 2, we have $\Pi_{\mathcal{Q}^{n+1}}((s,y)) = (s_0, s_0y_0) = \frac{s+\|y\|_2}{2\|y\|_2}(\|y\|_2, y)$.

Problem 4.

(a) The dual of (T) is given by

$$\begin{array}{ll}
\inf & y_1 + y_2 + 4y_3 + 2y_4 + 2y_5 \\
\text{subject to} & S = \begin{bmatrix} y_1+y_2-1 & -y_1 & -y_2 \\ -y_1 & y_1+y_3+y_4 & -y_3 \\ -y_2 & -y_3 & y_2+y_3+y_5 \end{bmatrix} \succeq \mathbf{0}.
\end{array}$$

(b) The constraints in (T) imply that Z takes the form

$$Z = \begin{bmatrix} Z_{11} & (1+Z_{11})/2 & (1+Z_{11})/2 \\ (1+Z_{11})/2 & 2 & 0 \\ (1+Z_{11})/2 & 0 & 2 \end{bmatrix}.$$

Since $Z \succeq 0$, we must have $\det(Z) \geq 0$. In other words, we have

$$0 \leq \det(Z) = 4Z_{11} - (1+Z_{11})^2 = -(1-Z_{11})^2,$$

which implies that $Z_{11} = 1$. To verify that the resulting Z is positive semidefinite, recall that if λ is an eigenvalue of Z, then

$$0 = \det(Z - \lambda I) = \lambda(2-\lambda)(\lambda - 3).$$

It follows that all the eigenvalues of Z are non–negative. Hence, the feasible region of (T) consists of the single point

$$\bar{Z} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix},$$

which implies that it is also the optimal solution to (T). The optimal value of (T) is then equal to 1.

(c) Clearly, if $AB = 0$, then we have $A \bullet B = \mathrm{tr}(AB) = 0$. Conversely, suppose that $A \bullet B = 0$. Since $A, B \succeq 0$, there exist $A^{1/2}, B^{1/2} \succeq 0$ such that $A = A^{1/2}A^{1/2}$ and $B = B^{1/2}B^{1/2}$. Hence, we have

$$\mathrm{tr}(AB) = \mathrm{tr}\left((A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}\right) = 0.$$

Let $M = (A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}$, and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Since $M \succeq 0$, we have $\lambda_i \geq 0$ for $i = 1, \ldots, n$. Since $\mathrm{tr}(M) = \sum_{i=1}^n \lambda_i = 0$, we see that $\lambda_1 = \cdots = \lambda_n = 0$, which implies that $M = 0$. In particular, we have $A^{1/2}B^{1/2} = 0$, from which it follows that $AB = 0$, as desired.

(d) Suppose that there exists a dual feasible solution $\bar{S}$ whose objective value is equal to 1. Then, we have

$$\bar{Z} \bullet \bar{S} = \bar{Z}_{11} - (\bar{y}_1 + \bar{y}_2 + 4\bar{y}_3 + 2\bar{y}_4 + 2\bar{y}_5) = 0.$$

Since $\bar{Z}, \bar{S} \succeq 0$, we have $\bar{Z}\bar{S} = 0$ by the result in (c). However, upon expanding this identity, we find that $(\bar{Z}\bar{S})_{11} = -1$, which is a contradiction.

$$Z = \begin{bmatrix} 1, & 1, & 1 \\ 1 & 2, & 0 \\ 1 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} y_1+y_2-1 & -y_1 & -y_2 \\ -y_1 & y_1+y_3+y_4 & -y_3 \\ -y_2 & -y_3 & y_2+y_3+y_5 \end{bmatrix}$$

$= y_1+y_2-1 + -y_1 - y_2 \quad -1$

$+ \; -y_1 + 2y_1+2y_3+2y_4 \; +$

$-y_2 \quad +2y_2+2y_3+2y_5$

$= y_1+y_2+4y_3+2y_4+\overset{3}{2}y_5 = 0$

-1

Problem 1] HW3 WANG Xiaolu 1155099202

System (I) has a solution $\Leftrightarrow \exists x$ s.t. $Ax \leq 0,\ Ax \neq 0,\ x \geq 0$

$\Leftrightarrow \exists x, z$ s.t. $Ax + z = 0,\quad x \geq 0,\ z \geq 0,\ z \neq 0$

$\Leftrightarrow \exists x, z$ s.t. $Ax + z = 0,\ e^T z > 0,\ (x, z) \geq 0$ $(e = (1, 1, \cdots, 1)^T$;

$\Leftrightarrow \exists x, z$ s.t. $Ax + z = 0,\ e^T z = 1,\ (x, z) \geq 0$

(since we can scale (x, z) with $\lambda > 0$ to make $e^T(\lambda z) = 1$)

$$\Leftrightarrow \exists (x, z) \text{ s.t. } \begin{bmatrix} A & I \\ 0 & e^T \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\quad (x, z) \geq 0 \qquad (\text{III})$$

By Farkas's Lemma, the following system (IV) corresponds to (III)

$$\begin{bmatrix} A^T & 0 \\ I & e \end{bmatrix} \begin{bmatrix} y \\ w \end{bmatrix} \leq 0,\quad (0, 1) \begin{pmatrix} y \\ w \end{pmatrix} > 0 \qquad (\text{IV})$$

$\Leftrightarrow \exists (y, w)$ s.t. $A^T y \leq 0,\ y + we \leq 0,\ w > 0$

$\Leftrightarrow \exists y$ s.t. $A^T y \leq 0,\ y < 0.$

10.

$\Leftrightarrow \exists y$ s.t. $A^T y \geq 0,\ y > 0$ which is exactly system (II)

Hence, exactly one of (I) and (II) has a solution.

Problem 2]

(a) Let $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$, then $C = \{x \in \mathbb{R}^n : Ax \geq 0\}$

C is pointed $\Leftrightarrow \forall x.\ x \in C$ and $-x \in C$ implies $x = 0$

$\Leftrightarrow \forall x.\ Ax \geq 0$ and $-Ax \geq 0$ implies

$\Leftrightarrow \forall x.\ Ax = 0$ implies $x = 0$

$\Leftrightarrow Ax = 0$ iff $x = 0$

15

$\Leftrightarrow \text{rank}(A) = n$

$\Leftrightarrow \dim \text{span}\{a_1 \cdots a_m\} = n$ (since $\text{rank} A^T = \text{rank} A = n$

$\Leftrightarrow \exists\ n$ vectors in $\{a_1 \cdots a_m\}$ which are independent (since $m \geq n$

(b) ① If $\exists d$, d is an extreme ... $\in$ ...

then $d \in C$. for $\forall \alpha \geq 0$, $\alpha d \in C$ (C is a cone)

$\therefore c^T(\alpha d) < 0$ when $\alpha \to \infty$, $c^T(\alpha d) \to -\infty$

$\therefore v^* = -\infty$

② If $v^* = -\infty$, then $\exists \bar{x} \in C$ s.t. $c^T x < 0$

i.e. $\exists \bar{x}$ s.t. $A\bar{x} \geq 0$, $c^T \bar{x} = -1$

Define Polyhedron $P = \{x \in \mathbb{R}^n : Ax \geq 0, c^T x = -1\}$

Obviously $\bar{x} \in P$, thus $P \neq \emptyset$

We claim that P doesn't contain a line.

15

Suppose not, we have $\exists y \in P$, $v \in \mathbb{R}^n \setminus \{0\}$ s.t. $y + \lambda v \in$

Thus, $A(y + \lambda v) \geq 0$, $c^T(y + \lambda v) = -1$ $(\forall \lambda \in \mathbb{R})$

$\Rightarrow Ay + \lambda Av \geq 0$ $(\forall \lambda \in \mathbb{R})$

Besides $y \in P \Rightarrow Ay \geq 0$.

Hence $Av = 0$.

Combined with $v \neq 0$, we get $\text{rank}(A) < n$. Since C is ... it con... what we have shown in (a). Hence the claim is true

By Theorem 3 in the lecture notes, we know that P ... a extreme point (vertex), which we denote as d.

By Theorem 2 in the lecture notes, we know that d is basic feasible solution. Thus there are n linearly independent constraints at d.

Since $c^T d = -1$, which is an active constraint, there must ... $n-1$ linearly independent active constraint among $a_1^T d \geq 0, \cdots$

$\therefore d$ is an extreme ray, and $c^T d < 0$.

Scanned with CamScanner

[Problem 4]

a) The Primal SDP can be written as (P)

$$(P)\quad \begin{aligned} \inf\ & C \cdot Z \\ \text{s.t. }\ & A_1 \cdot Z = b_1 \\ & A_2 \cdot Z = b_2 \\ & A_3 \cdot Z = b_3 \\ & A_4 \cdot Z = b_4 \\ & A_5 \cdot Z = b_5 \\ & Z \succeq 0 \end{aligned}$$

where $C = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_1 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$A_2 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$

$A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$Z = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{12} & z_{22} & z_{23} \\ z_{13} & z_{23} & z_{33} \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 2 \\ 2 \end{bmatrix}$

Hence the dual of (T) is

$$(D)\quad \begin{aligned} \sup\ & b^T y \\ \text{s.t. }\ & C - \sum_{i=1}^{5} y_i A_i \succeq 0 \end{aligned} \iff \begin{aligned} \sup\ & y_1 + y_2 + 4y_3 + 2y_4 + 2y_5 \\ \text{s.t. }\ & \begin{bmatrix} -y_1 - y_2 - 1 & y_1 & y_2 \\ y_1 & -y_1 - y_3 - y_4 & y_3 \\ y_2 & y_3 & -y_2 - y_3 - y_5 \end{bmatrix} \succeq 0 \end{aligned}$$

b)
$$\begin{cases} z_{11} - 2z_{12} + z_{22} = 1 \\ z_{11} - 2z_{13} + z_{33} = 1 \\ z_{22} - 2z_{23} + z_{33} = 4 \\ z_{22} = z_{33} = 2 \\ Z \succeq 0 \end{cases} \Rightarrow \begin{cases} z_{11} = 2z_{12} - 1 \\ z_{12} = z_{13} \\ z_{22} = z_{33} = 2 \\ z_{23} = 0 \\ Z \succeq 0 \end{cases} \Rightarrow \begin{bmatrix} 2z_{12} - 1 & z_{12} & z_{12} \\ z_{12} & 2 & 0 \\ z_{12} & 0 & 2 \end{bmatrix} \succeq 0$$

$$\Rightarrow \begin{cases} 2z_{12} - 1 \geq 0 \\ 2(2z_{12} - 1) - z_{12}^2 \geq 0 \\ 4(2z_{12} - 1) - z_{12} \cdot 2z_{12} + z_{12}(-2z_{12}) \geq 0 \end{cases} \Rightarrow z_{12} = 1 \Rightarrow Z^* = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

is the only element on the feasible region of (P). $V_P^* = -z_{11} = -1$.

$\therefore V_T^* = z_{11} = 1$

(c) 1° $AB = 0 \Rightarrow \operatorname{tr} AB = 0 \Rightarrow A \cdot B = 0$

2° $A \cdot B = 0 \Rightarrow \operatorname{tr} AB = 0 \Rightarrow \operatorname{tr}(A^{\frac12} A^{\frac12} B^{\frac12} B^{\frac12}) = 0$

$\Rightarrow \operatorname{tr}(A^{\frac12} A^{\frac12} B^{\frac12} B^{\frac12}) = \operatorname{tr}(B^{\frac12} A^{\frac12} A^{\frac12} B^{\frac12}) = \operatorname{tr}(A^{\frac12} B^{\frac12})^T (A^{\frac12} B^{\frac12})$ (since $(A^{\frac12})^T = A^{\frac12}, (B^{\frac12})^T$)

$\triangleq \operatorname{tr} M = \sum \lambda_i = 0$ (λ_i is the eigenvalue of M)

Scanned with CamScanner

$M = (A^{\frac{1}{2}} B^{\frac{1}{2}})^T (A^{\frac{1}{2}} B^{\frac{1}{2}}) \triangleq X^T X \succeq 0$ (... $\|X v\|_2 \geq 0$,

$\Rightarrow \lambda_i \geq 0.\ \forall i$

Hence $\lambda_i = 0\ (\forall i)$, then $M = 0$.

i.e. $(A^{\frac{1}{2}} B^{\frac{1}{2}})^T (A^{\frac{1}{2}} B^{\frac{1}{2}}) = 0 \Rightarrow \| A^{\frac{1}{2}} B^{\frac{1}{2}} \|_F^2 = 0 \Rightarrow A^{\frac{1}{2}} B^{\frac{1}{2}} = 0$

$\therefore\ AB = A^{\frac{1}{2}} (A^{\frac{1}{2}} B^{\frac{1}{2}}) B^{\frac{1}{2}} = 0$

(d) It's obvious that (P) is bounded and strictly feasible (since

Suppose $\bar{y}$ is a feasible solution to (D).

$$Z^* \bullet (C - \sum_{i=1}^{5} \bar{y}_i A_i) = 0 \quad (*) \quad \text{if and only if}$$

$$Z^* (C - \sum_{i=1}^{5} \bar{y}_i A_i) \triangleq Y = 0$$

By simple calculation, we can get that the element in first row and first column of Y is $-1 \neq 0$

$\therefore\ Y \neq 0$.

$\therefore$ complementary slackness (*) does not hold

$\therefore$ The duality gap of (P) and (D) is not 0. (Str

i.e. $b^T \bar{y} < C \bullet Z^*$ for $\forall \bar{y}$ feasible to (D).

$\therefore$ No dual feasible solution can attain the optimal va

WANG Xiaolu 1155099202

100

ENGG 5501: Foundations of Optimization — 2017–18 First Term

Homework Set 4

Instructor: Anthony Man–Cho So — Due: November 24, 2017

INSTRUCTIONS: Problem 1 is compulsory. For Problems 2 to 4, two of them will be graded. Nevertheless, you are advised to solve all the problems.

Problem 1. Let $A_1,\ldots,A_m \in \mathcal{S}^n_+$, $\alpha_1,\ldots,\alpha_m > 0$, and $\beta_1,\ldots,\beta_m > 0$ be given. Consider the following SDP:

$$\begin{array}{lll} \inf & \displaystyle\sum_{i=1}^m \operatorname{tr}(Z_i) & \\ \text{subject to} & \operatorname{tr}\left[A_i\left(\alpha_i Z_i - \displaystyle\sum_{j\neq i} Z_j\right)\right] \geq \beta_i & \text{for } i=1,\ldots,m, \\ & Z_1,\ldots,Z_m \in \mathcal{S}^n_+. & \end{array} \tag{Q}$$

(a) Write down the dual of (Q).

(b) Using the result in (a), show that the dual is strictly feasible. ✓

Problem 2. Let $B \in \mathcal{S}^n$ and $U \in \mathcal{S}^n_+$ be given. Consider the following SDP:

$$\begin{array}{rcll} v^* & = & \text{maximize} & \operatorname{tr}(B(Y-Z)) \\ & & \text{subject to} & Y+Z=U, \\ & & & Y, Z \succeq 0. \end{array}$$

(a) Show that

$$\begin{array}{rcll} v^* & = & \text{maximize} & \operatorname{tr}(VU^{1/2}BU^{1/2}) \\ & & \text{subject to} & \|V\|_2 \leq 1, \\ & & & V \in \mathcal{S}^n, \end{array}$$

where $\|V\|_2$ denotes the spectral norm of V.

(b) Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of $U^{1/2}BU^{1/2}$. Using the result in (a), show that $v^* = \sum_{i=1}^n |\lambda_i|$.

✓ **Problem 3.** We say that a set $X \subset \mathbb{R}^n$ is *SOC–representable* if there exist matrices $A^j \in \mathbb{R}^{(n_j+1)\times(n+\ell)}$ and vectors $b^j \in \mathbb{R}^{n_j+1}$ for $j=1,\ldots,m$ such that

$$x \in X \iff \exists u \in \mathbb{R}^\ell \text{ such that } A^j \begin{bmatrix} x \\ u \end{bmatrix} - b^j \in \mathcal{Q}^{n_j+1} \text{ for } j=1,\ldots,m.$$

(a) Show that if $X \subset \mathbb{R}^n$ is SOC–representable and $A : \mathbb{R}^n \to \mathbb{R}^k$ is an affine mapping by $A(x) = Qx + w$, where $Q \in \mathbb{R}^{k \times n}$ and $w \in \mathbb{R}^k$ are given, then the set $A(X) \subset$ SOC–representable.

(b) Show that if $X \subset \mathbb{R}^n$ is SOC–representable and $A' : \mathbb{R}^k \to \mathbb{R}^n$ is an affine mapping by $A'(y) = My + p$, where $M \in \mathbb{R}^{n \times k}$ and $p \in \mathbb{R}^n$ are given, then the set $(A')^{-1}$ $\{y \in \mathbb{R}^k : A'(y) \in X\} \subset \mathbb{R}^k$ is SOC–representable.

Problem 4. Consider the ℓ_1–regularized ℓ_2–regression problem

$$\min_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} \left\{ \|b - Ax - te\|_2 + \lambda \|x\|_1 \right\},$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ are given, and $e \in \mathbb{R}^m$ is the vector of all–ones. A co interpretation of the ℓ_1–regularizer $x \mapsto \|x\|_1$, which is based on heuristic arguments, is t promotes sparsity in the optimal solution to (1). In this problem, we will show in a rig manner that problem (1) is actually equivalent to a robust ℓ_2–regression problem.

To begin, consider the following robust optimization problem:

$$\min_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} \ \max_{\Delta A \in \mathcal{U}_\lambda} \ \|b - (A + \Delta A)x - te\|_2.$$

Here, the uncertainty set $\mathcal{U}_\lambda$ is defined as

$$\mathcal{U}_\lambda = \left\{ X \in \mathbb{R}^{m \times n} : \|X\|_{1,2} \le \lambda \right\},$$

where $\|X\|_{1,2} = \max_{\|v\|_1 = 1} \|Xv\|_2$. Note that $\|\cdot\|_{1,2}$ defines a matrix norm (see Handout B).

(a) Show that for any $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ and $\Delta A \in \mathcal{U}_\lambda$,

$$\|b - (A + \Delta A)x - te\|_2 \le \|b - Ax - te\|_2 + \lambda \|x\|_1.$$

(b) Show that for any $(x, t) \in \mathbb{R}^n \times \mathbb{R}$, there exists a $\Delta A^* \in \mathcal{U}_\lambda$ such that the inequality in holds as equality; i.e.,

$$\|b - (A + \Delta A^*)x - te\|_2 = \|b - Ax - te\|_2 + \lambda \|x\|_1.$$

Hence, conclude that problems (1) and (2) are equivalent.

ENGG 5501: Foundations of Optimization **2017–18 First Term**

Homework Set 4 Solution

Instructor: Anthony Man-Cho So November 24, 2017

Problem 1.

(a) Observe that (Q) is equivalent to

$$\begin{array}{lll} \inf & D \bullet Z & \\ \text{subject to} & H_i \bullet Z = \beta_i & \text{for } i = 1, \ldots, m, \\ & Z \in \mathcal{S}_+^{m(n+1)}, & \end{array} \tag{1}$$

where

$$\begin{aligned} D &= \operatorname{BlkDiag}(I, \ldots, I, 0) \in \mathcal{S}^{m(n+1)}, \\ H_i &= \operatorname{BlkDiag}(-A_i, \ldots, -A_i, \underbrace{\alpha_i A_i}_{i^{th}}, -A_i, \ldots, -A_i, 0, \ldots, 0, -1, 0, \ldots, 0) \in \mathcal{S}^{m(n+1)}, \end{aligned}$$

and $\operatorname{BlkDiag}(Q_1, \ldots, Q_l)$ denotes the block diagonal matrix whose i-th diagonal block is Q_i, for $i = 1, \ldots, l$. Indeed, since $D, H_1, \ldots, H_m$ are block diagonal, every feasible solution $\bar{Z}$ to (1) gives rise to a block diagonal feasible solution $\bar{Z}' = \operatorname{BlkDiag}(\bar{Z}_1', \ldots, \bar{Z}_m', \bar{s}_1', \ldots, \bar{s}_m')$ to (1) whose objective value is equal to that of $\bar{Z}$. Now, the dual of (1) is given by

$$\begin{array}{ll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i \\ \text{subject to} & D - \displaystyle\sum_{i=1}^m y_i H_i \in \mathcal{S}_+^{m(n+1)}, \end{array}$$

which, using the structure of $D, H_1, \ldots, H_m$, is equivalent to

$$\begin{array}{lll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i & \\ \text{subject to} & I - \alpha_i y_i A_i + \displaystyle\sum_{j \neq i} y_j A_j \in \mathcal{S}_+^n & \text{for } i = 1, \ldots, m, \\ & y \geq 0. & \end{array} \tag{2}$$

(b) Since $A_1, \ldots, A_m \in \mathcal{S}_+^n$ and $y \geq 0$, we have

$y_i > 0$.

$$I + \sum_{j \neq i} y_j A_j \succeq I$$

for $i = 1, \ldots, m$. Hence, to show that (2) is strictly feasible, it suffices to choose $y_i > 0$ such that $I \succ \alpha_i y_i A_i$ for $i = 1, \ldots, m$. Since $\alpha_i > 0$ and we may assume that $A_i \neq 0$, we have $\alpha_i \lambda_{\max}(A_i) > 0$. Hence, any $y \in \mathbb{R}^m$ satisfying

$$0 < y_i < \frac{1}{\alpha_i \lambda_{\max}(A_i)} \quad \text{for } i = 1, \ldots, m$$

is a strictly feasible solution to (2).

Problem 2.

(a) We first claim that

$$\max_{Y,Z\succeq 0:\, Y+Z=U} \operatorname{tr}(B(Y-Z)) = \max_{P,Q\succeq 0:\, P+Q=I} \operatorname{tr}(B(U^{1/2}(P-Q)U^{1/2})).$$

This is clearly true for $U \succ 0$ (simply set $P = U^{-1/2}YU^{-1/2}$ and $Q = U^{-1/2}ZU^{-1/2}$). For general $U \succeq 0$, we can argue as follows. Let $Y, Z \succeq 0$ be such that $Y + Z = U$. For any ϵ define

$$U_\epsilon = U + \epsilon I \succ 0, \quad P_\epsilon = U_\epsilon^{-1/2}\left(Y + \frac{\epsilon}{2}I\right)U_\epsilon^{-1/2}, \quad Q_\epsilon = U_\epsilon^{-1/2}\left(Z + \frac{\epsilon}{2}I\right)U_\epsilon^{-1/2}.$$

Then, we have $P_\epsilon + Q_\epsilon = I$ for any $\epsilon > 0$. Since $P_\epsilon, Q_\epsilon \succeq 0$ for any $\epsilon > 0$, the sequences $\{P_\epsilon\}_\epsilon$ and $\{Q_\epsilon\}_\epsilon$ lie in the compact set $\{X \in \mathcal{S}^n : I \succeq X \succeq 0\}$. Hence, by letting $\epsilon \searrow 0$ and taki[ng] subsequences if necessary, we have $P_\epsilon \to P$ and $Q_\epsilon \to Q$ for some $P, Q \succeq 0$. Now, obse[rve] that $P + Q = I$, and that

$$U^{1/2}PU^{1/2} = \lim_{\epsilon\searrow 0} U_\epsilon^{1/2}P_\epsilon U_\epsilon^{1/2} = Y, \quad U^{1/2}QU^{1/2} = \lim_{\epsilon\searrow 0} U_\epsilon^{1/2}Q_\epsilon U_\epsilon^{1/2} = Z.$$

This establishes the claim.

Next, we compute

$$\begin{aligned}
\max_{Y,Z\succeq 0:\, Y+Z=U} \operatorname{tr}(B(Y-Z)) &= \max_{P,Q\succeq 0:\, P+Q=I} \operatorname{tr}(B(U^{1/2}(P-Q)U^{1/2})) \\
&= \max_{P,Q\succeq 0:\, P+Q=I} \operatorname{tr}((U^{1/2}BU^{1/2})(P-Q)) \\
&\le \max_{P,Q\succeq 0:\, P+Q\preceq I} \operatorname{tr}((U^{1/2}BU^{1/2})(P-Q)) \\
&= \max_{V\in\mathcal{S}^n:\, \|V\|_2\le 1} \operatorname{tr}(VU^{1/2}BU^{1/2}),
\end{aligned}$$

where the inequality in (3) follows from the fact that the maximization is performed ove[r a] larger feasible set. To prove (4), observe that

$$\{P - Q \in \mathcal{S}^n : P, Q \succeq 0,\ P + Q \preceq I\} = \{V \in \mathcal{S}^n : \|V\|_2 \le 1\}.$$

Indeed, for any $x \in \mathbb{R}^n$ and $P, Q \succeq 0$ such that $P + Q \preceq I$, we have $|x^T(P-Q)x| \le 1$, fro[m] which it follows that $\|P - Q\|_2 \le 1$. Conversely, let $V \in \mathcal{S}^n$ be such that $\|V\|_2 \le 1$. Le[t] $V = H\Gamma H^T$ be its spectral decomposition, and define $P = H\Gamma_+H^T$, $Q = H\Gamma_-H^T$, whe[re] Γ_+, Γ_- are diagonal matrices with

$$(\Gamma_+)_{ii} = \max\{0, \Gamma_{ii}\}, \quad (\Gamma_-)_{ii} = \max\{0, -\Gamma_{ii}\} \qquad \text{for } i = 1, \ldots, n.$$

Then, we have $P, Q \succeq 0$ and $V = P - Q$. Moreover, since $\|V\|_2 \le 1$, we have $\|\Gamma_+\|_2, \|\Gamma_-\|_2 \le$ 1, from which it follows that $P + Q \preceq I$, as desired.

Finally, to complete the proof, we claim that the inequality in (3) holds as equality. This [is] obvious if $U^{1/2}BU^{1/2}$ is diagonal, for then we can restrict our attention to diagonal matri[ces]

P, Q with $P, Q \succeq 0$ and $P + Q \preceq I$. Otherwise, let $U^{1/2}BU^{1/2} = W\Lambda W^T$ be the spectral decomposition of $U^{1/2}BU^{1/2}$. Observe that

$$\mathrm{tr}((U^{1/2}BU^{1/2})(P - Q)) = \mathrm{tr}(\Lambda(W^T(P - Q)W)),$$

and $W^TPW, W^TQW \succeq 0$ are such that $W^T(P+Q)W \preceq W^TW = I$. Hence, we have reduced the problem to the diagonal case, and the proof is completed.

(b) Let $U^{1/2}BU^{1/2} = W\Lambda W^T$ be the spectral decomposition of $U^{1/2}BU^{1/2}$. Then, we have

$$\mathrm{tr}(VU^{1/2}BU^{1/2}) = \mathrm{tr}((W^TVW)\Lambda) \quad \text{and} \quad \|W^TVW\|_2 = \|V\|_2 \leq 1.$$

Γ

Since Λ is diagonal, it follows from the result in (a) that

$$\begin{aligned} v^* \quad = \quad & \text{maximize} \quad \mathrm{tr}(\Gamma\Lambda) \\ & \text{subject to} \quad \|\Gamma\|_2 \leq 1, \\ & \qquad\qquad\quad \Gamma \text{ diagonal.} \end{aligned}$$

However, this is equivalent to

$$v^* = \max\left\{\sum_{i=1}^{n} \gamma_i\lambda_i : \max_{1\leq i\leq n} |\gamma_i| \leq 1\right\},$$

from which it is clear that $v^* = \sum_{i=1}^{n} |\lambda_i|$.

Problem 3.

(a) By considering appropriate bases in $\mathbb{R}^k$ and $\mathbb{R}^n$, we may assume the existence of a non-singular $\bar{Q} \in \mathbb{R}^{p\times p}$ (where $p \in \{0, 1, \ldots, n\}$) such that for $x = (x^1, x^2) \in \mathbb{R}^n$ with $x^1 \in \mathbb{R}^p$ and $x^2 \in \mathbb{R}^{n-p}$, we have

$$Qx = \begin{bmatrix} \bar{Q}x^1 \\ 0 \end{bmatrix}.$$

Consequently,

$$\begin{bmatrix} y^1 \\ y^2 \end{bmatrix} = A(x) \iff x^1 = \bar{Q}^{-1}(y^1 - w^1),\ y^2 = w^2, \tag{5}$$

where $y^1, w^1 \in \mathbb{R}^p$; $y^2, w^2 \in \mathbb{R}^{k-p}$; $w = (w^1, w^2) \in \mathbb{R}^k$. Now, suppose that $X \subset \mathbb{R}^n$ is SOC-representable; i.e., there exist matrices $A^j \in \mathbb{R}^{(n_j+1)\times(n+\ell)}$ and vectors $b^j \in \mathbb{R}^{n_j+1}$ for $j = 1, \ldots, m$ such that

$$x \in X \iff \exists u \in \mathbb{R}^\ell \text{ such that } A^j \begin{bmatrix} x \\ u \end{bmatrix} - b^j \in \mathcal{Q}^{n_j+1} \text{ for } j = 1, \ldots, m. \tag{6}$$

Using (5) and (6), we see that

$$\begin{aligned}
\begin{bmatrix} y^1 \\ y^2 \end{bmatrix} \in A(X) &\iff y^2 = w^2 \text{ and } \exists v \in \mathbb{R}^{n-p} \text{ such that} \\
&\qquad x = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} \bar{Q}^{-1}(y^1 - w^1) \\ v \end{bmatrix} \in X \\
&\iff \exists u \in \mathbb{R}^\ell, v \in \mathbb{R}^{n-p} \text{ such that} \\
&\qquad \text{(i) } A^j \begin{bmatrix} \bar{Q}^{-1}(y^1 - w^1) \\ v \\ u \end{bmatrix} - b^j \in \mathcal{Q}^{n_j+1} \text{ for } j = 1, \ldots, m, \\
&\qquad \text{(ii) } (0, y^2 - w^2) \in \mathcal{Q}^{k-p+1}.
\end{aligned}$$

Since both (i) and (ii) stipulate that certain affine maps of (y^1, y^2, u, v) belong to second-order cones of appropriate dimensions, we conclude that $A(X)$ is SOC–representable.

(b) To show that $(A')^{-1}(X)$ is SOC-representable, it suffices to use (6) and observe that

$$\begin{aligned}
y \in (A')^{-1}(X) &\iff A'(y) \in X \\
&\iff \exists u \in \mathbb{R}^\ell \text{ such that } A^j \begin{bmatrix} My + p \\ u \end{bmatrix} - b^j \in \mathcal{Q}^{n_j+1} \text{ for } j = 1, \ldots, m
\end{aligned}$$

Problem 4.

(a) By the triangle inequality and the definition of $\|\cdot\|_{1,2}$, for any $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ and $\Delta A \in$ we have

$$\begin{aligned}
\|b - (A + \Delta A)x - te\|_2 &\leq \|b - Ax - te\|_2 + \|\Delta A x\|_2 \\
&\leq \|b - Ax - te\|_2 + \|\Delta A\|_{1,2}\|x\|_1 \\
&\leq \|b - Ax - te\|_2 + \lambda\|x\|_1,
\end{aligned}$$

as desired.

(b) We need to construct a matrix $\Delta A^* \in \mathcal{U}_\lambda$ such that all the inequalities in (a) hold as equalities. Towards that end, let $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ be fixed. Define the matrix $\Delta A^* \in \mathbb{R}^{m \times n}$ by

$$\Delta A^* = \begin{cases} -\dfrac{\lambda(b - Ax - te)(\operatorname{sgn}(x))^T}{\|b - Ax - te\|_2} & \text{if } b - Ax - te \neq 0, \\ \lambda u (\operatorname{sgn}(x))^T & \text{otherwise,} \end{cases}$$

where $u \in \mathbb{R}^m$ is an arbitrary unit vector (i.e., $\|u\|_2 = 1$) and $\operatorname{sgn}(x) \in \mathbb{R}^n$ is given by

$$[\operatorname{sgn}(x)]_j = \begin{cases} 1 & \text{if } x_j \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

We first verify that

$$\|\Delta A^*\|_{1,2} = \max_{\|v\|_1 = 1} \|\Delta A^* v\|_2 = \lambda \max_{\|v\|_1=1} \left|(\operatorname{sgn}(x))^T v\right| = \lambda;$$

i.e., $\Delta A^* \in \mathcal{U}_\lambda$. Next, observe that when $b - Ax - te \neq 0$, we have

$$\begin{aligned}
\|b - (A + \Delta A^*)x - te\|_2 &= \left\| b - Ax - te + \frac{\lambda(b - Ax - te)}{\|b - Ax - te\|_2} (\operatorname{sgn}(x))^T x \right\|_2 \\
&= \left\| (b - Ax - te) \left(1 + \frac{\lambda \|x\|_1}{\|b - Ax - te\|_2} \right) \right\|_2 \checkmark \\
&= \|b - Ax - te\|_2 + \lambda \|x\|_1.
\end{aligned}$$

On the other hand, when $b - Ax - te = 0$, we have ✓

$$\|b - (A + \Delta A^*)x - te\|_2 = \|\Delta A^* x\|_2 = \lambda \|x\|_1 = \|b - Ax - te\|_2 + \lambda \|x\|_1.$$

This completes the proof. In particular, we conclude that

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \max_{\Delta A \in \mathcal{U}_\lambda} \|b - (A + \Delta A)x - te\|_2 = \min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \{\|b - Ax - te\|_2 + \lambda \|x\|_1\}. \checkmark$$

$$\max_{\|v\|_1 = 1} \|\Delta A^* v\|_2 = \max_{\|v\|_1 = 1} \left\| - \frac{\lambda(b - Ax - te)(\operatorname{sgn}(x))^T v}{\|b - Ax - te\|_2} \right\|_2$$

ENGG 5501 Homework 4 (WANG Xiaolu 1155099202)

[Problem 1]

(a) The original problem (Q) can be written as

$$\inf \ \mathrm{tr}\begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_m \end{bmatrix}$$

$$\text{s.t.} \quad \mathrm{tr}\left(\begin{bmatrix} -A_i & & & & \\ & \ddots & & & \\ & & \alpha_i A_i & & \\ & & & \ddots & \\ & & & & -A_i \end{bmatrix}\begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_m \end{bmatrix}\right) \geq \beta_i \quad (i=1,2,\cdots,m) \qquad (Q')$$

$$\begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_m \end{bmatrix} \in S_+^{mn}$$

Denote $Z = \begin{bmatrix} Z_1 & & \\ & \ddots & \\ & & Z_m \end{bmatrix}$, $\tilde{A}_i = \begin{bmatrix} -A_i & & & & \\ & \ddots & & & \\ & & \alpha_i A_i & & \\ & & & \ddots & \\ & & & & -A_i \end{bmatrix}$

Then, problem (Q') is equivalent to

$$\inf \ I_{mn} \bullet Z$$

$$\text{s.t.} \quad \tilde{A}_i \bullet Z - x_i = \beta_i \quad (i=1,2,\cdots,m) \qquad (T)$$

$$Z \in S_+^{mn}, \ x_i \geq 0 \quad (i=1,2,\cdots,m)$$

(T) is also equivalent to

$$\inf \begin{bmatrix} I_{mn} & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \bullet \begin{bmatrix} Z & & & \\ & x_1 & & \\ & & \ddots & \\ & & & x_m \end{bmatrix}$$

$$\text{s.t.} \quad \begin{bmatrix} \tilde{A}_i & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \ddots \ 0 \end{bmatrix} \bullet \begin{bmatrix} Z & & & & \\ & x_1 & & & \\ & & \ddots & & \\ & & & x_i & \\ & & & & \ddots \ x_m \end{bmatrix} = \beta_i \quad (i=1,2,\cdots,m)$$

$$\begin{bmatrix} Z & & & \\ & x_1 & & \\ & & \ddots & \\ & & & x_m \end{bmatrix} \in S_+^{(m+1)n}$$

Hence, the dual of (Q) is the dual of (T), which is

$$\sup \ \beta^T y$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i \begin{bmatrix} -A_i & & & & & \\ & \ddots & & & & \\ & & \alpha_i A_i & & & \\ & & & \ddots \ -A_i & & \\ & & & & 0 & \\ & & & & & \ddots \ 1 \end{bmatrix} + S = \begin{bmatrix} I_{mn} & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}, \quad S \in S_+^{(m+1)n}, \quad y \in \mathbb{R}^m.$$

$(\beta = (\beta_1, \cdots, \beta_m)^T)$

(b) We want to find $y \in R^m$, s.t $y > 0$ ... is equ...

$$I + \sum_{j \neq i} y_j A_j - \alpha_i y_i A_i > 0 \quad (i = 1, 2, \cdots, m)$$

and $y_i > 0 \quad (i = 1, 2, \cdots, m)$

Since $A_i \succeq 0$, we just need to find $y \in R^m$ s.t.

$$I - \alpha_i y_i A_i > 0 \text{ and } y_i > 0 \quad (i = 1, 2, \ldots$$

Apply spectral decomposition to A_i, we get

$$I - \alpha_i y_i Q \Sigma Q^T = Q(I - \alpha_i y_i \Sigma) Q^T > 0.$$

$$\Leftrightarrow \quad 1 - \alpha_i y_i \lambda_k > 0 \quad \text{for } \forall k = 1, 2, \cdots, n$$

$$\Leftrightarrow \quad y_i < \frac{1}{\alpha_i \lambda_k} \quad \text{for } \forall k = 1, 2, \cdots, r$$

$$\Leftrightarrow \quad y_i < \frac{1}{\alpha_i \lambda_{max}(A_i)}$$

Hence, we just need to choose $y \in R^m$, s.t.

$$0 < y_i < \frac{1}{\alpha_i \lambda_{max}(A_i)} \quad (i = 1, 2, \ldots$$

[Problem 2]

(a) Let $Y - Z = X$, combined with $Y + Z = U$, we get

$$Y = \frac{U + X}{2}, \quad Z = \frac{U - X}{2}$$

Then, we obtain the equivalent problem (1):

$$\begin{aligned} \max \ & tr(BX) \\ s.t. \ & U + X \succeq 0 \qquad (1) \\ & U - X \succeq 0 \end{aligned}$$

Apply spectral decomposition to U, we know $U = Q \Sigma$...

Hence. $Q \Sigma Q^T + X \succeq 0$ and $Q \Sigma Q^T - X \succeq 0$

$$\Leftrightarrow \quad \Sigma + Q^T X Q \succeq 0 \text{ and } \Sigma - Q^T X Q \succeq 0$$

(Since Q is orthogonal, thus it is full rank)

Suppose $\Sigma = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}$. $Q^T X Q = \bar{X} = \begin{bmatrix} \bar{X}_{11} & \bar{X}_{12} \\ \bar{X}_{12}^T & \bar{X}_{22} \end{bmatrix}$

Then, (2) can be written as

$$\begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \bar{X}_{11} & \bar{X}_{12} \\ \bar{X}_{12}^T & \bar{X}_{22} \end{bmatrix} \succeq 0 \text{ and } \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \bar{X}_{11} & \bar{X}_{12} \\ \bar{X}_{12}^T & X_{22} \end{bmatrix} \succeq 0$$

$\Leftrightarrow \quad \Lambda \pm \bar{X}_{11} \succeq 0, \quad \pm \bar{X}_{12} \succeq 0, \quad \pm \bar{X}_{22} \succeq 0$

$\Leftrightarrow \quad \Lambda \pm \bar{X}_{11} \succeq 0, \quad \bar{X}_{12} = 0, \quad X_{22} = 0$

$\Leftrightarrow \quad \begin{bmatrix} \Lambda \pm \bar{X}_{11} & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$

$\Leftrightarrow \quad \begin{bmatrix} \Lambda^{\frac{1}{2}} (I \pm \Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}}) \Lambda^{\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$

$\Leftrightarrow \quad \Lambda^{\frac{1}{2}} (I \pm \Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}}) \Lambda^{\frac{1}{2}} \succeq 0$

$\Leftrightarrow \quad I \pm \Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}} \succeq 0$ (since $\Lambda^{\frac{1}{2}}$ is symmetric and full ran

Apply spectral decomposition to $\Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}}$, we get

$$I \pm Q D Q^T \succeq 0 \quad \Leftrightarrow \quad Q(I \pm D) Q^T \succeq 0$$

Thus, the eigenvalues of $\Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}}$ are all between $[-1, 1]$. (3)

We want to find V, s.t. $U^{\frac{1}{2}} V U^{\frac{1}{2}} = X$,

Hence, $V = (U^{\frac{1}{2}})^{-1} X (U^{\frac{1}{2}})^{-1}$

$$= Q \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} Q^T Q \begin{bmatrix} \bar{X}_{11} & 0 \\ 0 & 0 \end{bmatrix} Q^T Q \begin{bmatrix} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= Q \begin{bmatrix} \Lambda^{-\frac{1}{2}} \bar{X}_{11} \Lambda^{-\frac{1}{2}} & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

By (3), we know that the eigenvalues of V are all between $[-1, 1]$

Hence, $\|V\|_2 = \sqrt{\lambda_{max}(V^T V)} = \sqrt{\lambda_{max}(V^2)} = \sqrt{\lambda_{max}^2(V)} = |\lambda_{max}(V)| \leq 1$

We write problem (1) as problem (4):

$$\max \ tr(BX) \quad (4)$$
$$\text{s.t. } X = U^{\frac{1}{2}} V U^{\frac{1}{2}}$$
$$\|V\|_2 \leq 1$$
$$V \in S^n$$

We have proved that the constraints of (1) implies the constraints

Reversely, if the constraints of (4) is satisfied, we can easily

$$U + X = U + U^{\frac{1}{2}} V U^{\frac{1}{2}} = U^{\frac{1}{2}}(I+V)U^{\frac{1}{2}} \geq 0$$ (since $\|V\|_2$

$$U - X = U - U^{\frac{1}{2}} V U^{\frac{1}{2}} = U^{\frac{1}{2}}(I-V)U^{\frac{1}{2}} \geq 0$$

Thus the constraints of (1) is fullfilled.

Hence, problem (1) $\Leftrightarrow$ problem (4)

(4) can be further written as

$$\begin{aligned} &\max \ tr(BU^{\frac{1}{2}} V U^{\frac{1}{2}}) \\ &\text{s.t. } \|V\|_2 \leq 1 \\ &V \in S^n \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\max \ tr(VU^{\frac{1}{2}} B U^{\frac{1}{2}}) \\ &\text{s.t. } \|V\|_2 \leq 1 \\ &V \in S^n \end{aligned}$$

Therefore, $V^* = \max \ tr(VU^{\frac{1}{2}} B U^{\frac{1}{2}})$
$$\text{s.t. } \|V\|_2 \leq 1$$
$$V \in S^n$$

(b) Apply spectral decomposition to $U^{\frac{1}{2}} B U^{\frac{1}{2}}$, we get

$$U^{\frac{1}{2}} B U^{\frac{1}{2}} = P H P^T = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} P^T$$

Then, $tr(VU^{\frac{1}{2}} B U^{\frac{1}{2}}) = tr(V P H P^T) = tr(P^T V P H) = P^T V P$

Since P is orthogonal matrix, then $P^T = P^{-1}$.

Thus, $P^T V P$ and V are similar matrices, and they have the same eigen values.

Denote $\tilde{V} = P^T V P$. Since $Q^T V Q$ is symmetric and $\|V\|_2 \leq 1$, we

$$\|\tilde{V}\|_2 = \|Q^T V Q\|_2 = \sqrt{\lambda^2_{max}(Q^T V Q)} = \sqrt{\lambda^2_{max}(V)} = |\lambda_{max}(V)| \leq 1$$

Then, $tr(VU^{\frac{1}{2}}BU^{\frac{1}{2}}) = P^T V P \circ H = \sum_{i=1}^{n} \tilde{V}_{ii} \lambda_i$.

Suppose the spectral decomposition of $\tilde{V}$ is $\tilde{V} = \Omega^T R \Omega$,

then we have $\tilde{V}_{ii} = e_i^T \tilde{V} e_i = e_i^T \Omega^T R \Omega e_i = (\Omega e_i)^T R (\Omega e_i)$

$$= \omega_i^T \begin{bmatrix} r_1 & & \\ & \ddots & \\ & & r_n \end{bmatrix} \omega_i = \sum_{k=1}^{n} r_k (\omega_i^{(k)})^2$$

$$\leq r_{max} \sum_{k=1}^{n} (\omega_i^{(k)})^2 = r_{max}\, \omega_i^T \omega_i = r_{max}$$

(where w_i is the ith column of Ω, $\omega_i^{(k)}$ is the kth entry of w_i.)

Since $\|\tilde{V}\|_2 \leq 1$, we have $\tilde{V}_{ii} \leq r_{max} = \lambda_{max}(\tilde{V}) \leq 1$

Hence, $tr(VU^{\frac{1}{2}}BU^{\frac{1}{2}}) = \sum_{i=1}^{n} \tilde{V}_{ii} \lambda_i \leq \sum_{i=1}^{n} \lambda_i \leq \sum_{i=1}^{n} |\lambda_i|$ (upper bound)

Take $\tilde{V}_{ii} = \tilde{V}_{ii}^* = \begin{cases} 1, & \text{if } \lambda_i > 0 \\ -1, & \text{if } \lambda_i < 0 \\ 0, & \text{if } \lambda_i = 0 \end{cases}$, and all other elements are arbitrarily chosen,

then the upper bound can be attained.

In such cace, check that the corresponding $\tilde{V}^*$ surely satisfies $\tilde{V}^* \in S^n$ and $\|\tilde{V}^*\|_2 \leq 1$. Since $\tilde{V}^* = P^T V^* P$, we know $V^* = P\tilde{V}^* P^T$. P is orthogonal, thus $V^* \in S^n$ and $\|V^*\|_2 \leq 1$.

Therefore, $v^* = \sum_{i=1}^{n} |\lambda_i|$, which is attained by optimal solution V^*.

oblem 3 (

Suppose $A^j = \begin{bmatrix} p_j^T \\ D_j \end{bmatrix}$, $b^j = \begin{bmatrix} q_j \\ d_j \end{bmatrix}$, where $p_j \in R^{(n+\ell)\times 1}$, $q_j \in R$.

Then $A^j \begin{bmatrix} X \\ u \end{bmatrix} - b^j \in Q^{n_j+1} \iff \|D_j \begin{bmatrix} X \\ u \end{bmatrix} - d_j\|_2 \leq p_j^T \begin{bmatrix} X \\ u \end{bmatrix} - q_j$. (1)

Apply singular-value decomposition to Q, and assume $rank(Q) = r$,

we have $Q = U \Sigma V^T = U \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} V^T$, where $U \in R^{k\times k}$, $V \in R^{n\times n}$.

For $\forall y \in A(X) \subseteq R^K$, (we write it as $y = \begin{bmatrix} y_r \\ y_{K-r} \end{bmatrix}$, where $y_r \in$

$\exists x = \begin{pmatrix} x_r \\ x_{n-r} \end{pmatrix} \in R^n$ s.t. $\begin{bmatrix} y_r \\ y_{K-r} \end{bmatrix} = Q \begin{bmatrix} x_r \\ x_{n-r} \end{bmatrix} + \begin{bmatrix} w_r \\ w_{K-r} \end{bmatrix}$

$$\Leftrightarrow \begin{bmatrix} y_r \\ y_{K-r} \end{bmatrix} = U \Sigma V^T \begin{bmatrix} x_r \\ x_{n-r} \end{bmatrix} + \begin{bmatrix} w_r \\ w_{K-r} \end{bmatrix}$$

$$\Leftrightarrow U^T \begin{bmatrix} y_r \\ y_{K-r} \end{bmatrix} = \Sigma \left(V^T \begin{bmatrix} x_r \\ x_{n-r} \end{bmatrix} \right) + U^T \begin{bmatrix} w_r \\ w_{K-r} \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} y_r' \\ y_{K-r}' \end{bmatrix} = \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_r' \\ x_{n-r}' \end{bmatrix} + U^T \begin{bmatrix} w_r \\ w_{K-r} \end{bmatrix} \quad \left(\text{let } U^T \begin{bmatrix} y_r \\ y_{K-r} \end{bmatrix} = \begin{bmatrix} y_r' \\ y_{K-r}' \end{bmatrix}, V^T \begin{bmatrix} x_r \\ x_{n-r} \end{bmatrix} \right.$$

$$\Leftrightarrow \begin{bmatrix} y_r' \\ y_{K-r}' \end{bmatrix} = \begin{bmatrix} \Lambda_r x_r' \\ 0 \end{bmatrix} + \begin{bmatrix} w_r' \\ w_{K-r}' \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} y_r' = \Lambda_r x_r' + w_r' \\ y_{K-r}' = w_{K-r}' \end{cases}$$

$$\Leftrightarrow \begin{cases} x_r' = \Lambda_r^{-1}(y_r' - w_r') \\ y_{K-r}' = w_{K-r} \end{cases}$$

$$\Leftrightarrow \begin{cases} \begin{bmatrix} x_r' \\ x_{n-r}' \end{bmatrix} = \begin{bmatrix} \Lambda_r^{-1}(y_r' - w_r') \\ \phi \end{bmatrix}, \quad \text{for some } \phi \in R^{n-r}. \\ y_{K-r}' = w_{K-r}' \end{cases}$$

$$\Leftrightarrow \begin{cases} V^T x = \begin{bmatrix} \Lambda_r^{-1}(y_r' - w_r') \\ \phi \end{bmatrix}, \quad \text{for some } \phi \in R^{n-r} \\ y_{K-r}' = w_{K-r}' \end{cases}$$

$$\Leftrightarrow \begin{cases} x = V \begin{bmatrix} \Lambda_r^{-1}(y_r' - w_r') \\ \phi \end{bmatrix}, \quad \text{for some } \phi \in R^{n-r} \\ y_{K-r}' = w_{K-r}' \end{cases} \qquad (2)$$

Since $X \subseteq R^n$ is SOC-representable, by (1.) we have

$\exists u \in R^l$ s.t. $p_j^T \begin{bmatrix} V \begin{pmatrix} \Lambda_r^{-1}(y_r' - w_r') \\ \phi \end{pmatrix} \\ u \end{bmatrix} - q_j \geq \| D_j \begin{bmatrix} V \begin{pmatrix} \Lambda_r^{-1}(y_r' - w_r') \\ \phi \end{pmatrix} \\ u \end{bmatrix}$

writing the matrices V, P_j, D_j in block form, and apply trivial matrix manipulations, the inequalities (3) can be transformed into the following form:

$$\alpha_j^T \begin{bmatrix} y_r' \\ y_{n-r}' \\ \varphi \end{bmatrix} - \beta_j \geq \| G_j \begin{bmatrix} y_r' \\ y_{n-r}' \\ \varphi \end{bmatrix} - g_j \|_2 \quad (j=1,2,\cdots,m) \quad (4)$$

replace $\begin{bmatrix} y_r' \\ y_{n-r}' \end{bmatrix}$ by $U^T \begin{bmatrix} y_r \\ y_{n-r} \end{bmatrix} = U^T y$, (4) can be farther transformed into

$$\gamma_j^T \begin{bmatrix} y \\ \theta \end{bmatrix} - \eta_j \geq \| H_j \begin{bmatrix} y \\ \theta \end{bmatrix} - h_j \|_2 \quad (j=1,2,\cdots,m) \quad (5)$$

Besides, $y_{k-r}' = w_{k-r}' \Leftrightarrow \| y_{k-r}' - w_{k-r}' \|_2 \leq 0$

$$\Leftrightarrow \| (0, I) \begin{bmatrix} y_r' \\ y_{k-r}' \end{bmatrix} - \begin{bmatrix} 0 \\ w_{k-r}' \end{bmatrix} \|_2 \leq 0$$

$$\Leftrightarrow \| (0, I) U y - \begin{bmatrix} 0 \\ w_{k-r}' \end{bmatrix} \|_2 \leq 0$$

$$\Leftrightarrow \| [(0, I) U, 0] \begin{bmatrix} y \\ \theta \end{bmatrix} - \begin{bmatrix} 0 \\ w_{k-r}' \end{bmatrix} \|_2 \leq 0^T \begin{bmatrix} y \\ \theta \end{bmatrix} - 0 \quad (6)$$

Therefore, we've shown that (2) $\Leftrightarrow$ (5) & (6)

Hence, $A(X)$ is SOC-representable.

b) $\forall y \in (A')^{-1}(X)$, we have $My + p \in X$

Since X is SOC-representable, we can get

$$\exists u \in R^l \text{ s.t. } \| D_j \begin{bmatrix} My+p \\ u \end{bmatrix} - d_j \|_2 \leq p_j^T \begin{bmatrix} My+p \\ u \end{bmatrix} - q_j \quad (j=1,2,\cdots,m)$$

By some matrix manipulations, the inequalities can be written as

$$\| G_j \begin{bmatrix} y \\ \varphi \end{bmatrix} - g_j \|_2 \leq \alpha_j^T \begin{bmatrix} y \\ \varphi \end{bmatrix} - \beta_j \quad (j=1,2,\cdots m),$$

which shows that $(A')^{-1}(X)$ is SOC-representable.

[Problem 4]

(a) $\|b-(A+\Delta A)x - te\|_2 = \|b-Ax-te-\Delta A x\|_2$

$\leq \|b-Ax-te\|_2 + \|\Delta A x\|_2 \leq \|b-Ax-te\|_2 + \lambda\|x\|_1$.

The last inequality holds because

$$\|\Delta A\|_{1,2} = \max_{\|x\|_1=1}\|\Delta A x\|_2 = \max_{\|x\|_1=1}\frac{\|\Delta A x\|_2}{\|x\|_1} \leq \max_{x\neq 0}\frac{\|\Delta A x\|}{\|x\|_1}$$

$$\leq \max_{x\neq 0}\left\|\Delta A\left(\frac{x}{\|x\|_1}\right)\right\|_2 \leq \max_{\|y\|_1=1}\|\Delta A y\|_2 = \max_{\|x\|_1=1}\|\Delta A x\|_2$$

$$\therefore \|\Delta A\|_{1,2} = \max_{x\neq 0}\frac{\|\Delta A x\|_2}{\|x\|_1} \geq \frac{\|\Delta A x\|_2}{\|x\|_1} \quad (\forall x\neq 0)$$

$$\therefore \|\Delta A x\|_2 \leq \|\Delta A\|_{1,2}\|x\|_1 \leq \lambda\|x\|_1 \quad (\forall x\neq 0)$$

For $x=0$, $\|\Delta A x\|_2 = \lambda\|x\|_1 = 0$ trivially holds.

(b) Define

$$\Delta A^* = \begin{cases} -\dfrac{\lambda(b-Ax-te)(\mathrm{sgn}(x))^T}{\|b-Ax-te\|_2}, & \text{if } b-Ax-te\neq 0 \\ \lambda u(\mathrm{sgn}(x))^T, & \text{otherwise} \end{cases}$$

where $u\in\mathbb{R}^m$, $\|u\|_2=1$, $\mathrm{sgn}(x)\in\mathbb{R}^n$, $[\mathrm{sgn}(x)]_j = \begin{cases} 1, & x_j \\ -1, & x_j \end{cases}$

Then, $\|\Delta A^*\|_{1,2} = \max_{\|v\|_1=1}\|\Delta A^* v\|_2 = \lambda \max_{\|v\|_1=1}|(\mathrm{sgn}(x))^T v| = \lambda\max_{\|v\|_1=1}\|v\|_1$

(The second equality is because both $\frac{b-Ax-te}{\|b-Ax-te\|_2}$ and u are unit vectors.

Thus $\Delta A^*\in\mathcal{U}_\lambda$.

1° If $b-Ax-te\neq 0$, we have

$$\|b-(A+\Delta A^*)x-te\|_2 = \left\|b-Ax-te+\frac{\lambda(b-Ax-te)}{\|b-Ax-te\|_2}(\mathrm{sgn}(x))^T x\right\|_2$$

$$=\left(1+\frac{\lambda\|x\|_1}{\|b-Ax-te\|_2}\right)\|b-Ax+te\|_2 = \|b-Ax-te\|_2+\lambda\|x\|_1$$

(The second equality is because $(\mathrm{sgn}(x))^T x = \sum_{i=1}^n |x_i| = \|x\|_1$)

2° If $b-Ax-te=0$, we have

$$\|b-(A+\Delta A^*)x-te\|_2 = \|b-Ax-te-\Delta A^* x\|_2 = \|\Delta A^* x\|_2$$

$$= \|\lambda u(\mathrm{sgn}(x))^T x\|_2 = \|\lambda u\cdot\|x\|_1\|_2 = \lambda\|x\|_1\|u\|_2 = \lambda\|x$$

$$= \|b-Ax-te\|_2 + \lambda\|x\|_1.$$

The proof is completed.

WANG Xiaolu 1155099202

100

ENGG 5501: Foundations of Optimization 2017–18 First Term

Homework Set 5

Instructor: Anthony Man–Cho So Due: December 5, 2017

SOLVE THE FOLLOWING PROBLEMS:

Problem 1. Let $m \geq n$ be given integers and $A \in \mathcal{S}^m$ be a given symmetric matrix. Consider the following optimization problem:

$$\begin{array}{ll} \min & \operatorname{tr}(X^T A X) \\ \text{subject to} & X^T X = I, \\ & X \in \mathbb{R}^{m \times n}. \end{array} \tag{S}$$

(a) Show that the first–order optimality condition of (S) can be expressed as

$$AX - XX^T AX = 0. \tag{1}$$

(b) Explain why condition (1) is necessary for optimality.

(c) Show that $X \in \mathbb{R}^{m \times n}$ satisfies (1) iff X can be written as $X = PQ^T$, where the columns of $P \in \mathbb{R}^{m \times n}$ are n orthonormal eigenvectors of A (in particular, $P^T P = I$) and $Q \in \mathbb{R}^{n \times n}$ is some orthogonal matrix.

Problem 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function and $C \subset \mathbb{R}^n$ be a non–empty closed convex set. Show that x^* is an optimal solution to the convex optimization problem

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

iff

$$x^* = \Pi_C(x^* - \nabla f(x^*)),$$

where $\Pi_C(\cdot)$ is the projection operator onto C.

Problem 3. Let $A^j \in \mathbb{R}^{m_j \times n}$ be given matrices, $C_j \subset \mathbb{R}^{m_j}$ be given non–empty convex sets, and $f_j : \mathbb{R}^{m_j} \to \mathbb{R} \cup \{+\infty\}$ be given functions that are convex on C_j, where $j = 1, \ldots, J$. Consider the following optimization problem:

$$\begin{array}{lll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j x) & \\ \text{subject to} & A^j x \in C_j & \text{for } j = 1, \ldots, J. \end{array} \tag{2}$$

(a) Show that (2) is equivalent to

$$\begin{array}{lll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j x^j) & \\ \text{subject to} & A^j x^j = A^j \bar{x} & \text{for } j = 1, \ldots, J, \\ & x^j \in (A^j)^{-1} C_j & \text{for } j = 1, \ldots, J, \\ & \bar{x} \in \mathbb{R}^n. & \end{array} \tag{3}$$

(Recall that $(A^j)^{-1}C_j = \{x \in \mathbb{R}^n : A^j x \in C_j\}$.) Hence, by letting $w^j \in \mathbb{R}^{m_j}$ be the Lagrangian multiplier associated with the constraint $A^j x^j = A^j \bar{x}$, where $j = 1, \dots, J$, show that the Lagrangian dual of (3) can be expressed as

$$\begin{array}{ll} \sup & \displaystyle\sum_{j=1}^{J} \inf_{x \in C_j} \{f_j(x) + (w^j)^T x\} \\ \text{subject to} & \displaystyle\sum_{j=1}^{J} (A^j)^T w^j = 0. \end{array}$$

(b) Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ be given. Using the result in (a), construct a dual of the following *ridge regression problem*:

$$\min_{x \in \mathbb{R}^n} \{\|Ax - b\|_2^2 + \lambda \|x\|_2^2\}.$$

Simplify your answer as much as possible.

ENGG 5501: Foundations of Optimization

Homework Set 5 Solution

2017–18 First Term

Instructor: Anthony Man–Cho So

December 5, 2017

Problem 1.

(a) Let $W \in \mathbb{R}^{n\times n}$ be the Lagrangian multiplier associated with the constraint $X^TX = I$. Then, the Lagrangian function is given by

$$L(X, W) = \text{tr}(X^TAX) + W \bullet (X^TX - I).$$

The first–order optimality conditions are

$$X^TX = I, \quad \nabla_X L(X, W) = 0. \tag{1}$$

Since $\nabla_X L(X, W) = 2AX + X(W + W^T)$, we see from (1) that $0 = X^T\nabla_X L(X, W) = 2X^TAX + (W + W^T)$, which implies that $W + W^T = -2X^TAX$. It follows that (1) is equivalent to

$$X^TX = I, \quad AX - XX^TAX = 0. \tag{2}$$

(b) Let $h : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times n}$ be the function defined by $h(X) = X^TX - I$. Then, the Jacobian of h is

$$Dh(X) = (I_{n^2} + K_{nn})(I_n \otimes X^T),$$

where I_n is the $n \times n$ identity matrix and K_{nn} is the $n^2 \times n^2$ *commutation matrix*; see, e.g., [1, Chapter 9, Section 13]. Note that $Dh(X)$ is an $n^2 \times (mn)$ matrix. Since $h(X) = 0$ contains only $n(n+1)/2$ different constraints, we show that $Dh(X)$ has rank $n(n+1)/2$, which would then imply that the linear independence regularity condition holds. By [1, Chapter 3, Section 7, Theorem 11], we have $\text{rank}(I_{n^2} + K_{nn}) = n(n+1)/2$. Moreover, the matrix $I_n \otimes X^T$ has full row rank, as $(I_n \otimes X^T)(I_n \otimes X^T)^T = (I_n \otimes X^T)(I_n \otimes X) = I_n \otimes (X^TX) = I_n \otimes I_n$ and $\text{rank}(I_n \otimes I_n) = n^2$ (recall that $m \geq n$ by assumption). It follows that $\text{rank}(Dh(X)) = n(n+1)/2$, as desired.

(c) Suppose that X satisfies the first–order optimality conditions (2). Let $X^TAX = Q\Sigma Q^T$ be a spectral decomposition of X^TAX and set $P = XQ \in \mathbb{R}^{m\times n}$. Observe that $P^TP = Q^TX^TXQ = I$ and $AP = P\Sigma$. Since Σ is diagonal, we conclude that each column of P is a unit eigenvector of A. Hence, $X = PQ^T$ has the required properties.

Conversely, suppose that $X = PQ^T$, where the columns of $P \in \mathbb{R}^{m\times n}$ are n orthonormal eigenvectors of A and $Q \in \mathbb{R}^{n\times n}$ is some orthogonal matrix. Clearly, we have $X^TX = QP^TPQ^T = I$. Moreover, since $AP = P\Sigma$ for some diagonal $\Sigma \in \mathbb{R}^{n\times n}$, we have

$$\begin{aligned} AX - XX^TAX &= APQ^T - PQ^TQP^TAPQ^T \\ &= P\Sigma Q^T - P\Sigma Q^T \\ &= 0. \end{aligned}$$

Hence, we conclude that X satisfies the first–order optimality conditions (2).

Problem 2. Suppose that $x^* = \Pi_C(x^* - \nabla f(x^*))$. Since C is a non-empty closed convex [...] any $x \in C$, we have

$$0 \geq (x - \Pi_C(x^* - \nabla f(x^*)))^T (x^* - \nabla f(x^*) - \Pi_C(x^* - \nabla f(x^*))) = -\nabla f(x^*)^T(x - x^*)$$

This, together with the continuous differentiability and convexity of f, implies that for all x

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

i.e., x^* is an optimal solution to the optimization problem

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

Conversely, suppose that $x^* \neq \Pi_C(x^* - \nabla f(x^*))$. Since

$$(x^* - \Pi_C(x^* - \nabla f(x^*)))^T (x^* - \nabla f(x^*) - \Pi_C(x^* - \nabla f(x^*))) \leq 0,$$

we have

$$\nabla f(x^*)^T (x^* - \Pi_C(x^* - \nabla f(x^*))) \geq \|x^* - \Pi_C(x^* - \nabla f(x^*))\|_2^2 > 0,$$

or equivalently,

$$\nabla f(x^*)^T (\Pi_C(x^* - \nabla f(x^*)) - x^*) < 0.$$

This implies that $d = \Pi_C(x^* - \nabla f(x^*)) - x^*$ is a ~~descent direction~~ of f at x^*. Moreover, si[nce] $x^*, \Pi_C(x^* - \nabla f(x^*)) \in C$, we see that $x^* + \alpha d \in C$ for all $\alpha \in [0, 1]$, which implies that d is al[so a] feasible direction at x^*. It follows that x^* is not an optimal solution to (3).

Problem 3.

(a) Upon eliminating the variables $x^1, \ldots, x^J \in \mathbb{R}^n$ from the problem

$$\begin{array}{lll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j x^j) & \\ \text{subject to} & A^j x^j = A^j \bar{x} & \text{for } j = 1, \ldots, J, \\ & x^j \in (A^j)^{-1} C_j & \text{for } j = 1, \ldots, J, \\ & \bar{x} \in \mathbb{R}^n, & \end{array}$$

we obtain

$$\begin{array}{lll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j \bar{x}) & \\ \text{subject to} & A^j \bar{x} \in C_j & \text{for } j = 1, \ldots, J, \\ & \bar{x} \in \mathbb{R}^n, & \end{array}$$

which is the same as the given problem. Note that we may assume $A^j(A^j)^{-1}C_j = C_j$ without loss of generality. Indeed, we always have $A^j(A^j)^{-1}C_j \subset C_j$. If $y \in C_j \setminus A^j(A^j)^{-1}C_j$, then there does not exist an $\bar{x} \in \mathbb{R}^n$ such that $y = A^j\bar{x}$. Hence, we can remove y from C_j without affecting the optimal value of and optimal solutions to problem (5). Consequently, we can derive the Lagrangian dual of (4) as follows:

$$\sup_{\substack{w^j \in \mathbb{R}^{m_j} \\ j=1,\ldots,J}} \theta(w^1, \ldots, w^J),$$

where

$$\begin{aligned}
\theta(w^1, \ldots, w^J) &= \inf_{\substack{x^j \in (A^j)^{-1}C_j \\ j=1,\ldots,J; \\ \bar{x} \in \mathbb{R}^n}} \sum_{j=1}^{J} \{f_j(A^j x^j) + (w^j)^T A^j (x^j - \bar{x})\} \\
&= \sum_{j=1}^{J} \inf_{x^j \in (A^j)^{-1}C_j} \{f_j(A^j x^j) + (w^j)^T A^j x^j\} - \sup_{\bar{x} \in \mathbb{R}^n} \sum_{j=1}^{J} (w^j)^T A^j \bar{x} \\
&= \begin{cases} \displaystyle\sum_{j=1}^{J} \inf_{x \in C_j} \{f_j(x) + (w^j)^T x\} & \text{if } \displaystyle\sum_{j=1}^{J} (A^j)^T (w^j) = 0, \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

This is equivalent to

$$\begin{array}{ll} \sup & \displaystyle\sum_{j=1}^{J} \inf_{x \in C_j} \{f_j(x) + (w^j)^T x\} \\ \text{subject to} & \displaystyle\sum_{j=1}^{J} (A^j)^T w^j = 0, \end{array}$$

as desired.

(b) Upon letting $f_1(y) = \|y - b\|_2^2$, $f_2(y) = \lambda \|y\|_2^2$, $A^1 = A$, $A^2 = I$, $C_1 = \mathbb{R}^m$, $C_2 = \mathbb{R}^n$ and using the result in (a), we obtain the following dual of the given problem:

$$\begin{array}{ll} \sup & \displaystyle\inf_{x \in \mathbb{R}^m} \{\|x - b\|_2^2 + (w^1)^T x\} + \inf_{x \in \mathbb{R}^n} \{\lambda \|x\|_2^2 + (w^2)^T x\} \\ \text{subject to} & A^T w^1 + w^2 = 0. \end{array} \tag{6}$$

Now, using the first-order optimality conditions, we have

$$\begin{aligned}
\inf_{x \in \mathbb{R}^m} \{\|x - b\|_2^2 + (w^1)^T x\} &= -\frac{1}{4}\|w^1\|_2^2 + b^T w^1, \\
\inf_{x \in \mathbb{R}^n} \{\lambda \|x\|_2^2 + (w^2)^T x\} &= -\frac{1}{4\lambda}\|w^2\|_2^2.
\end{aligned}$$

It follows that (6) is equivalent to

$$\sup_{w \in \mathbb{R}^m} \left\{ -\frac{1}{4}\|w\|_2^2 + b^T w - \frac{1}{4\lambda}\|A^T w\|_2^2 \right\}.$$

Using again the first-order optimality condition and noting that $I + (1/\lambda)AA^T \succ 0$ for all $\lambda > 0$, we can express the optimal solution to the above problem in (

$$w^* = 2\left(I + \frac{1}{\lambda}AA^T\right)^{-1} b.$$

References

[1] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics* *Econometrics.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Chiche England, revised edition, 1999.

[Problem 1]

(a) $X^TX = I$ can be written as

$$X^TX \cdot E_{ij} = \mathbb{1}_{i=j} \quad (1 \le i \le j \le n)$$

where E_{ij} is an $n\times n$ matrix with $(E_{ij})_{ij} = 1$, and all other elements are 0. $\mathbb{1}_{i=j} = \begin{cases} 1, & i=j \\ 0, & o/w. \end{cases}$

The KKT condition of (S) is

$$\nabla \operatorname{tr}(X^TAX) + \sum_{1\le i\le j\le n} w_{ij} \cdot \nabla (X^TX \cdot E_{ij} - \mathbb{1}_{i=j}) = 0 \qquad (*)$$

Since $\nabla \operatorname{tr} X^TAX = 2AX$

$$\nabla(X^TX\cdot E_{ij} - \mathbb{1}_{i=j}) = \nabla \operatorname{tr}(XE_{ij}X^T) = X(E_{ij}^T + E_{ij}).$$

we have $(*)$ results in

$$2AX + \sum_{1\le i\le j\le n} w_{ij} X(E_{ij}^T + E_{ij}) = 0$$

$$\Leftrightarrow \quad AX + X(W^T+W) = 0 \qquad (\text{denote } (w_{ij}) = W)$$

$$\Rightarrow \quad X^TAX + X^TX(W^T+W) = 0$$

$$\Rightarrow \quad W^T + W = -X^TAX \qquad (\text{since } X^TX = I)$$

$$\Rightarrow \quad AX - XX^TAX = 0$$

(b) Denote $X_{ij} = X(E_{ij}^T + E_{ij}) = \begin{array}{c} \quad\ i\text{th} \quad\ j\text{th column} \\ [0 \cdots x_j \cdots x_i \cdots 0] \end{array} \quad (1 \le i \le j \le n)$

Suppose $\sum_{1\le i\le j\le n} \lambda_{ij} X_{ij} = 0 \ (\in \mathbb{R}^{n\times n})$ (where $\lambda_{ij} \in \mathbb{R}$)

Then $X^T \sum_{1\le i\le j\le n} \lambda_{ij} X_{ij} = 0$

$$\Rightarrow \sum_{1\le i\le j\le n} \lambda_{ij} \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} [0 \cdots x_j \cdots x_i \cdots 0] = 0$$

$$\Rightarrow \sum \begin{array}{c} \begin{array}{cc} i\text{th} & j\text{th} \end{array} \\ \begin{bmatrix} \cdots & \cdots & \lambda_{ij} & \cdots \\ \cdots & \lambda_{ij} & \cdots & \cdots \end{bmatrix} \begin{array}{l} i\text{th} \\ j\text{th} \end{array} \end{array} = 0$$

$\Rightarrow \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{n1} & \cdots & \lambda_{nn} \end{bmatrix} = 0$ i.e. $\lambda_{ij} = 0$ $(\forall\ 1 \le i \le j \le n)$

Hence, $X(E_{ij}^T + E_{ij})$ $(1 \le i \le j \le n)$ are independent at all feasible point

Suppose $\bar{X}$ is an optimal point, then it is a local minimum, and $\bar{X}$ is

$\therefore$ $\bar{X}$ satisfies KKT condition, then there must be $AX - XX^TAX = 0$.

(c) [$\Rightarrow$] Apply spectral decomposition to X^TAX, we have

$$X^TAX = Q\Sigma Q^T.$$

$$\Rightarrow X^TAXQ = Q\Sigma$$

Besides, $AX = XX^TAX \Rightarrow AXQ = X(X^TAXQ)$

$\therefore$ $AXQ = XQ\Sigma$

Let $P = XQ$, we have $AP = P\Sigma$.

$\therefore$ $\Sigma = \text{diag}(\lambda_1 \cdots \lambda_n)$ is the eigenvalue of A

Since $P^TP = (XQ)^T XQ = Q^TX^TXQ = I$

we get the column of P are n orthonormal eigenvectors of

Write X as $X = PQ^T$, we get the result.

[$\Leftarrow$] If $X = PQ^T$, the

$$AX - XX^TAX = APQ^T - PQ^TQP^TAPQ = APQ - PP^TAPQ$$

$$= APQ - (P_1 \cdots P_n)\begin{pmatrix} P_1^T \\ \vdots \\ P_n^T \end{pmatrix} APQ$$

$$= APQ - (P_1 \cdots P_n)\begin{pmatrix} P_1^TA \\ \vdots \\ P_n^TA \end{pmatrix} PQ$$

$$= APQ - (P_1 \cdots P_n)\begin{pmatrix} \lambda_1 P_1^T \\ \vdots \\ \lambda_n P_n^T \end{pmatrix} (P_1 \cdots P_n)\, Q$$

$$= APQ - (P_1 \cdots P_n)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} Q$$

$= (Ap_1 \cdots Ap_n) Q - (\lambda_1 p_1 \cdots \lambda_n p_n) Q$

$= (Ap_1 - \lambda_1 p_1, \cdots, Ap_n - \lambda_n p_n) Q$

$= 0.$ (since $Ap_i = \lambda_i p_i$ $i = 1, \cdots, n$)

i.e. $AX - XX^T AX = 0$

roblem 2]

1° $\Pi_C(x^* - \nabla f(x^*)) = x^*$

$\Leftrightarrow (x^* - \nabla f(x^*) - x^*)^T (z - x^*) \leq 0 \quad \forall z \in C$ (since C is convex & closed)

$\Leftrightarrow (\nabla f(x^*))^T (z - x^*) \geq 0 \quad \forall z \in C$

$\Rightarrow f(z) \geq f(x^*) + (\nabla f(x^*))^T (z - x^*) \geq f(x^*) \ (\forall z \in C)$ (since f is convex)

$\Rightarrow$ x^* is an optimal solution

2° If x^* is an optimal solution, we claim that

$\nabla f(x^*)(z - x^*) \geq 0 \quad \forall z \in C$

Suppose not, $\exists\, x \in C$ s.t. $\nabla f(x^*)(x - x^*) < 0$.

Then, $d = x - x^*$ is a descent direction of f at x^*.

Thus $\exists\, \alpha_0 > 0$ s.t. $\forall \alpha \in [0, \alpha_0)$, $f(x^* + \alpha d) < f(x^*)$

Note that $x^* + \alpha d = x^* + \alpha(x - x^*) = \alpha x + (1-\alpha) x^* \in C$

(because of the convexity of C.)

We obtain that x^* is not the optimal solution, which is a contradiction.

Hence, the claim is proved.

By the equivalence relation shown in 1°, we have

$\Pi_C(x^* - \nabla f(x^*)) = x^*$.

Scanned with CamScanner

oblem 3]

(a)
$$\begin{array}{l} \inf \sum_{j=1}^{J} f_j(A^j x^j) \\ \text{s.t. } A^j x^j = A^j \bar{x} \\ \quad A^j x^j \in G_j \\ \quad \bar{x} \in \mathbb{R}^n \\ \quad (j=1,\cdots,J) \end{array} \quad (\#) \quad \Longleftrightarrow \quad \begin{array}{l} \inf \sum_{j=1}^{J} f_j(A^j \bar{x}) \\ \text{s.t. } A^j \bar{x} \in G_j \\ \quad \bar{x} \in \mathbb{R}^n \\ \quad (j=1,\cdots,J) \end{array} \quad \Longleftrightarrow \quad \begin{array}{l} \inf \sum_{j=1}^{J} f_j(A^j x) \\ \text{s.t. } A^j x \in G_j \\ \quad (j=1,\cdots,J) \end{array}$$

The Lagrangran dual of (#) is $\sup\limits_{\substack{\omega^j \in \mathbb{R}^{m_j} \\ j=1,\cdots,J}} \theta(\omega^1,\cdots,\omega^J)$

where
$$\theta(\omega^1,\cdots,\omega^J) = \inf_{\substack{A^j x^j \in G_j \\ j=1,\cdots,J \\ \bar{x} \in \mathbb{R}^n}} \sum_{j=1}^{J} \left[f_j(A^j x^j) + (\omega^j)^T A^j (x^j - \bar{x}) \right]$$

$R(A^j) \cap G_j = G_j$

$$= \sum_{j=1}^{J} \inf_{A^j x^j \in G_j} \left[f_j(A^j x^j) + (\omega^j)^T A^j x^j \right] - \inf_{\bar{x} \in \mathbb{R}^n} \sum_{j=1}^{J} (\omega^j)^T A^j \bar{x}$$

$+$

$$= \begin{cases} \sum_{j=1}^{J} \inf_{x \in G_j} \left[f_j(x) + (\omega^j)^T x \right], & \text{if } \sum_{j=1}^{J} (A^j)^T (\omega^j) = 0 \\ -\infty \quad y \in G_j \setminus A^j x, & \text{otherwise.} \end{cases}$$

G

Hence, the Lagrangian dual can be written as an optimization problem
$$\begin{array}{ll} \sup & \sum_{j=1}^{J} \inf_{x \in G} \left[f_j(x) + (\omega^j)^T x \right] \\ \text{s.t.} & \sum_{j=1}^{J} (A^j)^T \omega^j = 0 \qquad j=1,\cdots,J \end{array}$$

(b) Let $f_1(Ax) = \|Ax-b\|_2^2$, $f_2(Ix) = \lambda \|x\|_2^2$.

Then $\|Ax-b\|_2^2 + \lambda \|x\|_2^2 = f_1(Ax) + f_2(Ix)$

Hence, the dual problem of the ridge regression problem is
$$\sup \ \inf_{y \in \mathbb{R}^m} \{ \|y-b\|_2^2 + \omega_1^T y \} + \inf_{y \in \mathbb{R}^m} \{ \lambda \|y\|_2^2 + \omega_2^T y \} \qquad (*)$$
$$\text{s.t. } A^T \omega_1 + I \omega_2 = 0$$

Let $\frac{\partial}{\partial y}\left(\|y-b\|_2^2 + w_1^T y\right) = 0$, we get $y = b - \frac{1}{2}w_1$

$\therefore \inf_{y\in\mathbb{R}^m}\{\|y-b\|_2^2 + w_1^T y\} = -\frac{1}{4}\|w_1\|_2^2 + b^T w_1$

Let $\frac{\partial}{\partial y}\left(\lambda\|y\|_2^2 + w_2^T y\right) = 0$, we get $y = -\frac{1}{2\lambda}w$

$\therefore \inf_{y\in\mathbb{R}^m}\left(\lambda\|y\|_2^2 + w_2^T y\right) = -\frac{1}{4\lambda}\|w_2\|_2^2$

Hence, $(*)$ is equivalent to

$$\sup_{w\in\mathbb{R}^m}\left\{-\frac{1}{4}\|w\|_2^2 + b^T w - \frac{1}{4\lambda}\|A^T w\|_2^2\right\} \triangleq \sup_{w\in\mathbb{R}^m}\{F(w)\}$$

Let $\frac{\partial}{\partial w}F(w) = 0$, we get

$$\left(\frac{1}{2\lambda}AA^T + \frac{1}{2}I\right)w = b$$

$$\Leftrightarrow \quad \left(\frac{1}{\lambda}AA^T + I\right)w = 2b$$

Apply spectral decomposition to AA^T. we have

$$I + \frac{1}{\lambda}AA^T = I + \frac{1}{\lambda}\Omega\Sigma\Omega^T = \Omega\left(I + \frac{1}{\lambda}\Sigma\right)\Omega^T$$

Since $AA^T \succeq 0$, then $\lambda_i \geq 0 \Rightarrow 1 + \frac{\lambda_i}{\lambda} > 0$

(where λ_i is the ith diagonal element of Σ).

$\therefore I + \frac{1}{\lambda}AA^T \succ 0$, then $I + \frac{1}{\lambda}AA^T$ is invertible

$\therefore w^* = 2\left(I + \frac{1}{\lambda}AA^T\right)^{-1}b$ ✓

ENGG 5501: Foundations of Optimization | 2016–17 First Term

Homework Set 1

Instructor: Anthony Man–Cho So

Due: September 30, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $P = \{x \in \mathbb{R}^n : a_i^T x \le b_i \text{ for } i = 1, \ldots, m\}$, where $a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$ are given. Recall that a ball with center $\bar{x} \in \mathbb{R}^n$ and radius $r > 0$ is defined as the set $B(\bar{x}, r) = \{x \in \mathbb{R}^n : \|x - y\|_2 \le r\}$. We are interested in finding a ball with the largest possible radius, subject to the condition that it is entirely contained within the set P (also known as the *largest inscribed ball* in P). Give a linear programming formulation of this problem.

Problem 2 (30pts). Let $S = \{x \in \mathbb{R}^n : x^T A x + b^T x + c \le 0\}$, where $A \in \mathcal{S}^n$, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$ are given.

(a) **(15pts).** Show that S is convex if $A \succeq \mathbf{0}$. Is the converse true? Explain.

(b) **(15pts).** Let $H = \{x \in \mathbb{R}^n : g^T x + h = 0\}$, where $g \in \mathbb{R}^n \backslash \{0\}$ and $h \in \mathbb{R}$. Show that $S \cap H$ is convex if $A + \lambda g g^T \succeq \mathbf{0}$ for some $\lambda \in \mathbb{R}$.

Problem 3 (50pts). Let S, T be closed convex sets in $\mathbb{R}^n$ such that $S \cap T \ne \emptyset$. A problem that arises frequently in optimization is that of finding a point $x \in S \cap T$. A natural algorithm is to start with an arbitrary $x_0 \in S$ and then alternately project onto S and T; i.e., compute the sequence

$$y_k = \Pi_T(x_k), \quad x_{k+1} = \Pi_S(y_k) \qquad \text{for } k = 0, 1, \ldots.$$

Clearly, we have $x_k \in S$ and $y_k \in T$ for $k = 0, 1, \ldots$. Our goal is to prove that the sequences $\{x_k\}_{k \ge 0}$ and $\{y_k\}_{k \ge 0}$ both converge to a point $x^* \in S \cap T$.

(a) **(20pts).** Let $\bar{x} \in S \cap T$ be arbitrary. Show that

$$\|y_k - \bar{x}\|_2^2 \le \|x_k - \bar{x}\|_2^2 - \|x_k - y_k\|_2^2$$

and

$$\|x_{k+1} - \bar{x}\|_2^2 \le \|y_k - \bar{x}\|_2^2 - \|x_{k+1} - y_k\|_2^2.$$

Hence, or otherwise, show that the sequences $\{x_k\}_{k \ge 0}$ and $\{y_k\}_{k \ge 0}$ are bounded.

(b) **(15pts).** Using the results in (a), show that the sequences $\{\|x_k - y_k\|_2\}_{k \ge 0}$ and $\{\|x_{k+1} - y_k\|_2\}_{k \ge 0}$ both converge to 0.

(c) **(15pts).** Let x^* be a limit point of the sequence $\{x_k\}_{k \ge 0}$, which exists because of the boundedness of $\{x_k\}_{k \ge 0}$. Using the result in (b), show that $x^* \in S \cap T$. Hence, or otherwise, show that $x_k \to x^*$ and $y_k \to x^*$. (*Hint: Note that x^* is defined as a limit point of $\{x_k\}_{k \ge 0}$, which means that there is a* subsequence *of $\{x_k\}_{k \ge 0}$ converging to x^*. Here, you are asked to show that in fact the* entire sequence *$\{x_k\}_{k \ge 0}$ converges to x^*.*)

ENGG 5501: Foundations of Optimization **2016–17 First Term**

Homework Set 1 Solution

Instructor: Anthony Man–Cho So October 6, 2016

Problem 1 (20pts). We begin by observing that the ball $B(y,r)$ with center y and radius r can be represented as

$$B(y,r) = \{x \in \mathbb{R}^n : \|x-y\|_2 \le r\} = \{y+u \in \mathbb{R}^n : \|u\|_2 \le r\}.$$

Now, consider a fixed $i \in \{1,\dots,n\}$. Observe that $B(y,r) \subset H_i = \{x \in \mathbb{R}^n : a_i^T x \le b_i\}$ iff $a_i^T(y+u) \le b_i$, where $\|u\|_2 \le r$. By the Cauchy–Schwarz inequality, we have

$$\sup_{u \in \mathbb{R}^n : \|u\|_2 \le r} \{a_i^T u\} = a_i^T \left(r \cdot \frac{a_i}{\|a_i\|_2} \right) = r\|a_i\|_2.$$

It follows that $B(y,r) \subset H_i$ iff

$$a_i^T y + r\|a_i\|_2 \le b_i, \tag{1}$$

which is a linear inequality in $y \in \mathbb{R}^n$ and $r \in \mathbb{R}_+$. In particular, we have $B(y,r) \subset P = \{x \in \mathbb{R}^n : a_i^T x \le b_i \text{ for } i = 1,\dots,m\}$ iff (1) holds for $i = 1,\dots,m$. Hence, the problem of finding the largest inscribed ball in P can be formulated as the following LP:

$$\begin{array}{ll} \text{maximize} & r \\ \text{subject to} & a_i^T y + r\|a_i\|_2 \le b_i \quad \text{for } i = 1,\dots,m, \\ & y \in \mathbb{R}^n,\ r \ge 0. \end{array}$$

Problem 2 (30pts).

(a) **(15pts).** Let $x_1, x_2 \in S$, and let $\alpha \in (0,1)$. Then, we have

$$\begin{aligned} x_1^T A x_1 + b^T x_1 + c &\le 0, & (2) \\ x_2^T A x_2 + b^T x_2 + c &\le 0. & (3) \end{aligned}$$

Now, we compute

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T(\alpha x_1 + (1-\alpha)x_2) + c \\ = \; & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + \alpha\left(b^T x_1 + c\right) + (1-\alpha)\left(b^T x_2 + c\right) \\ \le \; & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) - \alpha x_1^T A x_1 - (1-\alpha) x_2^T A x_2 & (4) \\ = \; & -\alpha(1-\alpha) x_1^T A x_1 - (1-\alpha)(1-(1-\alpha)) x_2^T A x_2 + 2\alpha(1-\alpha) x_1^T A x_2 \\ = \; & -\alpha(1-\alpha)\left(x_1^T A x_1 - 2x_1^T A x_2 + x_2^T A x_2\right) \\ = \; & -\alpha(1-\alpha) \cdot (x_1 - x_2)^T A (x_1 - x_2) \\ \le \; & 0, & (5) \end{aligned}$$

where (4) follows from the fact that $b^T x_i + c \le -x_i^T A x_i$ for $i = 1, 2$ (by (2) and (3)), and (5) follows from the assumption that $A \succeq 0$. This proves that S is convex if $A \succeq 0$.

Note that the converse of the claim need not be true. Indeed, let $n = 1$, and let $A = -1$, $b = c = 0$. Then, we have $S = \{x \in \mathbb{R} : -x^2 \le 0\} = \mathbb{R}$, which is trivially convex.

(b) **(15pts).** Let $x_1, x_2 \in S \cap H$, and let $\alpha \in (0,1)$. From the calculations in part (a), we have

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T (\alpha x_1 + (1-\alpha)x_2) + c \\ \le\; & -\alpha(1-\alpha)\cdot(x_1 - x_2)^T A (x_1 - x_2). \end{aligned} \tag{6}$$

Since $A + \gamma g g^T \succeq 0$, we have

$$0 \le (x_1 - x_2)^T \left(A + \gamma g g^T\right)(x_1 - x_2) = (x_1 - x_2)^T A (x_1 - x_2) + \gamma \left(g^T (x_1 - x_2)\right)^2.$$

It follows from (6) that

$$\begin{aligned} & (\alpha x_1 + (1-\alpha)x_2)^T A (\alpha x_1 + (1-\alpha)x_2) + b^T (\alpha x_1 + (1-\alpha)x_2) + c \\ \le\; & -\alpha(1-\alpha)\cdot(x_1 - x_2)^T A (x_1 - x_2) \\ \le\; & \alpha(1-\alpha)\cdot\gamma\left(g^T (x_1 - x_2)\right)^2 \\ =\; & 0, \end{aligned}$$

where the last equality follows from the fact that $g^T x_1 + h = g^T x_2 + h = 0$.

Problem 3 (50pts).

(a) **(20pts).** We compute

$$\begin{aligned} \|x_k - \bar{x}\|_2^2 &= \|x_k - y_k + y_k - \bar{x}\|_2^2 \\ &= \|x_k - y_k\|_2^2 + \|y_k - \bar{x}\|_2^2 + 2(x_k - y_k)^T (y_k - \bar{x}). \end{aligned}$$

Since $y_k = \Pi_T(x_k)$ and $\bar{x} \in S \cap T$, by Theorem 5 of Handout 2, we have $(x_k - y_k)^T (y_k - \bar{x}) \ge 0$. It follows that

$$\|x_k - \bar{x}\|_2^2 \ge \|x_k - y_k\|_2^2 + \|y_k - \bar{x}\|_2^2,$$

or equivalently,

$$\|y_k - \bar{x}\|_2^2 \le \|x_k - \bar{x}\|_2^2 - \|x_k - y_k\|_2^2, \tag{7}$$

as desired. By applying a similar argument to $\|y_k - \bar{x}\|_2^2$, we obtain

$$\|x_{k+1} - \bar{x}\|_2^2 \le \|y_k - \bar{x}\|_2^2 - \|x_{k+1} - y_k\|_2^2. \tag{8}$$

Now, observe that (7) and (8) imply

$$\|y_k - \bar{x}\|_2^2 \le \|x_k - \bar{x}\|_2^2 \le \|y_{k-1} - \bar{x}\|_2^2 \le \|x_{k-1} - \bar{x}\|_2^2 \quad \text{for } k = 1, 2, \ldots. \tag{9}$$

In particular, we have $\|x_k - \bar{x}\|_2^2 \le \|x_0 - \bar{x}\|_2^2$ and $\|y_k - \bar{x}\|_2^2 \le \|y_0 - \bar{x}\|_2^2$ for $k = 0, 1, \ldots$, which implies that the sequences $\{x_k\}_{k \ge 0}$ and $\{y_k\}_{k \ge 0}$ are bounded.

(b) **(15pts).** From (9), we see that the sequences $\{\|x_k - \bar{x}\|_2^2\}_{k\geq 0}$ and $\{\|y_k - \bar{x}\|_2^2\}_{k\geq 0}$ are monotonically decreasing and hence converge. Now, by (7) and (8), we have

$$\|x_k - y_k\|_2^2 \leq \|x_k - \bar{x}\|_2^2 - \|y_k - \bar{x}\|_2^2 \leq \|y_{k-1} - \bar{x}\|_2^2 - \|y_k - \bar{x}\|_2^2 \quad \text{for } k = 1, 2, \ldots.$$

Since the sequence $\{\|y_k - \bar{x}\|_2^2\}_{k\geq 0}$ converges, the rightmost expression in the above chain of inequalities converges to zero. It follows that the sequence $\{\|x_k - y_k\|_2\}_{k\geq 0}$ converges to zero. A similar argument shows that the sequence $\{\|x_{k+1} - y_k\|_2\}_{k\geq 0}$ converges to zero.

(c) **(15pts).** Since $x_k \in S$ for $k = 0, 1, \ldots$, we have $x^* \in S$ by the closedness of S. Moreover, since $\|x_k - y_k\|_2 \to 0$ and $y_k \in T$ for $k = 0, 1, \ldots$, we have $x^* \in T$ by the closedness of T. It follows that $x^* \in S \cap T$.

Now, by taking $\bar{x} = x^*$ in (b), we see that the sequence $\{\|x_k - x^*\|_2^2\}_{k\geq 0}$ is monotonically decreasing and hence converges. Since a subsequence of this sequence converges to zero (recall that x^* is a limit point of $\{x_k\}_{k\geq 0}$), it follows that the entire sequence converges to zero; i.e., $x_k \to x^*$. A similar argument shows that $y_k \to x^*$.

(iii) $\Rightarrow$ (ii): Just need to prove $\neg$(ii) $\Rightarrow$ $\neg$(iii) Using Gordan's theorem

If $\forall y.\ A^T y \le 0 \Rightarrow A^T \mathbb{1} \le 0. \Rightarrow b^T \mathbb{1} = x^T A^T \mathbb{1} \le 0 \quad (x \ge 0)$

$\therefore \sum_{i=1}^{n} x_i a_i^T \mathbb{1} = b^T \mathbb{1}$ ~~Suppose $a_i^T \mathbb{1}$ $a_k^T \mathbb{1}$~~

$\therefore \sum_{i=1}^{n} \max_i (a_i^T \mathbb{1}) \cdot x_i = \max_i (a_i^T \mathbb{1}) \sum_{i=1}^{n} x_i \ge b^T \mathbb{1}$

bounded.

ENGG 5501: Foundations of Optimization | **2016–17 First Term**

Homework Set 2

Instructor: Anthony Man–Cho So | Due: October 14, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (15pts). Let $f : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ be the function given by

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } \|(x_1, x_2)\|_2 < 1, \\ \in [0, +\infty] & \text{if } \|(x_1, x_2)\|_2 = 1, \\ +\infty & \text{if } \|(x_1, x_2)\|_2 > 1. \end{cases}$$

Show that f is convex. Hence, or otherwise, give an example of a convex function whose epigraph is not closed.

Problem 2 (20pts). Let $C \in \mathcal{S}^n_+$ be given. Show that the function $f : \mathcal{S}^n_{++} \to \mathbb{R}_+$ given by $f(X) = \mathrm{tr}(CX^{-1})$ is convex.

Problem 3 (25pts). Given a set $C \subset \mathbb{R}^n$, define the indicator function $i_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of C by

$$i_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Compute i_C^*, the conjugate of i_C, for the following sets. Show your calculations.

(a) **(15pts).** $C = \{x \in \mathbb{R}^n : a^T x \le b\}$, where $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ are given.

(b) **(10pts).** $C = \{x \in \mathbb{R}^n_+ : \|x\|_2 \le 1\}$.

Problem 4 (20pts). Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be given. Consider the following systems:

(I) $Ax = b$.

(II) $A^T y = 0,\ b^T y = 1$.

4(a) If $\begin{cases} Ax=b \\ A^Ty=0 \\ b^Ty=1 \end{cases} \Leftrightarrow \begin{bmatrix} A & 0 \\ 0 & A^T \\ 0^T & b^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 1 \end{bmatrix} \Leftarrow$

$\tilde{A} z =$

rank $\tilde{A} = 2\,\mathrm{rank}\,A <$ rank $(\tilde{A}$

$\|$

2rank

(a) **(5pts).** Show that (I) and (II) cannot both have solutions.

(b) **(15pts).** Suppose that (I) has no solution. Show, without using any theorems of alternatives, that (II) has a solution.

Problem 5 (20pts). Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, consider the polyhedron $P = \{x \in \mathbb{R}^n : Ax = b,\ x \ge 0\}$. Suppose that $P \ne \emptyset$. We say that P contains a *recession direction* $d \in \mathbb{R}^n \setminus \{0\}$ if for any $x_0 \in P$, we have $\{x \in \mathbb{R}^n : x = x_0 + \lambda d,\ \lambda \ge 0\} \subset P$. Show that the following statements are equivalent:

or: $b^T y = x^T A y = x^T 0 = 0$

(i) P contains a recession direction $d \in \mathbb{R}^n$.

(ii) There exists a vector $d \in \mathbb{R}^n$ satisfying

$$Ad = 0,\ d \ge 0,\ d \ne 0.$$

4(b) $Ax=b$ has no sol. $\Leftrightarrow$ rank $A <$ rank

$\Leftrightarrow \mathrm{rank}(A)+1 \le \mathrm{rank}(A, b$

(iii) P is unbounded.

Then, $A^Ty=0$, $b^Ty=1$

$\Leftrightarrow \begin{bmatrix} A^T \\ b^T \end{bmatrix} y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ (*)

$\mathrm{rank}(A)+1 \le \mathrm{rank}(A, b) = \mathrm{rank}\begin{bmatrix} A^T \\ b^T \end{bmatrix} = \mathrm{rank}(A, b) \le \mathrm{rank}\begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} = \mathrm{rank}($

$\therefore$ rank $\begin{bmatrix} A^T \\ b^T \end{bmatrix} =$ rank $\begin{bmatrix} A^T & 0 \\ b^T & 1 \end{bmatrix}$ $\therefore$ (*) has a sol.

ENGG 5501: Foundations of Optimization	2016–17 First Term
Homework Set 2 Solution	
Instructor: Anthony Man–Cho So	October 30, 2016

Problem 1 (15pts). Let $x, y \in \mathbb{R}^2$ and $\alpha \in (0,1)$ be arbitrary with $x \neq y$. Consider the following cases:

Case 1: $\|x\|_2 \leq 1$ and $\|y\|_2 \leq 1$

In this case, since $x \neq y$ and $\alpha \in (0,1)$, we have

$$\|\alpha x + (1-\alpha)y\|_2^2 = \alpha\|x\|_2^2 + (1-\alpha)\|y\|_2^2 - \alpha(1-\alpha)\|x-y\|_2^2 < 1.$$

It follows that $f(\alpha x + (1-\alpha)y) = 0 \leq \alpha f(x) + (1-\alpha)f(y)$.

Case 2: $\|x\|_2 > 1$

In this case, since $f(x) = +\infty$, we trivially have $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$.

Since x and y are interchangeable, we conclude from the above that f is convex, as desired.

Next, let us take $f : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ to be the function given in the problem and specify

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } \|(x_1,x_2)\|_2 = 1 \text{ and } x_2 < 0, \\ +\infty & \text{if } \|(x_1,x_2)\|_2 = 1 \text{ and } x_2 \geq 0. \end{cases}$$

We claim that this furnishes an example of a convex function whose epigraph is not closed. Indeed, observe that

$$\begin{aligned} \text{epi}(f) &= \{(x,t) \in \mathbb{R}^2 \times \mathbb{R} : f(x) \leq t\} \\ &= \{(x,t) \in \mathbb{R}^2 \times \mathbb{R} : \|x\|_2 < 1,\ t \geq 0\} \cup \{(x,t) \in \mathbb{R}^2 \times \mathbb{R} : \|x\|_2 = 1,\ x_2 < 0,\ t \geq 0\}. \end{aligned}$$

In particular, we see that the sequence $\{(0, 1-1/k, 0)\}_{k \geq 1}$ belongs to epi(f), but its limit $(0,1,0)$ does not belong to epi(f).

Problem 2 (20pts). Let $X_0 \in \mathcal{S}^n_{++}$ and $H \in \mathcal{S}^n$ be fixed. For any $t \in \mathbb{R}$ such that $X_0 + tH \in \mathcal{S}^n_{++}$, we have

$$g(t) \equiv \text{tr}\left(C(X_0+tH)^{-1}\right) = \text{tr}\left(CX_0^{-1/2}(I + tX_0^{-1/2}HX_0^{-1/2})^{-1}X_0^{-1/2}\right).$$

Since $X_0^{-1/2}HX_0^{-1/2} \in \mathcal{S}^n$, there exist an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\Lambda \in \mathcal{S}^n$ such that $X_0^{-1/2}HX_0^{-1/2} = U\Lambda U^T$. Hence, we have

$$g(t) = \text{tr}\left(CX_0^{-1/2}U(I+t\Lambda)^{-1}U^TX_0^{-1/2}\right) = \text{tr}\left(U^TX_0^{-1/2}CX_0^{-1/2}U(I+t\Lambda)^{-1}\right).$$

Now, let $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Since $X_0 + tH \in \mathcal{S}^n_{++}$, we have $1 + t\lambda_i > 0$ for $i = 1, \ldots, n$. Thus, we may write $(I+t\Lambda)^{-1} = \text{diag}((1+t\lambda_1)^{-1}, \ldots, (1+t\lambda_n)^{-1})$. It then follows that

$$g(t) = \sum_{i=1}^n \left(U^TX_0^{-1/2}CX_0^{-1/2}U\right)_{ii} \cdot \frac{1}{1+t\lambda_i}.$$

It is easy to verify that for $i = 1, \ldots, n$, the function $t \mapsto (1 + t\lambda_i)^{-1}$ is convex over the region $\{t \in \mathbb{R} : t > \lambda_i^{-1}$ for $i = 1, \ldots, n\}$. Moreover, since $U^T X_0^{-1/2} C X_0^{-1/2} U \in \mathcal{S}_+^n$, we have $\left(U^T X_0^{-1/2} C X_0^{-1/2} U\right)_{ii} \geq 0$ for $i = 1, \ldots, n$. It follows that g is a non-negative linear combination of convex functions, which implies that g is convex. This in turn implies that f is convex, as desired.

Problem 3 (25pts). We begin by observing that for any $C \subset \mathbb{R}^n$,

$$i_C^*(y) = \sup_{x \in \mathbb{R}^n} \left\{ y^T x - i_C(x) \right\} = \sup_{x \in C} y^T x.$$

(a) **(15pts).** For $C = \left\{ x \in \mathbb{R}^n : a^T x \leq b \right\}$, where $a \in \mathbb{R}^n \backslash \{0\}$ and $b \in \mathbb{R}$ are given, suppose that $y \notin \operatorname{span}(a)$; i.e., $y \neq \alpha a$ for any $\alpha \in \mathbb{R}$. We claim that there exists an $u \in \mathbb{R}^n$ such that $a^T u = 0$ and $y^T u > 0$. Indeed, let u be the projection of y onto the subspace $H = \{x \in \mathbb{R}^n : a^T x = 0\}$; i.e., $u = (I - aa^T / \|a\|_2^2) y$. It is easy to verify that $a^T u = 0$. Moreover, since $y \notin \operatorname{span}(a)$, by the Cauchy–Schwarz inequality, we have

$$y^T u = \|y\|_2^2 - \frac{(a^T y)^2}{\|a\|_2^2} > 0.$$

This establishes the claim. Now, let $z \in C$ be arbitrary. Note that $z + \beta u \in C$ for all $\beta \in \mathbb{R}$. Moreover, we have

$$y^T (z + \beta u) = y^T z + \beta y^T u \to +\infty \quad \text{as } \beta \to +\infty.$$

Hence, $i_C^*(y) = +\infty$ whenever $y \notin \operatorname{span}(a)$.

On the other hand, suppose that $y \in \operatorname{span}(a)$; i.e., $y = \alpha a$ for some $\alpha \in \mathbb{R}$. If $\alpha \geq 0$, then $i_C^*(y) = \sup_{x \in C} y^T x = \alpha b$. If $\alpha < 0$, then by noting that $-\beta a \in C$ for all sufficiently large $\beta > 0$, we have $i_C^*(y) \geq -\alpha \beta \|a\|_2^2 \to +\infty$ as $\beta \to +\infty$. To summarize, we obtain

$$i_C^*(y) = \begin{cases} \dfrac{a^T y}{\|a\|_2^2} b & \text{if } y \in \{\alpha a : \alpha \geq 0\}, \\ +\infty & \text{otherwise.} \end{cases}$$

(b) **(10pts).** For $C = \left\{ x \in \mathbb{R}_+^n : \|x\|_2 \leq 1 \right\}$, observe that for any $x \in C$ and $y \in \mathbb{R}^n$,

$$y^T x = \sum_{i: y_i \geq 0} x_i y_i + \sum_{i: y_i < 0} x_i y_i \leq \sum_{i: y_i \geq 0} x_i y_i = y^T x' = y_+^T x,$$

where $x' \in C$ is given by

$$x_j' = \begin{cases} x_j & \text{if } y_j \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

and $y_+ \in \mathbb{R}^n$ is given by $(y_+)_j = \max\{y_j, 0\}$, for $j = 1, \ldots, n$. It follows from the Cauchy–Schwarz inequality that $i_C^*(y) = \|y_+\|_2$.

Problem 4 (20pts).

(a) (5pts). Suppose that $\bar{x} \in \mathbb{R}^n$ is feasible for (I) and $\bar{y} \in \mathbb{R}^m$ is feasible for (II). Then, we have

$$0 = \bar{x}^T A^T \bar{y} = b^T \bar{y} = 1,$$

which is a contradiction.

(b) (15pts). Let $a_i \in \mathbb{R}^m$ be the i-th column of A, where $i = 1, \ldots, n$. Without loss of generality, let $\{a_1, \ldots, a_r\}$ be a basis of $\text{span}\{a_1, \ldots, a_n\}$. Since (I) has no solution, we have $b \notin \text{span}\{a_1, \ldots, a_n\}$, which implies that the collection $\{a_1, \ldots, a_r, b\}$ is linearly independent. Thus, the system

$$a_i^T y = 0 \quad \text{for } i = 1, \ldots, r, \quad b^T y = 1$$

has a solution $\bar{y} \in \mathbb{R}^m$. Moreover, since a_i is a linear combination of $a_1, \ldots, a_r$ for $i = r+1, \ldots, n$, we see that $a_i^T \bar{y} = 0$ for $i = r+1, \ldots, n$ as well. It follows that (II) has a solution, as desired.

Problem 5 (20pts). We show that (i)$\Rightarrow$(iii)$\Rightarrow$(ii)$\Rightarrow$(i).

Step 1: (i)$\Rightarrow$(iii). Suppose that P contains a recession direction $d \in \mathbb{R}^n$. Then, for any $x_0 \in P$, we have $x(\lambda) = x_0 + \lambda d \in P$ for all $\lambda \geq 0$. Let $i \in \{1, \ldots, n\}$ be an index such that $d_i \neq 0$. Then, we have

$$\|x(\lambda)\|_2 \geq |(x_0)_i + \lambda d_i| \to \infty \quad \text{as } \lambda \to \infty.$$

It follows that P is unbounded.

Step 2: (iii)$\Rightarrow$(ii). We establish the contrapositive. Suppose that the system

$$Ad = 0, \ d \geq 0, \ d \neq 0$$

has no solution. Then, by Gordan's theorem (Corollary 2 of Handout 3), there exists a $v \in \mathbb{R}^m$ such that $A^T v \geq e$. Now, let $x \in P$ be arbitrary. Since $Ax = b$ and $x \geq 0$, we have

$$b^T v = x^T A^T v \geq e^T x = \|x\|_1.$$

In particular, we see that P is bounded.

Step 3: (ii)$\Rightarrow$(i). Suppose that $d \in \mathbb{R}^n$ satisfies

$$Ad = 0, \ d \geq 0, \ d \neq 0.$$

Let $x_0 \in P$ and $\lambda \geq 0$ be arbitrary. Consider the point $x(\lambda) = x_0 + \lambda d \in \mathbb{R}^n$. Since $Ax_0 = b$ and $x_0 \geq 0$, we have

$$Ax(\lambda) = Ax_0 + \lambda Ad = b, \ x(\lambda) = x_0 + \lambda d \geq 0;$$

i.e., $x(\lambda) \in P$. It follows that P contains the recession direction d.

ENGG 5501: Foundations of Optimization | **2016–17 First Term**

Homework Set 3

Instructor: Anthony Man–Cho So | Due: October 31, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (25pts). Consider a *production game* defined as follows. Let $\mathcal{N} = \{1,\ldots,n\}$ be the set of players, each of whom is given a vector $b^i = (b^i_1,\ldots,b^i_m)$ $(i = 1,\ldots,n)$ of resources. These resources can be used to produce goods, which in turn can be sold at a given market price. Specifically, we assume the following production model: For player $i \in \mathcal{N}$, a unit of the j–th good $(j = 1,\ldots,p)$ requires a^i_{kj} units of the k–th resource $(k = 1,\ldots,m)$ to produce. Furthermore, the j–th good can be sold at a price c_j, where $j = 1,\ldots,p$.

Now, let $S \subset \mathcal{N}$ be a coalition of players. Such a coalition will possess

$$b_k(S) = \sum_{i\in S} b^i_k$$

units of the k–th resource, where $k = 1,\ldots,m$. Using all of their resources, the coalition S can produce any vector $x = (x_1,\ldots,x_p) \in \mathbb{R}^p_+$ of goods that satisfies $A(S)x \le b(S)$, where

$$A(S)_{kj} = \min_{i\in S}\left\{a^i_{kj}\right\} \quad \text{for } k = 1,\ldots,m;\ j = 1,\ldots,p \quad \text{and} \quad b(S) = (b_1(S),\ldots,b_m(S)).$$

Naturally, a coalition S would like to maximize its revenue. The optimization problem it faces can be formulated as the following LP:

$$\begin{array}{rcll} v(S) & = & \text{maximize} & c^T x \\ & & \text{subject to} & A(S)x \le b(S), \\ & & & x \ge 0. \end{array} \tag{1}$$

The function v will be the value function of this game. Recall that an allocation vector $z = (z_1,\ldots,z_n) \in \mathbb{R}^n$ is in the core iff $\sum_{i\in\mathcal{N}} z_i = v(\mathcal{N})$ and $\sum_{i\in S} z_i \ge v(S)$ for all $S \subset \mathcal{N}$.

(a) **(10pts).** Consider the LP faced by the grand coalition $\mathcal{N}$. Write down its dual.

(b) **(15pts).** Suppose that the dual LP given in (a) is feasible. Let y^* be one of its optimal solutions. Show that the allocation vector

$$z^* = \left(\left(b^1\right)^T y^*, \left(b^2\right)^T y^*, \ldots, \left(b^n\right)^T y^*\right) \in \mathbb{R}^n$$

belongs to the core.

Problem 2 (15pts). Show that for $n \ge 2$, the Lorentz cone

$$\mathcal{Q}^{n+1} = \{(t,x) \in \mathbb{R} \times \mathbb{R}^n : t \ge \|x\|_2\}$$

is non–polyhedral.

Problem 3 (45pts). Consider the set $\mathcal{C}^n = \{X \in \mathcal{S}^n : v^T X v \ge 0 \text{ for all } v \ge 0\}$.

(a) **(15pts).** Show that $\mathcal{C}^n$ is a non–empty closed convex cone.

(b) **(15pts).** Show that $\text{conv}(\{vv^T : v \geq 0\})$ is a non–empty closed convex set.

(c) **(15pts).** Let

$$(\mathcal{C}^n)^* = \{Y \in \mathcal{S}^n : X \bullet Y \geq 0 \text{ for all } X \in \mathcal{C}^n\}$$

be the dual cone of $\mathcal{C}^n$. Show that $(\mathcal{C}^n)^* = \text{conv}(\{vv^T : v \geq 0\})$.

Problem 4 (15pts). Let $Q \in \mathcal{S}_+^n$ be given. Give an equivalent SOCP formulation of the follo problem:

$$\begin{array}{ll} \inf & t \\ \text{subject to} & x^T Q x \leq t. \end{array}$$

Justify your answer.

ENGG 5501: Foundations of Optimization 2016–17 First Term

Homework Set 3 Solution

Instructor: Anthony Man–Cho So

November 11, 2016

Problem 1 (25pts).

(a) **(10pts).** The dual of the LP that is faced by the grand coalition is given by

$$\begin{array}{rl} \text{minimize} & b(\mathcal{N})^T y \\ \text{subject to} & A(\mathcal{N})^T y \ge c, \\ & y \ge \mathbf{0}. \end{array}$$

(b) **(15pts).** By the LP Strong Duality Theorem, we have

$$\sum_{i \in \mathcal{N}} z_i^* = \left(\sum_{i \in \mathcal{N}} b^i \right)^T y^* = b(\mathcal{N})^T y^* = v(\mathcal{N}).$$

Now, let $S \subset \mathcal{N}$ be an arbitrary coalition. By definition, $A(S)$ satisfies the following component–wise inequality:

$$A(S) \ge A(\mathcal{N}).$$

Since y^* satisfies $y \ge \mathbf{0}$ and $A(\mathcal{N})^T y^* \ge c$, we conclude that

$$A(S)^T y^* \ge A(\mathcal{N})^T y^* \ge c.$$

In other words, y^* is feasible for the dual of the LP that is faced by S. Hence, by the LP Weak Duality Theorem, we have

$$\sum_{i \in S} z_i^* = b(S)^T y^* \ge v(S),$$

as desired.

Problem 2 (15pts). Observe that if P is a polyhedron and H is a hyperplane, then $P \cap H$ is also a polyhedron, because $P \cap H$ can be expressed as a finite intersection of halfspaces. Now, consider the hyperplane $H = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t = 1\}$, where $n \ge 2$. Clearly, we have

$$Q^{n+1} \cap H = \{(1, x) \in \mathbb{R} \times \mathbb{R}^n : 1 \ge \|x\|_2\},$$

which shows that $Q^{n+1} \cap H$ is an n–dimensional unit ball lying on the plane H. In particular, it is straightforward to verify that every point $(1, x) \in Q^{n+1} \cap H$ satisfying $\|x\|_2 = 1$ is an extreme point of $Q^{n+1} \cap H$; see Example 4 of Handout 2. This implies that $Q^{n+1} \cap H$ is not polyhedral, and hence Q^{n+1} is not polyhedral.

Problem 3 (45pts).

(a) **(10pts).** We verify the four required properties:

(1) (Non-emptiness) This is clear, since $0 \in \mathcal{C}^n$.

(2) (Closedness) Let $\{X^k\}_{k\geq 1}$ be a sequence in $\mathcal{C}^n$ such that $X^k \to X$. Then, for any $v \geq 0$, we have $v^T X^k v \geq 0$ for all $k \geq 1$, which, by the continuity of the function $Y \mapsto v^T Y v$, implies that $0 \leq \lim_{k\to\infty} v^T X^k v = v^T X v$. Since this holds for an arbitrary $v \geq 0$, we conclude that $X \in \mathcal{C}^n$, as desired.

(3) (Convexity) Let $X_1, X_2 \in \mathcal{C}^n$ be arbitrary. Then, for any $v \geq 0$ and $\alpha \in (0,1)$, we have

$$v^T(\alpha X_1 + (1-\alpha)X_2)v = \alpha \cdot v^T X_1 v + (1-\alpha) \cdot v^T X_2 v \geq 0,$$

which implies that $\alpha X_1 + (1-\alpha)X_2 \in \mathcal{C}^n$ for all $\alpha \in (0,1)$, as desired.

(4) (Conic Set) Let $X \in \mathcal{C}^n$ be arbitrary. Then, for any $\alpha > 0$ and $v \geq 0$, we have $v^T(\alpha X)v = \alpha \cdot v^T X v \geq 0$.

(b) **(15pts).** Let $\mathcal{U} = \operatorname{conv}(\{vv^T : v \geq 0\})$. It is clear that $0 \in \mathcal{U}$. Moreover, by definition, $\mathcal{U}$ is convex. Thus, it remains to show that $\mathcal{U}$ is closed. Towards that end, let $\{X^k\}_k$ be a sequence in $\mathcal{U}$ such that $X^k \to X$. By Carathéodory's theorem, for each $k \geq 1$, we can find $N = n(n+1)/2 + 1$ vectors $v_1^k, v_2^k, \ldots, v_N^k \geq 0$ such that $X^k = \sum_{j=1}^N v_j^k (v_j^k)^T = A^k (A^k)^T$. Here, A^k is an $n \times N$ matrix whose j–th column is v_j^k. Now, let $(a_i^k)^T \in \mathbb{R}^N$ be the i–th row of A^k. Observe that $a_i^k \geq 0$ and

$$X_{ii} = \lim_{k\to\infty} X_{ii}^k = \lim_{k\to\infty} \|a_i^k\|_2^2 \qquad \text{for } i = 1, \ldots, n. \tag{1}$$

This implies that for each $i = 1, \ldots, n$, the sequence $\{a_i^k\}_k$ is bounded. By considering subsequences if necessary, we may assume that for each $i = 1, \ldots, n$, the sequence $\{a_i^k\}_k$ has a limit, say, $a_i \geq 0$. It follows from (1) that $X_{ii} = \|a_i\|_2^2$. In a similar fashion, we have

$$X_{ij} = \lim_{k\to\infty} (a_i^k)^T a_j^k = a_i^T a_j.$$

Hence, we can write $X = AA^T$, where the i–th row of A is a_i^T. This implies that $X \in \mathcal{U}$, as desired.

(c) **(15pts).** Consider a fixed $Y \in \mathcal{U} = \operatorname{conv}(\{vv^T : v \geq 0\})$. By Carathéodory's theorem, there exist $N \leq n(n+1)/2 + 1$ vectors $v_1, \ldots, v_N \geq 0$ such that $Y = \sum_{i=1}^N v_i v_i^T$. Now, for any $X \in \mathcal{C}^n$, we compute

$$X \bullet Y = X \bullet \left(\sum_{i=1}^N v_i v_i^T \right) = \sum_{i=1}^N v_i^T X v_i \geq 0,$$

where the last inequality follows from the fact that $X \in \mathcal{C}^n$ and $v_1, \ldots, v_N \geq 0$. This implies that $Y \in (\mathcal{C}^n)^*$.

Conversely, suppose that $Y \notin \mathcal{U}$. By the result in (b), $\mathcal{U}$ is a non-empty closed convex set. Thus, by the separation theorem, there exists a matrix $W \in \mathcal{S}^n$ such that $W \bullet Y < W \bullet Z$ for all $Z \in \mathcal{U}$. In particular, we have $W \bullet Y < 0$, since $0 \in \mathcal{U}$.

Now, we claim that $W \in \mathcal{C}^n$. Suppose to the contrary that this is not the case. Then, there exists a vector $u \geq 0$ such that $u^T W u < 0$. Since $\alpha \cdot uu^T \in \mathcal{U}$ for all $\alpha > 0$, we see that $W \bullet Y < W \bullet (\alpha \cdot uu^T) = \alpha \cdot u^T W u$ for all $\alpha > 0$, which is impossible because $W \bullet Y$ is a fixed negative number. It follows that $W \in \mathcal{C}^n$ as claimed.

To complete the proof, it suffices to observe that $W \in \mathcal{C}^n$ and $W \bullet Y < 0$ imply $Y \notin (\mathcal{C}^n)^*$.

REMARK: The sets $\mathcal{C}^n$ and $(\mathcal{C}^n)^*$ are known as the *copositive cone* and *completely positive cone*, respectively. They play a fundamental role in *copositive optimization*, an area that has received some attention in recent years. We refer the interested reader to [2, 1] for a survey of the field.

?**Problem 4 (15pts).** Since $Q \in \mathcal{S}^n_{++}$, there exists a $Q^{1/2} \in \mathcal{S}^n_+$ such that $Q = Q^{1/2}Q^{1/2}$. Observe that

$$x^TQx \le t \iff \left(t - \frac{1}{4}\right)^2 + \|Q^{1/2}x\|_2^2 \le \left(t + \frac{1}{4}\right)^2$$

$$\iff \left\| \begin{bmatrix} 1 & 0 \\ 0 & Q^{1/2} \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} - \begin{bmatrix} 1/4 \\ 0 \end{bmatrix} \right\|_2 \le t + \frac{1}{4}.$$

It follows that

$$x^TQx \le t \iff \left(t + \frac{1}{4}, \begin{bmatrix} 1 & 0 \\ 0 & Q^{1/2} \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} - \begin{bmatrix} 1/4 \\ 0 \end{bmatrix}\right) \in \mathcal{Q}^{n+2}.$$

References

[1] I. M. Bomze. Copositive Optimization — Recent Developments and Applications. *European Journal on Operational Research*, 216(3):509–520, 2012.

[2] M. Dür. Copositive Programming — A Survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20. Springer–Verlag, Berlin/Heidelberg, 2010.

ENGG 5501: Foundations of Optimization 2016–17 First Term

Homework Set 4

Instructor: Anthony Man–Cho So

Due: November 21, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (25pts). Let $A_1, \ldots, A_m \in \mathcal{S}_+^n$, $\alpha_1, \ldots, \alpha_m > 0$, and $\beta_1, \ldots, \beta_m > 0$ be given. Consider the following SDP:

$$\begin{array}{lll} \inf & \displaystyle\sum_{i=1}^{m} \operatorname{tr}(Z_i) & \\ \text{subject to} & \operatorname{tr}\left[A_i\left(\alpha_i Z_i - \displaystyle\sum_{j \neq i} Z_j\right)\right] \geq \beta_i \quad \text{for } i = 1, \ldots, m, & \\ & Z_1, \ldots, Z_m \in \mathcal{S}_+^n. & \end{array} \tag{Q}$$

(a) **(15pts).** Write down the dual of (Q).

(b) **(10pts).** Using the result in (a), show that the dual is strictly feasible.

Problem 2 (45pts).

(a) **(10pts).** Let $u \in \mathbb{R}^n$ and $U \in \mathcal{S}^n$ be given. Define

$$W = \begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix}.$$

Show that $\operatorname{rank}(W) = 1$ if and only if $W = (1, u)(1, u)^T$.

Let $\mathcal{CP}_n = \operatorname{conv}\left(\left\{vv^T : v \in \mathbb{R}_+^n\right\}\right)$. The results in Homework 3, Problem 3 imply that $\mathcal{CP}_n$ is a closed pointed cone with non–empty interior. Now, given $Q \in \mathcal{S}^n$ and $c \in \mathbb{R}^n$, consider the following problem:

$$\begin{array}{lll} \text{minimize} & \dfrac{1}{2}x^T Q x + c^T x & \\ \text{subject to} & x_i \in \{0, 1\} & \text{for } i = 1, \ldots, n. \end{array} \tag{1}$$

Our goal now is to show that the discrete optimization problem (1) can be reformulated as a linear optimization problem over the pointed cone $\mathcal{CP}_{2n+1}$. In other words, problem (1) has a convex reformulation![1]

[1]Since problem (1) is NP–hard in general, we see that not all convex optimization problems are easy to solve.

(b) **(5pts).** Using the result in (a), show that problem (1) is equivalent to

$$
\begin{array}{lll}
\text{minimize} & \frac{1}{2}Q \bullet X + c^T x & \\
\text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & \text{(I)} \\
& x_i + s_i = 1 \qquad \text{for } i = 1, \dots, n, & \text{(II)} \\
& X_{ii} + 2Y_{ii} + S_{ii} = 1 \qquad \text{for } i = 1, \dots, n, & \text{(III)} \\
& x_i = X_{ii},\ s_i = S_{ii} \qquad \text{for } i = 1, \dots, n, & \text{(IV)} \\
& \text{rank}(Z) = 1. &
\end{array}
$$

(2)

(c) **(20pts).** Let

$$
\begin{aligned}
S_1 &= \text{conv}\left(\{(1,x,s)(1,x,s)^T : x_i, s_i \in \{0,1\},\ x_i + s_i = 1 \text{ for } i = 1,\dots,n\}\right), \\
S_2 &= \left\{Z \in \mathcal{S}^{2n+1} : Z \text{ satisfies (2–I) — (2–IV)}\right\}.
\end{aligned}
$$

Show that $S_1 = S_2$.

(d) **(10pts).** By relaxing the rank constraint in (2), we obtain a convex relaxation of problem (1) which is a CLP. Using the result in (c), show that this convex relaxation is in fact equivalent to problem (1); i.e., (i) the optimal values of both problems are equal, and (ii) if

$$
Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}
$$

is an optimal solution to the convex relaxation, then x^* is in the convex hull of the optimal solutions to problem (1).

Problem 3 (30pts). Consider the ℓ_1–regularized ℓ_2–regression problem

$$
\min_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} \left\{ \|b - Ax - te\|_2 + \lambda \|x\|_1 \right\}, \tag{3}
$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ are given, and $e \in \mathbb{R}^m$ is the vector of all–ones. A common interpretation of the ℓ_1–regularizer $x \mapsto \|x\|_1$, which is based on heuristic arguments, is that it promotes sparsity in the optimal solution to (3). In this problem, we will show in a rigorous manner that problem (3) is actually equivalent to a robust ℓ_2–regression problem.

To begin, consider the following robust optimization problem:

$$
\min_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} \ \max_{\Delta A \in \mathcal{U}_\lambda} \|b - (A + \Delta A)x - te\|_2. \tag{4}
$$

Here, the uncertainty set $\mathcal{U}_\lambda$ is defined as

$$
\mathcal{U}_\lambda = \left\{X \in \mathbb{R}^{m\times n} : \|X\|_{1,2} \le \lambda\right\},
$$

where $\|X\|_{1,2} = \max_{\|v\|_1 = 1} \|Xv\|_2$. Note that $\|\cdot\|_{1,2}$ defines a matrix norm (see Handout B).

(a) **(10pts).** Show that for any $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ and $\Delta A \in \mathcal{U}_\lambda$,

$$\|b - (A + \Delta A)x - te\|_2 \leq \|b - Ax - te\|_2 + \lambda\|x\|_1.$$

(b) **(20pts).** Show that for any $(x, t) \in \mathbb{R}^n \times \mathbb{R}$, there exists a $\Delta A^* \in \mathcal{U}_\lambda$ such that the inequality in (a) holds as equality; i.e.,

$$\|b - (A + \Delta A^*)x - te\|_2 = \|b - Ax - te\|_2 + \lambda\|x\|_1.$$

Hence, conclude that problems (3) and (4) are equivalent.

ENGG 5501: Foundations of Optimization | 2016–17 First Term

Homework Set 4 Solution

Instructor: Anthony Man–Cho So | November 29, 2016

Problem 1 (25pts).

(a) **(15pts).** Observe that (Q) is equivalent to

$$\begin{array}{lll} \inf & D \bullet Z & \\ \text{subject to} & H_i \bullet Z = \beta_i & \text{for } i = 1, \ldots, m, \\ & Z \in \mathcal{S}_+^{m(n+1)}, & \end{array} \tag{1}$$

where

$$\begin{aligned} D &= \mathrm{BlkDiag}(I, \ldots, I, \mathbf{0}) \in \mathcal{S}^{m(n+1)}, \\ H_i &= \mathrm{BlkDiag}(-A_i, \ldots, -A_i, \underbrace{\alpha_i A_i}_{i^{th}}, -A_i, \ldots, -A_i, 0, \ldots, 0, -1, 0, \ldots, 0) \in \mathcal{S}^{m(n+1)}, \end{aligned}$$

and $\mathrm{BlkDiag}(Q_1, \ldots, Q_l)$ denotes the block diagonal matrix whose i–th diagonal block is Q_i, for $i = 1, \ldots, l$. Indeed, since $D, H_1, \ldots, H_m$ are block diagonal, every feasible solution $\bar{Z}$ to (1) gives rise to a block diagonal feasible solution $\bar{Z}' = \mathrm{BlkDiag}(\bar{Z}_1', \ldots, \bar{Z}_m', \bar{s}_1', \ldots, \bar{s}_m')$ to (1) whose objective value is equal to that of $\bar{Z}$. Now, the dual of (1) is given by

$$\begin{array}{ll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i \\ \text{subject to} & D - \displaystyle\sum_{i=1}^m y_i H_i \in \mathcal{S}_+^{m(n+1)}, \end{array}$$

which, using the structure of $D, H_1, \ldots, H_m$, is equivalent to

$$\begin{array}{lll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i & \\ \text{subject to} & I - \alpha_i y_i A_i + \displaystyle\sum_{j \neq i} y_j A_j \in \mathcal{S}_+^n & \text{for } i = 1, \ldots, m, \\ & y \geq \mathbf{0}. & \end{array} \tag{2}$$

(b) **(10pts).** Since $A_1, \ldots, A_m \in \mathcal{S}_+^n$ and $y \geq \mathbf{0}$, we have

$$I + \sum_{j \neq i} y_j A_j \succeq I$$

for $i = 1, \ldots, m$. Hence, to show that (2) is strictly feasible, it suffices to choose $y_i > 0$ such that $I \succ \alpha_i y_i A_i$ for $i = 1, \ldots, m$. Since $\alpha_i > 0$ and we may assume that $A_i \neq \mathbf{0}$, we have $\alpha_i \lambda_{\max}(A_i) > 0$. Hence, any $y \in \mathbb{R}^m$ satisfying

$$0 < y_i < \frac{1}{\alpha_i \lambda_{\max}(A_i)} \quad \text{for } i = 1, \ldots, m$$

is a strictly feasible solution to (2).

$\boxed{1 - \alpha_i y_i \lambda_{max}} > 0$

$\boxed{I - \alpha_1 y_1 A_1} \succ 0$

$0 < y_i < \dfrac{1}{\alpha_i \lambda_{max}}$

$\cancel{Q(I - \Sigma}$

$Q\,(I - \alpha_i y_i \Sigma)\,Q^T \succ 0$

Problem 2 (45pts).

(a) **(10pts).** Clearly, if $W = (1,u)(1,u)^T$, then $\text{rank}(W) = 1$. Conversely, suppose that $\text{rank}(W) = 1$. Since $W \in \mathcal{S}^{n+1}$, it admits a spectral decomposition $W = \lambda \bar{v}\bar{v}^T$, where $\bar{v} = (v_0, v) \in \mathbb{R}^{n+1}$ and $\lambda \neq 0$. Thus, we have

$$\begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix} = \begin{bmatrix} \lambda v_0^2 & \lambda v_0 v^T \\ \lambda v_0 v & \lambda v v^T \end{bmatrix},$$

which gives $\lambda v_0^2 = 1$ and $\lambda v_0 v = u$. In particular, we see that $\lambda > 0$ and $v_0 \neq 0$. This implies that

$$U = \lambda v v^T = \lambda \left(\frac{u}{\lambda v_0}\right)\left(\frac{u}{\lambda v_0}\right)^T = uu^T,$$

from which we obtain $W = (1,u)(1,u)^T$.

(b) **(5pts).** The given problem is equivalent to

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx & \\ \text{subject to} & x_i + s_i = 1 & \text{for } i = 1,\ldots,n, \\ & (x_i + s_i)^2 = 1 & \text{for } i = 1,\ldots,n, \\ & x_i = x_i^2,\ s_i = s_i^2 & \text{for } i = 1,\ldots,n. \end{array} \qquad (3)$$

Now, set $Z = (1,x,s)(1,x,s)^T \in \mathcal{S}_+^{2n+1}$. Since $x, s \geq 0$, we have $Z \in \mathcal{CP}_{2n+1}$. Moreover, by the result in (b), we have

$$Z = (1,x,s)(1,x,s)^T \iff Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix},\ \text{rank}(Z) = 1.$$

Since $X = xx^T$, $Y = xs^T$, and $S = ss^T$, we conclude that problem (3) is equivalent to

$$\begin{array}{llll} \text{minimize} & \frac{1}{2}Q \bullet X + c^Tx & & \\ \text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & & \text{(I)} \\ & x_i + s_i = 1 & \text{for } i = 1,\ldots,n, & \text{(II)} \\ & X_{ii} + 2Y_{ii} + S_{ii} = 1 & \text{for } i = 1,\ldots,n, & \text{(III)} \\ & x_i = X_{ii},\ s_i = S_{ii} & \text{for } i = 1,\ldots,n, & \text{(IV)} \\ & \text{rank}(Z) = 1, & & \end{array} \qquad (4)$$

as desired.

(c) **(20pts).** Let $x, s \in \{0,1\}^n$ be such that $x_i + s_i = 1$ for $i = 1,\ldots,n$. Then, we have $(1,x,s)(1,x,s)^T \in S_2$. Since S_2 is convex, we see that $S_1 \subset S_2$.

Conversely, let $Z \in S_2$ be given. By (4–I), there exist vectors $(\alpha_k, u^k, v^k) \in \mathbb{R}_+^{2n+1}$, where $k = 1, \ldots, K$, such that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$. Upon equating the entries of both sides, we have

$$\sum_{k=1}^K \alpha_k^2 = 1, \quad x = \sum_{k=1}^K \alpha_k u^k, \quad s = \sum_{k=1}^K \alpha_k v^k,$$

$$X = \sum_{k=1}^K u^k (u^k)^T, \quad Y = \sum_{k=1}^K u^k (v^k)^T, \quad S = \sum_{k=1}^K (v^k)(v^k)^T.$$

By (4–II) and (4–III), we have

$$x_i + s_i = \sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = 1 = \sum_{k=1}^K \left[(u_i^k)^2 + 2u_i^k v_i^k + (v_i^k)^2 \right] = X_{ii} + 2Y_{ii} + S_{ii},$$

or equivalently,

$$\sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = \sum_{k=1}^K (u_i^k + v_i^k)^2 = 1 \quad \text{for } i = 1, \ldots, n.$$

Hence, by applying the Cauchy–Schwarz inequality to the leftmost expression in the above chain of equalities, we see that there exist scalars $\beta_1, \ldots, \beta_n$ such that

$$\beta_i \alpha_k = u_i^k + v_i^k \quad \text{for } i = 1, \ldots, n;\ k = 1, \ldots, K.$$

Now, observe that if $\alpha_k = 0$ for some $k \in \{1, \ldots, K\}$, then $u^k + v^k = \mathbf{0}$. Since $u^k, v^k \geq \mathbf{0}$, this implies that $u^k = v^k = \mathbf{0}$. In view of the fact that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$, we may thus assume without loss that $\alpha_k > 0$ for $k = 1, \ldots, K$. In particular, we may write

$$Z = \sum_{k=1}^K \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T,$$

where $\lambda_k = \alpha_k^2$, $x^k = u^k/\alpha_k$, and $s^k = v^k/\alpha_k$, for $k = 1, \ldots, K$. Note that $\lambda_k \geq 0$ for $k = 1, \ldots, K$, and $\sum_{k=1}^K \lambda_k = 1$. Hence, in order to show that $Z \in S_1$, it suffices to show that $x^k, s^k \in \{0,1\}^n$ for $k = 1, \ldots, K$, and $x_i^k + s_i^k = 1$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Towards that end, we use (4–II) to compute

$$1 = \sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = \sum_{k=1}^K \beta_i \alpha_k^2 = \beta_i \quad \text{for } i = 1, \ldots, n.$$

It follows that $\alpha_k = u_i^k + v_i^k$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. In particular, we have

$$x_i^k + s_i^k = \frac{u_i^k + v_i^k}{\alpha_k} = 1 \quad \text{for } i = 1, \ldots, n;\ k = 1, \ldots, K. \tag{5}$$

Moreover, using (4–IV), we have

$$\sum_{k=1}^K \alpha_k^2 x_i^k = \sum_{k=1}^K \alpha_k u_i^k = x_i = X_{ii} = \sum_{k=1}^K (u_i^k)^2 = \sum_{k=1}^K \alpha_k^2 (x_i^k)^2$$

and
$$\sum_{k=1}^{K} \alpha_k^2 s_i^k = \sum_{k=1}^{K} \alpha_k v_i^k = s_i = S_{ii} = \sum_{k=1}^{K} (v_i^k)^2 = \sum_{k=1}^{K} \alpha_k^2 (s_i^k)^2.$$

It follows that
$$\sum_{k=1}^{K} \alpha_k^2 \left[x_i^k - (x_i^k)^2 \right] = 0 \quad \text{for } i = 1, \ldots, n \tag{6}$$

and
$$\sum_{k=1}^{K} \alpha_k^2 \left[s_i^k - (s_i^k)^2 \right] = 0 \quad \text{for } i = 1, \ldots, n. \tag{7}$$

Using (5) and the fact that $x^k, s^k \geq 0$ for $k = 1, \ldots, K$, we have $0 \leq x_i^k, s_i^k \leq 1$, which implies that $x_i^k - (x_i^k)^2 \geq 0$ and $s_i^k - (s_i^k)^2 \geq 0$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. This, together with (6) and (7), implies that $x_i^k = (x_i^k)^2$ and $s_i^k = (s_i^k)^2$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, or equivalently, $x^k, s^k \in \{0,1\}^n$ for $k = 1, \ldots, K$. This completes the proof.

(d) **(10pts).** Let v^* and v_R^* denote the optimal values of the original problem and its convex relaxation, respectively. Furthermore, let
$$Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}$$
be an optimal solution to the convex relaxation. By the result in (d), we can find $x^k, s^k \in \{0,1\}^n$ and $\lambda_k \geq 0$, where $k = 1, \ldots, K$, such that $x_i^k + s_i^k = 1$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, $\sum_{k=1}^{K} \lambda_k = 1$, and $Z^* = \sum_{k=1}^{K} \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T$. In particular, we have
$$v_R^* = \frac{1}{2} Q \bullet X^* + c^T x^* = \sum_{k=1}^{K} \lambda_k \left[\frac{1}{2} (x^k)^T Q x^k + c^T x^k \right]. \tag{8}$$
Now, note that for $k = 1, \ldots, K$, the vector x^k is feasible for the original problem, which implies that
$$\frac{1}{2} (x^k)^T Q x^k + c^T x^k \geq v^*.$$
This, together with (8) and the fact that $v^* \geq v_R^*$, implies that $v^* = v_R^*$. Thus, for $k = 1, \ldots, K$, the vector x^k is an optimal solution to the original problem. Moreover, we have $x^* \in \text{conv}\{x^1, \ldots, x^K\}$, as desired.

Problem 3 (30pts).

(a) **(10pts).** By the triangle inequality and the definition of $\|\cdot\|_{1,2}$, for any $(x,t) \in \mathbb{R}^n \times \mathbb{R}$ and $\Delta A \in \mathcal{U}_\lambda$, we have
$$\begin{aligned} \|b - (A + \Delta A)x - te\|_2 &\leq \|b - Ax - te\|_2 + \|\Delta A x\|_2 \\ &\leq \|b - Ax - te\|_2 + \|\Delta A\|_{1,2} \|x\|_1 \\ &\leq \|b - Ax - te\|_2 + \lambda \|x\|_1, \end{aligned}$$
as desired.

$\|\Delta A x\|_{2} \leq \|\Delta A\|_{1,2} \, \|x\|_1$

$\frac{\|\Delta A x\|_2}{\|x\|_1} \leq \|\Delta A\|_{1,2}.$

$\|\eta\|_2 = 1$ $\quad \| \lambda \eta (\mathrm{sgn}(x))^T v \|_2$

$= \lambda \left\| \begin{matrix} \eta_1 \|v\|_1 \\ \vdots \\ \eta_n \|v\|_2 \end{matrix} \right\|_2 = \lambda \|v\|_1 \cdot 1 = \lambda.$

(b) (20pts). We need to construct a matrix $\Delta A^* \in \mathcal{U}_\lambda$ such that all the inequalities in (a) hold as equalities. Towards that end, let $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ be fixed. Define the matrix $\Delta A^* \in \mathbb{R}^{m \times n}$ by

$$\Delta A^* = \begin{cases} -\dfrac{\lambda(b - Ax - te)(\mathrm{sgn}(x))^T}{\|b - Ax - te\|_2} & \text{if } b - Ax - te \neq \mathbf{0}, \\ \lambda u (\mathrm{sgn}(x))^T & \text{otherwise,} \end{cases}$$

where $u \in \mathbb{R}^m$ is an arbitrary unit vector (i.e., $\|u\|_2 = 1$) and $\mathrm{sgn}(x) \in \mathbb{R}^n$ is given by

$$[\mathrm{sgn}(x)]_j = \begin{cases} 1 & \text{if } x_j \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

We first verify that

$$\|\Delta A^*\|_{1,2} = \max_{\|v\|_1 = 1} \|\Delta A^* v\|_2 = \lambda \max_{\|v\|_1 = 1} |(\mathrm{sgn}(x))^T v| = \lambda;$$

i.e., $\Delta A^* \in \mathcal{U}_\lambda$. Next, observe that when $b - Ax - te \neq \mathbf{0}$, we have

$$\begin{aligned} \|b - (A + \Delta A^*)x - te\|_2 &= \left\| b - Ax - te + \frac{\lambda(b - Ax - te)}{\|b - Ax - te\|_2} (\mathrm{sgn}(x))^T x \right\|_2 \\ &= \left\| (b - Ax - te) \left(1 + \frac{\lambda \|x\|_1}{\|b - Ax - te\|_2} \right) \right\|_2 = \left(1 + \frac{\lambda\|x\|_1}{\|b - Ax - te\|_2}\right)\|b-A \\ &= \|b - Ax - te\|_2 + \lambda \|x\|_1. \end{aligned}$$

$1 + \dfrac{\lambda [\mathrm{sgn}(x)]^T x}{\|b - Ax - te\|_2}$

$= \sum_{i=1}^{n} |x_i| = \|x\|_1$

$= \|b - Ax - te\|_2 + \lambda \|x\|_1.$

On the other hand, when $b - Ax - te = \mathbf{0}$, we have

$$\|b - (A + \Delta A^*)x - te\|_2 = \|\Delta A^* x\|_2 = \lambda \|x\|_1 = \|b - Ax - te\|_2 + \lambda \|x\|_1.$$

$= \|b - Ax - te - \Delta A^* x\|_2 = \| \lambda u (\mathrm{sgn}(x))^T x \|_2 = \lambda \| u \, \mathrm{sgn}(x)^T x \|.$

This completes the proof. In particular, we conclude that

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \max_{\Delta A \in \mathcal{U}_\lambda} \|b - (A + \Delta A)x - te\|_2 = \min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \{ \|b - Ax - te\|_2 + \lambda \|x\|_1 \}.$$

$= \lambda \|x\|_1.$

$\lambda \max_{\|v\|_1 = 1} |(\mathrm{sgn}(x))^T v| = \lambda.$

$(u_1 \cdots u_m) \begin{bmatrix} y_1 \\ \vdots \end{bmatrix}$

$\begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} [y_1 \cdots y_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $= \|x\|_1$

$\begin{bmatrix} u_1 \|x\|_1 \\ u_2 \|x\|_1 \\ \vdots \\ u_m \|x\|_1 \end{bmatrix}$

ENGG 5501: Foundations of Optimization

2016–17 First Term

Homework Set 5

Instructor: Anthony Man–Cho So

Due: December 7, 2016

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (30pts). Let $u \in \mathbb{R}^n$ be a given unit vector (i.e., $\|u\|_2 = 1$). The orthogonal projection of the identity matrix I onto the linear subspace

$$\mathcal{L}_u = \left\{ ux^T + xu^T \in \mathcal{S}^n : x \in \mathbb{R}^n \right\} \subset \mathcal{S}^n$$

is, by definition, the solution to the following optimization problem:

$$\min_{x \in \mathbb{R}^n} \left(I - (ux^T + xu^T)\right) \bullet \left(I - (ux^T + xu^T)\right). \tag{1}$$

(a) **(10pts).** Write down the first–order necessary optimality condition of (1).

(b) **(10pts).** Show that the condition found in (a) is also sufficient for optimality of (1).

(c) **(10pts).** Using the results in (a) and (b), determine the optimal solution x^* to (1).

Problem 2 (20pts). Let $A \in \mathcal{S}^n$ and $b \in \mathbb{R}^n$ be given. Consider the optimization problem

$$\begin{array}{rcll} v_p^* & = & \text{minimize} & x^T A x + 2b^T x \\ & & \text{subject to} & x^T x \leq 1. \end{array} \tag{2}$$

Let $u \geq 0$ be the Lagrangian multiplier associated with the constraint $x^T x \leq 1$. Show that the Lagrangian dual of (2) is equivalent to the following SDP:

$$\begin{array}{rcll} v_d^* & = & \text{maximize} & -t - u \\ & & \text{subject to} & \begin{bmatrix} A + uI & b \\ b^T & t \end{bmatrix} \succeq 0, \\ & & & u \geq 0. \end{array} \tag{3}$$

(*Hint: Consider the Moore–Penrose pseudoinverse of* $A + uI$.)

Remark: By the weak duality theorem, we always have $v_p^* \geq v_d^*$. Thus, we can view problem (3) as a convex relaxation of problem (2). In fact, one can use the Shapiro–Barvinok–Pataki theorem (Theorem 1 of Handout 6) and the SDP strong duality theorem to show that such a relaxation is *tight*; i.e., $v_p^* = v_d^*$. We leave this as an optional exercise to the reader.

Problem 3 (20pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable convex function and $C \subset \mathbb{R}^n$ be a non–empty closed convex set. Show that x^* is an optimal solution to the convex optimization problem

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

iff
$$x^* = \Pi_C(x^* - \nabla f(x^*)),$$
where $\Pi_C(\cdot)$ is the projection operator onto C.

Problem 4 (30pts). Let $A^j \in \mathbb{R}^{m_j \times n}$ be given matrices, $C_j \subset \mathbb{R}^{m_j}$ be given non-empty convex sets, and $f_j : \mathbb{R}^{m_j} \to \mathbb{R} \cup \{+\infty\}$ be given functions that are convex on C_j, where $j = 1, \ldots, J$. Consider the following optimization problem:

$$\begin{array}{ll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j x) \\ \text{subject to} & A^j x \in C_j \qquad \text{for } j = 1, \ldots, J. \end{array} \tag{4}$$

(a) **(15pts).** Show that (4) is equivalent to

$$\begin{array}{lll} \inf & \displaystyle\sum_{j=1}^{J} f_j(A^j x^j) & \\ \text{subject to} & A^j x^j = A^j \bar{x} & \text{for } j = 1, \ldots, J, \\ & x^j \in (A^j)^{-1} C_j & \text{for } j = 1, \ldots, J, \\ & \bar{x} \in \mathbb{R}^n. & \end{array} \tag{5}$$

(Recall that $(A^j)^{-1} C_j = \{x \in \mathbb{R}^n : A^j x \in C_j\}$.) Hence, by letting $w^j \in \mathbb{R}^{m_j}$ be the Lagrangian multiplier associated with the constraint $A^j x^j = A^j \bar{x}$, where $j = 1, \ldots, J$, show that the Lagrangian dual of (5) can be expressed as

$$\begin{array}{ll} \sup & \displaystyle\sum_{j=1}^{J} \inf_{x \in C_j} \left\{ f_j(x) + (w^j)^T x \right\} \\ \text{subject to} & \displaystyle\sum_{j=1}^{J} (A^j)^T w^j = 0. \end{array}$$

(b) **(15pts).** Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ be given. Using the result in (a), construct a dual of the following *ridge regression problem*:

$$\min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \lambda \|x\|_2^2 \right\}.$$

Simplify your answer as much as possible.

ENGG 5501: Foundations of Optimization 2016–17 First Term

Homework Set 5 Solution

Instructor: Anthony Man–Cho So December 10, 2016

Problem 1 (30pts).

(a) **(10pts).** Let $f : \mathbb{R}^n \to \mathbb{R}$ be the function given by

$$f(x) = \left(I - (ux^T + xu^T)\right) \bullet \left(I - (ux^T + xu^T)\right) = n - 4u^T x + 2(u^T x)^2 + 2\|x\|_2^2.$$

Since the given problem is an unconstrained optimization problem, its first–order necessary optimality condition is

$$0 = \nabla f(x) = 4\left(x - u + (u^T x)u\right). \in \mathbb{R}^n.$$

(b) **(10pts).** Since $\nabla^2 f(x) = 4(I + uu^T) \succ 0$ for all $x \in \mathbb{R}^n$, the function f is convex. It follows that the first–order ~~necessary optimality~~ condition is also sufficient.

(c) **(10pts).** By the results in (a) and (b), a solution $x^* \in \mathbb{R}^n$ is optimal for the given problem if and only if

$$x^* - u + (u^T x^*)u = 0.$$

The above equation implies that x^* and u must be parallel; i.e., $x^* = \alpha u$ for some $\alpha \in \mathbb{R}$. It then follows that $\alpha = 1/2$; i.e., $x^* = u/2$ is a solution to the above equation.

Problem 2 (20pts). The Lagrangian dual of the given problem is

$$v_d^* = \max_{u \geq 0} \theta(u), \tag{1}$$

where

$$\theta(u) = \min_{x \in \mathbb{R}^n} \left\{ x^T A x + 2b^T x + u(x^T x - 1) \right\}. \tag{2}$$

For a given $u \geq 0$, let $A + uI = V\Sigma V^T$ be the spectral decomposition of $A + uI$. Suppose that $A + uI \not\succeq 0$. Then, there exists an index $i \in \{1, \ldots, n\}$ such that $\Sigma_{ii} < 0$. Upon defining $x(\alpha) = \alpha V e_i$, where $\alpha \in \mathbb{R}$ is a parameter and $e_i \in \mathbb{R}^n$ is the i–th basis vector, we see that

$$x(\alpha)^T (A + uI) x(\alpha) + 2b^T x(\alpha) - u = \alpha^2 \Sigma_{ii} + 2\alpha b^T V e_i - u \to -\infty \quad \text{as } |\alpha| \to +\infty.$$

On the other hand, suppose that $A + uI \succeq 0$. We then have two possibilities. If $b \in \text{range}(A + uI)$, then there exists an $x^* \in \mathbb{R}^n$ such that $(A + uI)x^* = -b$, which implies that Problem (2) has an optimal solution. Otherwise, we have $\theta(u) = -\infty$. ~~Thus, it~~ suffices to focus on the former case. Let $(A + uI)^\dagger$ be the Moore–Penrose pseudoinverse of $A + uI$. We claim that $\bar{x} = -(A + uI)^\dagger b$ is an optimal solution to (2). Indeed, consider an arbitrary solution x^* to the linear equation $(A + uI)x = -b$. Since $(A + uI)^\dagger$ satisfies the identity

$$(A + uI)(A + uI)^\dagger (A + uI) = A + uI,$$

we have

$$-b = (A+uI)x^* = (A+uI)(A+uI)^\dagger(A+uI)x^* = (A+uI)\left[-(A+uI)^\dagger b\right] = (A+uI)\bar{x}.$$

This establishes the claim. In particular, this implies that

$$\theta(u) = \bar{x}^T(A+uI)\bar{x} + 2b^T\bar{x} - u = -b^T(A+uI)^\dagger b - u.$$

Summarizing the above discussion, we obtain

$$\theta(u) = \begin{cases} -b^T(A+uI)^\dagger b - u & \text{if } A+uI \succeq \mathbf{0} \text{ and } b \in \text{range}(A+uI), \\ -\infty & \text{otherwise,} \end{cases}$$

which shows that Problem (1) is equivalent to

$$\begin{array}{rcll} v_d^* & = & \text{maximize} & -b^T(A+uI)^\dagger b - u \\ & & \text{subject to} & A+uI \succeq \mathbf{0}, \\ & & & b \in \text{range}(A+uI), \\ & & & u \geq 0. \end{array} \quad (3)$$

Now, using the Schur complement, we have the following equivalence (see [1] for a nice exposition of this result):

$$t \geq b^T(A+uI)^\dagger b,\ A+uI \succeq \mathbf{0},\ b \in \text{range}(A+uI) \quad \Longleftrightarrow \quad \begin{bmatrix} A+uI & b \\ b^T & t \end{bmatrix} \succeq \mathbf{0}.$$

It follows that Problem (3) is equivalent to

$$\begin{array}{rcll} v_d^* & = & \text{maximize} & -t-u \\ & & \text{subject to} & \begin{bmatrix} A+uI & b \\ b^T & t \end{bmatrix} \succeq \mathbf{0}, \\ & & & u \geq 0, \end{array}$$

as desired.

Problem 3 (20pts). Suppose that $x^* = \Pi_C(x^* - \nabla f(x^*))$. Since C is a non-empty closed convex set, for any $x \in C$, we have

$$0 \geq (x - \Pi_C(x^* - \nabla f(x^*)))^T(x^* - \nabla f(x^*) - \Pi_C(x^* - \nabla f(x^*))) = -\nabla f(x^*)^T(x - x^*).$$

This, together with the continuous differentiability and convexity of f, implies that for all $x \in C$,

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

i.e., x^* is an optimal solution to the optimization problem

$$\min_{x \in C} f(x). \quad (4)$$

Scanned with CamScanner

Conversely, suppose that $x^* \neq \Pi_C(x^* - \nabla f(x^*))$. Since

$$(x^* - \Pi_C(x^* - \nabla f(x^*)))^T (x^* - \nabla f(x^*) - \Pi_C(x^* - \nabla f(x^*))) \leq 0,$$

we have

$$\nabla f(x^*)^T (x^* - \Pi_C(x^* - \nabla f(x^*))) \geq \|x^* - \Pi_C(x^* - \nabla f(x^*))\|_2^2 > 0,$$

or equivalently,

$$\nabla f(x^*)^T (\Pi_C(x^* - \nabla f(x^*)) - x^*) < 0.$$

This implies that $d = \Pi_C(x^* - \nabla f(x^*)) - x^*$ is a descent direction of f at x^*. Moreover, since $x^*, \Pi_C(x^* - \nabla f(x^*)) \in C$, we see that $x^* + \alpha d \in C$ for all $\alpha \in [0,1]$, which implies that d is also a feasible direction at x^*. It follows that x^* is not an optimal solution to (4).

Problem 4 (30pts).

(a) **(15pts).** Upon eliminating the variables $x^1, \ldots, x^J \in \mathbb{R}^n$ from the problem

$$\begin{array}{lll} \inf & \sum_{j=1}^J f_j(A^j x^j) = f_1(A^1x^1) + \cdots + f_J(A^J x^J), & \\ \text{subject to} & A^j x^j = A^j \bar{x} & \text{for } j = 1, \ldots, J, \\ & x^j \in (A^j)^{-1} C_j & \text{for } j = 1, \ldots, J, \quad A^j x^j \in C^j \\ & \bar{x} \in \mathbb{R}^n, & \end{array} \qquad \underline{(5)}$$

we obtain

$$\begin{array}{lll} \inf & \sum_{j=1}^J f_j(A^j \bar{x}) & L(x, v, w) = f(x) + \cancel{v^T G(x)} + w^T H(x) \\ \text{subject to} & A^j \bar{x} \in C_j & \text{for } j = 1, \ldots, J, \\ & \bar{x} \in \mathbb{R}^n, & \end{array}$$

which is the same as the given problem. The Lagrangian dual of (5) is given by

$$\sup_{\substack{w^j \in \mathbb{R}^{m_j} \\ j=1,\ldots,J}} \underline{\theta(w^1, \ldots, w^J)},$$

where

$$(w^j)^T (A^j x^j - A^j \bar{x})$$

$$\begin{aligned} \theta(w^1, \ldots, w^J) &= \inf_{\substack{x^j \in (A^j)^{-1} C_j \\ j=1,\ldots,J; \\ \bar{x} \in \mathbb{R}^n}} \sum_{j=1}^J \{ f_j(A^j x^j) + (w^j)^T A^j (x^j - \bar{x}) \} \\ &= \sum_{j=1}^J \inf_{x^j \in (A^j)^{-1} C_j} \{ f_j(A^j x^j) + (w^j)^T A^j x^j \} - \inf_{\bar{x} \in \mathbb{R}^n} \sum_{j=1}^J (w^j)^T A^j \bar{x} \\ &\qquad\qquad\qquad\qquad \left(\sum_{j=1}^J (w^j)^T A^j \right) \bar{x} \\ &= \begin{cases} \sum_{j=1}^J \inf_{x \in C_j} \{ f_j(x) + (w^j)^T x \} & \text{if } \sum_{j=1}^J (A^j)^T (w^j) = 0, \\ -\infty & \text{otherwise.} \quad \sum_{j=1}^J (A^j)^T w^j \neq 0 \end{cases} \end{aligned}$$

Scanned with CamScanner

This is equivalent to

$$\begin{aligned} \sup \quad & \sum_{j=1}^{J} \inf_{x\in C_j} \{f_j(x) + (w^j)^T x\} \\ \text{subject to} \quad & \sum_{j=1}^{J} (A^j)^T w^j = 0, \end{aligned}$$

as desired.

(b) **(15pts).** Upon letting $f_1(y) = \|y - b\|_2^2$, $f_2(y) = \lambda\|y\|_2^2$, $A^1 = A$, $A^2 = I$, $C_1 = \mathbb{R}^m$, $C_2 = \mathbb{R}^n$ and using the result in (a), we obtain the following dual of the given problem:

$$\begin{aligned} \sup \quad & \inf_{x\in\mathbb{R}^m} \{\|x - b\|_2^2 + (w^1)^T x\} + \inf_{x\in\mathbb{R}^n} \{\lambda\|x\|_2^2 + (w^2)^T x\} \\ \text{subject to} \quad & A^T w^1 + w^2 = 0. \end{aligned} \tag{6}$$

Now, using the first-order optimality conditions, we have

$$\begin{aligned} \inf_{x\in\mathbb{R}^m} \{\|x - b\|_2^2 + (w^1)^T x\} &= -\frac{1}{4}\|w^1\|_2^2 + b^T w^1, \\ \inf_{x\in\mathbb{R}^n} \{\lambda\|x\|_2^2 + (w^2)^T x\} &= -\frac{1}{4\lambda}\|w^2\|_2^2. \end{aligned}$$

It follows that (6) is equivalent to

w^1 w^1 w^2

$$\sup_{w\in\mathbb{R}^m} \left\{ -\frac{1}{4}\|w\|_2^2 + b^T w - \frac{1}{4\lambda}\|A^T w\|_2^2 \right\}.$$

Using again the first-order optimality condition and noting that $I + (1/\lambda)AA^T \succ 0$ for all $\lambda > 0$, we can express the optimal solution to the above problem in closed form:

$$w^* = 2\left(I + \frac{1}{\lambda}AA^T\right)^{-1} b.$$

References

[1] J. Gallier. The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices. Available at http://www.cis.upenn.edu/~jean/schur-comp.pdf, 2010.

ENGG 5501: Foundations of Optimization | **2015–16 First Term**

Homework Set 1

Instructor: Anthony Man–Cho So | Due: October 5, 2015

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $c, f \in \mathbb{R}^n$, $d, g \in \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$ be given. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \dfrac{c^T x + d}{f^T x + g} \\ \text{subject to} & Ax \le b, \\ & f^T x + g \ge 0. \end{array}$$

Here, we assume that $a/0 = +\infty$ if $a > 0$, and $a/0 = -\infty$ if $a \le 0$. Give an equivalent linear programming formulation of the above problem. Justify your answer.

Problem 2 (35pts).

(a) **(20pts).** Let $A \subset \mathbb{R}^n$ be a symmetric convex set; i.e., A is convex, and $-x \in A$ whenever $x \in A$. Furthermore, let $t > 1$ be given. Show that

$$\frac{2}{t+1}(\mathbb{R}^n \backslash (tA)) + \frac{t-1}{t+1} A \subset \mathbb{R}^n \backslash A.$$

(Recall that for any $A, B \subset \mathbb{R}^n$, $A + B = \{x + y : x \in A, y \in B\}$ and $tA = \{tx : x \in A\}$.)

Let μ be a probability measure on $\mathbb{R}^n$. We say that μ is *logconcave* if

$$\mu(\alpha S_1 + (1-\alpha) S_2) \ge \mu(S_1)^\alpha \mu(S_2)^{1-\alpha}$$

holds for any measurable sets $S_1, S_2 \subset \mathbb{R}^n$ and scalar $\alpha \in (0,1)$.

(b) **(15pts).** Let μ be a logconcave probability measure on $\mathbb{R}^n$. Furthermore, let $A \subset \mathbb{R}^n$ be a measurable symmetric convex set such that $\mu(A) = \theta > 1/2$. Use the result in (a) to show that for any $t > 1$,

$$\mu(\mathbb{R}^n \backslash (tA)) \le \theta \left(\frac{1-\theta}{\theta} \right)^{(t+1)/2}.$$

REMARK: The result in (b) is known as the *concentration of measure* phenomenon. Specifically, it says that if A has sufficiently large measure, then a slight enlargement of A will have overwhelming measure. Such a phenomenon has numerous applications and far-reaching consequences. We refer the interested reader to the recent book [1] for a detailed treatment of the topic.

Problem 3 (15pts). Consider the space $\mathcal{S}^n$ of $n \times n$ real symmetric matrices equipped with the inner product $\bullet$, where

$$A \bullet B = \sum_{i,j=1}^n A_{ij} B_{ij} \quad \text{for any } A, B \in \mathcal{S}^n.$$

Let $A \in S^n$ be arbitrary and $A = U\Lambda U^T$ be its spectral decomposition. Prove that $\Pi_{S^n_+}(A) = U\Lambda^+ U^T$, where $\Pi_{S^n_+}(A)$ is the projection of A on S^n_+ and Λ^+ is the $n \times n$ diagonal matrix given by

$$\Lambda^+_{ii} = \max\{\Lambda_{ii}, 0\} \quad \text{for } i = 1, \ldots, n.$$

Problem 4 (30pts). Let $A \in \mathbb{R}^{m \times n}$ be a matrix and $C \subset \mathbb{R}^m$ be a closed convex cone. Define

$$K = \{x \in \mathbb{R}^n : Ax \in C\}.$$

(a) **(10pts).** Show that K is a closed convex cone.

(b) **(20pts).** Show that

$$K^* = \mathrm{cl}\left(\{A^T y : y \in C^*\}\right).$$

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013.

ENGG 5501: Foundations of Optimization

2015–16 First Term

Homework Set 2

Instructor: Anthony Man–Cho So

Due: October 16, 2015

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $f : \mathbb{R}^2 \to \mathbb{R}_+$ be the function defined by

$$f(x_1, x_2) = \frac{1}{1 + x_1^2 + x_2^2}.$$

Show that f is not convex on *any* non–empty open convex set $S \subset \mathbb{R}^2$.

Problem 2 (25pts). For any given $k \geq 1$, let $\lambda_1^k : \mathcal{S}^n \to \mathbb{R}$ be the function that returns the sum of the k largest eigenvalues of its argument.

(a) **(20pts).** Show that

$$\begin{array}{rcll} \lambda_1^k(A) & = & \text{maximize} & A \bullet X \\ & & \text{subject to} & \operatorname{tr}(X) = k, \\ & & & I \succeq X \succeq 0. \end{array}$$

(b) **(5pts).** Using the result in (a), show that λ_1^k is convex for each $k \geq 1$.

Problem 3 (25pts). Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ be an arbitrary norm on $\mathbb{R}^n$. The *dual norm* of $\|\cdot\|$, which is denoted by $\|\cdot\|_*$, is defined as

$$\|x\|_* = \sup_{\|d\|=1} d^T x.$$

Show that

$$\partial \|x\| = \left\{ s \in \mathbb{R}^n : \|s\|_* \leq 1,\ s^T x = \|x\| \right\}.$$

Problem 4 (30pts). Let $A \in \mathbb{R}^{m \times n}$ be a given matrix with $\operatorname{null}(A) = \{0\}$. Define $K = \{x \in \mathbb{R}^n : Ax \leq 0\}$. Suppose that K is non–trivial; i.e., $K \neq \{0\}$.

(a) **(20pts).** Show that K^*, the dual cone of K, has a non–empty interior.

(b) **(10pts).** Using the result in (a), or otherwise, show that there exists a $v \in \mathbb{R}^n$ such that $v^T x > 0$ for all $x \in K \setminus \{0\}$.

ENGG 5501: Foundations of Optimization

2015–16 First Term

Homework Set 3

Instructor: Anthony Man–Cho So

Due: November 2, 2015

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $A \in \mathbb{R}^{m\times n}$ be given. Use the Farkas lemma to show that exactly one of the following systems has a solution:

$$\text{(I)} \quad Ax \le 0,\ Ax \ne 0,\ x \ge 0.$$
$$\text{(II)} \quad A^T y \ge 0,\ y > 0.$$

问题背景描述.

Problem 2 (25pts). Consider a rooted tree[1] whose nodes are numbered $0, 1, \ldots, n$, where node 0 is the root. Let $A \in \{0,1\}^{n\times n}$ be a binary matrix whose (i,j)–th entry equals 1 if and only if node i is on the path from the root to node j, where $i, j \in \{1, \ldots, n\}$. In particular, we have $A_{ii} = 1$ for $i = 1, \ldots, n$. Consider now the following problem:

$$\begin{array}{ll} \text{maximize} & \min\limits_{j\in\{1,\ldots,n\}} [Ax]_j \\ \text{subject to} & e^T x = 1, \\ & x \ge 0. \end{array} \qquad (1)$$

$\Rightarrow \|Ax\|_\infty$

Reformulation

Here, $[Ax]_j$ denotes the j–th component of the vector Ax.

(a) (10pts). Formulate Problem (1) as an LP and write down its dual.

(b) (15pts). Let a_i^T be the i–th row of A, where $i = 1, \ldots, n$. Define $I = \{i \in \{1, \ldots, n\} : e^T a_i = 1\}$. Using the result in (a), or otherwise, show that the solution $x^* \in \mathbb{R}^n$, where

$$x_i^* = \begin{cases} 1/|I| & \text{if } i \in I, \\ 0 & \text{otherwise}, \end{cases}$$

$$A = \begin{pmatrix} I_e & 0 \\ \tilde{A}_1 & \tilde{A}_2 \end{pmatrix}$$

is optimal for Problem (1).

Problem 3 (25pts). Consider a *production game* defined as follows. Let $\mathcal{N} = \{1, \ldots, n\}$ be the set of players, each of whom is given a vector $b^i = (b_1^i, \ldots, b_m^i)$ $(i = 1, \ldots, n)$ of resources. These resources can be used to produce goods, which in turn can be sold at a given market price. Specifically, we assume the following production model: For player $i \in \mathcal{N}$, a unit of the j–th good $(j = 1, \ldots, p)$ requires a_{kj}^i units of the k–th resource $(k = 1, \ldots, m)$ to produce. Furthermore, the j–th good can be sold at a price c_j, where $j = 1, \ldots, p$.

Now, let $S \subset \mathcal{N}$ be a coalition of players. Such a coalition will possess

$$b_k(S) = \sum_{i\in S} b_k^i$$

[1] A tree is a simple connected graph with no cycles. A rooted tree is a tree in which a particular node (known as the root) is distinguished from others.

units of the k-th resource, where $k = 1, \ldots, m$. Using all of their resources, the coalition S can produce any vector $x = (x_1, \ldots, x_p) \in \mathbb{R}^p_+$ of goods that satisfies $A(S)x \leq b(S)$, where

$$A(S)_{kj} = \min_{i \in S} \{a^i_{kj}\} \quad \text{for } k = 1, \ldots, m;\ j = 1, \ldots, p \quad \text{and} \quad b(S) = (b_1(S), \ldots, b_m(S)).$$

Naturally, a coalition S would like to maximize its revenue. The optimization problem it faces can be formulated as the following LP:

$$\begin{array}{rcll} v(S) & = & \text{maximize} & c^T x \\ & & \text{subject to} & A(S)x \leq b(S), \\ & & & x \geq \mathbf{0}. \end{array} \tag{2}$$

The function v will be the value function of this game. Recall that an allocation vector $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ is in the core iff $\sum_{i \in \mathcal{N}} z_i = v(\mathcal{N})$ and $\sum_{i \in S} z_i \geq v(S)$ for all $S \subset \mathcal{N}$.

(a) **(10pts).** Consider the LP faced by the grand coalition $\mathcal{N}$. Write down its dual.

(b) **(15pts).** Suppose that the dual LP given in (a) is feasible. Let y^* be one of its optimal solutions. Show that the allocation vector

$$z^* = \left((b^1)^T y^*, (b^2)^T y^*, \ldots, (b^n)^T y^* \right) \in \mathbb{R}^n$$

belongs to the core.

Problem 4 (15pts). Show that for $n \geq 2$, the Lorentz cone

$$Q^{n+1} = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t \geq \|x\|_2\}$$

is non-polyhedral.

Problem 5 (15pts). Let $Q \in \mathcal{S}^n_{++}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$ be given. Give an equivalent SOCP formulation of the following quadratic programming problem:

$$\begin{array}{ll} \text{minimize} & x^T Q x + c^T x \\ \text{subject to} & Ax = b, \\ & x \geq \mathbf{0}. \end{array}$$

Justify your answer.

ENGG 5501: Foundations of Optimization

Homework Set 4

2015–16 First Term

Instructor: Anthony Man–Cho So

Due: November 13, 2015

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (25pts). Let $C, A_1, \ldots, A_m \in \mathcal{S}^n$ and $b_1, \ldots, b_m > 0$ be given. Let $X^* \in \mathcal{S}^n_+$ be an optimal solution to the following SDP:

$$\begin{array}{rcll} v^*_{\mathsf{SDP}} & = & \text{minimize} & C \bullet X \\ & & \text{subject to} & A_i \bullet X = b_i \quad \text{for } i = 1, \ldots, m, \\ & & & X \succeq 0. \end{array}$$

Since $X^* \succeq 0$, we can write $X^* = UU^T$ for some $U \in \mathbb{R}^{n \times r}$, where $r = \text{rank}(X^*)$. Consider now the following SDP:

$$\begin{array}{rcll} v^*_{\mathsf{ASDP}} & = & \text{minimize} & (U^T C U) \bullet Z \\ & & \text{subject to} & (U^T A_i U) \bullet Z = b_i \quad \text{for } i = 1, \ldots, m, \\ & & & Z \succeq 0. \end{array} \tag{1}$$

(a) **(10pts).** Show that $v^*_{\mathsf{SDP}} = v^*_{\mathsf{ASDP}}$.

(b) **(15pts).** Show that every feasible solution to (1) is an optimal solution to (1).

Problem 2 (60pts).

(a) **(15pts).** Show that the set $\mathcal{CP}_n = \text{conv}\left(\{vv^T : v \in \mathbb{R}^n_+\}\right)$ is a pointed cone.

(b) **(10pts).** Let $u \in \mathbb{R}^n$ and $U \in \mathcal{S}^n$ be given. Define

$$W = \begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix}.$$

Show that $\text{rank}(W) = 1$ if and only if $W = (1, u)(1, u)^T$.

Let $Q \in \mathcal{S}^n$ and $c \in \mathbb{R}^n$ be given. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} x^T Q x + c^T x \\ \text{subject to} & x_i \in \{0, 1\} \quad \text{for } i = 1, \ldots, n. \end{array} \tag{2}$$

Our goal now is to show that the discrete optimization problem (2) can be reformulated as a linear optimization problem over the pointed cone $\mathcal{CP}_{2n+1}$. In other words, problem (2) has a convex reformulation![1]

[1]Since problem (2) is NP–hard in general, we see that not all convex optimization problems are easy to solve.

(c) **(5pts).** Using the result in (b), show that problem (2) is equivalent to

$$
\begin{array}{lll}
\text{minimize} & \frac{1}{2}Q \bullet X + c^T x & \\
\text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & \text{(I)} \\
& x_i + s_i = 1 \quad \text{for } i = 1, \dots, n, & \text{(II)} \\
& X_{ii} + 2Y_{ii} + S_{ii} = 1 \quad \text{for } i = 1, \dots, n, & \text{(III)} \\
& x_i = X_{ii},\ s_i = S_{ii} \quad \text{for } i = 1, \dots, n, & \text{(IV)} \\
& \text{rank}(Z) = 1. &
\end{array} \tag{3}
$$

(d) **(20pts).** Let

$$
\begin{aligned}
S_1 &= \text{conv}\left(\left\{(1, x, s)(1, x, s)^T : x_i, s_i \in \{0, 1\},\ x_i + s_i = 1 \text{ for } i = 1, \dots, n\right\}\right), \\
S_2 &= \left\{Z \in \mathcal{S}^{2n+1} : Z \text{ satisfies (3–I) — (3–IV)}\right\}.
\end{aligned}
$$

Show that $S_1 = S_2$.

(e) **(10pts).** By relaxing the rank constraint in (3), we obtain a convex relaxation of problem (2), which is a CLP. Using the result in (d), show that this convex relaxation is in fact equivalent to problem (2); i.e., (i) the optimal values of both problems are equal, and (ii) if

$$
Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}
$$

is an optimal solution to the convex relaxation, then x^* is in the convex hull of the optimal solutions to problem (2).

Problem 3 (15pts). Let $A \in \mathcal{S}^n$, $P, Q \in \mathbb{R}^{n \times n}$, and $\rho \geq 0$ be given. Show that

$$
A \succeq P^T Z Q + Q^T Z^T P \quad \text{for all } Z \in \mathbb{R}^{n \times n} \text{ with } \|Z\|_F \leq \rho
$$

if and only if there exists a $\lambda \geq 0$ such that

$$
\begin{bmatrix} A - \lambda Q^T Q & -\rho P^T \\ -\rho P & \lambda I \end{bmatrix} \succeq 0.
$$

ENGG 5501: Foundations of Optimization | **2015–16 First Term**

Homework Set 5

Instructor: Anthony Man–Cho So | **Due: December 14, 2015**

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (40pts). Let $u \in \mathbb{R}^n$ be a given unit vector (i.e., $\|u\|_2 = 1$), and define the linear subspace $\mathcal{L}_u \subset \mathcal{S}^n$ via

$$\mathcal{L}_u = \left\{ux^T + xu^T \in \mathcal{S}^n : x \in \mathbb{R}^n\right\}.$$

Now, the orthogonal projection of the identity matrix I onto the subspace $\mathcal{L}_u$ can be found by solving the following optimization problem:

$$\min_{x\in\mathbb{R}^n} \left(I - (ux^T + xu^T)\right) \bullet \left(I - (ux^T + xu^T)\right). \qquad (P)$$

(a) **(10pts).** Write down the first–order necessary optimality condition of (P).

(b) **(10pts).** Show that the condition found in (a) is also a sufficient optimality condition of (P).

(c) **(20pts).** Using the results in (a) and (b), determine the optimal solution x^* to (P).

Problem 2 (45pts). Let $A \in \mathbb{R}^{m\times n}$ and $c \in \mathbb{R}^n$ be given. Suppose that A has full row rank. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = 0, \\ & \|x\|_2^2 \le 1. \end{array} \qquad (Q)$$

(a) **(15pts).** Write down the first–order optimality conditions of (Q) and explain why they are necessary and sufficient for optimality.

(b) **(15pts).** Using the result in (a), express the optimal solution x^* to (Q) in terms of A and c. Show your derivation.

(c) **(15pts).** Let $u \in \mathbb{R}^m$ be the Lagrange multiplier associated with the constraint $Ax = 0$. By dualizing only the constraint $Ax = 0$, show that the dual of (Q) is given by

$$\sup_{u\in\mathbb{R}^m} -\|c + A^T u\|_2. \qquad (DQ)$$

Hence, show that there is no duality gap between (Q) and (DQ).

Problem 3 (15pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \ge 0. \end{array} \qquad (1)$$

Show that $\bar{x} \in \mathbb{R}^n$ is an optimal solution to (1) iff $\bar{x} \in \mathbb{R}^n$ is a solution to the following system:

$$\begin{aligned} \nabla f(x) &\ge 0, \\ x &\ge 0, \\ x^T \nabla f(x) &= 0. \end{aligned}$$

ENGG 5501: Foundations of Optimization 2015–16 First Term

Instructor: Anthony Man–Cho So

Homework Set 2 Solution

October 24, 2015

Problem 1 (20pts). Note that f is twice continuously differentiable on $\mathbb{R}^2$ with

$$\nabla f(x_1, x_2) = -\left(\frac{2x_1}{(1+x_1^2+x_2^2)^2}, \frac{2x_2}{(1+x_1^2+x_2^2)^2} \right)$$

and

$$\nabla^2 f(x_1, x_2) = \frac{1}{(1+x_1^2+x_2^2)^3} \begin{bmatrix} 2\left(3x_1^2 - x_2^2 - 1\right) & 8x_1x_2 \\ 8x_1x_2 & 2\left(-x_1^2 + 3x_2^2 - 1\right) \end{bmatrix}.$$

Suppose that $\nabla^2 f(x_1, x_2) \succeq \mathbf{0}$ for some $(x_1, x_2) \in \mathbb{R}^2$. Then, we must have

$$\begin{aligned} 3x_1^2 - x_2^2 - 1 &\geq 0, & (1) \\ -x_1^2 + 3x_2^2 - 1 &\geq 0, & (2) \\ 4(3x_1^2 - x_2^2 - 1)(-x_1^2 + 3x_2^2 - 1) - 64x_1^2x_2^2 &\geq 0. & (3) \end{aligned}$$

From (3), we have

$$\begin{aligned} 0 &\leq 4(3x_1^2 - x_2^2 - 1)(-x_1^2 + 3x_2^2 - 1) - 64x_1^2x_2^2 \\ &= -12(x_1^2 + x_2^2)^2 - 8(x_1^2 + x_2^2) + 4 \\ &= -4(1 + x_1^2 + x_2^2)\left(-1 + 3(x_1^2 + x_2^2)\right), \end{aligned}$$

which implies that $3(x_1^2+x_2^2) \leq 1$. On the other hand, upon adding (1) and (2), we get $x_1^2+x_2^2 \geq 1$. It follows that (1)–(3) are inconsistent. Thus, we conclude that $\nabla^2 f(x_1, x_2) \not\succeq \mathbf{0}$ for all $(x_1, x_2) \in \mathbb{R}^2$; i.e., f is not convex on any non–empty open convex set $S \subset \mathbb{R}^2$.

Problem 2 (25pts).

(a) **(20pts).** Let $A = U\Lambda U^T$ be the spectral decomposition of A. Then, we have $A \bullet X = (U\Lambda U^T) \bullet X = \Lambda \bullet (U^TXU)$. Since

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

and

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

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

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

In particular, the given optimization problem is equivalent to

$$\begin{array}{ll} \text{maximize} & \Lambda \bullet X \\ \text{subject to} & \mathrm{tr}(X) = k, \\ & I \succeq X \succeq 0. \end{array} \qquad (4)$$

Now, we claim that there exists an optimal solution to (4) that is diagonal. To see this, observe that $\Lambda \bullet X = \sum_{i=1}^n \Lambda_{ii} X_{ii}$, and $I \succeq X \succeq 0$ implies that $X_{ii} \in [0,1]$ for $i = 1, 2, \ldots, n$. In particular, if X^* is an optimal solution to (4), then the diagonal matrix $\bar{X}^* = \operatorname{diag}(X_{11}^*, X_{22}^*, \ldots, X_{nn}^*)$ is feasible for (4) and has the same objective value as X^*. This establishes the claim. Consequently, Problem (4) is equivalent to the following linear program:

$$
\begin{array}{ll}
\text{maximize} & \displaystyle\sum_{i=1}^n \Lambda_{ii} x_i \\
\text{subject to} & \displaystyle\sum_{i=1}^n x_i = k, \\
& 0 \le x \le e.
\end{array} \qquad (5)
$$

It is easy to verify that the optimal value of (5) 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) **(5pts).** For each $X \in \mathcal{S}^n$, define the function $f_X : \mathcal{S}^n \to \mathbb{R}$ by $f_X(A) = A \bullet X$. 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 3 (25pts). Consider a fixed $x \in \mathbb{R}^n$. Let $s \in \mathbb{R}^n$ be such that $\|s\|_* \le 1$ and $s^T x = \|x\|$. By definition of the dual norm, for any $\bar{x} \in \mathbb{R}^n \setminus \{0\}$, we have

$$1 \ge \|s\|_* = \sup_{d \ne 0} \frac{d^T s}{\|d\|} \ge \frac{s^T \bar{x}}{\|\bar{x}\|}.$$

It follows that

$$\|\bar{x}\| \ge s^T \bar{x} = s^T x + s^T(\bar{x} - x) = \|x\| + s^T(\bar{x} - x).$$

Note that the above inequality is also valid at $\bar{x} = 0$, because $s^T x = \|x\|$ by assumption. Hence, we conclude that $s \in \partial\|x\|$.

Conversely, suppose that $s \in \partial\|x\|$. Consider first the case where $x \ne 0$. We have

$$
\begin{aligned}
2\|x\| &= \|x + x\| \ge \|x\| + s^T x, \\
0 &= \|x - x\| \ge \|x\| - s^T x,
\end{aligned}
$$

which together imply that $s^T x = \|x\|$. Since $x \ne 0$, it follows that

$$\|s\|_* = \sup_{\|d\|=1} d^T x \ge \frac{s^T x}{\|x\|} = 1.$$

We claim that $\|s\|_* = 1$. Suppose that this is not the case. Then, we have $\|s\|_* > 1$, which implies that $s^T d > 1$ for some $d \in \mathbb{R}^n$ with $\|d\| = 1$. We compute

$$\|x\| + 1 = \|x\| + \|d\| \ge \|x + d\| \ge \|x\| + s^T d > \|x\| + 1,$$

which is a contradiction. This establishes the claim.

Now, consider the case where $x = \mathbf{0}$. The condition $s^T x = \|x\|$ is automatically satisfied. On the other hand, let $f : \mathbb{R}^n \to \mathbb{R}_+$ be given by $f(x) = \|x\|$. The directional derivative of f at $\mathbf{0}$ in the direction $d \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ is, by definition,

$$f'(\mathbf{0}, d) = \lim_{t \searrow 0} \frac{\|td\|}{t} = \|d\|.$$

Hence, by Theorem 16(a) of Handout 2, we have $f'(\mathbf{0}, d) = \|d\| \geq s^T d$ for all $s \in \partial\|\mathbf{0}\|$. This, together with the definition of the dual norm, yields

$$\|s\|_* = \sup_{d \neq \mathbf{0}} \frac{s^T d}{\|d\|} \leq 1,$$

as desired.

Problem 4 (30pts).

(a) **(20pts).** Suppose that $\mathrm{int}(K^*)$ is empty. Since K^* is a non-empty convex set (note that $\mathbf{0} \in K^*$), by Theorem 3 of Handout 2, it has a non-empty relative interior. In particular, we have $\dim(\mathrm{aff}(K^*)) < n$. This, together with the fact that $\mathbf{0} \in \mathrm{aff}(K^*)$, implies the existence of a vector $u \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $u^T x = 0$ for all $x \in \mathrm{aff}(K^*)$. Consequently, we have $u, -u \in (K^*)^*$. Since K is non-empty, closed, and convex, we have $K = (K^*)^*$ by Proposition 3 of Handout 2. It follows that $u \in \mathrm{null}(A)$, which contradicts the fact that $\mathrm{null}(A) = \{\mathbf{0}\}$.

(b) **(10pts).** Since $\mathrm{int}(K^*) \neq \emptyset$, there exist a vector $v \in \mathbb{R}^n$ and a constant $\epsilon > 0$ such that $B(v, \epsilon) \subset K^*$. Hence, for any $x \in K \setminus \{\mathbf{0}\}$ and $w \in B(0, \epsilon)$, we have $(v - w)^T x \geq 0$. This implies that

$$v^T x \geq \sup_{w \in B(\mathbf{0}, \epsilon)} w^T x = \epsilon \|x\|_2 > 0,$$

as desired.

Homework Set 3 Solution

Instructor: Anthony Man–Cho So

November 10, 2015

Problem 1 (20pts). First, let us show that systems (I) and (II) cannot both have solutions. Suppose to the contrary that there exist vectors $\bar{x} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^m$ satisfying (I) and (II), respectively. Then, since $\bar{y} > 0$, $A\bar{x} \le 0$ and $A\bar{x} \ne 0$, we have $\bar{y}^T A\bar{x} < 0$. On the other hand, since $\bar{x} \ge 0$ and $A^T\bar{y} \ge 0$, we have $\bar{y}^T A\bar{x} \ge 0$. This results in a contradiction.

Now, suppose that system (I) does not have a solution. Then, by a simple scaling argument, the system

$$(\mathrm{I}') \quad Ax \le 0,\ e^T Ax = -1,\ x \ge 0$$ 重点：维度对齐.

does not have a solution either. By introducing slack variables, we see that system (I′) is equivalent to

$$\begin{bmatrix} A & I \\ e^T A & 0^T \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix},\ (x, s) \ge 0.$$

$A \in R^{m\times n}$.
$x \in R^{n\times 1}$
$Ax \in R^{m\times 1}$ ones(m,1)
$e \in R^{m\times 1}$ 全 1.
$A^T e \in R^{n\times 1}$

Hence, by Farkas' lemma, there exists a $\bar{z} = (\bar{u}, \bar{t}) \in \mathbb{R}^{m+1}$ such that

$$\begin{bmatrix} A^T & A^T e \\ I & 0 \end{bmatrix} \begin{bmatrix} \bar{u} \\ \bar{t} \end{bmatrix} \ge 0,\ \bar{t} > 0,$$

or equivalently,

$$A^T(\bar{u} + \bar{t}e) \ge 0,\ \bar{u} \ge 0,\ \bar{t} > 0.$$

Now, let $\bar{y} = \bar{u} + \bar{t}e \in \mathbb{R}^m$. Clearly, we have $A^T\bar{y} \ge 0$. Moreover, since $\bar{u} \ge 0$ and $\bar{t} > 0$, we have $\bar{y} \ge \bar{t}e > 0$. This completes the proof.

Problem 2 (25pts).

(a) **(10pts).** By introducing an extra variable $t \in \mathbb{R}$ and observing that the structure of A implies $Ax \ge 0$ for any $x \ge 0$, the given problem is equivalent to the following LP:

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & Ax \ge te, \\ & e^T x = 1, \\ & x \ge 0,\ t \ge 0. \end{array} \tag{1}$$

The dual of Problem (1) is then given by

$$\begin{array}{ll} \text{minimize} & \theta \\ \text{subject to} & A^T z \le \theta e, \\ & e^T z \ge 1, \\ & z \ge 0. \end{array} \tag{2}$$

(b) **(15pts).** Without loss of generality, we may assume that the leaf nodes of the rooted tree are numbered $1, \ldots, \ell$. We claim that the matrix A can be decomposed as

$$A = \begin{bmatrix} I_\ell & 0_{\ell\times(n-\ell)} \\ \bar{A}_1 & \bar{A}_2 \end{bmatrix}, \tag{3}$$

where I_ℓ is the $\ell \times \ell$ identity matrix, $0_{\ell\times(n-\ell)}$ is the $\ell \times (n-\ell)$ zero matrix, $\bar{A}_1 \in \{0,1\}^{(n-\ell)\times\ell}$, and $\bar{A}_2 \in \{0,1\}^{(n-\ell)\times(n-\ell)}$. Indeed, since a leaf node has no children, it cannot be on the path from the root to any other node. This yields the desired structure of A. Next, we claim that each row of $\bar{A}_1$ has at least one entry equal to 1. Indeed, each row of $\bar{A}_1$ corresponds to a non-leaf node. Since a rooted tree is connected, a non-leaf node must lie on the path from the root to some leaf node. This establishes the claim. Note that the diagonal entries of $\bar{A}_2$ are all ones. Hence, a corollary of the claims above is that $I = \{i \in \{1,\ldots,n\} : e^T a_i = 1\} = \{1,\ldots,\ell\}$.

From the above discussion and the structure of A in (3), we have

$$\min_{j\in\{1,\ldots,n\}} [Ax^*]_j \geq \frac{1}{|I|}.$$

This implies that the optimal value of Problem (1) is at least $1/|I|$. On the other hand, set $\bar{z} = x^* \in \mathbb{R}^n$. It is clear that $\bar{z} \geq \mathbf{0}$ and $e^T\bar{z} = e^T x^* = 1$. Moreover, we have

$$A^T\bar{z} = \sum_{i=1}^{n} a_i x_i^* = \frac{1}{|I|}\sum_{i\in I} a_i \leq \frac{1}{|I|}e,$$

which implies that the optimal value of Problem (2) is at most $1/|I|$. It then follows from the LP Weak Duality Theorem that x^* is optimal for Problem (1).

Problem 3 (25pts).

(a) **(10pts).** The dual of the LP faced by the grand coalition is given by

$$\begin{array}{rl} \text{minimize} & b(\mathcal{N})^T y \\ \text{subject to} & A(\mathcal{N})^T y \geq c, \\ & y \geq \mathbf{0}. \end{array}$$

(b) **(15pts).** By the LP Strong Duality Theorem, we have

$$\sum_{i\in\mathcal{N}} z_i^* = \left(\sum_{i\in\mathcal{N}} b^i\right)^T y^* = b(\mathcal{N})^T y^* = v(\mathcal{N}).$$

Now, let $S \subset \mathcal{N}$ be an arbitrary coalition. By definition, $A(S)$ satisfies the following component-wise inequality:

$$A(S) \geq A(\mathcal{N}).$$

Since y^* satisfies $y \geq \mathbf{0}$ and $A(\mathcal{N})^T y^* \geq c$, we conclude that

$$A(S)^T y^* \geq A(\mathcal{N})^T y^* \geq c.$$

In other words, y^* is feasible for the dual of the LP that is faced by S. Hence, by the LP Weak Duality Theorem, we have

$$\sum_{i \in S} z_i^* = b(S)^T y^* \geq v(S),$$

as desired.

Problem 4 (15pts). Observe that if P is a polyhedron and H is a hyperplane, then $P \cap H$ is also a polyhedron, because $P \cap H$ can be expressed as a finite intersection of halfspaces. Now, consider the hyperplane $H = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t = 1\}$, where $n \geq 2$. Clearly, we have

$$Q^{n+1} \cap H = \{(1, x) \in \mathbb{R} \times \mathbb{R}^n : 1 \geq \|x\|_2\},$$

which shows that $Q^{n+1} \cap H$ is an n-dimensional unit ball lying on the plane H. In particular, it is straightforward to verify that every point $(1, x) \in Q^{n+1} \cap H$ satisfying $\|x\|_2 = 1$ is an extreme point of $Q^{n+1} \cap H$; cf. p. 5 of Handout 2. This implies that $Q^{n+1} \cap H$ is not polyhedral, and hence Q^{n+1} is not polyhedral.

Problem 5 (15pts). Since $Q \in \mathcal{S}^n_{++}$, Q is invertible and $Q^{-1} \in \mathcal{S}^n_{++}$. Let

$$u = Q^{1/2}x + \frac{1}{2}Q^{-1/2}c \in \mathbb{R}^n.$$

Then, we have

$$x^T Q x + c^T x = u^T u - \frac{1}{4}c^T Q^{-1} c.$$

Note that $(1/4)c^T Q^{-1} c$ is a constant. Hence, the given problem is equivalent to the following problem:

$$\begin{array}{ll} \text{minimize} & u^T u \\ \text{subject to} & Q^{1/2}x - u = -\frac{1}{2}Q^{-1/2}c, \\ & Ax = b, \\ & x \geq 0. \end{array}$$

It is now easy to see that the problem above is equivalent to the following SOCP:

$$\begin{array}{ll} \text{minimize} & u_0 \\ \text{subject to} & Q^{1/2}x - u = -\frac{1}{2}Q^{-1/2}c, \\ & Ax = b, \\ & x \geq 0, \ (u_0, u) \succeq_{Q^{n+1}} 0. \end{array}$$

ENGG 5501: Foundations of Optimization 2015–16 First Term

Homework Set 4 Solution

Instructor: Anthony Man–Cho So November 26, 2015

Problem 1 (25pts).

(a) **(10pts).** Recall that

$$v^*_{\mathsf{SDP}} = \begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & A_i \bullet X = b_i \quad \text{for } i = 1, \ldots, m, \\ & X \succeq 0 \end{array} \tag{SDP}$$

and

$$v^*_{\mathsf{ASDP}} = \begin{array}{ll} \text{minimize} & (U^T C U) \bullet Z \\ \text{subject to} & (U^T A_i U) \bullet Z = b_i \quad \text{for } i = 1, \ldots, m, \\ & Z \succeq 0. \end{array} \tag{ASDP}$$

Let $\bar{C} = U^T C U$ and $\bar{A}_i = U^T A_i U$ for $i = 1, \ldots, m$. Observe that the solution $W = I$ is feasible for (ASDP) and has an objective value

$$\bar{C} \bullet I = C \bullet LL^H = C \bullet Z^* = v^*_{\mathsf{SDP}}.$$

This implies that $v^*_{\mathsf{ASDP}} \le v^*_{\mathsf{SDP}}$. On the other hand, every feasible solution Z to (ASDP) corresponds to a feasible solution $X(Z) = UZU^T$ to (SDP), which implies that $v^*_{\mathsf{ASDP}} \ge v^*_{\mathsf{SDP}}$.

(b) **(15pts).** Consider the dual of (ASDP):

$$\begin{array}{ll} \sup & b^T y \\ \text{subject to} & \bar{C} - \sum_{i=1}^m y_i \bar{A}_i \succeq 0, \\ & y \in \mathbb{R}^m. \end{array} \tag{ASDD}$$

Since (ASDP) is bounded below and strictly feasible, by the CLP Strong Duality Theorem, (ASDD) has an optimal solution y^*. Moreover, since $Z = I$ is optimal for (ASDP), the CLP Strong Duality Theorem yields

$$I \bullet \left(\bar{C} - \sum_{i=1}^m y_i^* \bar{A}_i \right) = 0.$$

This, together with the fact that $\bar{C} - \sum_{i=1}^m y_i^* \bar{A}_i \succeq 0$, implies

$$\bar{C} - \sum_{i=1}^m y_i^* \bar{A}_i = 0.$$

It follows that every feasible solution $\bar{Z}$ to (ASDP) satisfies the complementarity condition

$$\bar{Z} \bullet \left(\bar{C} - \sum_{i=1}^m y_i^* \bar{A}_i \right) = 0.$$

Hence, by the CLP Strong Duality Theorem, we conclude that $\bar{Z}$ is optimal for (ASDP).

Problem 2 (60pts).

(a) **(15pts).** The non-emptiness of $\mathcal{CP}_n$ follows from the fact that $\mathbf{0} \in \mathcal{CP}_n$. Now, for any $\alpha > 0$ and $v \geq 0$, we have $\sqrt{\alpha}v \geq 0$, which implies that $\alpha vv^T = (\sqrt{\alpha}v)(\sqrt{\alpha}v)^T \in \mathcal{CP}_n$. For an arbitrary $C \in \mathcal{CP}_n$, there exist vectors $v_1, \ldots, v_K \geq 0$ and scalars $\lambda_1, \ldots, \lambda_K \geq 0$ such that $C = \sum_{k=1}^{N} \lambda_k v_k v_k^T$ and $\sum_{k=1}^{K} \lambda_k = 1$. It follows that $\alpha C = \sum_{k=1}^{K} \lambda_k (\sqrt{\alpha}v_k)(\sqrt{\alpha}v_k)^T \in \mathcal{CP}_n$ for any $\alpha > 0$. Hence, we conclude that $\mathcal{CP}_n$ is a cone. As a corollary, we see that $\mathcal{CP}_n$ is closed under addition. Indeed, since $\mathcal{CP}_n$ is convex by definition and is a cone, for arbitrary $C_1, C_2 \in \mathcal{CP}_n$, we have

$$C_1 + C_2 = 2\left(\frac{C_1 + C_2}{2}\right) \in \mathcal{CP}_n.$$

Finally, it is clear that every matrix in $\mathcal{CP}_n$ is non-negative. It follows that if $C, -C \in \mathcal{CP}_n$, then $C = 0$; i.e., $\mathcal{CP}_n$ is pointed.

(b) **(10pts).** Clearly, if $W = (1, u)(1, u)^T$, then $\text{rank}(W) = 1$. Conversely, suppose that $\text{rank}(W) = 1$. Since $W \in \mathcal{S}^{n+1}$, it admits a spectral decomposition $W = \lambda \bar{v}\bar{v}^T$, where $\bar{v} = (v_0, v) \in \mathbb{R}^{n+1}$ and $\lambda \neq 0$. Thus, we have

$$\begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix} = \begin{bmatrix} \lambda v_0^2 & \lambda v_0 v^T \\ \lambda v_0 v & \lambda vv^T \end{bmatrix},$$

which gives $\lambda v_0^2 = 1$ and $\lambda v_0 v = u$. In particular, we see that $\lambda > 0$ and $v_0 \neq 0$. This implies that

$$U = \lambda vv^T = \lambda \left(\frac{u}{\lambda v_0}\right)\left(\frac{u}{\lambda v_0}\right)^T = uu^T,$$

from which we obtain $W = (1, u)(1, u)^T$.

(c) **(5pts).** The given problem is equivalent to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^T Q x + c^T x \\ \text{subject to} \quad & x_i + s_i = 1 && \text{for } i = 1, \ldots, n, \\ & (x_i + s_i)^2 = 1 && \text{for } i = 1, \ldots, n, \\ & x_i = x_i^2,\ s_i = s_i^2 && \text{for } i = 1, \ldots, n. \end{aligned} \tag{1}$$

Now, set $Z = (1, x, s)(1, x, s)^T \in \mathcal{S}_+^{2n+1}$. Since $x, s \geq 0$, we have $Z \in \mathcal{CP}_{2n+1}$. Moreover, by the result in (b), we have

$$Z = (1, x, s)(1, x, s)^T \iff Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix},\ \text{rank}(Z) = 1.$$

Since $X = xx^T$, $Y = xs^T$, and $S = ss^T$, we conclude that problem (1) is equivalent to

$$
\begin{array}{llll}
\text{minimize} & \frac{1}{2}Q \bullet X + c^T x & & \\
\text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & & \text{(I)} \\
& x_i + s_i = 1 & \text{for } i = 1, \dots, n, & \text{(II)} \\
& X_{ii} + 2Y_{ii} + S_{ii} = 1 & \text{for } i = 1, \dots, n, & \text{(III)} \\
& x_i = X_{ii},\ s_i = S_{ii} & \text{for } i = 1, \dots, n, & \text{(IV)} \\
& \operatorname{rank}(Z) = 1, & &
\end{array} \tag{2}
$$

as desired.

(d) **(20pts).** Let $x, s \in \{0, 1\}^n$ be such that $x_i + s_i = 1$ for $i = 1, \dots, n$. Then, we have $(1, x, s)(1, x, s)^T \in S_2$. Since S_2 is convex, we see that $S_1 \subset S_2$.

Conversely, let $Z \in S_2$ be given. By (2–I), there exist vectors $(\alpha_k, u^k, v^k) \in \mathbb{R}_+^{2n+1}$, where $k = 1, \dots, K$, such that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$. Upon equating the entries of both sides, we have

$$\sum_{k=1}^K \alpha_k^2 = 1, \quad x = \sum_{k=1}^K \alpha_k u^k, \quad s = \sum_{k=1}^K \alpha_k v^k,$$

$$X = \sum_{k=1}^K u^k (u^k)^T, \quad Y = \sum_{k=1}^K u^k (v^k)^T, \quad S = \sum_{k=1}^K (v^k)(v^k)^T.$$

By (2–II) and (2–III), we have

$$x_i + s_i = \sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = 1 = \sum_{k=1}^K \left[(u_i^k)^2 + 2u_i^k v_i^k + (v_i^k)^2 \right] = X_{ii} + 2Y_{ii} + S_{ii},$$

or equivalently,

$$\sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = \sum_{k=1}^K (u_i^k + v_i^k)^2 = 1 \quad \text{for } i = 1, \dots, n.$$

Hence, by applying the Cauchy–Schwarz inequality to the leftmost expression in the above chain of equalities, we see that there exist scalars $\beta_1, \dots, \beta_n$ such that

$$\beta_i \alpha_k = u_i^k + v_i^k \quad \text{for } i = 1, \dots, n;\ k = 1, \dots, K.$$

Now, observe that if $\alpha_k = 0$ for some $k \in \{1, \dots, K\}$, then $u^k + v^k = 0$. Since $u^k, v^k \ge 0$, this implies that $u^k = v^k = \mathbf{0}$. In view of the fact that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$, we may thus assume without loss that $\alpha_k > 0$ for $k = 1, \dots, K$. In particular, we may write

$$Z = \sum_{k=1}^K \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T,$$

where $\lambda_k = \alpha_k^2$, $x^k = u^k/\alpha_k$, and $s^k = v^k/\alpha_k$, for $k = 1, \ldots, K$. Note that $\lambda_k \geq 0$ for $k = 1, \ldots, K$, and $\sum_{k=1}^K \lambda_k = 1$. Hence, in order to show that $Z \in S_1$, it suffices to show that $x^k, s^k \in \{0,1\}^n$ for $k = 1, \ldots, K$, and $x_i^k + s_i^k = 1$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Towards that end, we use (2–II) to compute

$$1 = \sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = \sum_{k=1}^K \beta_i \alpha_k^2 = \beta_i \quad \text{for } i = 1, \ldots, n.$$

It follows that $\alpha_k = u_i^k + v_i^k$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. In particular, we have

$$x_i^k + s_i^k = \frac{u_i^k + v_i^k}{\alpha_k} = 1 \quad \text{for } i = 1, \ldots, n;\ k = 1, \ldots, K. \tag{3}$$

Moreover, using (2–IV), we have

$$\sum_{k=1}^K \alpha_k^2 x_i^k = \sum_{k=1}^K \alpha_k u_i^k = x_i = X_{ii} = \sum_{k=1}^K (u_i^k)^2 = \sum_{k=1}^K \alpha_k^2 (x_i^k)^2$$

and

$$\sum_{k=1}^K \alpha_k^2 s_i^k = \sum_{k=1}^K \alpha_k v_i^k = s_i = S_{ii} = \sum_{k=1}^K (v_i^k)^2 = \sum_{k=1}^K \alpha_k^2 (s_i^k)^2.$$

It follows that

$$\sum_{k=1}^K \alpha_k^2 \left[x_i^k - (x_i^k)^2 \right] = 0 \quad \text{for } i = 1, \ldots, n \tag{4}$$

and

$$\sum_{k=1}^K \alpha_k^2 \left[s_i^k - (s_i^k)^2 \right] = 0 \quad \text{for } i = 1, \ldots, n. \tag{5}$$

Using (3) and the fact that $x^k, s^k \geq 0$ for $k = 1, \ldots, K$, we have $0 \leq x_i^k, s_i^k \leq 1$, which implies that $x_i^k - (x_i^k)^2 \geq 0$ and $s_i^k - (s_i^k)^2 \geq 0$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. This, together with (4) and (5), implies that $x_i^k = (x_i^k)^2$ and $s_i^k = (s_i^k)^2$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, or equivalently, $x^k, s^k \in \{0,1\}^n$ for $k = 1, \ldots, K$. This completes the proof.

(e) **(10pts).** Let v^* and v_R^* denote the optimal values of the original problem and its convex relaxation, respectively. Furthermore, let

$$Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}$$

be an optimal solution to the convex relaxation. By the result in (d), we can find $x^k, s^k \in \{0,1\}^n$ and $\lambda_k \geq 0$, where $k = 1, \ldots, K$, such that $x_i^k + s_i^k = 1$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, $\sum_{k=1}^K \lambda_k = 1$, and $Z^* = \sum_{k=1}^K \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T$. In particular, we have

$$v_R^* = \frac{1}{2} Q \bullet X^* + c^T x^* = \sum_{k=1}^K \lambda_k \left[\frac{1}{2} (x^k)^T Q x^k + c^T x^k \right]. \tag{6}$$

Now, note that for $k = 1, \ldots, K$, the vector x^k is feasible for the original problem, which implies that

$$\frac{1}{2}(x^k)^T Q x^k + c^T x^k \geq v^*.$$

This, together with (6) and the fact that $v^* \geq v_R^*$, implies that $v^* = v_R^*$. Thus, for $k = 1, \ldots, K$, the vector x^k is an optimal solution to the original problem. Moreover, we have $x^* \in \text{conv}\{x^1, \ldots, x^K\}$, as desired.

Problem 3 (15pts). Observe that

$$
\begin{aligned}
& A \succeq P^T Z Q + Q^T Z^T P \quad \text{for all } Z \in \mathbb{R}^{n\times n} \text{ with } \|Z\|_F \leq \rho \\
\iff \quad & v^T A v \geq v^T (P^T Z Q + Q^T Z^T P) v \quad \text{for all } v \in \mathbb{R}^n \text{ and } Z \in \mathbb{R}^{n\times n} \text{ with } \|Z\|_F \leq \rho \\
\iff \quad & v^T A v \geq \max_{\|Z\|_F \leq \rho} \left\{ v^T \left(P^T Z Q + Q^T Z^T P \right) v \right\} \quad \text{for all } v \in \mathbb{R}^n. \qquad (7)
\end{aligned}
$$

Since

$$v^T \left(P^T Z Q + Q^T Z^T P \right) v = 2(Qv)^T Z^T (Pv) = 2(Pv)(Qv)^T \bullet Z,$$

it follows from the Cauchy–Schwarz inequality that

$$\max_{\|Z\|_F \leq \rho} \left\{ v^T \left(P^T Z Q + Q^T Z^T P \right) v \right\} = 2\rho (Pv)(Qv)^T \bullet \frac{(Pv)(Qv)^T}{\|(Pv)(Qv)^T\|_F} = 2\rho \|Pv\|_2 \|Qv\|_2.$$

Using the Cauchy–Schwarz inequality again, we have

$$\|Pv\|_2 \|Qv\|_2 = \max_{\|y\|_2 \leq \|Qv\|_2} y^T P v.$$

Hence, the inequality (7) is equivalent to

$$v^T A v \geq 2\rho y^T P v \quad \text{for all } (v, y) \in \mathbb{R}^n \times \mathbb{R}^n \text{ satisfying } y^T y \leq v^T Q^T Q v. \qquad (8)$$

Now, let $u = (v, y) \in \mathbb{R}^{2n}$,

$$S = \begin{bmatrix} A & -\rho P^T \\ -\rho P & 0 \end{bmatrix}, \quad T = \begin{bmatrix} Q^T Q & 0 \\ 0 & -I \end{bmatrix}.$$

It is easy to verify that

$$v^T A v \geq 2\rho y^T P v \iff u^T S u \geq 0, \qquad y^T y \leq v^T Q^T Q v \iff u^T T u \geq 0.$$

Since $Q \neq 0$, we have $\lambda_{\max}(Q^T Q) > 0$. It follows that if we let $\bar{v} \in \mathbb{R}^n$ to be the unit eigenvector corresponding to the largest eigenvalue of $Q^T Q$ and $\bar{y} = \sqrt{\lambda_{\max}(Q^T Q)/2} \cdot \bar{v} \in \mathbb{R}^n$, then $\bar{y}^T \bar{y} < \bar{v}^T Q^T Q \bar{v}$, or equivalently, $\bar{u}^T T \bar{u} > 0$, where $\bar{u} = (\bar{v}, \bar{y}) \in \mathbb{R}^{2n}$. Thus, by the S–Lemma, we conclude that (8) is equivalent to $S - \lambda T \succeq 0$ for some $\lambda \geq 0$. This completes the proof.

ENGG 5501: Foundations of Optimization

Homework Set 1

2014–15 First Term

Instructor: Anthony Man–Cho So

Due: September 19, 2014

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (15pts). Suppose that Z is a random variable taking values in the set $\{0,1,\ldots,K\}$, and that the first two moments of Z, namely, $\mathbb{E}[Z]=\sum_{i=0}^{K} ip_i$ and $\mathbb{E}[Z^2]=\sum_{i=0}^{K} i^2p_i$, are given, where $\Pr(Z=i)=p_i$ for $i=0,1,\ldots,K$. We are interested in finding upper and lower bounds on the fourth moment $\mathbb{E}[Z^4]=\sum_{i=0}^{k} i^4p_i$ of Z. Give a linear programming formulation of this problem. Justify your answer.

Problem 2 (30pts). Consider the affine subspace

Doubly Stochastic Matrices

$$V_n=\left\{M\in\mathbb{R}^{n\times n}:\sum_{j=1}^{n} m_{ij}=1 \text{ for } i=1,\ldots,n;\ \sum_{i=1}^{n} m_{ij}=1 \text{ for } j=1,\ldots,n\right\}$$

of the space of $n\times n$ matrices and the set

$$\Sigma_n=\{M\in V_n: m_{ij}\geq 0 \text{ for } i,j=1,\ldots,n\}$$

of $n\times n$ doubly stochastic matrices.

(a) **(15pts).** Show that $\dim(V_n)=(n-1)^2$.

(b) **(15pts).** Show that $\mathrm{aff}(\Sigma_n)=V_n$.

Problem 3 (55pts). Let S,T be closed convex sets in $\mathbb{R}^n$ such that $S\cap T\neq\emptyset$. A problem that arises frequently in optimization is that of finding a point $x\in S\cap T$. A natural algorithm is to start with an arbitrary $x_0\in S$ and then alternately project onto S and T; i.e., compute the sequence

$$y_k=\Pi_T(x_k),\quad x_{k+1}=\Pi_S(y_k)\qquad \text{for } k=0,1,\ldots.$$

Clearly, we have $x_k\in S$ and $y_k\in T$ for $k=0,1,\ldots$. Our goal is to prove that the sequences $\{x_k\}_{k\geq 0}$ and $\{y_k\}_{k\geq 0}$ both converge to a point $x^*\in S\cap T$.

(a) **(20pts).** Let $\bar{x}\in S\cap T$ be arbitrary. Show that

$$\|y_k-\bar{x}\|_2^2\leq\|x_k-\bar{x}\|_2^2-\|x_k-y_k\|_2^2$$

and

$$\|x_{k+1}-\bar{x}\|_2^2\leq\|y_k-\bar{x}\|_2^2-\|x_{k+1}-y_k\|_2^2.$$

Hence, or otherwise, show that the sequences $\{x_k\}_{k\geq 0}$ and $\{y_k\}_{k\geq 0}$ are bounded.

(b) **(15pts).** Using the results in (a), show that the sequences $\{\|x_k-y_k\|_2\}_{k\geq 0}$ and $\{\|x_{k+1}-y_k\|_2\}_{k\geq 0}$ both converge to 0.

(c) **(20pts).** Let x^* be a limit point of the sequence $\{x_k\}_{k\geq 0}$, which exists because of the boundedness of $\{x_k\}_{k\geq 0}$. Using the result in (b), show that $x^*\in S\cap T$. Hence, or otherwise, show that $x_k\to x^*$ and $y_k\to x^*$. (*Hint: Note that x^* is defined as a limit point of $\{x_k\}_{k\geq 0}$, which means that there is a subsequence of $\{x_k\}_{k\geq 0}$ converging to x^*. Here, you are asked to show that in fact the entire sequence $\{x_k\}_{k\geq 0}$ converges to x^*.*)

Homework Set 1 Solutions

September 19, 2014

Instructor: Anthony Man–Cho So

Problem 1 (15pts). The problem is equivalent to finding a probability distribution $\{p_i\}_{i=0}^K$ on $\{0,1,\ldots,K\}$ that optimizes the fourth moment while respecting the constraints on the first and second moments. Thus, we can formulate the problem as the following LP:

$$\begin{array}{ll} \text{maximize/} & \displaystyle\sum_{i=0}^{K} i^4 p_i \\ \text{minimize} & \\ \text{subject to} & \displaystyle\sum_{i=0}^{K} i p_i = \mathbb{E}[Z], \\ & \displaystyle\sum_{i=0}^{K} i^2 p_i = \mathbb{E}\left[Z^2\right], \\ & \displaystyle\sum_{i=0}^{K} p_i = 1, \\ & p \geq 0. \end{array}$$

Here, $p = (p_0, p_1, \ldots, p_K) \in \mathbb{R}^{K+1}$ is the decision vector. The last two constraints simply state that p must be a probability distribution on $\{0,1,\ldots,K\}$. The first two constraints state that the probability distribution p must have the given first and second moments.

Problem 2 (30pts).

(a) **(15pts).** Let J be the $n \times n$ matrix whose entries are all equal to $1/n$. Clearly, we have $J \in \Sigma_n$. Now, observe that $V_n = \{J\} + L_n$, where

$$L_n = \left\{ M \in \mathbb{R}^{n \times n} : \sum_{j=1}^{n} m_{ij} = 0 \text{ for } i = 1,\ldots,n;\ \sum_{i=1}^{n} m_{ij} = 0 \text{ for } j = 1,\ldots,n \right\}$$

is a linear subspace of the space of $n \times n$ matrices. Thus, it suffices to determine $\dim(L_n)$. Towards that end, let

$$X \bullet Y = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} y_{ij} = \operatorname{tr}(X^T Y)$$

be the inner product on the space of $n \times n$ matrices. Moreover, for $i = 1,\ldots,n$, let E_i be the $n \times n$ matrix whose i–th row is all ones and all other entries are zero. Then, it is easy to verify that $M \in L_n$ if and only if M is orthogonal to the matrices $E_1,\ldots,E_n,E_1^T,\ldots,E_n^T$; i.e.,

$$E_i \bullet M = 0,\ E_i^T \bullet M = 0 \quad \text{for } i = 1,\ldots,n.$$

In particular, we have

$$\dim(V_n) = \dim(L_n) = n^2 - \dim(\mathrm{span}(E_1, \ldots, E_n, E_1^T, \ldots, E_n^T)). \qquad (1$$

We claim that

$$d = \dim(\mathrm{span}(E_1, \ldots, E_n, E_1^T, \ldots, E_n^T)) = 2n - 1. \qquad (2$$

Indeed, on one hand, we have

$$E_n^T = \sum_{i=1}^{n} E_i - \sum_{i=1}^{n-1} E_i^T$$

and hence $d \le 2n - 1$. On the other hand, let $\{\alpha_i\}_{i=1}^n$ and $\{\beta_i\}_{i=1}^{n-1}$ be such that

$$\sum_{i=1}^{n} \alpha_i E_i + \sum_{i=1}^{n-1} \beta_i E_i^T = 0.$$

By comparing both sides of the above equation, we deduce that $\alpha_i = 0$ for $i = 1, \ldots, n$ and $\beta_i = 0$ for $i = 1, \ldots, 2n-1$. It follows that the matrices $E_1, \ldots, E_n, E_1^T, \ldots, E_{n-1}^T$ are linearly independent and hence $d \ge 2n - 1$. This completes the proof of the claim.

Upon putting (1) and (2) together, we conclude that $\dim(V_n) = n^2 - (2n-1) = (n-1)^2$, as required.

(b) **(15pts).** Since V_n is an affine subspace and $\Sigma_n \subset V_n$, we clearly have $\mathrm{aff}(\Sigma_n) \subset V_n$. To prove the converse, we need to show that every $M \in V_n$ belongs to $\mathrm{aff}(\Sigma_n)$. Towards that end, let $M \in V_n$ be arbitrary. Consider the matrix

$$Q = (1 - \alpha)J + \alpha M, \qquad (3)$$

where

$$\alpha = \frac{1}{n(\max_{1 \le i,j \le n} |m_{ij} - 1/n| + 1)} > 0.$$

We claim that $Q \in \Sigma_n$. Indeed, since $M \in V_n$, we have

$$\sum_{j=1}^{n} q_{ij} = (1 - \alpha) \sum_{j=1}^{n} \frac{1}{n} + \alpha \sum_{j=1}^{n} m_{ij} = 1 \quad \text{for } i = 1, \ldots, n,$$

and similarly, $\sum_{i=1}^{n} q_{ij} = 1$ for $j = 1, \ldots, n$. Moreover, using the definition of α, we have

$$q_{ij} = \frac{1 - \alpha}{n} + \alpha \cdot m_{ij} = \frac{1}{n} + \alpha\,(m_{ij} - 1/n) > 0 \quad \text{for } i, j = 1, \ldots, n.$$

This establishes the claim. In particular, we have

$$M = \frac{\alpha - 1}{\alpha} J + \frac{1}{\alpha} Q,$$

which shows that M is an affine combination of $J, Q \in \Sigma_n$. It follows that $M \in \mathrm{aff}(\Sigma_n)$, as desired.

Remark: To understand the intuition behind the above argument, it would be helpful to visualize (3) geometrically.

Problem 3 (55pts).

(a) **(20pts).** We compute

$$\begin{aligned}\|x_k - \bar{x}\|_2^2 &= \|x_k - y_k + y_k - \bar{x}\|_2^2 \\ &= \|x_k - y_k\|_2^2 + \|y_k - \bar{x}\|_2^2 + 2(x_k - y_k)^T(y_k - \bar{x}).\end{aligned}$$

Since $y_k = \Pi_T(x_k)$ and $\bar{x} \in S \cap T$, by Theorem 8 of Handout 2, we have $(x_k - y_k)^T(y_k - \bar{x}) \geq 0$. It follows that

$$\|x_k - \bar{x}\|_2^2 \geq \|x_k - y_k\|_2^2 + \|y_k - \bar{x}\|_2^2.$$

or equivalently,

$$\|y_k - \bar{x}\|_2^2 \leq \|x_k - \bar{x}\|_2^2 - \|x_k - y_k\|_2^2. \tag{4}$$

as desired. By applying a similar argument to $\|y_k - \bar{x}\|_2^2$, we obtain

$$\|x_{k+1} - \bar{x}\|_2^2 \leq \|y_k - \bar{x}\|_2^2 - \|x_{k+1} - y_k\|_2^2. \tag{5}$$

Now, observe that (4) and (5) imply

$$\|y_k - \bar{x}\|_2^2 \leq \|x_k - \bar{x}\|_2^2 \leq \|y_{k-1} - \bar{x}\|_2^2 \leq \|x_{k-1} - \bar{x}\|_2^2 \quad \text{for } k = 1, 2, \ldots. \tag{6}$$

In particular, we have $\|x_k - \bar{x}\|_2^2 \leq \|x_0 - \bar{x}\|_2^2$ and $\|y_k - \bar{x}\|_2^2 \leq \|y_0 - \bar{x}\|_2^2$ for $k = 0, 1, \ldots$, which implies that the sequences $\{x_k\}_{k \geq 0}$ and $\{y_k\}_{k \geq 0}$ are bounded.

(b) **(15pts).** From (6), we see that the sequences $\{\|x_k - \bar{x}\|_2^2\}_{k \geq 0}$ and $\{\|y_k - \bar{x}\|_2^2\}_{k \geq 0}$ are monotonically decreasing and hence converge. Now, by (4) and (5), we have

$$\|x_k - y_k\|_2^2 \leq \|x_k - \bar{x}\|_2^2 - \|y_k - \bar{x}\|_2^2 \leq \|y_{k-1} - \bar{x}\|_2^2 - \|y_k - \bar{x}\|_2^2 \quad \text{for } k = 1, 2, \ldots.$$

Since the sequence $\{\|y_k - \bar{x}\|_2^2\}_{k \geq 0}$ converges, the rightmost expression in the above chain of inequalities converges to zero. It follows that the sequence $\{\|x_k - y_k\|_2\}_{k \geq 0}$ converges to zero. A similar argument shows that the sequence $\{\|x_{k+1} - y_k\|_2\}_{k \geq 0}$ converges to zero.

(c) **(20pts).** Since $x_k \in S$ for $k = 0, 1, \ldots$, we have $x^* \in S$ by the closedness of S. Moreover, since $\|x_k - y_k\|_2 \to 0$ and $y_k \in T$ for $k = 0, 1, \ldots$, we have $x^* \in T$ by the closedness of T. It follows that $x^* \in S \cap T$.

Now, by taking $\bar{x} = x^*$ in (b), we see that the sequence $\{\|x_k - x^*\|_2^2\}_{k \geq 0}$ is monotonically decreasing and hence converges. Since a subsequence of this sequence converges to zero (recall that x^* is a limit point of $\{x_k\}_{k \geq 0}$), it follows that the entire sequence converges to zero, i.e., $x_k \to x^*$. A similar argument shows that $y_k \to x^*$.

ENGG 5501: Foundations of Optimization | 2014–15 First Term

Homework Set 2

Instructor: Anthony Man–Cho So | Due: October 3, 2014

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (45pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function with Lipschitz continuous gradient; i.e.,

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2 \quad \text{for all } x, y \in \mathbb{R}^n$$

for some constant $L > 0$.

(a) **(15pts).** Show that

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2}\|x - y\|_2^2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

(Hint: Fix $x, y \in \mathbb{R}^n$ and apply the Fundamental Theorem of Calculus to the function $t \mapsto f(x + t(y - x))$.)

(b) **(20pts).** Using the result in (a), show that

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

(c) **(10pts).** Using the result in (b), show that

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

Problem 2 (20pts). Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ be an arbitrary norm on $\mathbb{R}^n$. The *dual norm* of $\|\cdot\|$, which is denoted by $\|\cdot\|_*$, is defined as

$$\|x\|_* = \sup_{\|d\|=1} d^T x.$$

Show that

$$\partial\|x\| = \left\{s \in \mathbb{R}^n : \|s\|_* \leq 1,\ s^T x = \|x\|\right\}.$$

Problem 3 (15pts). Given $C \in \mathcal{S}_+^n$, consider the function $f : \mathcal{S}_+^n \to \mathbb{R}$ defined by $f(X) = \text{tr}(CX^2)$. Prove that f is convex.

Problem 4 (20pts). Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, consider the polyhedron $P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$. Suppose that $P \neq \emptyset$. We say that P contains a *recession direction* $d \in \mathbb{R}^n\backslash\{0\}$ if for any $x_0 \in P$, we have $\{x \in \mathbb{R}^n : x = x_0 + \lambda d,\ \lambda \geq 0\} \subset P$. Show that the following statements are equivalent:

(i) P contains a recession direction $d \in \mathbb{R}^n$.

(ii) There exists a vector $d \in \mathbb{R}^n$ satisfying

$$Ad = 0,\ d \geq 0,\ d \neq 0.$$

(iii) P is unbounded.

ENGG 5501: Foundations of Optimization — 2014–15 First Term

Homework Set 2 Solutions

October 6, 2014

Instructor: Anthony Man-Cho So

Problem 1 (45pts).

(a) **(15pts).** 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

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

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

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

as desired.

(b) **(20pts).** Consider a fixed $x \in \mathbb{R}^n$. By the convexity of f, for any $z \in \mathbb{R}^n$, we have

$$f(z) \ge f(x) + \nabla f(x)^T(z-x),$$

or equivalently,

$$f(z) - \nabla f(x)^T z \ge f(x) - \nabla f(x)^T x. \tag{1}$$

In particular, given any $y \in \mathbb{R}^n$, by letting $z = y - (1/L)(\nabla f(y) - \nabla f(x))$ in (1) and using the result in (a), we have

$$\begin{aligned} f(x) - \nabla f(x)^T x &\le f\left(y - \frac{1}{L}(\nabla f(y) - \nabla f(x))\right) - \nabla f(x)^T \left(y - \frac{1}{L}(\nabla f(y) - \nabla f(x))\right) \\ &= f\left(y - \frac{1}{L}(\nabla f(y) - \nabla f(x))\right) - f(y) + \frac{1}{L}\nabla f(y)^T(\nabla f(y) - \nabla f(x)) \\ &\quad + f(y) - \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|_2^2 - \nabla f(x)^T y \\ &\le f(y) - \nabla f(x)^T y - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2. \end{aligned}$$

Upon rearranging, we obtain the desired conclusion.

(c) **(10pts).** Let $x, y \in \mathbb{R}^n$ be given. By the result in (b), we have

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2. \tag{2}$$

Since $x, y \in \mathbb{R}^n$ are arbitrary, we can interchange the role of x and y to get

$$f(x) \geq f(y) + \nabla f(y)^T(x - y) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2. \tag{3}$$

Now, by adding (2) and (3), we obtain

$$0 \geq (\nabla f(x) - \nabla f(y))^T(y - x) + \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2,$$

which is equivalent to

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2,$$

as desired.

Problem 2 (20pts). Consider a fixed $x \in \mathbb{R}^n$. Let $s \in \mathbb{R}^n$ be such that $\|s\|_* \leq 1$ and $s^T x = \|x\|$. By definition of the dual norm, for any $\bar{x} \in \mathbb{R}^n \backslash \{0\}$, we have

$$1 \geq \|s\|_* = \sup_{d \neq 0} \frac{d^T s}{\|d\|} \geq \frac{s^T \bar{x}}{\|\bar{x}\|}.$$

It follows that

$$\|\bar{x}\| \geq s^T \bar{x} = s^T x + s^T(\bar{x} - x) = \|x\| + s^T(\bar{x} - x).$$

Note that the above inequality is also valid at $\bar{x} = 0$, because $s^T x = \|x\|$ by assumption. Hence, we conclude that $s \in \partial\|x\|$.

Conversely, suppose that $s \in \partial\|x\|$. Consider first the case where $x \neq 0$. We have

$$\begin{aligned} 2\|x\| &= \|x + x\| \geq \|x\| + s^T x, \\ 0 &= \|x - x\| \geq \|x\| - s^T x, \end{aligned}$$

which together imply that $s^T x = \|x\|$. Since $x \neq 0$, it follows that

$$\|s\|_* = \sup_{\|d\|=1} d^T x \geq \frac{s^T x}{\|x\|} = 1.$$

We claim that $\|s\|_* = 1$. Suppose that this is not the case. Then, we have $\|s\|_* > 1$, which implies that $s^T d > 1$ for some $d \in \mathbb{R}^n$ with $\|d\| = 1$. We compute

$$\|x\| + 1 = \|x\| + \|d\| \geq \|x + d\| \geq \|x\| + s^T d > \|x\| + 1,$$

which is a contradiction. This establishes the claim.

Now, consider the case where $x = 0$. The condition $s^T x = \|x\|$ is automatically satisfied. On the other hand, let $f : \mathbb{R}^n \to \mathbb{R}_+$ be given by $f(x) = \|x\|$. The directional derivative of f at 0 in the direction $d \in \mathbb{R}^n \backslash \{0\}$ is, by definition,

$$f'(0, d) = \lim_{t \searrow 0} \frac{\|td\|}{t} = \|d\|.$$

Hence, by Theorem 16(a) of Handout 2, we have $f'(0,d) = \|d\| \geq s^T d$ for all $s \in \partial\|0\|$. This, together with the definition of the dual norm, yields

$$\|s\|_* = \sup_{d \neq 0} \frac{s^T d}{\|d\|} \leq 1,$$

as desired.

Problem 3 (15pts). Let $X_0 \in \mathcal{S}^n_+$ and $H \in \mathcal{S}^n$ be arbitrary. Define the function $g_{X_0,H} : T \to \mathbb{R}$ by

$$g_{X_0,H}(t) = f(X_0 + tH) = \operatorname{tr}\left(C(X_0 + tH)^2\right),$$

where $T = \{t \in \mathbb{R} : X_0 + \cancel{tH \in \mathcal{S}^n_+}\}$. To prove that f is convex, it suffices to prove that $g_{X_0,H}$ is convex. Towards that end, we compute

$$g_{X_0,H}(t) = \operatorname{tr}(CX_0^2) + t \cdot \operatorname{tr}(CX_0H + CHX_0) + t^2 \cdot \operatorname{tr}(CH^2).$$

In particular, $g_{X_0,H}$ is a quadratic function in t and hence is continuously differentiable. Now, observe that $\operatorname{tr}(CH^2) \geq 0$, since $C, H^2 \in \mathcal{S}^n_+$. Thus, we have

$$\frac{d^2 g_{X_0,H}(t)}{dt^2} = 2 \cdot \operatorname{tr}(CH^2) \geq 0,$$

which shows that $g_{X_0,H}$ is convex. In fact, the same proof shows that f is convex on $\mathcal{S}^n$.

Alternatively, one can prove the convexity of f on $\mathcal{S}^n$ directly by first principles. Indeed, observe that for any $X, Y \in \mathcal{S}^n$, the following identity holds:

$$\frac{X^2 + Y^2}{2} - \left(\frac{X+Y}{2}\right)^2 = \frac{1}{4}(X^2 + Y^2 - XY - YX) = \frac{1}{4}(X - Y)^2 \succeq 0.$$

Since $C \succeq 0$, it follows that

$$f\left(\frac{X+Y}{2}\right) = \operatorname{tr}\left[C\left(\frac{X+Y}{2}\right)^2\right] \leq \operatorname{tr}\left[\frac{C(X^2+Y^2)}{2}\right] = \frac{f(X) + f(Y)}{2},$$

as desired.

Yet another approach is to use calculus, as the function f is twice continuously differentiable. Using the results in [1] and the symmetry of C, we compute

$$\nabla f(X) = CX + XC, \quad \nabla^2 f(X) = I \otimes C + C \otimes I.$$

Since $C, I \succeq 0$, we have $I \otimes C, C \otimes I \succeq 0$. It follows that $\nabla^2 f(X) \succeq 0$ for all $X \in \mathcal{S}^n$, as desired.

Problem 4 (20pts). We show that (i)$\Rightarrow$(iii)$\Rightarrow$(ii)$\Rightarrow$(i).

Step 1: (i)$\Rightarrow$(iii). Suppose that P contains a recession direction $d \in \mathbb{R}^n$. Then, for any $x_0 \in P$, we have $x(\lambda) = x_0 + \lambda d \in P$ for all $\lambda \geq 0$. Let $i \in \{1, \ldots, n\}$ be an index such that $d_i \neq 0$. Then, we have

$$\|x(\lambda)\|_2 \geq |(x_0)_i + \lambda d_i| \to \infty \quad \text{as } \lambda \to \infty.$$

It follows that P is unbounded.

Step 2: (iii)$\Rightarrow$(ii). We establish the contrapositive. Suppose that the system

$$Ad = 0, \; d \geq 0, \; d \neq 0$$

has no solution. Then, by Gordan's theorem (Corollary 2 of Handout 3), there exists a $v \in \mathbb{R}^m$ such that $A^T v \geq e$. Now, let $x \in P$ be arbitrary. Since $Ax = b$ and $x \geq 0$, we have

$$b^T v = x^T A^T v \geq e^T x = \|x\|_1.$$

In particular, we see that P is bounded.

Step 3: (ii)$\Rightarrow$(i). Suppose that $d \in \mathbb{R}^n$ satisfies

$$Ad = 0, \; d \geq 0, \; d \neq 0.$$

Let $x_0 \in P$ and $\lambda \geq 0$ be arbitrary. Consider the point $x(\lambda) = x_0 + \lambda d \in \mathbb{R}^n$. Since $Ax_0 = b$ and $x_0 \geq 0$, we have

$$Ax(\lambda) = Ax_0 + \lambda A d = b, \; x(\lambda) = x_0 + \lambda d \geq 0;$$

i.e., $x(\lambda) \in P$. It follows that P contains the recession direction d.

References

[1] M. Brookes. The Matrix Reference Manual. Available online at `http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html`, 2011.

ENGG 5501: Foundations of Optimization — 2014–15 First Term

Homework Set 3

Instructor: Anthony Man-Cho So — Due: October 27, 2014

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (30pts). Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$ be given. Consider the following LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \geq b, \\ & x \geq 0. \end{array} \tag{P}$$

(a) **(10pts).** Write down the dual of (P).

(b) **(20pts).** Suppose that (P) has an optimal solution. Using the result in (a), or otherwise, show that (P) can be formulated as a standard form LP whose optimal value is zero.

Problem 2 (15pts). Show that for $n \geq 2$, the Lorentz cone

$$Q^{n+1} = \{(t,x) \in \mathbb{R}\times\mathbb{R}^n : t \geq \|x\|_2\}$$

is non-polyhedral.

Problem 3 (25pts). Let $A_1,\ldots,A_m \in \mathcal{S}^n_+$, $\alpha_1,\ldots,\alpha_m > 0$, and $\beta_1,\ldots,\beta_m > 0$ be given. Consider the following SDP:

$$(Q) \quad \begin{array}{lll} \inf & \displaystyle\sum_{i=1}^m \operatorname{tr}(Z_i) & \\ \text{subject to} & \operatorname{tr}\left[A_i\left(\alpha_i Z_i - \displaystyle\sum_{j\neq i} Z_j\right)\right] \geq \beta_i & \text{for } i = 1,\ldots,m, \\ & Z_1,\ldots,Z_m \in \mathcal{S}^n_+. & \end{array}$$

(a) **(15pts).** Write down the dual of (Q).

(b) **(10pts).** Using the result in (a), show that the dual is strictly feasible.

Problem 4 (30pts). Consider the problem

$$\begin{array}{ll} \text{minimize} & g_0(x) \\ \text{subject to} & g_i(x) \leq 0 \quad \text{for } i = 1,\ldots,m, \end{array} \tag{C}$$

where $g_0, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are *convex* functions. Suppose that (C) is feasible.

(a) **(15pts).** Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary convex function. Suppose there exists an $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x}) \leq 0$. Define

$$C = \{(t, x) \in \mathbb{R}_{++} \times \mathbb{R}^n : t \cdot f(x/t) \leq 0\}.$$

Show that C is a pointed cone.

(b) **(15pts).** Hence, or otherwise, show that (C) can be formulated as a CLP problem. Justify your answer.

REMARKS: Problem (C) is a so-called **convex optimization problem**. As shown in (b), any convex optimization problem of the form (C) can be formulated as a CLP problem.

ENGG 5501: Foundations of Optimization | 2014–15 First Term

Homework Set 3 Solutions

Instructor: Anthony Man-Cho So | October 27, 2014

Problem 1 (30pts).

(a) **(10pts).** By introducing a slack variable $s \in \mathbb{R}^m$, we see that (P) is equivalent to

$$\begin{array}{ll} \text{minimize} & q^T z \\ \text{subject to} & Qz = b, \\ & z \geq 0, \end{array}$$

where $q = (c, 0) \in \mathbb{R}^n \times \mathbb{R}^m$, $Q = [A \ \ -I] \in \mathbb{R}^{m \times (m+n)}$, and $z = (x, s) \in \mathbb{R}^n \times \mathbb{R}^m$. The dual of the above LP is given by

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & Q^T y \leq q. \end{array}$$

Using the definition of Q and q, we conclude that the dual of (P) is given by

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \leq c, \\ & y \geq 0. \end{array} \tag{D}$$

(b) **(20pts).** Consider the following LP:

$$\begin{array}{ll} \text{minimize} & c^T x - b^T y \\ \text{subject to} & Ax \geq b, \\ & A^T y \leq c, \\ & x, y \geq 0. \end{array} \tag{1}$$

Since (P) has an optimal solution by assumption, the LP strong duality theorem implies that (D) also has an optimal solution, and the optimal values of (P) and (D) are equal. Thus, if $x^* \in \mathbb{R}^n$ and $y^* \in \mathbb{R}^m$ are optimal for (P) and (D), respectively, then (x^*, y^*) is feasible for (1) and has zero objective value. We claim that (x^*, y^*) is in fact optimal for (1). Indeed, it is clear that $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is feasible for (1) if and only if $\bar{x}$ and $\bar{y}$ are feasible for (P) and (D), respectively. Hence, by the feasibility of (1) and the LP weak duality theorem, the optimal value of (1) is at least zero. This establishes the claim. Conversely, if $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ is optimal for (1), then x^* and y^* are feasible for (P) and (D), respectively, and $c^T x^* - b^T y^* = 0$. It follows again from the LP strong duality theorem that x^* and y^* are optimal for (P) and (D), respectively.

From the above discussion, we see that (P) is equivalent to (1), and the optimal value of (1) is zero. Now, it is straightforward to put problem (1) into standard form:

$$\begin{array}{ll} \text{minimize} & h^T w \\ \text{subject to} & Hw = \theta, \\ & w \geq 0. \end{array}$$

Here, $h = (c, -b, 0, 0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$,

$$H = \begin{bmatrix} A & 0 & -I & 0 \\ 0 & A^T & 0 & I \end{bmatrix} \in \mathbb{R}^{(m+n)\times(2m+2n)},$$

$\theta = (b, c) \in \mathbb{R}^m \times \mathbb{R}^n$, and $w = (x, y, u, v) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$.

Problem 2 (15pts). Observe that if P is a polyhedron and H is a hyperplane, then $P \cap H$ is also a polyhedron, because $P \cap H$ can be expressed as a finite intersection of halfspaces. Now, consider the hyperplane $H = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n : t = 1\}$, where $n \geq 2$. Clearly, we have

$$Q^{n+1} \cap H = \{(1, x) \in \mathbb{R} \times \mathbb{R}^n : 1 \geq \|x\|_2\},$$

which shows that $Q^{n+1} \cap H$ is an n-dimensional unit ball lying on the plane H. In particular, it is straightforward to verify that every point $(1, x) \in Q^{n+1} \cap H$ satisfying $\|x\|_2 = 1$ is an extreme point of $Q^{n+1} \cap H$; cf. p. 5 of Handout 2. This implies that $Q^{n+1} \cap H$ is not polyhedral, and hence Q^{n+1} is not polyhedral.

Problem 3 (25pts).

(a) **(15pts).** Observe that (Q) is equivalent to

$$\begin{array}{lll} \inf & D \bullet Z & \\ \text{subject to} & H_i \bullet Z = \beta_i & \text{for } i = 1, \dots, m, \\ & Z \in \mathcal{S}_+^{m(n+1)}, & \end{array} \tag{2}$$

where

$$\begin{aligned} D &= \text{BlkDiag}(I, \dots, I, 0) \in \mathcal{S}^{m(n+1)}, \\ H_i &= \text{BlkDiag}(-A_i, \dots, -A_i, \underbrace{\alpha_i A_i}_{i^{th}}, -A_i, \dots, -A_i, 0, \dots, 0, -1, 0, \dots, 0) \in \mathcal{S}^{m(n+1)}, \end{aligned}$$

and $\text{BlkDiag}(Q_1, \dots, Q_l)$ denotes the block diagonal matrix whose i-th diagonal block is Q_i, for $i = 1, \dots, l$. Indeed, since $D, H_1, \dots, H_m$ are block diagonal, every feasible solution $\bar{Z}$ to (2) gives rise to a block diagonal feasible solution $\bar{Z}' = \text{BlkDiag}(\bar{Z}'_1, \dots, \bar{Z}'_m, \bar{s}'_1, \dots, \bar{s}'_m)$ to (2) whose objective value is equal to that of $\bar{Z}$. Now, the dual of (2) is given by

$$\begin{array}{ll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i \\ \text{subject to} & D - \displaystyle\sum_{i=1}^m y_i H_i \in \mathcal{S}_+^{m(n+1)}, \end{array}$$

which, using the structure of $D, H_1, \dots, H_m$, is equivalent to

$$\begin{array}{lll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i & \\ \text{subject to} & I - \alpha_i y_i A_i + \displaystyle\sum_{j \neq i} y_j A_j \in \mathcal{S}_+^n & \text{for } i = 1, \dots, m, \\ & y \geq 0. & \end{array} \tag{3}$$

(b) **(10pts).** Since $A_1, \dots, A_m \in \mathcal{S}^n_+$ and $y \geq 0$, we have

$$I + \sum_{j \neq i} y_j A_j \succeq I$$

for $i = 1, \dots, m$. Hence, to show that (3) is strictly feasible, it suffices to choose $y_i > 0$ such that $I \succ \alpha_i y_i A_i$ for $i = 1, \dots, m$. Since $\alpha_i > 0$ and we may assume that $A_i \neq 0$, we have $\alpha_i \lambda_{\max}(A_i) > 0$. Hence, any $y \in \mathbb{R}^m$ satisfying

$$0 < y_i < \frac{1}{\alpha_i \lambda_{\max}(A_i)} \quad \text{for } i = 1, \dots, m$$

is a strictly feasible solution to (3).

Problem 4 (20pts).

(a) **(15pts).** By assumption, we have $(1, \bar{x}) \in C$, and hence C is non-empty. Moreover, given the pairs $(t_1, x_1), (t_2, x_2) \in C$, we have

$$\begin{aligned}
(t_1 + t_2) f\left(\frac{x_1 + x_2}{t_1 + t_2}\right) &= (t_1 + t_2) f\left(\frac{t_1(x_1/t_1) + t_2(x_2/t_2)}{t_1 + t_2}\right) \\
&\leq t_1 f(x_1/t_1) + t_2 f(x_2/t_2) \\
&\leq 0,
\end{aligned}$$

where the first inequality follows from the convexity of f. Thus, C is closed under addition. Now, if $(t, x) \in C$, then for any $\alpha > 0$, we have

$$\alpha t f\left(\frac{\alpha x}{\alpha t}\right) = \alpha t f(x/t) \leq 0,$$

i.e., $\alpha(t, x) \in C$, which implies that C is a cone. Finally, C is vacuously pointed, since there does not exist any (t, x) with $(t, x) \in C$ and $-(t, x) \in C$. Hence, we conclude that C is a pointed cone.

(b) **(15pts).** Note that the given optimization problem can be written as

$$\begin{array}{lll}
\text{minimize} & \theta & \\
\text{subject to} & g_0(x) \leq \theta, & \\
 & g_i(x) \leq 0 & \text{for } i = 1, \dots, m.
\end{array} \tag{4}$$

Now, define

$$\begin{aligned}
C_0 &= \{(t, \theta, x) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n : t > 0,\ t(g_0(x/t) - \theta/t) \leq 0\}, \\
C_i &= \{(t, \theta, x) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n : t > 0,\ t g_i(x/t) \leq 0\} \qquad \text{for } i = 1, \dots, m.
\end{aligned}$$

Then, problem (4) is equivalent to

$$\begin{array}{ll}
\text{minimize} & (0, 1, 0)^T (t, \theta, x) \\
\text{subject to} & (1, 0, 0)^T (t, \theta, x) = 1, \\
 & (t, \theta, x) \in C_0 \cap C_1 \cap \cdots \cap C_m.
\end{array} \tag{5}$$

Since problem (4) is feasible, by the result in (a), $C_0, C_1, \dots, C_m$ are pointed cones. It follows that (5) is a CLP problem.

ENGG 5501: Foundations of Optimization

2014–15 First Term

Homework Set 4

Instructor: Anthony Man-Cho So

Due: November 7, 2014

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (45pts). Let $P, Q \in \mathcal{S}^n$ be given. Suppose there exists a $\bar{u} \in \mathbb{R}^n$ such that $\bar{u}^T P \bar{u} > 0$. Consider the following statements:

(I) $u^T Q u \geq 0$ whenever $u^T P u \geq 0$, where $u \in \mathbb{R}^n$.

(II) There exists a $\lambda \geq 0$ such that $Q \succeq \lambda P$.

(a) **(5pts).** Show that (II) implies (I).

To prove the converse, consider the semidefinite program

$$\text{(SDP)} \quad \theta^* = \max_{\mu,\lambda} \{\mu : Q - \lambda P \succeq \mu I, \lambda \geq 0\}.$$

(b) **(10pts).** Write down the dual (SDD) of (SDP) and explain why (SDD) has an optimal solution.

(c) **(15pts).** Let $A, B \in \mathcal{S}^n$ be such that $\mathrm{tr}(A) \geq 0$ and $\mathrm{tr}(B) < 0$. Show that there exists a vector $w \in \mathbb{R}^n$ that satisfies $w^T A w \geq 0$ and $w^T B w < 0$.

(d) **(15pts).** Using the results in (b) and (c), show that (I) implies (II).

Problem 2 (55pts).

(a) **(10pts).** Show that the set $\mathcal{CP}_n = \mathrm{conv}\left(\{vv^T : v \in \mathbb{R}^n_+\}\right)$ is a pointed cone.

(b) **(10pts).** Let $u \in \mathbb{R}^n$ and $U \in \mathcal{S}^n$ be given. Define

$$W = \begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix}.$$

Show that $\mathrm{rank}(W) = 1$ if and only if $W = (1, u)(1, u)^T$.

Let $Q \in \mathcal{S}^n$ and $c \in \mathbb{R}^n$ be given. Consider the following problem:

$$\begin{array}{lll} \text{minimize} & \frac{1}{2} x^T Q x + c^T x & \\ \text{subject to} & x_i \in \{0, 1\} & \text{for } i = 1, \ldots, n. \end{array} \qquad (1)$$

Our goal now is to show that the discrete optimization problem (1) can be reformulated as a linear optimization problem over the pointed cone $\mathcal{CP}_{2n+1}$. In other words, problem (1) has a convex reformulation![1]

[1] Since problem (1) is NP-hard in general, we see that not all convex optimization problems are easy to solve.

(c) **(5pts).** Using the result in (b), show that problem (1) is equivalent to

$$\begin{array}{lll} \text{minimize} & \frac{1}{2}Q \bullet X + c^T x & \\ \text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & & \text{(I)} \\ & x_i + s_i = 1 & \text{for } i = 1, \ldots, n, & \text{(II)} \\ & X_{ii} + 2Y_{ii} + S_{ii} = 1 & \text{for } i = 1, \ldots, n, & \text{(III)} \\ & x_i = X_{ii},\ s_i = S_{ii} & \text{for } i = 1, \ldots, n, & \text{(IV)} \\ & \text{rank}(Z) = 1. & & \end{array} \quad (2$$

(d) **(20pts).** Let

$$\begin{aligned} S_1 &= \text{conv}\left(\{(1, x, s)(1, x, s)^T : x_i, s_i \in \{0, 1\},\ x_i + s_i = 1 \text{ for } i = 1, \ldots, n\}\right), \\ S_2 &= \left\{Z \in \mathcal{S}^{2n+1} : Z \text{ satisfies (2–I) — (2–IV)}\right\}. \end{aligned}$$

Show that $S_1 = S_2$.

(e) **(10pts).** By relaxing the rank constraint in (2), we obtain a convex relaxation of problem (1 which is a CLP. Using the result in (d), show that this convex relaxation is in fact equivaler to problem (1): i.e., (i) the optimal values of both problems are equal, and (ii) if

$$Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}$$

is an optimal solution to the convex relaxation, then x^* is in the convex hull of the optim solutions to problem (1).

ENGG 5501: Foundations of Optimization — 2014–15 First Term

Homework Set 4 Solutions

Instructor: Anthony Man–Cho So — November 7, 2014

Problem 1 (45pts).

(a) **(5pts).** Suppose that there exists a $\lambda \geq 0$ such that $Q \succeq \lambda P$. Then, for any $u \in \mathbb{R}^n$, we have $u^T Q u \geq \lambda \cdot u^T P u$. In particular, if $u^T P u \geq 0$, then $u^T Q u \geq 0$.

(b) **(10pts).** Observe that (SDP) is equivalent to

$$
\begin{array}{ll}
\text{maximize} & \mu \\
\text{subject to} & \begin{bmatrix} Q & \\ & 0 \end{bmatrix} - \lambda \begin{bmatrix} P & \\ & -1 \end{bmatrix} - \mu \begin{bmatrix} I & \\ & 0 \end{bmatrix} \succeq 0,
\end{array}
$$

which is a semidefinite program in dual standard form. Hence, the dual of (SDP) is given by

$$
\text{(SDD)} \quad \begin{array}{ll}
\text{minimize} & \mathrm{tr}(QX) \\
\text{subject to} & \mathrm{tr}(X) = 1, \\
& \mathrm{tr}(PX) \geq 0, \\
& X \succeq 0.
\end{array}
$$

Since there exists a $\bar{u} \in \mathbb{R}^n$ such that $\bar{u}^T P \bar{u} > 0$, the matrix $\bar{X} = \bar{u}\bar{u}^T / \|\bar{u}\|_2^2$ is feasible for (SDD). In particular, by the CLP weak duality theorem, we have $\theta^* \leq \mathrm{tr}(Q\bar{X}) < \infty$. Moreover, observe that $(\lambda, \mu) = (1, \lambda_{\min}(Q - P) - 1)$ is a strictly feasible solution to (SDP). Hence, by the CLP strong duality theorem, we conclude that (SDD) has an optimal solution.

(c) **(15pts).** Let $B = U \Lambda U^T$ be the spectral decomposition of B. Then, we have $\mathrm{tr}(B) = \mathrm{tr}(\Lambda) < 0$. Now, let $\xi \in \mathbb{R}^n$ be a random vector whose entries are independent and take the values ± 1 with equal probability. Upon setting $\eta = U\xi \in \mathbb{R}^n$, we have

$$
\eta^T B \eta = \xi^T U^T B U \xi = \xi^T \Lambda \xi = \mathrm{tr}(\Lambda) < 0,
$$

$$
\mathbb{E}\left[\eta^T A \eta\right] = \mathbb{E}\left[\xi^T U A U^T \xi\right] = \mathbb{E}\left[\mathrm{tr}(U A U^T \xi \xi^T)\right] = \mathrm{tr}(U A U^T) = \mathrm{tr}(A) \geq 0.
$$

In particular, this implies that there exists a $\hat{\xi} \in \{-1, 1\}^n$ such that $w = U\hat{\xi} \in \mathbb{R}^n$ satisfies $w^T A w \geq 0$ and $w^T B w < 0$. This completes the proof.

(d) **(15pts).** Suppose that $u^T Q u \geq 0$ whenever $u^T P u \geq 0$, where $u \in \mathbb{R}^n$. Let $X^* \succeq 0$ be an optimal solution to (SDD), whose existence is guaranteed by the result in (b). Then, we have $\theta^* = \mathrm{tr}(QX^*)$ by the CLP strong duality theorem. Now, we claim that $\theta^* \geq 0$. Suppose to the contrary that $\theta^* < 0$. Let

$$
A = (X^*)^{1/2} P (X^*)^{1/2}, \quad B = (X^*)^{1/2} Q (X^*)^{1/2}.
$$

Note that $A, B \in \mathcal{S}^n$. Moreover, we have $\text{tr}(A) = \text{tr}(PX^*) \geq 0$ and $\text{tr}(B) = \text{tr}(Q$
$\theta^* < 0$. Hence, by the result in (c), there exists a vector $w \in \mathbb{R}^n$ such that $w^T Aw \geq$
$w^T Bw < 0$. However, this implies that the vector $u = (X^*)^{1/2} w \in \mathbb{R}^n$ satisfies $u^T P$
and $u^T Qu < 0$, which is a contradiction. Hence, we have $\theta^* \geq 0$, which, together wi
construction of (SDP), implies that $Q \succeq \lambda P$ for some $\lambda \geq 0$, as desired.

Problem 2 (55pts).

(a) **(10pts).** The non-emptiness of $\mathcal{CP}_n$ follows from the fact that $0 \in \mathcal{CP}_n$. Now, for any
and $v \geq 0$, we have $\sqrt{\alpha} v \geq 0$, which implies that $\alpha vv^T = (\sqrt{\alpha} v)(\sqrt{\alpha} v)^T \in \mathcal{CP}_n$.
arbitrary $C \in \mathcal{CP}_n$, there exist vectors $v_1, \ldots, v_K \geq 0$ and scalars $\lambda_1, \ldots, \lambda_K \geq 0$ suc
$C = \sum_{k=1}^N \lambda_k v_k v_k^T$ and $\sum_{k=1}^K \lambda_k = 1$. It follows that $\alpha C = \sum_{k=1}^K \lambda_k (\sqrt{\alpha} v_k)(\sqrt{\alpha} v_k)^T$
for any $\alpha > 0$. Hence, we conclude that $\mathcal{CP}_n$ is a cone. As a corollary, we see that
closed under addition. Indeed, since $\mathcal{CP}_n$ is convex by definition and is a cone, for ar
$C_1, C_2 \in \mathcal{CP}_n$, we have

$$C_1 + C_2 = 2\left(\frac{C_1 + C_2}{2}\right) \in \mathcal{CP}_n.$$

Finally, it is clear that every matrix in $\mathcal{CP}_n$ is non-negative. It follows that if $C, -C$
then $C = 0$; i.e., $\mathcal{CP}_n$ is pointed.

(b) **(10pts).** Clearly, if $W = (1, u)(1, u)^T$, then $\text{rank}(W) = 1$. Conversely, suppos
$\text{rank}(W) = 1$. Since $W \in \mathcal{S}^{n+1}$, it admits a spectral decomposition $W = \lambda \bar{v}\bar{v}^T$,
$\bar{v} = (v_0, v) \in \mathbb{R}^{n+1}$ and $\lambda \neq 0$. Thus, we have

$$\begin{bmatrix} 1 & u^T \\ u & U \end{bmatrix} = \begin{bmatrix} \lambda v_0^2 & \lambda v_0 v^T \\ \lambda v_0 v & \lambda vv^T \end{bmatrix},$$

which gives $\lambda v_0^2 = 1$ and $\lambda v_0 v = u$. In particular, we see that $\lambda > 0$ and $v_0 \neq 0$. This i
that

$$U = \lambda vv^T = \lambda \left(\frac{u}{\lambda v_0}\right)\left(\frac{u}{\lambda v_0}\right)^T = uu^T,$$

from which we obtain $W = (1, u)(1, u)^T$.

(c) **(5pts).** The given problem is equivalent to

$$\begin{array}{lll} \text{minimize} & \frac{1}{2} x^T Q x + c^T x & \\ \text{subject to} & x_i + s_i = 1 & \text{for } i = 1, \ldots, n, \\ & (x_i + s_i)^2 = 1 & \text{for } i = 1, \ldots, n, \\ & x_i = x_i^2, \ s_i = s_i^2 & \text{for } i = 1, \ldots, n. \end{array}$$

Now, set $Z = (1, x, s)(1, x, s)^T \in \mathcal{S}_+^{2n+1}$. Since $x, s \geq 0$, we have $Z \in \mathcal{CP}_{2n+1}$. Moreo
the result in (b), we have

$$Z = (1, x, s)(1, x, s)^T \iff Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix}, \ \text{rank}(Z) = 1.$$

Since $X = xx^T$, $Y = xs^T$, and $S = ss^T$, we conclude that problem (1) is equivalent to

$$
\begin{array}{lll}
\text{minimize} & \frac{1}{2}Q \bullet X + c^T x & \\
\text{subject to} & Z = \begin{bmatrix} 1 & x^T & s^T \\ x & X & Y \\ s & Y^T & S \end{bmatrix} \in \mathcal{CP}_{2n+1}, & \text{(I)} \\
& x_i + s_i = 1 \qquad \text{for } i = 1, \ldots, n, & \text{(II)} \\
& X_{ii} + 2Y_{ii} + S_{ii} = 1 \qquad \text{for } i = 1, \ldots, n, & \text{(III)} \\
& x_i = X_{ii},\ s_i = S_{ii} \qquad \text{for } i = 1, \ldots, n, & \text{(IV)} \\
& \operatorname{rank}(Z) = 1, &
\end{array}
\tag{2}
$$

as desired.

(d) (20pts). Let $x, s \in \{0,1\}^n$ be such that $x_i + s_i = 1$ for $i = 1, \ldots, n$. Then, we have $(1, x, s)(1, x, s)^T \in S_2$. Since S_2 is convex, we see that $S_1 \subset S_2$.

Conversely, let $Z \in S_2$ be given. By (2–I), there exist vectors $(\alpha_k, u^k, v^k) \in \mathbb{R}_+^{2n+1}$, where $k = 1, \ldots, K$, such that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$. Upon equating the entries of both sides, we have

$$\sum_{k=1}^K \alpha_k^2 = 1, \quad x = \sum_{k=1}^K \alpha_k u^k, \quad s = \sum_{k=1}^K \alpha_k v^k,$$

$$X = \sum_{k=1}^K u^k (u^k)^T, \quad Y = \sum_{k=1}^K u^k (v^k)^T, \quad S = \sum_{k=1}^K (v^k)(v^k)^T.$$

By (2–II) and (2–III), we have

$$x_i + s_i = \sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = 1 = \sum_{k=1}^K \left[(u_i^k)^2 + 2u_i^k v_i^k + (v_i^k)^2 \right] = X_{ii} + 2Y_{ii} + S_{ii},$$

or equivalently,

$$\sum_{k=1}^K \alpha_k (u_i^k + v_i^k) = \sum_{k=1}^K (u_i^k + v_i^k)^2 = 1 \quad \text{for } i = 1, \ldots, n.$$

Hence, by applying the Cauchy–Schwarz inequality to the leftmost expression in the above chain of equalities, we see that there exist scalars $\beta_1, \ldots, \beta_n$ such that

$$\beta_i \alpha_k = u_i^k + v_i^k \quad \text{for } i = 1, \ldots, n;\ k = 1, \ldots, K.$$

Now, observe that if $\alpha_k = 0$ for some $k \in \{1, \ldots, K\}$, then $u^k + v^k = 0$. Since $u^k, v^k \geq 0$, this implies that $u^k = v^k = 0$. In view of the fact that $Z = \sum_{k=1}^K (\alpha_k, u^k, v^k)(\alpha_k, u^k, v^k)^T$, we may thus assume without loss that $\alpha_k > 0$ for $k = 1, \ldots, K$. In particular, we may write

$$Z = \sum_{k=1}^K \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T,$$

where $\lambda_k = \alpha_k^2$, $x^k = u^k/\alpha_k$, and $s^k = v^k/\alpha_k$, for $k = 1,\ldots,K$. Note that $\lambda_k \geq 0$ for $k = 1,\ldots,K$, and $\sum_{k=1}^K \lambda_k = 1$. Hence, in order to show that $Z \in S_1$, it suffices to show that $x^k, s^k \in \{0,1\}^n$ for $k = 1,\ldots,K$, and $x_i^k + s_i^k = 1$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$. Towards that end, we use (2–II) to compute

$$1 = \sum_{k=1}^K \alpha_k(u_i^k + v_i^k) = \sum_{k=1}^K \beta_i \alpha_k^2 = \beta_i \quad \text{for } i = 1,\ldots,n.$$

It follows that $\alpha_k = u_i^k + v_i^k$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$. In particular, we have

$$x_i^k + s_i^k = \frac{u_i^k + v_i^k}{\alpha_k} = 1 \quad \text{for } i = 1,\ldots,n;\ k = 1,\ldots,K. \tag{3}$$

Moreover, using (2–IV), we have

$$\sum_{k=1}^K \alpha_k^2 x_i^k = \sum_{k=1}^K \alpha_k u_i^k = x_i = X_{ii} = \sum_{k=1}^K (u_i^k)^2 = \sum_{k=1}^K \alpha_k^2 (x_i^k)^2$$

and

$$\sum_{k=1}^K \alpha_k^2 s_i^k = \sum_{k=1}^K \alpha_k v_i^k = s_i = S_{ii} = \sum_{k=1}^K (v_i^k)^2 = \sum_{k=1}^K \alpha_k^2 (s_i^k)^2.$$

It follows that

$$\sum_{k=1}^K \alpha_k^2 \left[x_i^k - (x_i^k)^2\right] = 0 \quad \text{for } i = 1,\ldots,n \tag{4}$$

and

$$\sum_{k=1}^K \alpha_k^2 \left[s_i^k - (s_i^k)^2\right] = 0 \quad \text{for } i = 1,\ldots,n. \tag{5}$$

Using (3) and the fact that $x^k, s^k \geq 0$ for $k = 1,\ldots,K$, we have $0 \leq x_i^k, s_i^k \leq 1$, which implies that $x_i^k - (x_i^k)^2 \geq 0$ and $s_i^k - (s_i^k)^2 \geq 0$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$. This, together with (4) and (5), implies that $x_i^k = (x_i^k)^2$ and $s_i^k = (s_i^k)^2$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$, or equivalently, $x^k, s^k \in \{0,1\}^n$ for $k = 1,\ldots,K$. This completes the proof.

(e) **(10pts).** Let v^* and v_R^* denote the optimal values of the original problem and its convex relaxation, respectively. Furthermore, let

$$Z^* = \begin{bmatrix} 1 & (x^*)^T & (s^*)^T \\ x^* & X^* & Y^* \\ s^* & (Y^*)^T & S^* \end{bmatrix}$$

be an optimal solution to the convex relaxation. By the result in (d), we can find $x^k, s^k \in \{0,1\}^n$ and $\lambda_k \geq 0$, where $k = 1,\ldots,K$, such that $x_i^k + s_i^k = 1$ for $i = 1,\ldots,n$ and $k = 1,\ldots,K$, $\sum_{k=1}^K \lambda_k = 1$, and $Z^* = \sum_{k=1}^K \lambda_k (1, x^k, s^k)(1, x^k, s^k)^T$. In particular, we have

$$v_R^* = \frac{1}{2} Q \bullet X^* + c^T x^* = \sum_{k=1}^K \lambda_k \left[\frac{1}{2}(x^k)^T Q x^k + c^T x^k\right]. \tag{6}$$

Now, note that for $k = 1, \ldots, K$, the vector x^k is feasible for the original problem, which implies that

$$\frac{1}{2}(x^k)^T Q x^k + c^T x^k \geq v^*.$$

This, together with (6) and the fact that $v^* \geq v_R^*$, implies that $v^* = v_R^*$. Thus, for $k = 1, \ldots, K$, the vector x^k is an optimal solution to the original problem. Moreover, we have $x^* \in \mathrm{conv}\{x^1, \ldots, x^K\}$, as desired.

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (35pts). Let $A, B \in \mathcal{S}_+^n$ be given.

(a) **(5pts).** Show that $A \bullet B \geq 0$.

(b) **(15pts).** Show that $A \bullet B = 0$ if and only if $AB = 0$.

(c) **(15pts).** Show that for any $x \in \mathbb{R}^n$, $x^T A x = 0$ if and only if $Ax = 0$.

Problem 2 (50pts). This problem is a continuation of Problem 3 of Homework Set 3, which concerns the following SDP:

$$
(Q) \qquad \begin{array}{ll} \inf & \displaystyle\sum_{i=1}^m \mathrm{tr}(Z_i) \\ \text{subject to} & \mathrm{tr}\left[A_i \left(\alpha_i Z_i - \displaystyle\sum_{j \neq i} Z_j \right) \right] \geq \beta_i \quad \text{for } i = 1, \ldots, m, \\ & Z_1, \ldots, Z_m \in \mathcal{S}_+^n. \end{array}
$$

Here, $A_1, \ldots, A_m \in \mathcal{S}_+^n$, $\alpha_1, \ldots, \alpha_m > 0$, and $\beta_1, \ldots, \beta_m > 0$ are given. Suppose that (Q) has an optimal solution $\{Z_i^*\}_{i=1}^m$. Our goal is to show that there always exists an optimal solution $\{\bar{Z}_i\}_{i=1}^m$ to (Q) such that $\mathrm{rank}(\bar{Z}_i) = 1$ for $i = 1, \ldots, m$.

(a) **(5pts).** Show that for $i = 1, \ldots, m$, there exists a column $\bar{u}_i \in \mathbb{R}^n$ of $Z_i^* \in \mathcal{S}_+^n$ such that $A_i \bar{u}_i \neq 0$.

(b) **(15pts).** Show that the dual of (Q), which is given by

$$
\begin{array}{ll} \sup & \displaystyle\sum_{i=1}^m \beta_i y_i \\ \text{subject to} & I - \alpha_i y_i A_i + \displaystyle\sum_{j \neq i} y_j A_j \in \mathcal{S}_+^n \quad \text{for } i = 1, \ldots, m, \\ & y \geq 0, \end{array}
$$

has an optimal solution $\{y_i^*\}_{i=1}^m$. Hence, show that

$$
\left(I - \alpha_i y_i^* A_i + \sum_{j \neq i} y_j^* A_j \right) \bar{u}_i = 0 \quad \text{for } i = 1, \ldots, m,
$$

where $\bar{u}_1, \ldots, \bar{u}_m \in \mathbb{R}^n$ are defined in (a).

(c) **(5pts).** Define $D = \mathrm{Diag}(\alpha_1 \bar{u}_1^T A_1 \bar{u}_1, \ldots, \alpha_m \bar{u}_m^T A_m \bar{u}_m) \in \mathcal{S}_+^m$. Furthermore, define $G \in \mathbb{R}^{m \times m}$ by

$$G_{ij} = \begin{cases} \bar{u}_i^T A_j \bar{u}_i & \text{for } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

Show that D is invertible, and that $D^{-1}G$ is a non-negative matrix (i.e., every element of $D^{-1}G$ is non-negative).

(d) **(15pts).** The theory of non-negative matrices asserts that the spectral radius of $D^{-1}G$, which is defined as

$$\rho(D^{-1}G) = \max_{1 \le i \le m} \left| \lambda_i(D^{-1}G) \right|,$$

can be computed via the Collatz-Wielandt formula:

$$\rho(D^{-1}G) = \inf_{x > 0} \max_{1 \le i \le m} \frac{\left[(D^{-1}G)x \right]_i}{x_i}.$$

(Here, $\lambda_1(D^{-1}G), \ldots, \lambda_m(D^{-1}G)$ are the eigenvalues of $D^{-1}G$.) Using the results in (b) and (c), show that $\rho(D^{-1}G) < 1$. Hence, or otherwise, show that the system

$$(D - G)^T p = \beta, \quad p \ge 0, \tag{1}$$

where $\beta = (\beta_1, \ldots, \beta_m) \in \mathbb{R}_{++}^m$, has a unique solution in $p \in \mathbb{R}^m$.

(e) **(10pts).** Let $\bar{p} \in \mathbb{R}^m$ be the solution to (1). Show that $\{\bar{Z}_i\}_{i=1}^m$, where

$$\bar{Z}_i = \bar{p}_i \bar{u}_i \bar{u}_i^T \quad \text{for } i = 1, \ldots, m,$$

is an optimal solution to (Q).

Problem 3 (15pts). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \ge 0. \end{array} \tag{2}$$

Show that $\bar{x} \in \mathbb{R}^n$ is an optimal solution to (2) iff $\bar{x} \in \mathbb{R}^n$ is a solution to the following system:

$$\begin{aligned} \nabla f(x) &\ge 0, \\ x &\ge 0, \\ x^T \nabla f(x) &= 0. \end{aligned}$$

ENGG 5501: Foundations of Optimization — 2014–15 First Term

Homework Set 5 Solutions

November 21, 2014

Instructor: Anthony Man-Cho So

Problem 1 (35pts).

(a) **(5pts).** Let $A = U\Lambda U^T$ be the spectral decomposition of A, and set $\bar{B} = U^T BU$. Since $B \succeq 0$, we have $\bar{B} \succeq 0$. Now, we compute

$$A \bullet B = \Lambda \bullet \bar{B} = \sum_{i=1}^{n} \Lambda_{ii}\bar{B}_{ii}.$$

Since $A, \bar{B} \succeq 0$, we have $\Lambda_{ii}, \bar{B}_{ii} \geq 0$ for $i = 1, \ldots, n$. It follows that $A \bullet B \geq 0$.

(b) **(15pts).** Clearly, if $AB = 0$, then $A \bullet B = \mathrm{tr}(AB) = 0$. Conversely, suppose that $A \bullet B = 0$. Since $A, B \succeq 0$, there exist $A^{1/2}, B^{1/2} \succeq 0$ such that $A = A^{1/2}A^{1/2}$ and $B = B^{1/2}B^{1/2}$. Hence, we have

$$\mathrm{tr}(AB) = \mathrm{tr}\left((A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}\right) = 0.$$

Let $M = (A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}$, and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Since $M \succeq 0$, we have $\lambda_i \geq 0$ for $i = 1, \ldots, n$. Since $\mathrm{tr}(M) = \sum_{i=1}^{n} \lambda_i = 0$, we see that $\lambda_1 = \cdots = \lambda_n = 0$, which implies that $M = 0$. In particular, we have $A^{1/2}B^{1/2} = 0$, from which it follows that $AB = 0$, as desired.

(c) **(15pts).** Clearly, if $Ax = 0$, then $x^T Ax = 0$. Conversely, suppose that $x^T Ax = 0$. Let $y \in \mathbb{R}^n$ be arbitrary, and define the quadratic function $p_y : \mathbb{R} \to \mathbb{R}$ by $p_y(t) = (x + ty)^T A(x + ty)$. Since $A \succeq 0$, we have $p_y(t) \geq 0$ for all $t \in \mathbb{R}$. This, together with the fact that $p_y(0) = 0$, implies that $t^* = 0$ minimizes p_y, which in turn implies that $p_y'(0) = 0$. As $p_y'(0) = 2y^T Ax$, we obtain $y^T Ax = 0$. Since y is arbitrary, we conclude that $Ax = 0$.

Problem 2 (50pts).

(a) **(5pts).** Suppose that $Z_i^* = 0$ for some $i \in \{1, \ldots, m\}$. Then, the i-th inequality constraint becomes

$$-\sum_{j \neq i} \mathrm{tr}(A_i Z_j^*) \geq \beta_i.$$

Since $A_i, Z_j^* \succeq 0$, we have $\mathrm{tr}(A_i Z_j^*) \geq 0$ by the result in Problem 1(a). However, since $\beta_i > 0$, the above inequality cannot hold. This contradicts the feasibility of the solution $\{Z_i^*\}_{i=1}^m$.

(b) **(15pts).** By the CLP weak duality theorem, the dual of (Q) is equivalent to the following problem:

$$\begin{array}{lll} \sup & \sum_{i=1}^{m} \beta_i y_i & \\ \text{subject to} & I - \alpha_i y_i A_i + \sum_{j\neq i} y_j A_j \in \mathcal{S}_+^n \quad \text{for } i = 1,\dots,m, & \\ & \sum_{i=1}^{m} \beta_i y_i \le \sum_{i=1}^{m} \operatorname{tr}(Z_i^*), & (D) \\ & y \ge 0. & \end{array}$$

It is easy to verify that the feasible region of (D) is closed. Moreover, since $\beta_i > 0$ for $i = 1,\dots,m$ and $y \ge 0$, the feasible region of (D) is bounded. It follows that (D), which involves maximizing a linear function over a compact set, has an optimal solution $\{y_i^*\}_{i=1}^m$. By the result in Problem 1(b) and the complementarity property of the optimal solutions to (Q) and its dual, we have

$$\left(I - \alpha_i y_i^* A_i + \sum_{j\neq i} y_j^* A_j \right) Z_i^* = 0 \quad \text{for } i = 1,\dots,m.$$

Since $\bar{u}_i$ is a column of Z_i^* for $i = 1,\dots,m$, this implies that

$$\left(I - \alpha_i y_i^* A_i + \sum_{j\neq i} y_j^* A_j \right) \bar{u}_i = 0 \quad \text{for } i = 1,\dots,m,$$

as desired.

(c) **(5pts).** By the choice of $\{\bar{u}_i\}_{i=1}^m$ and the result in Problem 1(c), we have $\bar{u}_i^T A_i \bar{u}_i > 0$ for $i = 1,\dots,m$. This, together with the assumption that $\alpha_i > 0$ for $i = 1,\dots,m$, implies that D is invertible, and D^{-1} is a non-negative matrix. Moreover, since $A_i \succeq 0$ for $i = 1,\dots,m$, we see that G is also a non-negative matrix. It follows that $D^{-1}G$ is a non-negative matrix.

(d) **(15pts).** The result in (b) implies that

$$\bar{u}_i^T \left(I - \alpha_i y_i^* A_i + \sum_{j\neq i} y_j^* A_j \right) \bar{u}_i = 0 \quad \text{for } i = 1,\dots,m,$$

or equivalently,

$$(D - G) y^* = \theta, \tag{1}$$

where $y^* = (y_1^*,\dots,y_m^*) \in \mathbb{R}^m$ and $\theta = (\|\bar{u}_1\|_2^2,\dots,\|\bar{u}_m\|_2^2) \in \mathbb{R}^m$. By the result in (c), D is invertible, and hence (1) is equivalent to

$$(I - D^{-1}G) y^* = D^{-1}\theta. \tag{2}$$

Since $D^{-1}G \ge 0$, $D^{-1}\theta > 0$, and $y^* \ge 0$, upon considering the i-th row of (2), we see that $y_i^* > 0$ and thus

$$0 \le \frac{[(D^{-1}G)y^*]_i}{y_i^*} = 1 - \frac{(D^{-1}\theta)_i}{y_i^*} < 1 \quad \text{for } i = 1,\dots,m.$$

It follows from the Collatz–Wielandt formula that $\rho(D^{-1}G) < 1$. In particular, this implies that $I - D^{-1}G$ is invertible, and we can write

$$(I - D^{-1}G)^{-1} = \sum_{i=0}^{\infty}(D^{-1}G)^i.$$

Now, the system $(D-G)^T p = \beta$, where $\beta > 0$, is equivalent to

$$(I - D^{-1}G)^T Dp = \beta.$$

This implies that p is uniquely given by

$$p = D^{-1}(I - D^{-1}G)^{-T}\beta = D^{-1}\left(\sum_{i=0}^{\infty}(D^{-1}G)^i\right)^T \beta.$$

Since $D^{-1} \geq 0$, $D^{-1}G \geq 0$, and $\beta > 0$, it follows that $p \geq 0$. Hence, we conclude that the system

$$(D-G)^T p = \beta, \quad p \geq 0 \tag{3}$$

has a unique solution in $p \in \mathbb{R}^m$.

(e) **(10pts).** Since $\bar{p}$ solves (3), for $i = 1, \dots, m$, we have $\bar{Z}_i = \bar{p}_i \bar{u}_i \bar{u}_i^T \succeq 0$ and

$$\begin{aligned}
\operatorname{tr}\left[A_i\left(\alpha_i \bar{Z}_i - \sum_{j\neq i}\bar{Z}_j\right)\right] &= \bar{p}_i\alpha_i\bar{u}_i^T A_i \bar{u}_i - \sum_{j\neq i}\bar{p}_j\bar{u}_j^T A_i \bar{u}_j \\
&= \bar{p}_i D_{ii} - (G^T\bar{p})_i \\
&= \left[(D-G)^T\bar{p}\right]_i \\
&= \beta_i.
\end{aligned}$$

It follows that $\{\bar{Z}_i\}_{i=1}^m$ is feasible for (Q). Moreover, observe that

$$\left(I - \alpha_i y_i^* A_i + \sum_{j\neq i} y_j^* A_j\right)\bar{Z}_i = \bar{p}_i\left(I - \alpha_i y_i^* A_i + \sum_{j\neq i} y_j^* A_j\right)\bar{u}_i\bar{u}_i^T = 0 \quad \text{for } i = 1, \dots, m$$

and

$$y_i^*\left(\operatorname{tr}\left[A_i\left(\alpha_i\bar{Z}_i - \sum_{j\neq i}\bar{Z}_j\right)\right] - \beta_i\right) = 0 \quad \text{for } i = 1, \dots, m.$$

Since $\{y_i^*\}_{i=1}^m$ is feasible for the dual of (Q), we see that $\{\bar{Z}_i\}_{i=1}^m$ and $\{y_i^*\}_{i=1}^m$ form a complementary pair of solutions to (Q) and its dual. Thus, we conclude from the CLP strong duality theorem that $\{\bar{Z}_i\}_{i=1}^m$ is optimal for (Q).

REMARK: This problem arises in the study of optimal transmit beamforming schemes in multiuser multi-antenna systems. We refer the interested reader to [1, 2] for further details.

Problem 3 (15pts). Recall that we have the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0, \end{array} \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function. Now, let $\bar{x} \in \mathbb{R}^n$ be an optimal solution to (4). Since problem (4) satisfies the Slater condition, the vector $\bar{x}$ satisfies the KKT conditions

$$\begin{aligned} \nabla f(\bar{x}) &\geq 0, \\ \bar{x} &\geq 0, \\ \bar{x}^T \nabla f(\bar{x}) &= 0. \end{aligned} \tag{5}$$

Conversely, suppose that $\bar{x} \in \mathbb{R}^n$ satisfies the conditions in (5). Since f is convex, for any $x \geq 0$, we have

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) \geq f(\bar{x}).$$

It follows that $\bar{x}$ is an optimal solution to (4).

References

[1] M. Bengtsson and B. Ottersten. Optimum and Suboptimum Transmit Beamforming. In L. C. Godara, editor, *Handbook of Antennas in Wireless Communications*, The Electrical Engineering and Applied Signal Processing Series. CRC Press LLC, Boca Raton, Florida, 2001.

[2] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten. Convex Optimization-Based Beamforming: From Receive to Transmit and Network Designs. *IEEE Signal Processing Magazine*, 27(3):62–75, 2010.

ENGG 5501: Foundations of Optimization | **2013–14 First Term**

Homework Set 1

Instructor: Anthony Man–Cho So | Due: September 18, 2013

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Let $c, f \in \mathbb{R}^n$, $d, g \in \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ be given. Consider the following problem:

$$\begin{array}{ll} \text{minimize} & \dfrac{c^T x + d}{f^T x + g} \\ \text{subject to} & Ax \le b, \\ & f^T x + g \ge 0. \end{array}$$

Here, we assume that $a/0 = +\infty$ if $a > 0$, and $a/0 = -\infty$ if $a \le 0$. Give an equivalent linear programming formulation of the above problem. Justify your answer.

Problem 2 (40pts).

(a) **(20pts).** Let $A \subset \mathbb{R}^n$ be a symmetric convex set; i.e., A is convex, and $-x \in A$ whenever $x \in A$. Furthermore, let $t > 1$ be given. Show that

$$\frac{2}{t+1}(\mathbb{R}^n \backslash (tA)) + \frac{t-1}{t+1} A \subset \mathbb{R}^n \backslash A.$$

(Recall that for any $A, B \subset \mathbb{R}^n$, $A + B = \{x + y : x \in A, y \in B\}$, and $tA = \{tx : x \in A\}$.)

Let μ be a probability measure on $\mathbb{R}^n$. We say that μ is *logconcave* if

$$\mu(\alpha S_1 + (1-\alpha)S_2) \ge \mu(S_1)^\alpha \mu(S_2)^{1-\alpha}$$

holds for any measurable sets $S_1, S_2 \subset \mathbb{R}^n$ and scalar $\alpha \in (0,1)$.

(b) **(10pts).** Let μ be the standard Gaussian measure on $\mathbb{R}^n$; i.e.,

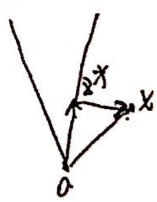
$$\mu(S) = \frac{1}{(2\pi)^{n/2}} \int_S \exp\left(-\|x\|_2^2/2\right) dx \qquad \text{for all measurable } S \subset \mathbb{R}^n.$$

Show that μ is logconcave.

(c) **(10pts).** Let μ be a logconcave probability measure on $\mathbb{R}^n$. Furthermore, let $A \subset \mathbb{R}^n$ be a measurable symmetric convex set such that $\mu(A) = \theta > 1/2$. Use the result in (a) to show that for any $t > 1$,

$$\mu(\mathbb{R}^n \backslash (tA)) \le \theta \left(\frac{1-\theta}{\theta}\right)^{(t+1)/2}.$$

Problem 3 (15pts). Let $K \subset \mathbb{R}^n$ be a non–empty closed convex cone, and let $x \in \mathbb{R}^n$ be arbitrary. Show that $z^* = \Pi_K(x)$ iff $z^* \in K$, $x - z^* \in K^\circ$, and $(x - z^*)^T z^* = 0$.



Problem 4 (25pts). Consider the set $C^n = \{X \in S^n : v^T X v \ge 0 \text{ for all } v \ge 0\}$.

(a) **(10pts).** Show that C^n is a non–empty closed convex cone.

(b) **(15pts).** Let

$$(C^n)^* = \{Y \in \mathcal{S}^n : X \bullet Y \geq 0 \text{ for all } X \in C^n\}$$

be the dual cone of C^n. Show that $(C^n)^* = \operatorname{conv}(\{vv^T : v \geq 0\})$.

$C^n = \{ Y \in S^n \quad v^T X v \geq 0 \quad \text{for } v \geq 0 \}$

ENGG 5501: Foundations of Optimization — 2013–14 First Term

Homework Set 1 Solutions

Instructor: Anthony Man–Cho So

September 18, 2013

Problem 1 (20pts). We claim that the given linear–fractional program

$$
\begin{aligned}
v_{\mathrm{lf}}^* \quad := \quad \text{minimize} \quad & \frac{c^T x + d}{f^T x + g} \\
\text{subject to} \quad & Ax \le b, \\
& f^T x + g \ge 0
\end{aligned}
\tag{1}
$$

is equivalent to the following linear program:

$$
\begin{aligned}
v_{\mathrm{lp}}^* \quad := \quad \text{minimize} \quad & c^T y + \theta d \\
\text{subject to} \quad & Ay \le \theta b, \\
& f^T y + \theta g = 1, \\
& \theta \ge 0
\end{aligned}
\tag{2}
$$

whenever the feasible set $\left\{ x \in \mathbb{R}^n : Ax \le b,\ f^T x + g > 0 \right\}$ is non–empty. Indeed, suppose that $\bar{x} \in \mathbb{R}^n$ is feasible for (1). Then, upon setting

$$
\bar{y} = \frac{\bar{x}}{f^T \bar{x} + g}, \quad \bar{\theta} = \frac{1}{f^T \bar{x} + g},
$$

we see that $(\bar{y}, \bar{\theta}) \in \mathbb{R}^n \times \mathbb{R}_+$ is feasible for (2) and has objective value $(c^T \bar{x} + d)/(f^T \bar{x} + g)$. It follows that $v_{\mathrm{lp}}^* \le v_{\mathrm{lf}}^*$.

Conversely, suppose that $(\bar{y}, \bar{\theta}) \in \mathbb{R}^n \times \mathbb{R}_+$ is feasible for (2). If $\bar{\theta} > 0$, then upon setting $\bar{x} = \bar{y}/\bar{\theta} \in \mathbb{R}^n$, we see that $\bar{x}$ is feasible for (1) and has objective value

$$
\frac{c^T \bar{x} + d}{f^T \bar{x} + g} = \frac{c^T (\bar{y}/\bar{\theta}) + d}{1/\bar{\theta}} = c^T \bar{y} + \bar{\theta} d,
$$

whence $v_{\mathrm{lp}}^* \ge v_{\mathrm{lf}}^*$. On the other hand, if $\bar{\theta} = 0$, then $A\bar{y} \le 0$ and $f^T \bar{y} = 1$. Now, let $x_0 \in \mathbb{R}^n$ be any feasible solution to (1). Then, the vector $\bar{x}(t) = x_0 + t\bar{y}$ is feasible for (1) for all $t \ge 0$. Moreover, we have

$$
\frac{c^T \bar{x}(t) + d}{f^T \bar{x}(t) + g} = \frac{c^T x_0 + t c^T \bar{y}}{f^T x_0 + t f^T \bar{y} + g} \to \frac{c^T \bar{y}}{f^T \bar{y}} = c^T \bar{y} + \bar{\theta} d \qquad \text{as } t \to \infty.
$$

It follows that there exists a sequence $\{\bar{x}(t)\}$ of feasible solutions to (1) whose objective values come arbitrarily close to that of (2). This again implies that $v_{\mathrm{lp}}^* \ge v_{\mathrm{lf}}^*$, and the claim is established.

We remark that if $f^T x + g = 0$ whenever $Ax \le b$, then our convention implies that the optimal value of (1) is either $-\infty$ or $+\infty$. This case occurs if the optimal value of the linear program

$$
\begin{aligned}
\text{maximize} \quad & f^T x \\
\text{subject to} \quad & Ax \le b, \\
& f^T x + g \ge 0
\end{aligned}
$$

is $-g$.

Problem 2 (40pts).

(a) **(20pts).** Suppose there exists an $x \in A$ such that

$$x = \frac{2}{t+1}y + \frac{t-1}{t+1}z,$$

where $y \in \mathbb{R}^n \backslash (tA)$ and $z \in A$. Since $x, z \in A$ and A is a symmetric convex set, we have

$$A \ni \frac{t-1}{t}z + \frac{1}{t}(-x) = \frac{t+1}{t}\left(x - \frac{2}{t+1}y\right) + \frac{1}{t}(-x) = x - \frac{2}{t}y.$$

In particular, by the symmetry of A, there exists a $w \in A$ such that $y = (t/2)x + (t/2)w$. However, since A is convex, we have $y = t((x+w)/2) \in tA$, which is a contradiction.

(b) **(10pts).** *This problem is more difficult than intended if one is not allowed to use the following result:*

Theorem 1 (Prékopa [4]; cf. Rinott [5]) Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a density function that is logconcave*; i.e., for any $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$, we have*

$$f(\alpha x + (1-\alpha)y) \ge f(x)^\alpha f(y)^{1-\alpha}.$$

Then, the probability measure μ given by

$$\mu(S) = \int_S f(x)\,dx \quad \textit{for all measurable } S \subset \mathbb{R}^n$$

is logconcave.

Hence, we shall assume the result of Theorem 1 in the sequel.

The desired result follows from the simple observation that for any $x, y \in \mathbb{R}^n$ and $\alpha \in (0,1)$,

$$\|\alpha x + (1-\alpha)y\|_2^2 \le \alpha\|x\|_2^2 + (1-\alpha)\|y\|_2^2,$$

which is equivalent to

$$\exp\left(-\frac{\|\alpha x + (1-\alpha)y\|_2^2}{2}\right) \ge \exp\left(-\frac{\alpha\|x\|_2^2}{2}\right) \cdot \exp\left(-\frac{(1-\alpha)\|y\|_2^2}{2}\right).$$

(c) **(10pts).** Since μ is a logconcave probability measure and $\mu(A) = \theta$, we have

$$\mu\left(\frac{2}{t+1}(\mathbb{R}^n \backslash (tA)) + \frac{t-1}{t+1}A\right) \ge \mu(\mathbb{R}^n \backslash (tA))^{2/(t+1)}\theta^{(t-1)/(t+1)}. \tag{3}$$

Now, the result in (a) implies that

$$\mu\left(\frac{2}{t+1}(\mathbb{R}^n \backslash (tA)) + \frac{t-1}{t+1}A\right) \le \mu(\mathbb{R}^n \backslash A) = 1 - \theta. \tag{4}$$

It follows from (3) and (4) that

$$\mu(\mathbb{R}^n \backslash (tA)) \le \theta^{-(t-1)/2}(1-\theta)^{(t+1)/2} = \theta\left(\frac{1-\theta}{\theta}\right)^{(t+1)/2},$$

as desired.

REMARK: The result in (c) is known as the *concentration of measure* phenomenon. Such a phenomenon has numerous applications and far-reaching consequences. We refer the interested reader to the recent book [2] for a detailed treatment of the topic.

Problem 3 (15pts). Let $z^* = \Pi_K(x)$. Then, we have $z^* \in K$. Since K is a cone, this implies that $\alpha z^* \in K$ for all $\alpha > 0$. It follows from Theorem 8 of Handout 2 that $(\alpha - 1)(x - z^*)^T z^* \leq 0$ for all $\alpha > 0$. Since $\alpha - 1$ can take either positive or negative value as $\alpha > 0$ varies, we must have $(x - z^*)^T z^* = 0$. This, together with Theorem 8 of Handout 2, implies that $z^T(x - z^*) \leq 0$ for all $z \in K$, which is equivalent to $x - z^* \in K^\circ$.

Conversely, suppose that $z^* \in K$, $x - z^* \in K^\circ$ and $(x - z^*)^T z^* = 0$. We claim that $\|x - z^*\|_2 \leq \|x - z\|_2$ for all $z \in K$. Indeed, for any $z \in K$, we have

$$\begin{aligned} \|x - z\|_2^2 &= \|x - z^* + z^* - z\|_2^2 \\ &= \|x - z^*\|_2^2 + \|z^* - z\|_2^2 + 2(x - z^*)^T(z^* - z) \\ &\geq \|z^* - z\|_2^2 - 2(x - z^*)^T z \\ &\geq \|z^* - z\|_2^2, \end{aligned}$$

where the last inequality follows from the fact that $x - z^* \in K^\circ$. This establishes the claim, and the proof is completed.

Problem 4 (25pts).

(a) **(10pts).** We verify the four required properties:

(1) (Non-emptiness) This is clear, since $\mathbf{0} \in \mathcal{C}^n$.

(2) (Closedness) Let $\{X^k\}_{k \geq 1}$ be a sequence in $\mathcal{C}^n$ such that $X^k \to X$. Then, for any $v \geq \mathbf{0}$, we have $v^T X^k v \geq 0$ for all $k \geq 1$, which, by the continuity of the function $Y \mapsto v^T Y v$, implies that $0 \leq \lim_{k \to \infty} v^T X^k v = v^T X v$. Since this holds for an arbitrary $v \geq \mathbf{0}$, we conclude that $X \in \mathcal{C}^n$, as desired.

(3) (Convexity) Let $X_1, X_2 \in \mathcal{C}^n$ be arbitrary. Then, for any $v \geq \mathbf{0}$ and $\alpha \in (0,1)$, we have

$$v^T(\alpha X_1 + (1 - \alpha)X_2)v = \alpha \cdot v^T X_1 v + (1 - \alpha) \cdot v^T X_2 v \geq 0,$$

which implies that $\alpha X_1 + (1 - \alpha)X_2 \in \mathcal{C}^n$ for all $\alpha \in (0,1)$, as desired.

(4) (Conic Set) Let $X \in \mathcal{C}^n$ be arbitrary. Then, for any $\alpha > 0$ and $v \geq \mathbf{0}$, we have $v^T(\alpha X)v = \alpha \cdot v^T X v \geq 0$.

(b) **(15pts).** Consider a fixed $Y \in \mathcal{U} = \operatorname{conv}(\{vv^T : v \geq \mathbf{0}\})$. By Carathéodory's theorem (Theorem 5 of Handout 2), there exist $N \leq n(n+1)/2 + 1$ vectors $v_1, \ldots, v_N \geq \mathbf{0}$ such that $Y = \sum_{i=1}^N v_i v_i^T$. Now, for any $X \in \mathcal{C}^n$, we compute

$$X \bullet Y = X \bullet \left(\sum_{i=1}^N v_i v_i^T \right) = \sum_{i=1}^N v_i^T X v_i \geq 0,$$

where the last inequality follows from the fact that $X \in \mathcal{C}^n$ and $v_1, \ldots, v_N \geq \mathbf{0}$. This implies that $Y \in (\mathcal{C}^n)^*$.

Conversely, suppose that $Y \notin \mathcal{U}$. It is clear that $\mathcal{U}$ is non-empty and convex. Moreover, as the following proposition shows, it is closed.

Proposition 1 *The set $\mathcal{U}$ is closed.*

Proof Let $\{X^k\}_k$ be a sequence in $\mathcal{U}$ such that $X^k \to X$. By Carathéodory's theorem (Theorem 5 of Handout 2), for each $k \geq 1$, there exist $N \leq n(n+1)/2+1$ vectors $v_1^k, v_2^k, \dots, v_N^k \geq 0$ such that $X^k = \sum_{j=1}^N v_j^k (v_j^k)^T = A^k (A^k)^T$. Here, A^k is an $n \times N$ matrix whose j–th column is v_j^k. Now, let $(a_i^k)^T \in \mathbb{R}^N$ be the i–th row of A^k. Observe that $a_i^k \geq \mathbf{0}$, and

$$X_{ii} = \lim_{k\to\infty} X_{ii}^k = \lim_{k\to\infty} \|a_i^k\|_2^2 \qquad \text{for } i = 1, \dots, n.$$

This implies that for each $i = 1, \dots, n$, the sequence $\{a_i^k\}_k$ is bounded. In particular, there is a convergent subsequence of $\{a_i^k\}_k$, whose limit is, say, $a_i \geq \mathbf{0}$. By the continuity of the function $v \mapsto \|v\|_2^2$, this yields $X_{ii} = \|a_i\|_2^2$.

Furthermore, we have

$$X_{ij} = \lim_{k\to\infty} (a_i^k)^T a_j^k,$$

which, by a similar argument as above, yields $X_{ij} = a_i^T a_j$. Hence, we can write $X = AA^T$, where the i–th row of A is a_i^T. This implies that $X \in \mathcal{U}$, as desired. $\square$

Since $\mathcal{U}$ is a non–empty closed convex set, by the separation theorem (Theorem 10 of Handout 2), there exists a matrix $W \in \mathcal{S}^n$ such that $W \bullet Y < W \bullet Z$ for all $Z \in \mathcal{U}$. In particular, we have $W \bullet Y < 0$, since $0 \in \mathcal{U}$.

Now, we claim that $W \in \mathcal{C}^n$. Suppose to the contrary that this is not the case. Then, there exists a vector $u \geq \mathbf{0}$ such that $u^T W u < 0$. Since $\alpha \cdot uu^T \in \mathcal{U}$ for all $\alpha > 0$, we see that $W \bullet Y < W \bullet (\alpha \cdot uu^T) = \alpha \cdot u^T W u$ for all $\alpha > 0$, which is impossible because $W \bullet Y$ is a fixed negative number. It follows that $W \in \mathcal{C}^n$ as claimed.

To complete the proof, it suffices to observe that $W \in \mathcal{C}^n$ and $W \bullet Y < 0$ imply $Y \notin (\mathcal{C}^n)^*$.

REMARK: The sets $\mathcal{C}^n$ and $(\mathcal{C}^n)^*$ are known as the *copositive cone* and *completely positive cone*, respectively. They play a fundamental role in *copositive optimization*, an area that has received much research interest recently. We refer the interested reader to [3, 1] for a survey of the field.

References

[1] I. M. Bomze. Copositive Optimization — Recent Developments and Applications. *European Journal on Operational Research*, 216(3):509–520, 2012.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, Oxford, 2013.

[3] M. Dür. Copositive Programming — A Survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering*, pages 3–20. Springer–Verlag, Berlin/Heidelberg, 2010.

[4] A. Prékopa. Logarithmic Concave Measures with Application to Stochastic Programming. *Acta Scientiarum Mathematicarum*, 32(3–4):301–316, 1971.

[5] Y. Rinott. On Convexity of Measures. *The Annals of Probability*, 4(6):1020–1026, 1976.

ENGG 5501: Foundations of Optimization **2013–14 First Term**

Homework Set 2

Instructor: Anthony Man–Cho So

Due: October 9, 2013

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (25pts). Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a given function that is not identically $+\infty$. The *Fenchel conjugate* of f is the function $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

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

(a) **(5pts).** Show that f^* is convex.

(b) **(20pts).** Given a set $C \subset \mathbb{R}^n$, define the indicator function $i_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of C by

$$i_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Compute i_C^* for the following sets. Show your calculations.

(i) $C = \left\{ x \in \mathbb{R}^n : a^T x \le b \right\}$, where $a \in \mathbb{R}^n \backslash \{0\}$ and $b \in \mathbb{R}$ are given.

(ii) $C = \left\{ x \in \mathbb{R}^n_+ : \|x\|_2 \le 1 \right\}$.

Problem 2 (30pts). Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ be an arbitrary norm on $\mathbb{R}^n$. The *dual norm* of $\|\cdot\|$, which is denoted by $\|\cdot\|_*$, is defined as

$$\|x\|_* = \sup_{\|d\|=1} d^T x.$$

(a) **(20pts).** Show that

$$\partial \|x\| = \left\{ s \in \mathbb{R}^n : \|s\|_* \le 1,\ s^T x = \|x\| \right\}.$$

(b) **(10pts).** Let $x \in \mathbb{R}^n$ be given. Using the result of (a), or otherwise, give an explicit expression for $\partial \|x\|_2$.

Problem 3 (20pts). Let $A \in \mathbb{R}^{m \times n}$ be a given matrix. Show that the following statements are equivalent:

(i) ⇒ (ii): $Ax \geq 0$, $x \geq 0$. $e_1^T x = 1$ has no solution

(i) If a vector $x \in \mathbb{R}^n$ satisfies

$$Ax \ge 0, \quad x \ge 0,$$

then $x_1 = 0$:

$\begin{bmatrix} A & -I \\ e_1^T & 0 \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $(x, s) \geq 0$ has no solution.

(ii) There exists a vector $y \in \mathbb{R}^m$ such that

(ii) ⇒ (i): $\exists\, y \geq 0$ s.t. $a_1^T y < 0$, $a_2^T y \leq 0 \cdots a_n^T y \leq 0$

$$A^T y \le 0, \quad y \ge 0, \quad a_1^T y < 0,$$

If $Ax \geq 0$. $x \geq 0 \Rightarrow y^T A x \geq 0$

where $a_1 \in \mathbb{R}^m$ is the first column of A.

i.e. $x_1(a_1^T y) + \sum_{i=2}^{n} x_i a_i^T y \geq 0 \Rightarrow x_1(a_1^T y) \geq -\sum_{i=2}^{n} (a_i^T y) x_i \geq$

$\Rightarrow x_1 \leq 0$

$\Rightarrow x_1 = 0$

Problem 4 (25pts). Let $A \in \mathbb{R}^{m \times n}$ be given. We are interested in finding a vector $x \in \mathbb{R}^n_+$ such that $Ax = 0$ and the number of positive components of x is maximized. Formulate this problem as a linear program. Justify your answer.

ENGG 5501: Foundations of Optimization — 2013–14 First Term

Homework Set 2 Solutions

Instructor: Anthony Man–Cho So — October 9, 2013

Problem 1 (25pts).

(a) **(5pts).** For a given $x \in \mathbb{R}^n$, let $f_x : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be given by $f_x(y) = y^T x - f(x)$. It is clear that f_x is convex for each $x \in \mathbb{R}^n$. Thus, the convexity of f^* follows from the fact that it is the pointwise supremum of the convex functions $\{f_x\}_{x \in \mathbb{R}^n}$.

(b) **(20pts).** We begin by observing that for any $C \subset \mathbb{R}^n$,

$$i_C^*(y) = \sup_{x \in \mathbb{R}^n} \{y^T x - i_C(x)\} = \sup_{x \in C} y^T x.$$

(i) **(12pts).** For $C = \{x \in \mathbb{R}^n : a^T x \le b\}$, where $a \in \mathbb{R}^n \backslash \{0\}$ and $b \in \mathbb{R}$ are given, suppose that $y \notin \operatorname{span}(a)$; i.e., $y \neq \alpha a$ for any $\alpha \in \mathbb{R}$. We claim that there exists an $u \in \mathbb{R}^n$ such that $a^T u = 0$ and $y^T u > 0$. Indeed, let u be the projection of y onto the subspace $H = \{x \in \mathbb{R}^n : a^T x = 0\}$; i.e., $u = (I - aa^T/\|a\|_2^2)y$. It is easy to verify that $a^T u = 0$. Moreover, since $y \notin \operatorname{span}(a)$, by the Cauchy–Schwarz inequality, we have

$$y^T u = \|y\|_2^2 - \frac{(a^T y)^2}{\|a\|_2^2} > 0.$$

This establishes the claim. Now, let $z \in C$ be arbitrary. Note that $z + \beta u \in C$ for all $\beta \in \mathbb{R}$. Moreover, we have

$$y^T(z + \beta u) = y^T z + \beta y^T u \to +\infty \quad \text{as } \beta \to +\infty.$$

Hence, $i_C^*(y) = +\infty$ whenever $y \notin \operatorname{span}(a)$.

On the other hand, suppose that $y \in \operatorname{span}(a)$; i.e., $y = \alpha a$ for some $\alpha \in \mathbb{R}$. If $\alpha \ge 0$, then $i_C^*(y) = \sup_{x \in C} y^T x = \alpha b$. If $\alpha < 0$, then by noting that $-\beta a \in C$ for all sufficiently large $\beta > 0$, we have $i_C^*(y) \ge -\alpha\beta\|a\|_2^2 \to +\infty$ as $\beta \to +\infty$. To summarize, we obtain

$$i_C^*(y) = \begin{cases} \dfrac{a^T y}{\|a\|_2^2} b & \text{if } y \in \{\alpha a : \alpha \ge 0\}, \\ +\infty & \text{otherwise.} \end{cases}$$

(ii) **(8pts).** For $C = \{x \in \mathbb{R}_+^n : \|x\|_2 \le 1\}$, observe that for any $x \in C$ and $y \in \mathbb{R}^n$,

$$y^T x = \sum_{i:y_i \ge 0} x_i y_i + \sum_{i:y_i < 0} x_i y_i \le \sum_{i:y_i \ge 0} x_i y_i = y^T x' = y_+^T x,$$

where $x' \in C$ is given by

$$x_j' = \begin{cases} x_j & \text{if } y_j \ge 0, \\ 0 & \text{otherwise} \end{cases}$$

and $y_+ \in \mathbb{R}^n$ is given by $(y_+)_j = \max\{y_j, 0\}$, for $j = 1, \dots, n$. It follows from the Cauchy–Schwarz inequality that $i_C^*(y) = \|y_+\|_2$.

Problem 2 (30pts).

(a) **(20pts).** Consider a fixed $x \in \mathbb{R}^n$. Let $s \in \mathbb{R}^n$ be such that $\|s\|_* \leq 1$ and $s^T x = \|x\|$. By definition of the dual norm, for any $\bar{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, we have

$$1 \geq \|s\|_* = \sup_{d \neq \mathbf{0}} \frac{d^T s}{\|d\|} \geq \frac{s^T \bar{x}}{\|\bar{x}\|}.$$

It follows that

$$\|\bar{x}\| \geq s^T \bar{x} = s^T x + s^T(\bar{x} - x) = \|x\| + s^T(\bar{x} - x).$$

Note that the above inequality is also valid at $\bar{x} = \mathbf{0}$, because $s^T x = \|x\|$ by assumption. Hence, we conclude that $s \in \partial\|x\|$.

Conversely, suppose that $s \in \partial\|x\|$. Consider first the case where $x \neq \mathbf{0}$. We have

$$\begin{aligned} 2\|x\| &= \|x + x\| \geq \|x\| + s^T x, \\ 0 &= \|x - x\| \geq \|x\| - s^T x, \end{aligned}$$

which together imply that $s^T x = \|x\|$. Since $x \neq \mathbf{0}$, it follows that

$$\|s\|_* = \sup_{\|d\|=1} d^T x \geq \frac{s^T x}{\|x\|} = 1.$$

We claim that $\|s\|_* = 1$. Suppose that this is not the case. Then, we have $\|s\|_* > 1$, which implies that $s^T d > 1$ for some $d \in \mathbb{R}^n$ with $\|d\| = 1$. We compute

$$\|x\| + 1 = \|x\| + \|d\| \geq \|x + d\| \geq \|x\| + s^T d > \|x\| + 1,$$

which is a contradiction. This establishes the claim.

Now, consider the case where $x = \mathbf{0}$. The condition $s^T x = \|x\|$ is automatically satisfied. On the other hand, let $f : \mathbb{R}^n \to \mathbb{R}_+$ be given by $f(x) = \|x\|$. The directional derivative of f at $\mathbf{0}$ in the direction $d \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ is, by definition,

$$f'(\mathbf{0}, d) = \lim_{t \searrow 0} \frac{\|td\|}{t} = \|d\|.$$

Hence, by Theorem 16(a) of Handout 2, we have $f'(\mathbf{0}, d) = \|d\| \geq s^T d$ for all $s \in \partial\|\mathbf{0}\|$. This, together with the definition of the dual norm, yields

$$\|s\|_* = \sup_{d \neq \mathbf{0}} \frac{s^T d}{\|d\|} \leq 1,$$

as desired.

(b) **(10pts).** Note that the norm $\|\cdot\|_2$ is self-dual; i.e., $(\|x\|_2)_* = \|x\|_2$ for all $x \in \mathbb{R}^n$. Hence, by the result in (a), we have

$$\partial\|x\|_2 = \left\{ s \in \mathbb{R}^n : \|s\|_2 \leq 1, \ s^T x = \|x\|_2 \right\}.$$

Now, by the Cauchy–Schwarz inequality, for any $s \in B(0,1)$,

$$s^T x \le \|s\|_2 \cdot \|x\|_2 \le \|x\|_2.$$

It follows that

$$\partial\|x\|_2 = \begin{cases} \left\{ \dfrac{x}{\|x\|_2} \right\} & \text{if } x \neq 0, \\ B(0,1) & \text{otherwise.} \end{cases}$$

Problem 3 (20pts). Suppose that (i) holds. Then, the system

$$Ax \ge 0, \quad x \ge 0, \quad x_1 = 1 \tag{1}$$

has no solution. By Farkas' lemma, the system

$$\begin{bmatrix} A^T & e_1 \\ -I & 0 \end{bmatrix} \begin{bmatrix} y \\ t \end{bmatrix} \le 0, \quad t > 0 \tag{2}$$

has a solution, say, $(\bar{y}, \bar{t}) \in \mathbb{R}^m \times \mathbb{R}$. In particular, we have

$$a_1^T \bar{y} + \bar{t} \le 0, \quad a_j^T \bar{y} \le 0 \quad \text{for } j = 2, \ldots, n,$$
$$\bar{y} \ge 0, \quad \bar{t} > 0,$$

which is equivalent to

$$A^T \bar{y} \le 0, \quad \bar{y} \ge 0, \quad a_1^T \bar{y} < 0;$$

i.e., (ii) holds. Conversely, suppose that (i) does not hold. Then, system (1) is solvable, and the Farkas lemma implies that system (2) has no solution. The latter is equivalent to the statement that (ii) does not hold, and the proof is completed.

Problem 4 (25pts). Consider the following system:

$$Ax = 0, \quad x \ge 0. \tag{3}$$

Suppose that $\bar{x} \in \mathbb{R}^n$ is feasible for (3). Then, $\alpha\bar{x}$ is also feasible for (3) for any $\alpha > 0$, and $\bar{x}$ and $\alpha\bar{x}$ have the same number of positive components. Thus, as far as the number of positive components is concerned, we may assume that all the non-zero entries of $\bar{x}$ have magnitude at least 1. Consequently, we can decompose $\bar{x}$ as $\bar{x} = \bar{y} + \bar{z}$, where $\bar{y}_i = \max\{\bar{x}_i - 1, 0\}$ and $\bar{z}_i = \bar{x}_i - \bar{y}_i \in \{0, 1\}$ for $i = 1, \ldots, n$. Note that $e^T \bar{z}$ is precisely the number of positive components in $\bar{x}$. The above observation motivates us to consider the following LP:

if $x_i = 0$.

$$\begin{aligned} \text{maximize} \quad & e^T z \\ \text{subject to} \quad & A(y+z) = 0, \\ & z_i \le 1 \qquad \text{for } i = 1, \ldots, n, \\ & y, z \ge 0. \end{aligned} \tag{4}$$

It is not hard to verify that if (y^*, z^*) is an optimal solution to (4), then we must have $z^* \in \{0,1\}^n$. Moreover, we have $z_i^* = 0$ if and only if every solution $\bar{x}$ to (3) satisfies $\bar{x}_i = 0$. Thus, the above LP indeed finds a solution to (3) that has the largest number of positive components.

ENGG 5501: Foundations of Optimization **2013–14 First Term**

Homework Set 3

Instructor: Anthony Man–Cho So

Due: October 30, 2013

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (20pts). Consider the following LP:

$$(P) \quad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \\ & x \geq \mathbf{0}, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$ are given. Suppose that (P) is feasible. In general, (P) need not be strictly feasible; i.e., there may not exist an $\bar{x} \in \mathbb{R}^n$ such that $A\bar{x} = b$ and $\bar{x} > \mathbf{0}$. However, it can be transformed into another LP that is strictly feasible. Indeed, consider the LP

$$(P') \quad \begin{array}{ll} \text{minimize} & c^T x + Mt \\ \text{subject to} & Ax + (b - Ae)t = b, \\ & x \geq \mathbf{0},\ t \geq 0, \end{array}$$

where $M > 0$ is a penalty parameter, and $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$ is the vector of all ones. Clearly, $(\bar{x}, \bar{t}) = (e, 1)$ is a strictly feasible solution to (P'). Now, to complete the transformation, prove the following theorem:

Theorem *There exists an $M_0 \in (0, \infty)$ such that if $M > M_0$ and $x^* \in \mathbb{R}^n$ is an optimal solution to (P), then $(x^*, 0) \in \mathbb{R}^n \times \mathbb{R}$ is an optimal solution to (P').*

Problem 2 (25pts). The goal of this problem is to prove a theorem of alternatives for linear matrix inequality systems. Let $A_1, \ldots, A_m \in \mathcal{S}^n$ and $b \in \mathbb{R}^m$ be given. Suppose that the set

$$C = \{(A_1 \bullet X, \ldots, A_m \bullet X) : X \succeq \mathbf{0}\} \subset \mathbb{R}^m$$

is closed. Show that exactly one of the following systems has a solution:

$$(\mathrm{I}) \quad \left\{ \begin{array}{rcll} A_i \bullet X & = & b_i & \text{for } i = 1, \ldots, m, \\ X & \succeq & \mathbf{0}. & \end{array} \right.$$

$$(\mathrm{II}) \quad \left\{ \begin{array}{rcl} \displaystyle\sum_{i=1}^m y_i A_i & \succeq & \mathbf{0}, \\ b^T y & = & -1. \end{array} \right.$$

REMARK: Compared with the Farkas lemma (Theorem 5 of Handout 3), the above theorem requires an extra condition, namely, the closedness of the set C. In fact, if such a condition does not hold, then the conclusion of the above theorem is false.

Problem 3 (55pts). Consider the following SDP:

$$
(T) \quad \begin{array}{ll} \sup & Z_{11} \\ \text{subject to} & Z_{11} - 2Z_{12} + Z_{22} = 1, \\ & Z_{11} - 2Z_{13} + Z_{33} = 1, \\ & Z_{22} - 2Z_{23} + Z_{33} = 4, \\ & Z_{22} = Z_{33} = 2, \\ & Z \in \mathcal{S}^3_+. \end{array}
$$

对称 $\underline{A_i}$

(a) **(10pts).** Write down the dual of (T).

(b) **(20pts).** Determine the feasible region of (T). Hence, or otherwise, determine the optimal solution and optimal value of (T).

ok $\Leftarrow$ $A \cdot B = \mathrm{tr}(AB)$

(c) **(15pts).** Let $A, B \in \mathcal{S}^n_+$. Show that $A \bullet B = 0$ if and only if $AB = \mathbf{0}$.

$\Rightarrow$ $AB_{ij} = \sum_{k=1}^{n} A$

(d) **(10pts).** Using the results in (a)–(c), show that there is no dual feasible solution that attains the optimal value computed in (b).

$= \sum_{k=1}^{n} A_{ik} \ldots$

sup Z_{11}

s.t.

(d) if $\exists$ then strictly feasible & bounded

$\Rightarrow V_p^* = V_d^*$

$A \cdot B = 0 \Rightarrow \sum_i \sum_j A_{ij} B_{ij} = 0$

$\Rightarrow (AB)_{ij} = \sum_K A_{ik} B_{kj}$

$= \sum_K A_{ik} B_{jk}$

$U\Lambda_1 U^T \cdot V\Lambda_2 V^T$

$A \cdot B = \mathrm{tr}(AB)$

$= \mathrm{tr}(U\Lambda_1 U^T V \Lambda_2 V^T)$

$=$

$b^T y$ $\quad \sum y_i A_i + S = C.$

ENGG 5501: Foundations of Optimization 2013–14 First Term

Homework Set 3 Solutions

Instructor: Anthony Man–Cho So November 18, 2013

Problem 1 (20pts). Since $x^* \in \mathbb{R}^n$ is an optimal solution to (P), by the strong duality theorem for LP, there exists a vector $y^* \in \mathbb{R}^m$ satisfying $c^T x^* = b^T y^*$ and $A^T y^* \leq c$. Now, let $(\bar{x}, \bar{t})$ be any feasible solution to (P') with $\bar{t} > 0$. We compute

$$\begin{aligned} c^T x^* &= b^T y^* \\ &= (A\bar{x} + (b - Ae)\bar{t})^T y^* && \text{(since } (\bar{x},\bar{t}) \text{ is feasible for } (P')) \\ &= \bar{x}^T A^T y^* + \bar{t}(b - Ae)^T y^* \\ &\leq c^T \bar{x} + \bar{t}(b - Ae)^T y^* && \text{(since } A^T y^* \leq c \text{ and } \bar{x} \geq 0). \end{aligned}$$

Set $M_0 = |(b - Ae)^T y^*| + 1$. If $M > M_0$, then the above calculation shows that

$$c^T x^* < c^T \bar{x} + M\bar{t}.$$

This, together with the fact that $Ax^* = b$ and $x^* \geq 0$, implies that $(x^*, 0)$ is optimal for (P').

Problem 2 (20pts). We first show that systems (I) and (II) cannot simultaneously have solutions. Suppose that this is not the case. Then, there exist a matrix $\bar{X} \in \mathcal{S}^n$ and a vector $\bar{y} \in \mathbb{R}^m$ satisfying (I) and (II), respectively. This implies that

$$\begin{aligned} 0 &\leq \left(\sum_{i=1}^m \bar{y}_i A_i \right) \bullet \bar{X} && \left(\text{since } \bar{X} \succeq 0 \text{ and } \sum_{i=1}^m \bar{y}_i A_i \succeq 0 \right) \\ &= \sum_{i=1}^m \bar{y}_i (A_i \bullet \bar{X}) && \text{(by linearity of the inner product } \bullet) \\ &= \sum_{i=1}^m \bar{y}_i b_i && \text{(since } A_i \bullet \bar{X} = b_i \text{ for } i = 1,2,\ldots,m) \\ &= -1, && \text{(since } b^T \bar{y} = -1) \end{aligned}$$

which is a contradiction.

Now, suppose that system (I) does not have a solution. Then, we have $b \notin C$. Clearly, the set C is non–empty and convex, and by assumption it is closed as well. Hence, by Theorem 10 of Handout 2, there exists a vector $s \in \mathbb{R}^m$ such that

$$\sup_{z \in C} s^T z < s^T b.$$

We claim that $\sup_{z \in C} s^T z = 0$. Indeed, since $0 \in C$, we have $\sup_{z \in C} s^T z \geq 0$. Suppose that $\sup_{z \in C} s^T z > 0$. Then, there exists a matrix $X' \succeq 0$ such that

$$\sum_{i=1}^m s_i (A_i \bullet X') > 0.$$

In particular, for any $\alpha > 0$, we have $\alpha X' \succeq 0$ and

$$0 < \alpha \sum_{i=1}^{m} s_i (A_i \bullet X') = \sum_{i=1}^{m} s_i (A_i \bullet (\alpha X')) \leq \sup_{x \in K} s^T x < s^T b.$$

However, since $s^T b$ is a constant, the above inequality cannot hold for all values of α. This contradiction shows that $\sup_{x \in K} s^T x \leq 0$, and hence the claim is established.

As a corollary of the claim, we have $s^T b > 0$. Thus, the vector $\bar{y} = -s / s^T b \in \mathbb{R}^m$ is well defined. It is immediate that $b^T \bar{y} = -1$. Moreover, the claim implies that

$$\left(\sum_{i=1}^{m} \bar{y}_i A_i \right) \bullet X \geq 0$$

for all $X \succeq 0$, which, by the self-duality of $\mathcal{S}^n_+$, is equivalent to

$$\sum_{i=1}^{m} \bar{y}_i A_i \succeq 0.$$

This shows that $\bar{y}$ is a solution to system (II).

Problem 3 (30pts).

(a) (10pts). The dual of (P) is given by

$$\begin{array}{ll} \inf & y_1 - y_2 - 4y_3 - 2y_4 - 2y_5 \\ \text{subject to} & S = \begin{bmatrix} y_1 - y_2 - 1 & -y_3 & -y_4 \\ -y_3 & y_1 - y_2 + 2y_4 & -y_5 \\ -y_4 & -y_5 & y_1 + y_2 + y_5 \end{bmatrix} \succeq 0. \end{array}$$

(b) (20pts). The constraints in (P) imply that Z takes the form

$$Z = \begin{bmatrix} Z_{11} & (1 - Z_{11})/2 & (1 - Z_{11})/2 \\ (1 - Z_{11})/2 & 2 & 0 \\ (1 - Z_{11})/2 & 0 & 2 \end{bmatrix}.$$

Since $Z \succeq 0$, we must have $\det(Z) \geq 0$. In other words, we have

$$0 \leq \det(Z) = 4Z_{11} - (1 - Z_{11})^2 = -(1 - Z_{11})^2,$$

which implies that $Z_{11} = 1$. To verify that the resulting Z is positive semidefinite, recall that if λ is an eigenvalue of Z, then

$$0 = \det(Z - \lambda I) = \lambda(2 - \lambda)(\lambda - 3).$$

It follows that all the eigenvalues of Z are non-negative. Hence, the feasible region of (P) consists of the single point

$$\bar{Z} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix},$$

which implies that it is also the optimal solution to (P). The optimal value of (P) is then equal to 5.

(15pts). Clearly, if $AB = 0$, then we have $A \bullet B = \mathrm{tr}(AB) = 0$. Conversely, suppose that $A \bullet B = 0$. Since $A, B \succeq 0$, there exist $A^{1/2}, B^{1/2} \succeq 0$ such that $A = A^{1/2}A^{1/2}$ and $B = B^{1/2}B^{1/2}$. Hence, we have

$$\mathrm{tr}(AB) = \mathrm{tr}\left((A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}\right) = 0.$$

Let $M = (A^{1/2}B^{1/2})^T A^{1/2}B^{1/2}$, and let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Since $M \succeq 0$, we have $\lambda_i \geq 0$ for $i = 1, \ldots, n$. Since $\mathrm{tr}(M) = \sum_{i=1}^n \lambda_i = 0$, we see that $\lambda_1 = \cdots = \lambda_n = 0$, which implies that $M = 0$. In particular, we have $A^{1/2}B^{1/2} = 0$, from which it follows that $AB = 0$, as desired.

d) (10pts). Suppose that there exists a dual feasible solution $\bar{S}$ whose objective value is equal to 1. Then, we have

$$\bar{Z} \bullet \bar{S} = \bar{Z}_{11} - (\bar{y}_1 + \bar{y}_2 + 4\bar{y}_3 + 2\bar{y}_4 + 2\bar{y}_5) = 0.$$

Since $\bar{Z}, \bar{S} \succeq 0$, we have $\bar{Z}\bar{S} = 0$ by the result in (c). However, upon expanding this identity, we find that $(\bar{Z}\bar{S})_{11} = -1$, which is a contradiction.

ENGG 5501: Foundations of Optimization | 2013–14 First Term

Homework Set 4

Instructor: Anthony Man–Cho So | Due: November 13, 2013

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (60pts). Let $P, Q \in \mathcal{S}^n$ be given. Suppose there exists a $\bar{u} \in \mathbb{R}^n$ such that $\bar{u}^T P \bar{u} > 0$. Consider the following statements:

(I) $u^T Q u \geq 0$ whenever $u^T P u \geq 0$, where $u \in \mathbb{R}^n$. (II) ⇒ (I) $\quad u^T(Q - \lambda P)u \geq 0$

(II) There exists a $\lambda \geq 0$ such that $Q \succeq \lambda P$. $\quad \Rightarrow u^T Q u \geq \lambda u^T P u \geq 0.$

(a) **(5pts).** Show that (II) implies (I).

To prove the converse, consider the semidefinite program

max μ + S =

s.t. $\lambda P + \mu I \le Q$

$\lambda \geq 0 \quad S \geq 0$

$$\text{(SDP)} \quad \theta^* = \max_{\mu, \lambda} \{ \mu : Q - \lambda P \succeq \mu I, \lambda \geq 0 \}.$$

$A_1 = I, A_2 = P \quad y = \begin{pmatrix} \mu \\ \lambda \end{pmatrix}, \quad b = (1, 0) \quad C = Q$

(b) **(10pts).** Write down the dual (SDD) of (SDP) and explain why (SDD) has an optimal solution.

min $Q \cdot X$

s.t. $X = 1$

$PX = 0$

(c) **(20pts).** Let $A, B \in \mathcal{S}^n$ be such that $\mathrm{tr}(A) \geq 0$ and $\mathrm{tr}(B) < 0$. Show that there exists a vector $w \in \mathbb{R}^n$ that satisfies $w^T A w \geq 0$ and $w^T B w < 0$.

(d) **(10pts).** Using the results in (b) and (c), show that (I) implies (II).

(e) **(15pts).** Now, let $P, Q \in \mathcal{S}^n$, $p, q \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$ be given. Define the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ by

$$\begin{aligned} f(u) &= u^T P u + 2p^T u + a, \\ g(u) &= u^T Q u + 2q^T u + b. \end{aligned}$$

Suppose that there exists a $\bar{u} \in \mathbb{R}^n$ such that $f(\bar{u}) > 0$. Using the equivalence of (I) and (II) established above, prove that the following statements are equivalent:

(III) $g(u) \geq 0$ whenever $f(u) \geq 0$, where $u \in \mathbb{R}^n$. (II)

(IV) There exists a $\lambda \geq 0$ such that

$$\underbrace{\begin{bmatrix} Q & q \\ q^T & b \end{bmatrix}}_{\tilde{Q}} - \lambda \underbrace{\begin{bmatrix} P & p \\ p^T & a \end{bmatrix}}_{\tilde{P}} \succeq \mathbf{0}.$$

$u^T \tilde{Q} u = (u^T \; 1) \begin{pmatrix} Q & q \\ q^T & b \end{pmatrix} \begin{pmatrix} u \\ 1 \end{pmatrix} \Rightarrow$

$= (u^T Q + q^T \quad u^T q + b) \begin{pmatrix} u \\ 1 \end{pmatrix}$

$= u^T Q u + 2 q^T u + b$

Problem 2 (20pts). Let $A \in \mathcal{S}^n$ and $P, Q \in \mathbb{R}^{n \times n}$ be given. Show that

$$A \succeq P^T Z Q + Q^T Z^T P \quad \text{for all } Z \in \mathbb{R}^{n \times n} \text{ with } \|Z\|_F \leq \rho$$

s.t

$\begin{pmatrix} A - \lambda Q^T Q & -\rho P^T \\ -\rho P & \lambda I \end{pmatrix} \succeq 0$ $\quad \|Z\|_F \le \rho$

$\Leftarrow$ ~~$\lambda A - \lambda Q^T Q$~~

$\lambda A - \lambda^2 Q^T Q \geq \dfrac{\rho^2 P^T P}{Q}$

$A \geq \dfrac{\rho^2 P^T P + \lambda^2 Q^T Q}{\lambda}$

if and only if there exists a $\lambda \geq 0$ such that

$$\begin{bmatrix} A - \lambda Q^T Q & -\rho P^T \\ -\rho P & \lambda I \end{bmatrix} \succeq 0.$$

0 **Problem 3 (25pts).** Let $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions. Consider the following problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0 \quad \text{for } i = 1, \ldots, m, \\ & x \in \mathbb{R}^n. \end{aligned} \tag{1}$$

Let $\bar{x}$ be a feasible solution, and let $I = \{i \in \{1, \ldots, m\} : g_i(\bar{x}) = 0\}$. Suppose that g_i, where $i \in I$, are concave at $\bar{x}$. Consider the following LP problem: min

$$\begin{aligned} \text{minimize} \quad & \nabla f(\bar{x})^T d \qquad -\nabla f(\bar{x}) \\ \text{subject to} \quad & \nabla g_i(\bar{x})^T d \leq 0 \quad \text{for } i \in I, \\ & -e \leq d \leq e, \end{aligned} \tag{2}$$

where $e = (1, 1, \ldots, 1)$. Let d^* be an optimal solution to (2), and set $v^* = \nabla f(\bar{x})^T d^*$.

(a) **(5pts).** Show that $v^* \leq 0$. $g_i(\bar x + \alpha d^*) \le 0$

(b) **(10pts).** Show that if $v^* < 0$, then there exists an $\alpha_0 > 0$ such that $\bar{x} + \alpha d^*$ is feasible and $f(\bar{x} + \alpha d^*) < f(\bar{x})$ for all $\alpha \in (0, \alpha_0)$. ok Propositional $-\bar G \subset F \subset \bar G$

Farkas (c) **(10pts).** Show that if $v^* = 0$, then $\bar{x}$ satisfies the KKT conditions of (1).

$g_i(\bar x + \alpha d^*)$

min c^Tx
s.t. $a_i^Tx \le b_i$

$$\begin{pmatrix} \nabla g_i(\bar x)^T \\ 1 \\ \quad 1 \\ \quad\quad 1 \\ -1 \\ \quad -1 \\ \quad\quad -1 \end{pmatrix} \begin{pmatrix} d \end{pmatrix} \le \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$d+e \ge 0$ $\quad -d-e \le 0$
$d-e \le 0$ $\quad d-e \le 0$

$-d \le e$
$d \le e$
$a_i^Td \le 0$

c^TX
$Ax \le b$
$Ax + s = b$
$c^T(x^+ - x^-)$
$A(x^+ - x^-) + s = b$
$x^+, x^-, s \ge 0$

$\begin{pmatrix} \nabla f(\bar x) \\ -\nabla f(\bar x) \\ 0 \end{pmatrix}$

$(A\ \ -A\ \ I)\begin{pmatrix} d^+ \\ d^- \\ s \end{pmatrix} =$

$\Rightarrow$ max b^Ty
s.t. $A^Ty \le c$

$(A^T\ \ -A^T\ \ I)\, y \le \begin{pmatrix} \nabla f(\bar x) \\ -\nabla f(\bar x) \\ 0 \end{pmatrix}$

$\nabla f(\bar x)^T d^* = b^T y^*$

$\sqrt{1+2\sqrt{1+3\sqrt{1+\cdots}}}$

ENGG 5501: Foundations of Optimization | **2013–14 First Term**

Homework Set 4 Solutions

Instructor: Anthony Man–Cho So | November 18, 2013

Problem 1 (60pts).

(a) **(5pts).** Suppose that there exists a $\lambda \geq 0$ such that $Q \succeq \lambda P$. Then, for any $u \in \mathbb{R}^n$, we have $u^T Q u \geq \lambda \cdot u^T P u$. In particular, if $u^T P u \geq 0$, then $u^T Q u \geq 0$.

(b) **(10pts).** Observe that (SDP) is equivalent to

$$\begin{array}{ll} \text{maximize} & \mu \\ \text{subject to} & \begin{bmatrix} Q & \\ & 0 \end{bmatrix} - \lambda \begin{bmatrix} P & \\ & -1 \end{bmatrix} - \mu \begin{bmatrix} I & \\ & 0 \end{bmatrix} \succeq \mathbf{0}, \end{array}$$

which is a semidefinite program in dual standard form. Hence, the dual of (SDP) is given by

$$\text{(SDD)} \qquad \begin{array}{ll} \text{minimize} & \text{tr}(QX) \\ \text{subject to} & \text{tr}(X) = 1, \\ & \text{tr}(PX) \geq 0, \\ & X \succeq \mathbf{0}. \end{array}$$

Now, since there exists a $\bar{u} \in \mathbb{R}^n$ such that $\bar{u}^T P \bar{u} > 0$, the matrix $\bar{X} = \bar{u}\bar{u}^T / \|\bar{u}\|_2^2$ is feasible for (SDD). In particular, by the CLP weak duality theorem, we have $\theta^* \leq \text{tr}(Q\bar{X}) < \infty$. Moreover, observe that $(\lambda, \mu) = (1, \lambda_{\min}(Q - P) - 1)$ is a strictly feasible solution to (SDP). Hence, by the CLP strong duality theorem, we conclude that (SDD) has an optimal solution.

(c) **(20pts).** Let $B = U \Lambda U^T$ be the spectral decomposition of B. Then, we have $\text{tr}(B) = \text{tr}(\Lambda) < 0$. Now, let $\xi \in \mathbb{R}^n$ be a random vector whose entries are independent and take the values ± 1 with equal probability. Upon setting $\eta = U\xi \in \mathbb{R}^n$, we have

$$\begin{aligned} \eta^T B \eta &= \xi^T U^T B U \xi = \xi^T \Lambda \xi = \text{tr}(\Lambda) < 0, \\ \mathbb{E}\left[\eta^T A \eta\right] &= \mathbb{E}\left[\xi^T U A U^T \xi\right] = \mathbb{E}\left[\text{tr}(U A U^T \xi \xi^T)\right] = \text{tr}(U A U^T) = \text{tr}(A) \geq 0. \end{aligned}$$

In particular, this implies that there exists a $\hat{\xi} \in \{-1, 1\}^n$ such that $w = U\hat{\xi} \in \mathbb{R}^n$ satisfies $w^T A w \geq 0$ and $w^T B w < 0$. This completes the proof.

(d) **(10pts).** Suppose that $u^T Q u \geq 0$ whenever $u^T P u \geq 0$, where $u \in \mathbb{R}^n$. Let $X^* \succeq \mathbf{0}$ be an optimal solution to (SDD), whose existence is guaranteed by the result in (b). Then, we have $\theta^* = \text{tr}(QX^*)$ by the CLP strong duality theorem. Now, we claim that $\theta^* \geq 0$. Suppose to the contrary that $\theta^* < 0$. Let

$$A = (X^*)^{1/2} P (X^*)^{1/2}, \quad B = (X^*)^{1/2} Q (X^*)^{1/2}.$$

Note that $A, B \in \mathcal{S}^n$. Moreover, we have $\operatorname{tr}(A) = \operatorname{tr}(PX^*) \geq 0$ and $\operatorname{tr}(B) = \operatorname{tr}(QX^*) = \theta^* < 0$. Hence, by the result in (c), there exists a vector $w \in \mathbb{R}^n$ such that $w^T Aw \geq 0$ and $w^T Bw < 0$. However, this implies that the vector $u = (X^*)^{1/2} w \in \mathbb{R}^n$ satisfies $u^T Pu \geq 0$ and $u^T Qu < 0$, which is a contradiction. Hence, we have $\theta^* \geq 0$, which, together with the construction of (SDP), implies that $Q \succeq \lambda P$ for some $\lambda \geq 0$, as desired.

(e) **(15pts).** Define the functions $\bar{f} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and $\bar{g} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ by

$$\begin{aligned} \bar{f}(u,t) &= u^T Pu + 2tp^T u + at^2, \\ \bar{g}(u,t) &= u^T Qu + 2tq^T u + bt^2. \end{aligned}$$

The function $\bar{f}$ is known as the *homogenization of f*, as it can be expressed as the homogeneous quadratic form $\bar{f}(u,t) = (u,t)^T \bar{P}(u,t)$, where

$$\bar{P} = \begin{bmatrix} P & p \\ p^T & a \end{bmatrix}.$$

In a similar fashion, we have $\bar{g}(u,t) = (u,t)^T \bar{Q}(u,t)$, where

$$\bar{Q} = \begin{bmatrix} Q & q \\ q^T & b \end{bmatrix}.$$

Now, observe that (IV) implies $(u,1)^T(\bar{Q} - \lambda\bar{P})(u,1) \geq 0$ for any $u \in \mathbb{R}^n$, from which it follows that (III) holds. To prove the converse, suppose that (III) holds. Let $(\tilde{u}, \tilde{t}) \in \mathbb{R}^n \times \mathbb{R}$ be such that $\bar{f}(\tilde{u}, \tilde{t}) \geq 0$. Consider the following two cases:

Case 1: $\tilde{t} \neq 0$. Then, we have $\bar{f}(\tilde{u}/\tilde{t}, 1) = f(\tilde{u}/\tilde{t}) \geq 0$. Since (III) holds, we have $g(\tilde{u}/\tilde{t}) \geq 0$, which implies that $\bar{g}(\tilde{u}, \tilde{t}) \geq 0$.

Case 2: $\tilde{t} = 0$. Then, we have $\bar{f}(\tilde{u}, 0) = \tilde{u}^T P\tilde{u} \geq 0$. We claim that this implies $\bar{g}(\tilde{u}, 0) = \tilde{u}^T Q\tilde{u} \geq 0$. Suppose to the contrary that $\tilde{u}^T Q\tilde{u} < 0$. Let $v(t) = t\tilde{u} + \bar{u}$, where $\bar{u} \in \mathbb{R}^n$ satisfies $f(\bar{u}) > 0$ by assumption. Observe that

$$f(v(t)) = v(t)^T Pv(t) + 2p^T v(t) + a = t^2 \tilde{u}^T P\tilde{u} + 2t(P\bar{u} + p)^T \tilde{u} + f(\bar{u}) > 0$$

for all $t \in \mathbb{R}$ such that $t(P\bar{u} + p)^T \tilde{u} \geq 0$, while

$$g(v(t)) = t^2 \tilde{u}^T Q\tilde{u} + 2t(Q\bar{u} + q)^T \tilde{u} + g(\bar{u}) < 0$$

when $|t|$ is sufficiently large. This contradicts (III), and hence the claim is established.

From the above considerations, we see that $\bar{g}(u,t) \geq 0$ whenever $\bar{f}(u,t) \geq 0$. It then follows from the equivalence of (I) and (II) that $\bar{Q} - \lambda\bar{P} \succeq 0$ for some $\lambda \geq 0$.

Problem 2 (15pts). Observe that

$$\begin{aligned} & A \succeq P^T ZQ + Q^T Z^T P \quad \text{for all } Z \in \mathbb{R}^{n\times n} \text{ with } \|Z\|_F \leq \rho \\ \iff\ & v^T Av \geq v^T(P^T ZQ + Q^T Z^T P)v \quad \text{for all } v \in \mathbb{R}^n \text{ and } Z \in \mathbb{R}^{n\times n} \text{ with } \|Z\|_F \leq \rho \\ \iff\ & v^T Av \geq \max_{\|Z\|_F \leq \rho} \left\{ v^T \left(P^T ZQ + Q^T Z^T P\right) v \right\} \quad \text{for all } v \in \mathbb{R}^n. \qquad (1) \end{aligned}$$

Since
$$v^T\left(P^T Z Q + Q^T Z^T P\right) v = 2(Qv)^T Z^T (Pv) = 2(Pv)(Qv)^T \bullet Z,$$
it follows from the Cauchy–Schwarz inequality that
$$\max_{\|Z\|_F \le \rho} \left\{ v^T\left(P^T Z Q + Q^T Z^T P\right) v \right\} = 2\rho (Pv)(Qv)^T \bullet \frac{(Pv)(Qv)^T}{\|(Pv)(Qv)^T\|_F} = 2\rho \|Pv\|_2 \|Qv\|_2.$$
Using the Cauchy–Schwarz inequality again, we have
$$\|Pv\|_2 \|Qv\|_2 = \max_{\|y\|_2 \le \|Qv\|_2} y^T P v.$$
Hence, the inequality (1) is equivalent to
$$v^T A v \ge 2\rho y^T P v \quad \text{for all } (v, y) \in \mathbb{R}^n \times \mathbb{R}^n \text{ satisfying } y^T y \le v^T Q^T Q v. \tag{2}$$
Now, let $u = (v, y) \in \mathbb{R}^{2n}$,
$$S = \begin{bmatrix} A & -\rho P^T \\ -\rho P & 0 \end{bmatrix}, \quad T = \begin{bmatrix} Q^T Q & 0 \\ 0 & -I \end{bmatrix}.$$
It is easy to verify that
$$v^T A v \ge 2\rho y^T P v \iff u^T S u \ge 0, \qquad y^T y \le v^T Q^T Q v \iff u^T T u \ge 0.$$
Thus, using the result in Problem 1, we conclude that (2) is equivalent to $S - \lambda T \succeq \mathbf{0}$ for some $\lambda \ge 0$. This completes the proof.

Problem 3 (25pts).

(a) **(5pts).** Since $d = \mathbf{0}$ is feasible for the given LP, we have $v^* \le 0$.

(b) **(10pts).** Suppose that $v^* < 0$. Then, by Proposition 1 of Handout 7, there exists an $\alpha_0' > 0$ such that $f(\bar{x} + \alpha d^*) < f(\bar{x})$ for all $\alpha \in (0, \alpha_0')$. Now, the concavity of g_i and the fact that $\nabla g_i(\bar{x})^T d^* \le 0$ for $i \in I$ imply that $g_i(x)$ concave.
$$g_i(\bar{x} + \alpha d^*) \le g_i(\bar{x}) + \alpha \nabla g_i(\bar{x})^T d^* \le g_i(\bar{x}) = 0$$
for all $\alpha > 0$ and $i \in I$. Moreover, since the functions $g_1, \ldots, g_m$ are continuous, for each $i \notin I$, we have $g_i(\bar{x} + \alpha d^*) < 0$ for sufficiently small $|\alpha| > 0$. It follows that there exists an $\alpha_0 > 0$ such that $\bar{x} + \alpha d^*$ is feasible *and* $f(\bar{x} + \alpha d^*) < f(\bar{x})$ for all $\alpha \in (0, \alpha_0)$.

(c) **(10pts).** Let A be the matrix whose i–th column is $\nabla g_i(\bar{x})$, where $i \in I$. Then, the fact that $v^* = 0$ implies that the homogeneous system
$$-\nabla f(\bar{x})^T d > 0, \quad A^T d \le 0$$
has no solution in $d \in \mathbb{R}^n$. By Farkas' lemma, there exists an $u \ge \mathbf{0}$ such that $Au = -\nabla f(\bar{x})$. Note that $Au = -\nabla f(\bar{x})$ is equivalent to
$$\nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) = \mathbf{0}.$$
This, together with the condition that $u \ge \mathbf{0}$ and the fact that $\bar{x}$ is feasible for (1), shows that $\bar{x}$ is a KKT point of (1).

ENGG 5501: Foundations of Optimization | **2013–14 First Term**

Homework Set 5

Instructor: Anthony Man–Cho So

Due: November 27, 2013

SOLVE THE FOLLOWING PROBLEMS:

Problem 1 (40pts). Let $p > 2$ be given. For $n \in \{2,3\}$, define the function $f_n : [0,3]^n \to \mathbb{R}$ by

$$f_n(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i^{1/p} \prod_{j \neq i} (3 - x_j)^{1/p}$$

and consider the problem

$$(Q_n) \quad \max \left\{ f_n(x_1,\ldots,x_n) : 0 \le x_i \le 3 \text{ for } i = 1,\ldots,n \right\}.$$

(a) **(20pts).** Let $\bar{x} = (\bar{x}_1, \bar{x}_2)$ be an optimal solution to (Q_2). Clearly, $\bar{x}$ lies either on the boundary or in the interior of the box $[0,3] \times [0,3]$. By considering appropriate optimality conditions, determine the optimal solution to and optimal value of (Q_2).

(b) **(20pts).** Let

$$S = \left\{ (x_1,x_2,x_3) \in [0,3]^3 : x_i \in \{0,3\} \text{ for some } i \in \{1,2,3\} \right\}.$$

Using the result in (a), show that

$$\sup_{(x_1,x_2,x_3) \in S} f_3(x_1,x_2,x_3) \le f_3(1,1,1) = 3 \cdot 2^{2/p}.$$

Problem 2 (20pts). Let $A \in \mathcal{S}^n$ and $b \in \mathbb{R}^n$ be given. Consider the optimization problem

$$\begin{array}{rcll} v_p^* & = & \text{minimize} & x^T A x + 2b^T x \\ & & \text{subject to} & x^T x \le 1. \end{array} \qquad (1)$$

Let $u \ge 0$ be the Lagrangian multiplier associated with the constraint $x^T x \le 1$. Show that the Lagrangian dual problem associated with (1) is equivalent to the following SDP:

$$\begin{array}{rcll} v_d^* & = & \text{maximize} & -t - u \\ & & \text{subject to} & \begin{bmatrix} A + uI & b \\ b^T & t \end{bmatrix} \succeq 0, \\ & & & u \ge 0. \end{array} \qquad (2)$$

(*Hint: Consider the Moore–Penrose pseudoinverse of $A + uI$.*)

REMARK: By the weak duality theorem, we always have $v_p^* \ge v_d^*$. Thus, we can view Problem (2) as a convex relaxation of Problem (1). In fact, one can show that the relaxation is *tight* in this case, i.e., $v_p^* = v_d^*$.

$$\frac{x_1^{\frac{1}{p}}(3-x_2)^{\frac{1}{p}}}{x_1} = \frac{(3-x_1)^{\frac{1}{p}} x_2^{\frac{1}{p}}}{(3-x_1)}$$

$$\frac{1}{p} x_1^{\frac{2-p}{p}}$$

$$\frac{x_1^{\frac{1}{p}}(3-x_2)^{\frac{1}{p}}}{(3-x_2)} = \frac{(3-x_1)^{\frac{1}{p}} x_2^{\frac{1}{p}}}{x_2}.$$

$$= \frac{1}{p} x_2^{\frac{2-p}{p}}$$

$$\frac{x_1}{3-x_1} = \frac{3-x_2}{x_2}.$$

$$\cancel{x_1 x_2} = 9 - 3(x_1+x_2)$$

Problem 3 (40pts). In this problem, we extend the gradient method covered in class to handle non-smooth optimization problems. To begin, let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function whose gradient is Lipschitz continuous with parameter $L > 0$, i.e.,

$$\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x - y\|_2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

Furthermore, let $g : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary convex function. Consider the optimization problem

$$\min_{x\in\mathbb{R}^n} \{F(x) \equiv f(x) + g(x)\}. \tag{3}$$

$f(\bar{x}) + g(\bar{x}) \geq -f(x^*) - g(x^*)$

We assume that the optimal solution set of Problem (3) is non-empty. Note that g need not be differentiable, and hence Problem (3) is a non-smooth convex optimization problem. To develop an algorithm for it, we proceed as follows. For any given $\tau > 0$, define the *proximity operator* $\text{prox}_\tau : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\text{prox}_\tau(x) = \arg\min_{y\in\mathbb{R}^n} \left\{ \tau g(y) + \frac{1}{2}\|x - y\|_2^2 \right\}.$$

$\leq f(\bar{x}) - f(x^*) + \frac{1}{2}\|\nabla f\|$

if $x^* \neq \bar{x}$

$g(x^*) + \frac{1}{2}\|x^* - \nabla f(x^*) - x^*\|_2^2 > g(\bar{x}) + \frac{1}{2}\|x^* - \nabla f(x^*) - \bar{x}\|$

(a) **(20pts).** Let $x^* \in \mathbb{R}^n$ be an optimal solution to Problem (3). Show that

$$x^* = \text{prox}_1(x^* - \nabla f(x^*)). \tag{4}$$

$\arg\min_y \{ g(y) + \frac{1}{2}\|x^* - \nabla f(x^*) - y\|_2^2 \}$

Hence, or otherwise, show that if $x^* \in \mathbb{R}^n$ is an optimal solution to the constrained convex optimization problem

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

let $g(x) = I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$

where $C \subset \mathbb{R}^n$ is a non-empty closed convex set, then

$$x^* = \Pi_C(x^* - \nabla f(x^*)).$$

$\text{Prox}_{\alpha_k}(x) = \arg\min \{ g(y) + \frac{1}{2}\|x - y\|_2^2 \} = \arg\min_{y\in C} \{ \frac{1}{2}\|x - y\|_2^2 \} = \Pi_C(x)$

(Recall that Π_C is the projection operator onto C.)

The optimality condition (4) is a fixed–point equation and suggests the following natural iterative scheme for solving Problem (3):

1. initialize $x^0 \in \mathbb{R}^n$, $k \longleftarrow 0$

$\min_{x\in C} f(x) \iff \min f(x) + g(x)$

2. repeat until convergence:

$x^* =$

$\arg\min \{ \alpha_k g(y) + \frac{1}{2}\|x^k - \alpha_k \nabla f(x^k) - y\|_2^2 \}$

$$x^{k+1} \longleftarrow \text{prox}_{\alpha_k}\left(x^k - \alpha_k \nabla f(x^k)\right),$$

where $\alpha_k > 0$ is the step size in the k–th iteration

Now, we are interested in understanding the convergence behavior of such a scheme.

(b) **(20pts).** Let $\{x^k\}_{k\ge0}$ be the sequence of iterates generated by the above scheme. Show that

$$F(x^{k+1}) - F(x^k) \le -\frac{1}{2}\left(\frac{1}{\alpha_k} - L\right)\|x^{k+1} - x^k\|_2^2$$

for all $k \ge 0$. Hence, or otherwise, show that if $\alpha_k < 1/L$ for all $k \ge 0$, then the sequence $\{F(x^k)\}_{k\ge0}$ converges.

$\frac{1}{\alpha_k} > L$

$F(x^{k+1}) - F(x^k)$

$= f(x^{k+1}) + g(x^{k+1}) - f(x^k) - g(x^k)$

$\leq f(x^{k+1}) - f(x^k) + \frac{1}{2}\|\alpha_k \nabla f(x^k)\|^2 - \frac{1}{2}\|x^k - \alpha_k \nabla f(x^k) - x^{k+1}\|^2$

α_k

$\leq f(x^{k+1}) - f(x^k) + \frac{\|x^k - x^{k+1}\|^2}{2\alpha_k}$

$< \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L}{2}\| \cdot - \cdot \|$

2.) $f_2(x_1, x_2) = x_1^{1/p}(3-x_2)^{1/p} + x_2^{1/p}(3-x_1)^{1/p}$

$\max f_2(x_1,x_2) \iff -\min -f_2(x_1,x_2)$

$0 \le x_i \le 3$

$$\begin{cases} -x_1 \le 0 \\ x_1 - 3 \le 0 \\ -x_2 \le 0 \\ x_2 - 3 \le 0 \end{cases} \quad (P)$$

$$Df_2(x_1,x_2) = \begin{pmatrix} \frac{1}{p} x_1^{\frac{1-p}{p}}(3-x_2)^{1/p} - \cdots \\ -\frac{1}{p} x_1^{1/p}(3-x_2)^{\frac{1-p}{p}} + \cdots \end{pmatrix}$$

$-2\left(\frac{3}{2}\right)^{\frac{2}{p}}$

$$\begin{pmatrix} \frac{1}{p}\frac{1-p}{p} x_1^{\frac{1-2p}{p}}(3-x_2)^{1/p} & -\frac{1}{p^2} x_1^{\frac{1-p}{p}}(3-x_2)\cdots \\ -\frac{1}{p^2} x_1^{\frac{1-p}{p}}(3-x_2)^{\frac{1-p}{p}} & \frac{1-p}{p^2} x_1^{1/p}\cdots \end{pmatrix}$$

$$\left(\; \right) + u_1\begin{pmatrix}-1\\0\end{pmatrix} + u_2\begin{pmatrix}0\\1\end{pmatrix} + u_3\begin{pmatrix}-1\\0\end{pmatrix} + u_4\begin{pmatrix}0\\1\end{pmatrix} = 0 \quad (D)$$

$u_i \ge 0$

$\frac{(1-p)^4}{p^4} x_1^{\frac{2-2p}{p}}(3-x_2)^{\frac{2-2p}{p}} - \frac{1}{p^4} x_1^{\frac{2-2p}{p}}$

$(1-p)^{4}\; x_1$

C)

$u_1 x_1 = 0$.

$u_2(x_1-3) = 0$.

$u_3(x_2) = 0$

$u_4(x_2-3) = 0$

$-D^2 f_2(x) \le 0$. concave

对称 ... $\underline{3^{2/p}}$

只可能在边上. opt =

1) $x_1 = 0 \quad u_2 = 0$.

$$\begin{pmatrix} \frac{1}{p} x_2^{1/p} 3^{\frac{1-p}{p}} \\ -\frac{1}{p} x_2^{\frac{1-p}{p}} 3^{1/p} \end{pmatrix} + \begin{pmatrix} -u_1 - u_3 \\ u_4 \end{pmatrix}$$

if $x_2 = 0$, $u_4 = 0$

$u_1 = u_3 = 0$

$f_3(x_1, x_2, x_3) = x_1^{\frac{1}{p}}(3-x_2) \cdots + x_2^{\frac{1}{p}}(3-x_1) \cdots + x_3^{\frac{1}{p}}(3-x_1) \cdots$

if $x_2 = 3$, $u_3 = 0$.

$$\begin{pmatrix} \frac{1}{p} 3^{\frac{2-p}{p}} - u_1 \\ -\frac{1}{p} 3^{\frac{2-p}{p}} + u_4 \end{pmatrix}$$

$u_1 = u_4 = \frac{1}{p} 3^{\frac{2-p}{p}}$

$-f_2(x_1, x_2) \ge -3^{2/p}$

$\Rightarrow f_2 \le 3^{2/p}$

$\le 3^{2/p}$

if $x_2 \ne 0$, $x_2 \ne 3$ $u_3 = u_4 = 0$

$\Rightarrow u_1 = 0$. $-\frac{1}{p} x_2^{\frac{1-p}{p}} 3^{1/p} = 0$. ×

$x_3 = 0$

$3^{\frac{1}{p}} f_2(x_1, x_2)$

$3^{\frac{2}{p}} \le$

$x_1 = 3 \quad u_1 = 0$

$$\begin{pmatrix} -\frac{1}{p} 3^{\frac{1-p}{p}}(3-x_2)^{1/p} + \cdots \\ \frac{1}{p} 3^{1/p}(3-x_2)^{\frac{1-p}{p}} \end{pmatrix} + \begin{pmatrix} -u_3 \\ u_2 + u_4 \end{pmatrix} = 0.$$

$x_3 = 3 \quad 3^{\frac{1}{p}}(3-x_1)^{\frac{1}{p}} \cdots$

$x_2 = 0$ $u_4 = 0$ $u_3 = 0$. $u_2 = 0$

$x_2 = 3$. $x_3 = 0$. $\le 3^{\frac{1}{p}} \cdot 2\left(\frac{3}{2}\right)^{\frac{1}{p}}$ $\quad 3 \cdot 2^{1-\frac{2}{p}}$

≤ 3.

$p > 2$

$1 - \frac{2}{p} < \frac{1}{p}$

$x_2 \ne 0, 3$ ×

2) $x_1 \ne 0, 3$

$$\begin{pmatrix} -\frac{1}{p} x_1^{\frac{1-p}{p}}(3-x_2)^{1/p} + \frac{1}{p} x_2^{1/p}(3-x_1)^{\frac{1-p}{p}} \\ \frac{1}{p} x_1^{1/p}(3-x_2)^{\frac{1-p}{p}} - \frac{1}{p} x_2^{\frac{1-p}{p}}(3-x_1)^{1/p} \end{pmatrix} = \begin{pmatrix} u_3 \\ -u_4 \end{pmatrix}$$

≥ 0

① = ②

2013–14 First Term

Homework Set 5 Solutions

Instructor: Anthony Man–Cho So

November 25, 2013

Problem 1 (40pts).

(a) **(20pts).** When $n = 2$, f_2 takes the form

$$f_2(x_1, x_2) = x_1^{1/p}(3 - x_2)^{1/p} + x_2^{1/p}(3 - x_1)^{1/p}.$$

Note that f_2 is not differentiable on the boundary of $[0,3] \times [0,3]$. Now, let $\bar{x}$ be an optimal solution to (Q_2). Since the constraints in (Q_2) are linear, if $\bar{x}$ lies in the interior of $[0,3]\times[0,3]$, then it satisfies the following first–order necessary conditions:

$$\left(\frac{\bar{x}_1}{3-\bar{x}_1}\right)^{1/p-1} = \left(\frac{\bar{x}_2}{3-\bar{x}_2}\right)^{1/p},$$
$$\left(\frac{\bar{x}_2}{3-\bar{x}_2}\right)^{1/p-1} = \left(\frac{\bar{x}_1}{3-\bar{x}_1}\right)^{1/p}.$$

Since $p > 2$, these equations imply that $\bar{x}_1 = \bar{x}_2 = 3/2$, and we have $f_2(3/2, 3/2) = 2^{1-2/p} \cdot 3^{2/p}$. On the other hand, observe that for any (x_1', x_2') that lies on the boundary of $[0,3]\times[0,3]$, the symmetry of f_2 and the fact that $p > 2$ imply

$$f_2(x_1', x_2') \le f_2(0,3) = 3^{2/p} < 2^{1-2/p} \cdot 3^{2/p}.$$

Hence, the optimal solution to (Q_2) is $(3/2, 3/2)$, and the optimal value is $2^{1-2/p} \cdot 3^{2/p}$.

(b) **(20pts).** Let $(x_1, x_2, x_3) \in S$ be arbitrary. By the symmetry of f_3, we may assume without loss that $x_1 \in \{0, 3\}$. Consider two cases:

Case 1: $x_1 = 0$. We compute

$$f_3(0, x_2, x_3) = 3^{1/p}\left(x_2^{1/p}(3 - x_3)^{1/p} + x_3^{1/p}(3 - x_2)^{1/p}\right) \le 3^{1/p} \cdot 2^{1-2/p} \cdot 3^{2/p},$$

where the last inequality follows from the result in (a). Thus, to prove that $f_3(0, x_2, x_3) < f_3(1,1,1) = 3 \cdot 2^{2/p}$, it suffices to show that $2^{4/p-1} \cdot 3^{1-3/p} > 1$, or equivalently,

$$\left(\frac{4}{p} - 1\right)\log 2 + \left(1 - \frac{3}{p}\right)\log 3 > 0.$$

Let us denote the left–hand side of the above inequality by $g(p)$. Note that

$$g(p) = (\log 3 - \log 2) + \frac{1}{p}(4\log 2 - 3\log 3).$$

Since $4\log 2 - 3\log 3 < 0$, it follows that $g(p) \ge g(2) = \log(2/\sqrt{3}) > 0$ for all $p > 2$.

Case 2: $x_1 = 3$. Then, we have

$$f_3(3, x_2, x_3) = 3^{1/p}(3 - x_2)^{1/p}(3 - x_3)^{1/p} \leq 3^{3/p}.$$

To prove that $f_3(3, x_2, x_3) < f_3(1, 1, 1) = 3 \cdot 2^{2/p}$, it suffices to show that $3^{1-3/p} \cdot 2^{2/p} > 1$, or equivalently,

$$\left(1 - \frac{3}{p}\right) \log 3 + \frac{2}{p} \log 2 > 0.$$

Let us denote the left–hand side of the above inequality by $h(p)$. Note that

$$h(p) = \log 3 + \frac{1}{p}(2 \log 2 - 3 \log 3).$$

Since $2 \log 2 - 3 \log 3 < 0$, it follows that $h(p) \geq h(2) = \log(2/\sqrt{3}) > 0$ for all $p > 2$. This completes the proof.

Problem 2 (20pts). The Lagrangian dual of the given problem is

$$v_d^* = \max_{u \geq 0} \theta(u), \tag{1}$$

where

$$\theta(u) = \min_{x \in \mathbb{R}^n} \left\{ x^T A x + 2b^T x + u(x^T x - 1) \right\}. \tag{2}$$

For a given $u \geq 0$, let $A + uI = V \Sigma V^T$ be the spectral decomposition of $A + uI$. Suppose that $A + uI \not\succeq \mathbf{0}$. Then, there exists an index $i \in \{1, \ldots, n\}$ such that $\Sigma_{ii} < 0$. Upon defining $x(\alpha) = \alpha V e_i$, where $\alpha \in \mathbb{R}$ is a parameter and $e_i \in \mathbb{R}^n$ is the i–th basis vector, we see that

$$x(\alpha)^T (A + uI) x(\alpha) + 2b^T x(\alpha) - u = \alpha^2 \Sigma_{ii} + 2\alpha b^T V e_i - u \to -\infty \quad \text{as } |\alpha| \to +\infty.$$

On the other hand, if $A + uI \succeq \mathbf{0}$, then every optimal solution x^* to (2) satisfies $(A + uI)x^* = -b$. Let $(A + uI)^\dagger$ be the Moore–Penrose pseudoinverse of $A + uI$. We claim that $\bar{x} = -(A + uI)^\dagger b$ is an optimal solution to (2). Indeed, consider an arbitrary solution x^* to the linear equation $(A + uI)x = -b$. Since $(A + uI)^\dagger$ satisfies the identity

$$(A + uI)(A + uI)^\dagger (A + uI) = A + uI,$$

we have

$$-b = (A + uI)x^* = (A + uI)(A + uI)^\dagger (A + uI)x^* = (A + uI)\left[-(A + uI)^\dagger b\right] = (A + uI)\bar{x}.$$

This establishes the claim. In particular, this implies that

$$\theta(u) = \bar{x}^T (A + uI)\bar{x} + 2b^T \bar{x} - u = -b^T (A + uI)^\dagger b - u.$$

Summarizing the above discussion, we obtain

$$\theta(u) = \begin{cases} -b^T (A + uI)^\dagger b - u & \text{if } A + uI \succeq \mathbf{0}, \\ -\infty & \text{otherwise,} \end{cases}$$

which shows that Problem (1) is equivalent to

$$\begin{aligned} v_d^* \quad = \quad & \text{maximize} \quad -b^T(A+uI)^\dagger b - u \\ & \text{subject to} \quad A + uI \succeq \mathbf{0}, \\ & \qquad\qquad\quad u \geq 0. \end{aligned} \tag{3}$$

Now, observe that the proof in Problem 1 of Tutorial 3 can be extended to obtain the following equivalence:

$$t \geq b^T(A+uI)^\dagger b \quad \Longleftrightarrow \quad \begin{bmatrix} A+uI & b \\ b^T & t \end{bmatrix} \succeq \mathbf{0}.$$

Thus, Problem (3) is equivalent to

$$\begin{aligned} v_d^* \quad = \quad & \text{maximize} \quad -t-u \\ & \text{subject to} \quad \begin{bmatrix} A+uI & b \\ b^T & t \end{bmatrix} \succeq \mathbf{0}, \\ & \qquad\qquad\quad u \geq 0. \end{aligned}$$

This completes the proof.

Problem 3 (40pts).

(a) **(20pts).** By Theorem 16(b) of Handout 2 and Propositions 2 and 3 of Handout 7, x^* is an optimal solution to the optimization problem

$$\min_{x \in \mathbb{R}^n} \{f(x) + g(x)\} \tag{4}$$

iff

$$0 \in \nabla f(x^*) + \partial g(x^*). \tag{5}$$

Now, let x^* be a solution to (5), and consider the optimization problem

$$\min_{y \in \mathbb{R}^n} \left\{ g(y) + \frac{1}{2}\|x^* - \nabla f(x^*) - y\|_2^2 \right\}. \tag{6}$$

Since the objective function of the above problem is strongly convex in y, it has a unique minimum $y^* = \text{prox}_1(x^* - \nabla f(x^*))$, which satisfies the following necessary and sufficient optimality condition:

$$0 \in \nabla f(x^*) - x^* + y^* + \partial g(y^*). \tag{7}$$

Since

$$\nabla f(x^*) + \partial g(x^*) = \nabla f(x^*) - x^* + x^* + \partial g(x^*),$$

it follows that x^* simultaneously solves (5) and (7). In particular, we have $x^* = \text{prox}_1(x^* - \nabla f(x^*))$, as desired.

Next, observe that the constrained convex optimization problem

$$\min_{x \in C} f(x) \tag{8}$$

is equivalent to the unconstrained convex optimization problem (4) when $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the indicator of C, i.e.,

$$g(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Hence, by our result above, x^* is an optimal solution to (8) iff it is the optimal solution to (6). However, Problem (6) is equivalent to

$$\min_{y \in C} \|x^* - \nabla f(x^*) - y\|_2^2,$$

which implies that

$$x^* = \text{prox}_1(x^* - \nabla f(x^*)) = \Pi_C(x^* - \nabla f(x^*)).$$

(b) **(20pts).** Since ∇f is Lipschitz continuous with parameter $L > 0$, we have

$$f(x^{k+1}) - f(x^k) \leq \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L}{2}\|x^{k+1} - x^k\|_2^2. \tag{9}$$

On the other hand, by definition of the iterative scheme, we have

$$x^{k+1} = \text{prox}_{\alpha_k}\left(x^k - \alpha_k \nabla f(x^k)\right).$$

This implies that for all $k \geq 0$,

$$\alpha_k g(x^{k+1}) + \frac{1}{2}\|x^k - \alpha_k \nabla f(x^k) - x^{k+1}\|_2^2 \leq \alpha_k g(x^k) + \frac{1}{2}\alpha_k^2 \|\nabla f(x^k)\|_2^2,$$

which, upon simplifying, is equivalent to

$$g(x^{k+1}) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{1}{2\alpha_k}\|x^k - x^{k+1}\|_2^2 \leq g(x^k). \tag{10}$$

By adding (9) and (10) together and noting that $F(x) = f(x) + g(x)$, we obtain

$$F(x^{k+1}) - F(x^k) \leq \frac{L}{2}\|x^{k+1} - x^k\|_2^2 - \frac{1}{2\alpha_k}\|x^k - x^{k+1}\|_2^2 = -\frac{1}{2}\left(\frac{1}{\alpha_k} - L\right)\|x^k - x^{k+1}\|_2^2,$$

as desired.

Now, if $\alpha_k < 1/L$ for all $k \geq 0$, then the above derivation shows that $\{F(x^k)\}_{k\geq 0}$ is a monotonically decreasing sequence. Since the optimal solution set of Problem (5) is assumed to be non–empty, the sequence $\{F(x^k)\}_{k\geq 0}$ has a finite lower bound. It follows that $\{F(x^k)\}_{k\geq 0}$ converges.