EECS 127/227AT Optimization Models in Engineering

Spring 2019

Midterm 2

1. **(1 Point) Tell us about something you are proud of.**

2. **(1 Point) Tell us about something interesting you learned in a class.**

Do not turn this page until the proctor tells you to do so. You may work on the questions above.

PRINT your name and student ID: ______________________________

3. (9 Points) Convexity of sets

Determine if the sets C given below are convex. Either prove that the set is convex or provide an example to show that it is not convex. Correctly guessing whether the set is convex or non-convex with no/incorrect justification will get 0 points. For each part in the first box write "Yes" or "No" depending on your answer and in the second box provide the justification for your answer. You may use any techniques used in class or discussion to demonstrate or disprove convexity.

(a) (3 points)

$$C = \{x \in \mathbb{R}^2 \mid x_1 x_2 \geq 0\},$$

where $x = [x_1, x_2]^\top$.

Solution: No. The set S is shown in Fig. 3.1 : From the figure it is clear that the set is

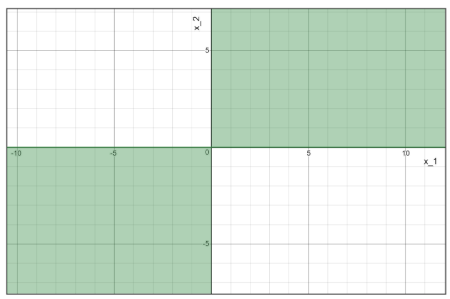

Figure 3.1: $S = \{(x_1, x_2) \mid x_1 x_2 \geq 0\}$.

non-convex. For a formal proof consider points $z_1 = (0, 1)$ and $z_2 = (-1, 0)$. We have $z_1 \in S$ and $z_2 \in S$. Then $z_3 = \frac{z_1+z_2}{2} = (-0.5, 0.5) \notin S$ since $(-0.5) * 0.5 < 0$.

(b) (3 points)

$$C = \{X \in \mathbb{S}^n \mid \lambda_{min}(X) \geq 2\},$$

where $\mathbb{S}^n$ is the set of symmetric matrices in $\mathbb{R}^{n\times n}$, and $\lambda_{min}(X)$ refers to the minimum eigenvalue of X.

Solution: Yes. Consider $X_1, X_2 \in C$. The minimum eigenvalue of X is given as,

$$\lambda_{min}(X) = \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} z^\top X z.$$

Thus we have

$$\min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} z^\top X_1 z \geq 2$$
$$\min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} z^\top X_2 z \geq 2$$

For $\theta \in [0,1]$ consider $X_\theta = \theta X_1 + (1-\theta)X_2$. Then,

$$\begin{aligned}
\lambda_{min}(X_\theta) &= \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} z^\top X_\theta z \\
&= \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} z^\top(\theta X_1 + (1-\theta)X_2)z \\
&= \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} (\theta z^\top X_1 z + (1-\theta) z^\top X_2 z) \\
&\geq \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} \theta z^\top X_1 z + \min_{z\in\mathbb{R}^n \,:\, \|z\|_2=1} (1-\theta) z^\top X_2 z \\
&\geq \theta 2 + (1-\theta)2 \\
&= 2.
\end{aligned}$$

(c) (3 points) Let $\mathcal{B} = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$. Let $\mathcal{H}(w)$ denote the hyperplane with normal direction $w \in \mathbb{R}^n$ i.e. $\mathcal{H}(w) = \{x \in \mathbb{R}^n \mid x^\top w = 0\}$. Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be given by,

$$P(x) = \operatorname*{argmin}_{y\in\mathcal{H}(w)} \|y - x\|_2 .$$

Let

$$C = \{P(x) \mid x \in \mathcal{B}\}.$$

Solution: Yes. **Method 1:**
Let $Q \in \mathbb{R}^{n\times n-1}$ denote the matrix with columns forming a basis for $H(w)$. Then the optimization problem for $P(x)$ can be written as,

$$P(x) = Q \operatorname*{argmin}_{w\in\mathbb{R}^{n-1}} \|Qw - x\|_2^2 ,$$

and has the closed form solution $P(x) = Q(Q^\top Q)^{-1}Q^\top x = Lx$ for $L = Q(Q^\top Q)^{-1}Q^\top$. Note that $P(x)$ is linear in x. $\mathcal{B}$ is a convex set and P is an affine operator and affine transformations of convex sets are convex so C is convex.
Method 2:
Let $Q \in \mathbb{R}^{n\times n-1}$ denote the matrix with columns forming a basis for $H(w)$. Then the optimization problem for $P(x)$ can be written as,

$$P(x) = Q \operatorname*{argmin}_{w\in\mathbb{R}^{n-1}} \|Qw - x\|_2^2 ,$$

and has the closed form solution $P(x) = Q(Q^\top Q)^{-1}Q^\top x = Lx$ for $L = Q(Q^\top Q)^{-1}Q^\top$. Note that $P(x)$ is linear in x.
Let $z_1, z_2 \in C$. This means there exist $x_1, x_2 \in \mathcal{B}$ such that $z_1 = Lx_1$ and $z_2 = Lx_2$. For $\lambda \in [0,1]$ consider $x_\theta = \theta x_1 + (1-\theta)x_2$. Since $\mathcal{B}$ is convex (since norm balls are convex), we have $x_\theta \in \mathcal{B}$. Then,

$$\begin{aligned}
z_\theta &= \theta z_1 + (1-\theta)z_2 \\
&= \theta L x_1 + (1-\theta)L x_2 \\
&= L(\theta x_1 + (1-\theta)x_2) \\
&= L x_\theta \\
&= P(x_\theta),
\end{aligned}$$

and thus belongs in C. This proves that C is convex.

PRINT your name and student ID: ____________________

4. (6 Points) Convexity of functions

Determine if the function f is convex in the following. Justify your answer. Correctly guessing with no/incorrect justification will get 0 points. For each part in the first box write "Yes" or "No" depending on your answer and in the second box provide the justification for your answer. You may use any techniques used in class or discussion to demonstrate or disprove convexity.

(a) (3 Points) $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = x_1 x_2 + 3x_1 + 4x_2 + 16,$$

where $x = [x_1, x_2]^\top, x_1 \in \mathbb{R}, x_2 \in \mathbb{R}$.

Solution: No. The gradient is given by $\nabla f(x) = \begin{bmatrix} x_2 + 3 \\ x_1 + 4 \end{bmatrix}$.

The Hessian is given by,

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The eigenvalues of the Hessian are $\lambda_1 = 1, \lambda_2 = -1$. Since one eigenvalue is negative, $\nabla^2 f(x)$ is not positive semidefinite and thus f is not convex.

(b) (3 Points) $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ with $f(X) = \sigma_{\max}(X)$, which is the largest singular value of X.

Solution: **Method 1:**

Note that $f(X) = \max\{u^\top X v : u \in \mathbb{R}^m, \|u\|_2 = 1, v \in \mathbb{R}^n, \|v\|_2 = 1\}$, which is the pointwise maximum of affine functions of X. Therefore f is convex.

Method 2:

$f(X) = \|X\|_{2,2}$, and any norm is a convex function, so $f(X)$ is convex.

PRINT your name and student ID: ______________________________

5. (12 Points) True or False

Consider the following primal optimization problem:

$$p^* = \inf_{x \in \mathbb{R}^n} f_0(x)$$
$$\text{s.t. } f_i(x) \leq 0, \quad i = 1, 2 \ldots, m,$$

where for all $i = 0, 1, \ldots, m$, the function $f_i : \mathbb{R}^n \to \mathbb{R}$ is differentiable and scalar-valued. **Note that we have made no assumption about the convexity of any of the f_i's.** The Lagrangian for this problem is given by,

$$\mathcal{L}(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i(f_i(x)),$$

where $\lambda \in \mathbb{R}^m, \lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]^\top$. The dual objective function is given by,

$$g(\lambda) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda).$$

The dual problem is,

$$d^* = \sup_{\lambda \geq 0} g(\lambda).$$

Classify each of the following statements as True or False. Justify your answer. Guessing correctly with no/incorrect justification will get 0 points. For each part, write "True" or "False" in the first box and provide the justification for your answer in the second box.

(a) (2 Points) The dual objective function $g(\lambda)$ is concave in λ.

Solution: True. Pointwise infimum of affine functions in λ is concave in λ.

(b) (2 Points) If $\tilde{x} \in \mathbb{R}^n$ and $\tilde{\lambda} \in \mathbb{R}^m$ satisfy the KKT conditions then we necessarily have $p^* = f_0(\tilde{x})$ and $d^* = g(\tilde{\lambda})$.

Solution: False. For an arbitrary optimization problem, the KKT conditions are not even necessary for optimality. Furthermore, if the problem is convex and satisfies a constraint qualification (such as slater's), then the KKT conditions completely characterize optimality.

(c) (2 Points) If $f_0(x)$ is a convex function, $\tilde{x}$ is a primal feasible point (i.e $\tilde{x}$ satisfies the inequality constraints $f_i(\tilde{x}) \leq 0$, for $i = 1, 2, \ldots, m$), and $d^* = 5$, then we necessarily have that $p^* = 5$.

Solution: False. Firstly, we do not even know if the problem is a convex optimization problem (no information is given about $f_i(x)$ being convex). Secondly, Slater's condition is only a sufficient condition for strong duality to hold but it is not necessary. Even if the problem is fully convex, it could have strong duality without Slater's condition being satisfied. Finally, since $\tilde{x}$ is primal feasible, that is not to say there does not exist another primal feasible point that strictly satisfies the inequality constraints. For all these reasons, strong duality may not hold.

(d) (2 Points) If there exists a pair of feasible $\tilde{x} \in \mathbb{R}^n$ and $\tilde{\lambda} \in \mathbb{R}^m$ such that $f_0(\tilde{x}) = g(\tilde{\lambda})$ then we necessarily have $p^* = d^*$.

Solution: True. We have that

$$f_0(\tilde{x}) \geq \inf_{x \in \mathbb{R}^n} f_0(x) = p^* \geq d^* = \sup_{\lambda \geq 0} g(\lambda) \geq g(\tilde{\lambda}),$$

and since $f_0(\tilde{x}) = g(\tilde{\lambda})$, the inequalities hold with equality, and $p^* = d^*$.

(e) Suppose that strong duality holds, the primal problem is convex and differentiable and has a unique optimizer, x^*. Let $n = 1$ and $m = 2$, so there are only two constraint functions $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$. Let these two constraint functions for the primal problem be:

$$f_1(x) = -x + a$$
$$f_2(x) = x - b,$$

with $a, b \in \mathbb{R}$ and $a < b$.

i. (2 Points) If $a < x^* < b$ then we necessarily have $d^* = g(\tilde{\lambda})$ for $\tilde{\lambda} = [0, 0]^\top$.
Solution: True. By complementary slackness, since $f_1(x^*) \neq 0$ and $f_2(x^*) \neq 0$, $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 0$.

ii. (2 Points) It is possible to have $d^* = g(\tilde{\lambda})$ for $\tilde{\lambda} = [1, 1]^\top$.
Solution: False. Complementary slackness would then imply that both inequality constraints are tight, which would require $x^* = a$ and $x^* = b$. Since $a \neq b$, this is impossible. Then $\tilde{\lambda} \neq [1, 1]^\top$.

PRINT your name and student ID: ______________________________

6. (9 Points) Cross-entropy minimization

Let $q_1, q_2, \dots, q_m$ be such that $q_i \geq 0$ for all $i \in \{1, \dots, m\}$ and $\sum_{i=1}^m q_i = 1$. Assume log means the natural log. Let $x = \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix}^\top$. Consider

$$\begin{aligned} \text{minimize} \quad & f(x) = -\sum_{i=1}^m q_i \log(x_i) \\ \text{subject to} \quad & \sum_{i=1}^m x_i \leq 1, \\ & x_i \geq 0 \quad \forall i \in \{1, \dots, m\}. \end{aligned}$$

(a) (2 Points) Is this a convex optimization problem? Justify.

Solution:

Logarithm is a concave function $\implies$ $-\log(x_i)$ is a convex function. The objective function is nonnegative combination of convex functions, so it is convex. The constraint functions are all affine, so they are also convex. As a result, the problem is convex.

(b) (4 Points) Write the dual problem by dualizing only the constraint $\sum_{i=1}^m x_i \leq 1$. Denote the corresponding dual variable by λ.

Solution:

We can write the Lagrangian as

$$\mathcal{L}(x, \lambda) = -\sum_{i=1}^m q_i \log(x_i) + \lambda(\sum_{i=1}^m x_i - 1).$$

The dual function is obtained by

$$g(\lambda) = \min_{x \geq 0} L(x, \lambda).$$

Setting derivative with respect to each x_i to zero, we obtain

$$-\frac{q_i}{x_i^*} + \lambda^* = 0 \quad \text{for } i = 1, \dots, m \implies x_i^* = \frac{q_i}{\lambda^*} \quad \text{for } i = 1, \dots, m. \tag{1}$$

Then the dual problem is given as

$$\max_{\lambda > 0} g(\lambda) = -\sum_{i=1}^m q_i \log(q_i) + \log(\lambda) + 1 - \lambda,$$

where we have used the fact that $\sum_{i=1}^m q_i = 1$.

(c) (3 Points) Find the primal optimal solution x^*. Justify.

Solution:

We can solve the dual problem by setting the derivative to zero:

$$\frac{1}{\lambda^*} - 1 = 0 \implies \lambda^* = 1.$$

In addition, since the primal problem is strictly feasible (e.g. $x_1 = \cdots = x_m = \frac{1}{m+1}$), Slater's condition is satisfied, and strong duality holds. Then, we can use (1) to obtain the optimal solution x^*:

$$x_i^* = q_i \quad \text{for } i = 1, \ldots, m.$$

PRINT your name and student ID: ______________________________

7. (11 Points) Gradient Descent Algorithm

Consider $g : \mathbb{R}^n \to \mathbb{R}$ with $g(x) = \frac{1}{2}x^\top Qx - x^\top b$ where Q is a symmetric positive definite matrix, i.e. $Q \in \mathbb{S}^n_{++}$.

(a) (4 Points) Write the update rule for the gradient descent algorithm

$$x_{k+1} = x_k - \eta \nabla g(x_k),$$

where η is the step size of the algorithm, and bring it into the form

$$(x_{k+1} - x^*) = P_\eta(x_k - x^*),$$

where $P_\eta \in \mathbb{R}^{n\times n}$ is a matrix that depends on η. Find x^* and P_η in terms of Q, b and η. *Note: x^* is a minimizer of g.*

Solution: We have $\nabla g(x) = Qx - b$ and

$$x_{k+1} = x_k - \eta(Qx_k - b) = x_k - \eta Q(x_k - Q^{-1}b).$$

We can write

$$x_{k+1} - Q^{-1}b = x_k - Q^{-1}b - \eta Q(x_k - Q^{-1}b) = (I - \eta Q)(x_k - Q^{-1}b).$$

This shows that $x^* = Q^{-1}b$ and $P_\eta = I - \eta Q$.

(b) (3 Points) Write a condition on the stepsize η and the matrix Q that ensures convergence of x_k to x^* for every initialization of x_0.

Solution: From part (a), we have

$$x_k - x^* = (I - \eta Q)^k(x_0 - x^*).$$

For every initialization x_0, $(x_k - x^*)$ converges to zero if (and only if) all eigenvalues of $(I - \eta Q)$ is in $(-1, 1)$:

$$-1 < 1 - \eta\lambda < 1 \quad \text{for all eigenvalue } \lambda \text{ of } Q.$$

Since Q is positive definite, all of its eigenvalues are positive, and the right hand side of the inequality is satisfied for all $\eta > 0$. For the left hand side of the inequality, we need

$$-1 < 1 - n\lambda \ \forall \lambda_Q \iff \eta < \frac{2}{\lambda_{\max}(Q)}.$$

(c) (4 Points) Assume all eigenvalues of Q are distinct. Let η_m denote the largest stepsize that ensures convergence for all initializations x_0, based on the condition computed in part (b). Does there exist an initialization $x_0 \neq x^*$ for which the algorithm converges to the minimum value of g for certain values of the step size η that are larger than η_m? Justify your answer. *Hint: The question asks if such initializations exist; not whether it is practical to find them.*

Solution: From part (a), we have

$$x_k - x^* = (I - \eta Q)^k (x_0 - x^*).$$

If we want

$$(I - \eta Q)^k (x_0 - x^*) \to \mathbf{0} \quad \text{as } k \to \infty$$

for a specific initialization x_0, the vector $(x_0 - x^*)$ must lie in the eigenspaces of $(I - \eta Q)$ corresponding to the eigenvalues in the range $(-1, 1)$. This explanation gets full credit.

For example, if $\frac{2}{\lambda_1} < \eta < \frac{2}{\lambda_2}$, where λ_1 and λ_2 are the largest two eigenvalues of Q, we have $(I - \eta Q)^k (x_0 - x^*) \to \mathbf{0}$ as long as $(x_0 - x^*)$ does not have any component in the eigenspace corresponding to the minimum eigenvalue of $(I - \eta Q)$.