Convex Optimization Assignment 5 Solutions

General Instructions

The following document contains the solutions to the questions for Assignment 5. Please note that the solutions provided may not be the only possible way to solve the questions. They indicate only one of the many (possibly) valid solutions. The solutions provided are relatively crisp and do not include all the steps that you must have. Your solution should be logical and contain all supporting arguments. Feel free to contact any of the TAs via email in case of any discrepancy you find in the solutions provided.

Question 1

Using the second order condition for convvexity, we can conclude a function is convex iff its Hessian is positive semi-definite.

Using Q1 from Assignment 4:

The Hessian is:

$$\begin{bmatrix} \dfrac{x_2^2}{(x_1^2 + x_2^2)^{(3/2)}} & \dfrac{-x_1 x_2}{(x_1^2 + x_2^2)^{(3/2)}} \\ \\ \dfrac{-x_1 x_2}{(x_1^2 + x_2^2)^{(3/2)}} & \dfrac{x_1^2}{(x_1^2 + x_2^2)^{(3/2)}} \end{bmatrix}$$

To check if the Hessian is positive semi-definite, we can use the following test:

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \dfrac{x_2^2}{(x_1^2 + x_2^2)^{(3/2)}} & \dfrac{-x_1 x_2}{(x_1^2 + x_2^2)^{(3/2)}} \\ \\ \dfrac{-x_1 x_2}{(x_1^2 + x_2^2)^{(3/2)}} & \dfrac{x_1^2}{(x_1^2 + x_2^2)^{(3/2)}} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \geq 0$$

The above inequality is true for all a_1 and a_2 if the Hessian is positive semi-definite.

The above inequality simplifies to:

$$\frac{(a_1 x_2 - a_2 x_1)^2}{(x_1^2 + x_2^2)^{(3/2)}} \geq 0 \ \forall a_1, a_2$$

Thus, the function is convex.

For $\sqrt{(x_1^n + x_2^n)}$, the Hessian is:

$$\begin{bmatrix} \dfrac{x_2^2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} & \dfrac{-x_1 x_2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} \\ \\ \dfrac{-x_1 x_2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} & \dfrac{x_1^2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} \end{bmatrix}$$

To check if the Hessian is positive semi-definite, we can use the following test:

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \dfrac{x_2^2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} & \dfrac{-x_1 x_2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} \\ \\ \dfrac{-x_1 x_2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} & \dfrac{x_1^2 n (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \geq 0$$

The above inequality is true for all a_1 and a_2 if the Hessian is positive semi-definite.

The above inequality simplifies to:

$$\frac{n(a_1 x_2 - a_2 x_1)^2 (x_1^n + x_2^n)^{(n-2)}}{(x_1^n + x_2^n)^2} \geq 0 \ \forall a_1, a_2$$

Thus, the function is convex.

Thus, the function is convex (whenever defined) for all $n \geq 2$

Question 2

Computing gradient of $f(x, \lambda) = \lambda e^{-\lambda x}$ where $x \in \mathbb{R}, \lambda \in \mathbb{R}$

$$\frac{\partial f}{\partial x} = -\lambda^2 e^{-\lambda x}$$
$$\frac{\partial f}{\partial \lambda} = -\lambda x e^{-\lambda x} + e^{-\lambda x}$$

Thus,

$$\nabla_{x,\lambda} f(x, \lambda) = e^{-\lambda x} \begin{bmatrix} -\lambda^2 \\ -\lambda x + 1 \end{bmatrix}$$

Computing the Hessian

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial(-\lambda^2 e^{-\lambda x})}{\partial x} = \lambda^3 e^{-\lambda x}$$

$$\frac{\partial^2 f}{\partial x \partial \lambda} = \frac{\partial(-\lambda x e^{-\lambda x} + e^{-\lambda x})}{\partial x} = (\lambda x - 2)\lambda e^{-\lambda x}$$

$$\frac{\partial^2 f}{\partial \lambda \partial x} = \frac{\partial(-\lambda^2 e^{-\lambda x})}{\partial \lambda} = (\lambda x - 2)\lambda e^{-\lambda x}$$

$$\frac{\partial^2 f}{\partial \lambda^2} = \frac{\partial(-\lambda x e^{-\lambda x} + e^{-\lambda x})}{\partial \lambda} = (\lambda x - 2)x e^{-\lambda x}$$

Thus,

$$\nabla^2_{x,\lambda} f = e^{-\lambda x} \begin{bmatrix} \lambda^3 & (\lambda x - 2)\lambda \\ (\lambda x - 2)\lambda & (\lambda x - 2)x \end{bmatrix}$$

Question 3

Given that $f(x)$ is a convex function, we need to show that $g(x) = f(Ax + b)$ is convex too where $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^m$

Case 1 : $m = n$

In this case, $\mathbf{A}$ is square. Suppose $f(\cdot)$ is convex. We need to show that

$$g(\mathbf{x}) := f(\mathbf{Ax} + \mathbf{b})$$

is also convex.

Let $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. We want to show that:

$$g(\lambda \mathbf{y}_1 + (1 - \lambda)\mathbf{y}_2) \le \lambda g(\mathbf{y}_1) + (1 - \lambda)g(\mathbf{y}_2)$$

Let $\mathbf{z}_1 = \mathbf{Ay}_1 + \mathbf{b}$ and $\mathbf{z}_2 = \mathbf{Ay}_2 + \mathbf{b}$.

By the convexity of $f(\cdot)$, we have:

$$f(\lambda \mathbf{z}_1 + (1 - \lambda)\mathbf{z}_2) \le \lambda f(\mathbf{z}_1) + (1 - \lambda)f(\mathbf{z}_2)$$

Now, $\mathbf{y}_1 = \mathbf{A}^{-1}(\mathbf{z}_1 - \mathbf{b})$ and $\mathbf{y}_2 = \mathbf{A}^{-1}(\mathbf{z}_2 - \mathbf{b})$.

Substituting back into $g(\cdot)$, we have:

$$g(\mathbf{y}_1) = f(\mathbf{A}\mathbf{A}^{-1}(\mathbf{z}_1 - \mathbf{b}) + \mathbf{b}) = f(\mathbf{z}_1)$$
$$g(\mathbf{y}_2) = f(\mathbf{A}\mathbf{A}^{-1}(\mathbf{z}_2 - \mathbf{b}) + \mathbf{b}) = f(\mathbf{z}_2)$$

$$g(\lambda \mathbf{y}_1 + (1 - \lambda)\mathbf{y}_2) = f(\lambda(\mathbf{A}\mathbf{A}^{-1}(\mathbf{z}_1 - \mathbf{b})) + (1 - \lambda)(\mathbf{A}\mathbf{A}^{-1}(\mathbf{z}_2 - \mathbf{b})) + \mathbf{b})$$
$$= f(\lambda \mathbf{z}_1 + (1 - \lambda)\mathbf{z}_2)$$

So,

$$g(\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) \le \lambda g(\mathbf{y}_1) + (1-\lambda)g(\mathbf{y}_2)$$

Thus, $g(\cdot)$ is convex.

Case 2: $m = 1$

The same result would hold for this case too, therefore, $g(\cdot)$ would still be convex.

Convexity of $J(\beta)$:

$$J(\beta) = \Sigma_{i=1}^{N} log(1 + e^{-y_i\beta^T x_i}) + \lambda||\beta||^2$$

Given the result that if $f(\cdot)$ is convex, then $g(\mathbf{x}) := f(\mathbf{Ax} + \mathbf{b})$ is also convex, we can show that $J(\beta)$ is convex by considering the function $f(z) = \log(1 + e^{-z})$, which is convex as logarithm of an affine function ($f(z) = -z$) is convex.

Thus,

$$g(\beta) = \sum_{i=1}^{N} f(-y_i\beta^\top x_i)$$

is convex. Additionally, $\lambda||\beta||^2$ is convex for $\lambda > 0$ as it is just the root of sum of squares of the vector β.

Since the sum of convex functions and a convex function is convex, $J(\beta)$ is convex.

Question 4

Let f be convex and $x_1, x_2 \in \text{domain}(f)$ and we defining $g(t)$ as above then Here $\alpha \in [0,1]$

$$\begin{aligned} g(\alpha t_1 + (1-\alpha)t_2) &= f(x_1 + \alpha t_1 x_2 + (1-\alpha)t_2 x_2) \\ &= f(\alpha(x_1 + t_1 x_2) + (1-\alpha)(x_1 + t_2 x_2)) \\ &\le \alpha f(x_1 + t_1 x_2) + (1-\alpha)f(x_1 + t_2 x_2) \\ &= \alpha g(t_1) + (1-\alpha)g(t_2) \end{aligned}$$

Hence g is a convex function.

Conversely, let $g(t)$ be a convex function and $t_1, t_2 \in \mathbb{R}$ and we define some x_1, x_2 without loss of generality.

$$\begin{aligned} g(\alpha t_1 + (1-\alpha)t_2) &\le \alpha g(t_1) + (1-\alpha)g(t_2) \\ f\left(\alpha(x_1 + t_1 x_2) + (1-\alpha)(x_1 + t_2 x_2)\right) &\le \alpha f(x_1 + t_1 x_2) + (1-\alpha)f(x_1 + t_2 x_2) \end{aligned}$$

As the above statement is true for all α, x_1, x_2 hence $f(x)$ is also convex.

Question 5

First, we prove that if a direction $\mathbf{d}$ is feasible, one can traverse ϵ distance along the direction and still be within the domain, and thus $\mathbf{Ad} = \mathbf{0}$

Thus,

$$\mathbf{A}(\mathbf{x}^* + \epsilon\mathbf{d}) = \mathbf{b}$$

Now, since

$$\mathbf{Ax}^* = \mathbf{b} \mathbf{A}\epsilon\mathbf{d} = \mathbf{0} \mathbf{Ad} = \mathbf{0}$$

Now, suppose $\mathbf{Ad} = \mathbf{0}$. We must show that $\mathbf{d}$ is feasible.

Now, consider a feasible point $\mathbf{x}^*$ such that

$$\mathbf{Ax}^* = \mathbf{b}$$

Now, $\mathbf{A}(\mathbf{x}^* + \epsilon\mathbf{d}) = \mathbf{b}$. Since $\mathbf{Ad} = \mathbf{0}$. Thus, $\mathbf{x}^* + \epsilon\mathbf{d}$ is feasible. Thus, $\mathbf{d}$ is a feasible direction.

Thus, $\mathbf{d}$ is feasible iff $\mathbf{Ad} = \mathbf{0}$.

Question 6

$a)$ Since $\mathbf{x} \in \mathbb{R}^2$ all points are interior points. Thus, the first order necessary condition for minima can be stated as follows:

If $\mathbf{x}^*$ is a point of local minima, then $\nabla f(\mathbf{x}^*) = 0$

$$f(\mathbf{x}) = x_1^2 + x_2^2$$

Thus, $\nabla f(\mathbf{x}) = 2\mathbf{x}$. So, $\nabla f(\mathbf{x}) = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$

$[0\ 0]^T$ is the only point that satisfies the first order necessary condition for minima.

$b)$ Since $\mathbf{x} \in \mathbb{R}^2$ all points are interior points. Thus, the second order necessary condition for minima can be stated as follows:

If $\mathbf{x}^*$ is a point of local minima, then for any $\mathbf{d} \in \mathbb{R}^2$

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} \geq 0 \quad \forall d \in \mathbb{R}^n$$

Since $f(\mathbf{x}) = x_1^2 + x_2^2$ and $\nabla f(\mathbf{x}) = 2\mathbf{x}$, $\nabla^2 f(\mathbf{x}) = 2I$

Thus, for any $\mathbf{d} \in \mathbb{R}^2$,

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*)\mathbf{d} = \mathbf{d}^T(2I)\mathbf{d}$$
$$= 2\mathbf{d}^T\mathbf{d}$$
$$= 2\|\mathbf{d}\|^2 \geq 0$$

Thus, $[0 \; 0]^T$ satisfies the second order necessary condition for minima.

Question 7

$a)$ Identifying Points where First-Order Necessary Condition is Satisfied:

To find the points $\mathbf{x}^*$ where the first-order necessary condition for a minimum is satisfied, we need to find the points where the gradient of the objective function $f(\mathbf{x})$ is zero.

Given $f(\mathbf{x}) = x_1^2 - x_2^2$, the gradient of $f(\mathbf{x})$ is:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & -2x_2 \end{bmatrix}$$

Setting $\nabla f(\mathbf{x})$ to zero, we get:

$$2x_1 = 0 \implies x_1 = 0 \text{—} 2x_2 = 0 \implies x_2 = 0$$

So, the point where the first-order necessary condition is satisfied is $\mathbf{x}^* = [0 \; 0]^T$.

$b)$ Checking Second-Order Necessary Condition:

To check if the point $x^* = [0 \; 0]^T$ satisfies the second-order necessary condition for a minimum, we need to check whether the Hessian matrix $H_f(x^*)$ at this point is PSD or not.

Given the Hessian matrix of $f(\mathbf{x})$ is:

$$H_f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

At $x^* = (0, 0)$, the Hessian matrix $H_f(x^*)$ is:

$$H_f(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

So, the eigenvalues of the Hessian are 2 and -2. $-2 < 0$, therefore, $H_f(x^*)$ is not Positive Semi Definite.

So, the point satisfying FONC, does not satisfy SONC.

Question 8

Given $f(x) = x^T Q x + b^T x + 3$

a) For internal points the optimality condition is $\nabla f(x^*) = 0$

$$\nabla f(x) = (Q + Q^T)x + b$$
$$= \begin{bmatrix} 4x_1 + 10x_2 + 1 \\ 10x_1 + 16x_2 + 3 \end{bmatrix}$$

Solving for $\nabla f(x) = 0$ we get

$$x^* = \frac{1}{18} \begin{bmatrix} -7 \\ 1 \end{bmatrix}$$

b) Calculating the Hessian

$$\nabla^2 f(x) = Q + Q^T = \begin{bmatrix} 4 & 10 \\ 10 & 16 \end{bmatrix}$$

Checking if $\nabla^2 f(x^*)$ is a positive semi-definite matrix or not. As $\det(\nabla f(x^*)) = -36$ which corresponds to multiplication of eigenvalues. Moreover, $\nabla^2 f(x)$ is a symmetric matrix hence it has only real eigenvalues. Thus, one eigenvalue of $\nabla^2 f(x)$ is positive and the other is negative. Hence, it is not a positive semi-definite matrix.

Question 9

The constraints can be written as:

$$\mathbf{Ax} \preceq \mathbf{b}$$

$$\begin{bmatrix} 2 & 4 \\ 6 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \preceq \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\implies 2x_1 + 4x_2 \leq 1 \implies 6x_1 + 8x_2 \leq 2 \implies x_1 + 5x_2 \leq 3$$

The objective function is:

$$f(x_1, x_2) = 2x_1 + 3x_2$$

Gradient of the objective function is:

$$\nabla f(x) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Thus, FONC isn't satisfied for any point in the domain subject to the constraints. It is trivial to show that there exists a feasible direction at each of the corner points of the feasible set where FONC isn't satisfied: i.e., $\exists d \in$ feasible $d \ s.t. \ d^T f(\mathbf{x}) < 0$

Question 10

Given $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ are differentiable. Let $h : \mathbb{R}^n \to \mathbb{R}$ such that

$$h(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T \mathbf{g}(\mathbf{x})$$

If $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ and $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots \ldots, g_m(\mathbf{x}))$, then

$$h(\mathbf{x}) = f_1(\mathbf{x})g_1(\mathbf{x}) + \ldots + f_m(\mathbf{x})g_m(\mathbf{x})$$

Note that,

$$\frac{\partial(f_i(\mathbf{x})g_i(\mathbf{x}))}{\partial x_j} = g_i(\mathbf{x})\frac{\partial f_i(\mathbf{x})}{\partial x_j} + f_i(\mathbf{x})\frac{\partial g_i(\mathbf{x})}{\partial x_j}$$

Thus,

$$\frac{\partial(h(\mathbf{x}))}{\partial x_j} = \sum_{i=1}^m (g_i(\mathbf{x})\frac{\partial f_i(\mathbf{x})}{\partial x_j} + f_i(\mathbf{x})\frac{\partial g_i(\mathbf{x})}{\partial x_j})$$

$$= \sum_{i=1}^m (g_i(\mathbf{x})\frac{\partial f_i(\mathbf{x})}{\partial x_j}) + \sum_{i=1}^m (f_i(\mathbf{x})\frac{\partial g_i(\mathbf{x})}{\partial x_j})$$

$$= \mathbf{g}(\mathbf{x})^T \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_j} + \mathbf{f}(\mathbf{x})^T \frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_j}$$

So,

$$D(\mathbf{f}(\mathbf{x})^T \mathbf{g}(\mathbf{x})) = D(h(\mathbf{x}))$$

$$= \mathbf{g}(\mathbf{x})^T (\frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n}) + \mathbf{f}(\mathbf{x})^T (\frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_n})$$

$$= \mathbf{g}(\mathbf{x})^T D(\mathbf{f}(\mathbf{x})) + \mathbf{f}(\mathbf{x})^T D(\mathbf{g}(\mathbf{x}))$$

$\square$