SI251 Convex Optimization
Quiz 1

Instructor: Professor Ye Shi

25 Nov 2024

Note:

- Ensure you provide a detailed calculation process to secure full marks.
- Complete this quiz in **English**, or points will be deducted.

Exercise 1. (Convex Set)(4 pts)

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be the solution set of a quadratic inequality,

$$\mathcal{C} \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \leq 0\} \tag{1}$$

where $\mathbf{A} \in \mathcal{S}^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$. Please use definition of convex set to show that $\mathcal{C}$ is a convex set <u>if</u> $\mathbf{A}$ is positive semi-definite.

Solution:
Let $\mathbf{x}, \mathbf{y} \in \mathcal{C}, \theta \in [0, 1]$, then,

$$\begin{aligned}
&(\theta\mathbf{x} + (1-\theta)\mathbf{y})^T\mathbf{A}(\theta\mathbf{x} + (1-\theta)\mathbf{y}) + \mathbf{b}^T(\theta\mathbf{x} + (1-\theta)\mathbf{y} + c) \\
&= \theta^2\mathbf{x}^T\mathbf{A}\mathbf{x} + 2\theta(1-\theta)\mathbf{x}^T\mathbf{A}\mathbf{y} + (1-\theta)^2\mathbf{y}^T\mathbf{A}\mathbf{y} + \theta(\mathbf{b}^T\mathbf{x} + c) + (1-\theta)(\mathbf{b}^T\mathbf{y} + c) \\
&\leq \theta^2\mathbf{x}^T\mathbf{A}\mathbf{x} + 2\theta(1-\theta)\mathbf{x}^T\mathbf{A}\mathbf{y} + (1-\theta)^2\mathbf{y}^T\mathbf{A}\mathbf{y} - \theta\mathbf{x}^T\mathbf{A}\mathbf{x} - (1-\theta)\mathbf{y}^T\mathbf{A}\mathbf{y} \\
&= \theta(\theta-1)\mathbf{x}^T\mathbf{A}\mathbf{x} + 2\theta(1-\theta)\mathbf{x}^T\mathbf{A}\mathbf{y} - \theta(1-\theta)\mathbf{y}^T\mathbf{A}\mathbf{y} \\
&= \theta(\theta-1)(\mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{y}^T\mathbf{A}\mathbf{y}) \\
&= \theta(\theta-1)(\mathbf{x}-\mathbf{y})^T\mathbf{A}(\mathbf{x}-\mathbf{y})
\end{aligned} \tag{2}$$

Since $\mathbf{A}$ is positive semi-definite, $\theta(\theta-1)(\mathbf{x}-\mathbf{y})^T\mathbf{A}(\mathbf{x}-\mathbf{y}) \leq 0$, which means $\mathcal{C}$ is a convex set.

Exercise 2. (Karush-Kuhn-Tucker Conditions) (5 pts)

Consider the equality constrained least-squares problem

$$\text{minimize} \quad \|\mathbf{A}x - b\|^2$$

$$\text{subject to } \mathbf{G}x = h$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank $\mathbf{A} = n$, and $\mathbf{G} \in \mathbb{R}^{p \times n}$ with rank $\mathbf{G} = p$.

Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution ν^*.

Solution:

(a) The Lagrangian is

$$L(x, \nu) = \|\mathbf{A}x - b\|^2 + \nu^T(\mathbf{G}x - h) = x^T\mathbf{A}^T\mathbf{A}x + (\nu^T\mathbf{G} - 2b^T\mathbf{A})x - \nu^T h,$$

with minimizer

$$x = -\frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{G}^T\nu - 2\mathbf{A}^T b).$$

The dual function is

$$g(\nu) = -\frac{1}{4}(\mathbf{G}^T\nu - 2\mathbf{A}^T b)^T(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{G}^T\nu - 2\mathbf{A}^T b) - \nu^T h.$$

(b) The optimality conditions are

$$2\mathbf{A}^T(\mathbf{A}x^* - b) + \mathbf{G}^T\nu^* = 0, \quad \mathbf{G}x^* = h.$$

(c) From the first equation,

$$x^* = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T b - \frac{1}{2}\mathbf{G}^T\nu^*).$$

Plugging this expression for x^* into the second equation gives

$$\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T b - \frac{1}{2}\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\nu^* = h,$$

i.e.,

$$\nu^* = -2(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T)^{-1}(h - \mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T b).$$

Substituting in the first expression gives an analytical expression for x^*.

Exercise 3. (Gradient Descent)(5 pts)

In this problem, you need to apply *gradient descent method* with *exact line search* to find the optimal value of smooth function $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$. The iteration formula of gradient descent is

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta^* \nabla_{\mathbf{x}} f(\mathbf{x}^{(t)}) \tag{3}$$

where $\eta^* = \arg\min_\eta \{f(\mathbf{x}^{(t)} - \eta \nabla_{\mathbf{x}} f(\mathbf{x}^{(t)}))\}$ is the stepsize determined by the exact line search.

Consider the problem of minimizing an unconstrained quadratic function:

$$\min_{\mathbf{x}} \quad f(x_1, x_2) := (x_1 - x_2)^2 + (x_2 + 1)^2 \tag{4}$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Please figure out $\mathbf{x}^{(1)}$, where the initial point is $\mathbf{x}^{(0)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

Solution:

Given $f(x_1, x_2) = (x_1 - x_2)^2 + (x_2 + 1)^2$, the gradient is:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2(x_1 - x_2) \\ -2(x_1 - x_2) + 2(x_2 + 1) \end{bmatrix}$$

At $\mathbf{x}^{(0)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$:

$$\frac{\partial f}{\partial x_1} = 2(3 - 1) = 4, \quad \frac{\partial f}{\partial x_2} = -2(3 - 1) + 2(1 + 1) = 0$$

$$\nabla f(\mathbf{x}^{(0)}) = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

New point along direction $-\nabla f$:

$$\mathbf{x}(\eta) = \begin{bmatrix} 3 - 4\eta \\ 1 \end{bmatrix}$$

Substitute into $f(x_1, x_2)$:

$$f(\eta) = (2 - 4\eta)^2 + 2^2 = 8 - 16\eta + 16\eta^2$$

Minimize $f(\eta)$:

$$\frac{df}{d\eta} = -16 + 32\eta, \quad 0 = -16 + 32\eta \implies \eta^* = \frac{1}{2}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \eta^* \nabla f(\mathbf{x}^{(0)}) = \begin{bmatrix} 3 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Exercise 4. (μ-strongly Convex Functions) (6 pts)

Using the definition of a μ-strongly convex function:

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|_2^2$$

prove that $\nabla^2 f(x) \succeq \mu I$, assuming f is twice differentiable.
(hint: is $g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2$ convex?)

Solution:

To show that $\nabla^2 f(x) \succeq \mu I$, we start from the definition of μ-strong convexity:

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|_2^2.$$

Consider the function $g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2$. We want to show that $g(x)$ is convex. For convexity, we require:

$$g(y) \geq g(x) + \nabla g(x)^T(y-x).$$

Substituting $g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2$, we have:

$$g(y) = f(y) - \frac{\mu}{2}\|y\|_2^2,$$
$$g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2,$$
$$\nabla g(x) = \nabla f(x) - \mu x.$$

Substituting these into the convexity condition:

$$f(y) - \frac{\mu}{2}\|y\|_2^2 \geq f(x) - \frac{\mu}{2}\|x\|_2^2 + (\nabla f(x) - \mu x)^T(y-x).$$

This simplifies to:

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|_2^2,$$

which is satisfied by the definition of μ-strong convexity. Thus, $g(x)$ is convex.
Since $g(x)$ is convex, its Hessian is positive semidefinite:

$$\nabla^2 g(x) \succeq 0.$$

The Hessian of $g(x)$ is:

$$\nabla^2 g(x) = \nabla^2 f(x) - \mu I.$$

Therefore:

$$\nabla^2 f(x) - \mu I \succeq 0,$$

which implies:

$$\nabla^2 f(x) \succeq \mu I.$$

This shows that the smallest eigenvalue of the Hessian $\nabla^2 f(x)$ is at least μ for all x, confirming the μ-strong convexity of f.