Oregon State University

Final Project

Hyuntaek Oh

ohhyun@oregonstate.edu

Due: Noon on June 15, Sunday

ECE599/ AI539 Nonlinear Optimization (Spring 2025)
Final Project
(Due: Noon on June 15, Sunday.)

Instruction: Students should provide enough detail of the logical procedure of deriving answers. Answers without sufficient justification will receive partial or no credit. For questions involving MATLAB experiments, provide codes with comments. You are allowed to use **Python** instead of MATLAB. The maximum possible score is 100.

1. (25 points) Consider a constrained optimization problem:

$$\begin{aligned} \text{minimize}_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0,\ i = 1, ..., m \\ & g_i(\mathbf{x}) \le 0,\ i = 1, ..., p \end{aligned} \tag{1}$$

Suppose that f, h_i's, and g_i's are twice continuously differentiable. In this question, we will derive a simple way to check whether a given feasible point $\mathbf{x}$ satisfies the KKT conditions.

Let $\bar{\mathbf{x}}$ be a feasible point of (1) and $\mathcal{W}$ denote the subset of indices of inequality constraints that are *active* at $\bar{\mathbf{x}}$, i.e., $\mathcal{W} \subset \{1, ..., p\}$ is such that $i \in \mathcal{W}$ if and only if $g_i(\bar{\mathbf{x}}) = 0$. Let q denote the cardinality of $\mathcal{W}$.

(a) Prove that **if** there exist real constants $\bar{\lambda}_1, ..., \bar{\lambda}_m$ and $\bar{\mu}_i, i \in \mathcal{W}$, such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{\mathbf{x}}) + \sum_{i\in\mathcal{W}} \bar{\mu}_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}^T \tag{2}$$

and $\bar{\mu}_i \ge 0$ for all $i \in \mathcal{W}$, **then** $\bar{\mathbf{x}}$ satisfies the KKT conditions of (1).

To verify that $\bar{x}$ satisfies all the original constraints of the optimization problem:

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

$$g_j(\bar{x}) \le 0 \ \text{ for all } j = 1, ..., p$$

The problem statement assumes that $\bar{x}$ is a feasible point. This means that $\bar{x}$ satisfies all equality constraints and all inequality constraints:

$$h_i(\bar{x}) = 0 \ \text{ for all } i$$

$$g_j(\bar{x}) \le 0 \ \text{ for all } j$$

So, primal feasibility is satisfied. Next, the KKT stationary condition is satisfied at $\bar{x}$ needs verification. Based on the given condition, there exist multipliers $\bar{\lambda}_1, ..., \bar{\lambda}_m \in \mathbb{R}$ and $\bar{\mu}_i \in \mathbb{R}$ for only the active set $i \in \mathcal{W}$, where:

$$\mathcal{W} := \{j \in \{1, ..., p\} | g_j(\bar{x}) = 0\}$$

These satisfy the relation:

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{x}) + \sum_{i \in \mathcal{W}} \bar{\mu}_i \nabla g_i(\bar{x}) = 0$$

Since the KKT stationary condition requires all $j = 1, ..., p$, a full set of multipliers μ_j is:

$$\mu_j := \begin{cases} \bar{\mu}_j & \text{if } j \in \mathcal{W} \\ 0 & \text{if } j \in \mathcal{W} \end{cases}$$

Then, we get:

$$\sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) = \sum_{j \in \mathcal{W}} \bar{\mu}_j \nabla g_j(\bar{x}) + \sum_{j \notin \mathcal{W}} 0 \cdot \nabla g_j(\bar{x}) = \sum_{j \in \mathcal{W}} \bar{\mu}_j \nabla g_j(\bar{x})$$

By applying it with $\lambda_i = \bar{\lambda}_i$ to the earlier equation:

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) = 0$$

This means the KKT stationary condition is satisfied.
What's more, that KKT dual feasibility condition needs to be verified:

$$\mu_j \geq 0 \quad \text{for all } j = 1, ..., p$$

For each case in μ_j, when $j \in W$, $\mu_j = \bar{\mu}_j \geq 0$ by assumption, which means Dual feasibility holds. When $j \notin W$, $\mu_j = 0$ by definition, Dual feasibility trivially holds. So, the condition guarantees that Dual feasibility is satisfied.
Finally, that KKT complementary slackness condition needs to be verified:

$$\mu_j \cdot g_j(\bar{x}) = 0 \quad \text{for all } j = 1, ..., p$$

For each case, if $j \in \mathcal{W}$, then $g_j(\bar{x}) = 0$, and $\mu_j = \bar{\mu}_j$, so:

$$\mu_j g_j(\bar{x}) = \bar{\mu}_j \cdot 0 = 0$$

On the other hand, if $j \notin \mathcal{W}$, then $g_j(\bar{x}) < 0$, and $\mu_j = 0$, so:

$$\mu_j g_j(\bar{x}) = 0 \cdot (\textit{negative}) = 0$$

These mean that complementary slackness holds for all j.
Thus, all four KKT conditions are satisfied at $\bar{x}$, using the given multipliers $\bar{\lambda}_1, ..., \bar{\lambda}_m$ and $\bar{\mu}_i, i \in \mathcal{W}$.

(b) Prove that **if** $\bar{\mathbf{x}}$ satisfies the KKT conditions of (1), **then** there exist real constants $\bar{\lambda}_1, ..., \bar{\lambda}_m$ and $\bar{\mu}_i, i \in \mathcal{W}$, such that

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{\mathbf{x}}) + \sum_{i \in \mathcal{W}} \bar{\mu}_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}^T$$

and $\bar{\mu}_i \geq 0$ for all $i \in \mathcal{W}$.

If $\bar{x}$ satisfies the KKT conditions, then the stationary condition (gradient condition) must hold. So the existence of the multipliers is built into the definition of the KKT conditions.
An example of the problem is:

$$\textit{objective function: } f(x) = x_1^2 + x_2^2$$

$$\textit{Equality constraint: } h(x) = x_1 + x_2 - 1 = 0$$

$$\textit{Inequality constraints: } g_1(x) = -x_1 \leq 0, g_2(x) = -x_2 \leq 0$$

So, this is:

$$\min_{x \in \mathbb{R}^2} \quad f(x) : x_1^2 + x_2^2 \tag{3}$$
$$\textit{subject to} \quad h(x) : x_1 + x_2 = 1 \tag{4}$$
$$g_i(x) : x_1 \geq 0, x_2 \geq 0 \tag{5}$$

Then, suppose $\bar{x} = (0.5, 0.5)$ and check the KKT conditions at $\bar{x}$. The gradient of objective is:

$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \Rightarrow \nabla f(\bar{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The gradient of equality constraint is:

$$\nabla h(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The gradients of inequality constraints are:

$$\nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Since $x_1 = 0.5 > 0$ and $x_2 = 0.5 > 0$, both inequality constraints are inactive $\Rightarrow W = \varnothing$, meaning that at the point $\bar{x}$, non of the inequality constraints are binding (i.e.$= 0$).
Next, let's find λ such that:

$$\nabla f(\bar{x}) + \lambda \nabla h(\bar{x}) = 0$$

That is:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \Rightarrow \lambda = -1$$

So, stationary condition satisfied.
Because $\bar{x}$ satisfies KKT, the definition tells there must exist $\bar{\lambda}, \bar{\mu}$ such that:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{\mathbf{x}}) + \sum_{i \in \mathcal{W}} \bar{\mu}_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}^T$$

In the example above, we found $\lambda = -1$, and there are no active inequality constraints $\Rightarrow \mu_i$ do not appear. Therefore, the multipliers exist.
*Thus, we conclude that **if** $\bar{\mathbf{x}}$ satisfies the KKT conditions of (1), **then** there exist real constants $\bar{\lambda}_1, ..., \bar{\lambda}_m$ and $\bar{\mu}_i, i \in \mathcal{W}$, such that*

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{\mathbf{x}}) + \sum_{i \in \mathcal{W}} \bar{\mu}_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}^T$$

and $\bar{\mu}_i \geq 0$ for all $i \in \mathcal{W}$.

(c) Let $\mathbf{A}$ denote the $(m+q)$ by n matrix with its first m rows being $\nabla h_1(\bar{\mathbf{x}}), \nabla h_2(\bar{\mathbf{x}}), ..., \nabla h_m(\bar{\mathbf{x}})$ and its remaining q rows being $\{\nabla g_i(\bar{\mathbf{x}}) : i \in \mathcal{W}\}$. Suppose that $\mathbf{A}$ has full row rank, i.e., its rows are linearly independent. Show that if there exist $\bar{\lambda}_1, ..., \bar{\lambda}_m$ and $\bar{\mu}_i, i \in \mathcal{W}$, satisfying the equation (2), then they can be obtained as follows:

$$\begin{bmatrix} \bar{\lambda} \\ \bar{\mu} \end{bmatrix} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A} \cdot (-\nabla f(\bar{\mathbf{x}})^T).$$

where $\bar{\lambda} := (\bar{\lambda}_1, ..., \bar{\lambda}_m)$, and $\bar{\mu}$ denotes the q-dimensional vector consisting of $\bar{\mu}_i, i \in \mathcal{W}$.

Firstly, we can rewrite the stationary condition in matrix form. $A \in \mathbb{R}^{(m+q)\times n}$ *is the matrix whose rows are gradients of the equality and active inequality constraints. This means that* A *is:*

$$A = \begin{bmatrix} \nabla h_1(\bar{x})^T \\ \nabla h_2(\bar{x})^T \\ \cdots \\ \nabla h_m(\bar{x})^T \\ \nabla g_{i_1}(\bar{x})^T \\ \cdots \\ \nabla g_{i_q}(\bar{x})^T \end{bmatrix}$$

, and z *is the matrix of equality and active inequality constraints:*

$$z := \begin{bmatrix} \bar{\lambda} \\ \bar{\mu} \end{bmatrix} \in \mathbb{R}^{m+q}$$

Based on the equation (2) in the problem (a), the KKT stationary condition would be:

$$\nabla f(\bar{x}) + A^T z = 0 \Rightarrow A^T z = -\nabla f(\bar{x})$$

, where $\nabla f(\bar{x}) \in \mathbb{R}^n$ *is the gradient of the objective function.*
In the problem (c), we suppose that A *has full row rank, meaning that the rows of* A *are linearly independent. This also implies that the matrix* $AA^T \in \mathbb{R}^{(m+q)\times(m+q)}$ *is invertible. To solve* $A^T z = -\nabla f(\bar{x})$ *for* z*, both sides are multiplying by* A*:*

$$AA^T z = -A\nabla f(\bar{x})^T$$

Now since AA^T *is invertible by the full row rank assumption, both sides can be multiplying by* $(AA^T)^{-1}$*:*

$$z = (AA^T)^{-1}A(-\nabla f(\bar{x})^T)$$

This gives us a closed-form solution for the Lagrange multipliers.
Thus, we finally get:

$$\begin{bmatrix} \bar{\lambda} \\ \bar{\mu} \end{bmatrix} = (AA^T)^{-1}A(-\nabla f(\bar{x})^T)$$

(d) Discuss how you can use the results of part (a)-(c) to verify the KKT conditions of (1) for any given point x̄. Provide concrete steps.

From the result of (a), $\bar{x}$ *is a feasible point, satisfying all equality constraints and all inequality constraints:*

$$h_i(\bar{x}) = 0 \text{ for all } i, \quad g_j(\bar{x}) \leq 0 \text{ for all } j$$

This means primal feasibility is satisfied in the KKT conditions.

In the KKT conditions, inequality constraints contribute Lagrange multipliers μ_j only when they are active at the point $\bar{x}$. So, constructing the full KKT system is to determine which inequality constraints are active. Based on the problem statement, the active set of inequality constraints:

$$\mathcal{W} := \{j \in \{1, ..., p\} | g_j(\bar{x}) = 0\}$$

Only the constraints in $\mathcal{W}$ can have nonzero Lagrange multipliers μ_j. For all inactive constraints like $g_j(\bar{x}) < 0$, the corresponding Lagrange multipliers must be zero:

$$\mu_j = 0 \quad \text{for all } j \notin \mathcal{W}$$

The result of (a) confirms that inactive constraints are excluded from the stationary condition, and $\mu_j = 0$ is assigned automatically for them.

To solve for the Lagrange multipliers λ and μ, the gradients of the constraints is represented in matrix form. This is a step of construction of matrix A, which will be used to solve the stationary condition of the KKT (we get this in problem 1-(c)).

From the result of (c), the matrix A has full row rank, and the stationary condition of the KKT can be a closed-form solution as described in (c):

$$z := \begin{bmatrix} \bar{\lambda} \\ \bar{\mu} \end{bmatrix} = (AA^T)^{-1} A(-\nabla f(\bar{x})^T)$$

, where $\bar{\lambda} = (\bar{\lambda}_1, ..., \bar{\lambda}_m) \in \mathbb{R}^m$ and $\bar{\mu} = (\bar{\mu}_j)_{j \in \mathcal{W}} \in \mathbb{R}^q$. If needed for verification, the full $\mu \in \mathbb{R}^p$ is defined by:

$$\mu_j := \begin{cases} \bar{\mu}_j & \text{if } j \in \mathcal{W} \\ 0 & \text{if } j \in \mathcal{W} \end{cases}$$

Then, to check whether it has dual feasibility:

$$\mu_j \geq 0 \quad \text{for all } j \in \{1, ..., p\}$$

This is for $j \in \mathcal{W}$, $\mu_j \geq 0$ is verified, and for for $j \notin \mathcal{W}$, $\mu_j = 0$. So, they hold, meaning KKT condition, as negative multipliers for inequality constraints are not allowed.

Next, we verify whether the complementary slackness condition holds at the feasible point $\bar{x}$, using the Lagrange multipliers μ_j constructed above.

For each inequality constraint $g_j(x) \leq 0$, the KKT condition requires:

$$\mu_j \cdot g_j(\bar{x}) = 0 \quad \text{for all } j = 1, ..., p$$

If $\mu_j > 0$, *then* $g_j(\bar{x}) = 0$. *This means that the constraint must be active. On the other hand, if* $g_j(\bar{x}) < 0$, *then* $\mu_j = 0$. *This means the constraint is inactive, so no force acts on it.*

As stated in the result of (a), complementary slackness holds in these cases:

case 1 (active constraint): $j \in \mathcal{W} \to (g_j(\bar{x}) = 0,\ \mu_j = \bar{\mu}_j) \Rightarrow \mu_j \cdot g_j(\bar{x}) = \bar{\mu}_j \cdot 0 = 0$

case 2 (inactive constraint): $j \notin \mathcal{W} \to (g_j(\bar{x}) < 0,\ \mu_j = 0) \Rightarrow \mu_j \cdot g_j(\bar{x}) = 0 \cdot (negative) = 0$

These show that complementary slackness hold.

Lastly, we confirm that the computed Lagrange multipliers $\bar{\lambda}, \bar{\mu}$ *satisfy the KKT stationary condition at the point* $\bar{x}$.

The condition requires that:

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) = 0$$

Based on the two cases of μ_j, *the equation above can be simplified to:*

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{x}) + \sum_{j \in \mathcal{W}} \bar{\mu}_j \nabla g_j(\bar{x}) = 0$$

Now, the construction we did above can be used:

$$A^T z = \sum_{i=1}^{m} \bar{\lambda}_i \nabla h_i(\bar{x}) + \sum_{j \in \mathcal{W}} \bar{\mu}_j \nabla g_j(\bar{x}) \in \mathbb{R}^n$$

So, $\nabla f(\bar{x}) + A^T z = 0$ *shows that its stationary holds.*

Thus, the KKT conditions at any feasible point $\bar{x}$ *is verified.*

2. (25 points) This is Exercise 2 in Chapter 13 of our textbook [Luenberger & Ye]. Consider the constrained problem:

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}\in\mathbb{R}^n} & f(x) \\ \text{subject to} & \mathbf{x} \in \mathcal{S} \end{array} \tag{6}$$

Let $\{c_k\}$ denote a sequence of positive real numbers satisfying (i) $c_{k+1} > c_k$ for all k, and (ii) $c_k \to \infty$. Let $P(\mathbf{x})$ be a penalty function satisfying (i) $P(\mathbf{x})$ is a continuous function of $\mathbf{x}$, (ii) $P(\mathbf{x}) = 0$ if $\mathbf{x} \in \mathcal{S}$, and (iii) $P(\mathbf{x}) > 0$ if $\mathbf{x} \notin \mathcal{S}$. In addition, let $q(c, \mathbf{x}) := f(\mathbf{x}) + cP(\mathbf{x})$.

Let ϵ be a positive real number, and let $\{\mathbf{x}_k\}$ be a sequence such that each $\mathbf{x}_k$ satisfies

$$q(c_k, \mathbf{x}_k) \le \left[\min_{\mathbf{x}\in\mathbb{R}^n} q(c_k, \mathbf{x})\right] + \epsilon.$$

In other words, $\mathbf{x}_k$ might not be a global minimum point of $q(c_k, \mathbf{x})$; however, it at least satisfies that $q(c_k, \mathbf{x}_k) - \min_{\mathbf{x}\in\mathbb{R}^n} q(c_k, \mathbf{x})$ is bounded above by ϵ (e.g., $\mathbf{x}_k$ could be a local minimum point of $q(c_k, \mathbf{x})$ instead of a global minimum point).

Let $\mathbf{x}^*$ denote a global minimum point of (3). Prove that any limit point $\bar{\mathbf{x}}$ of $\{\mathbf{x}_k\}$ is (i) feasible and (ii) satisfies $f(\bar{\mathbf{x}}) \le f(\mathbf{x}^*) + \epsilon$.

(Hint: Using the fact that $\mathbf{x}^*$ is in $\mathcal{S}$, first show that

$$f(\mathbf{x}_k) + c_k P(\mathbf{x}_k) \le f(\mathbf{x}^*) + \epsilon.$$

Let $\bar{\mathbf{x}}$ be an arbitrary limit point of $\{\mathbf{x}_k\}$. Then, by definition, there exists $\mathcal{K} \subset \{1, 2, ...,\}$ such that the subsequence $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ converges to $\bar{\mathbf{x}}$. The above inequality can be used together with the subsequence $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ to show that $P(\bar{\mathbf{x}}) = 0$.)

Given the hint in the problem, to show the inequality:

$$f(\mathbf{x}_k) + c_k P(\mathbf{x}_k) \le f(x^*) + \epsilon.$$

Since x^* is a feasible point ($x^* \in \mathcal{S}$), the penalty function vanishes at the point. So, we get $P(x^*) = 0$. The penalized objective at x^* is:

$$q(c_k, x^*) = f(x^*) + c_k P(x^*) = f(x^*) + c_k \cdot 0 = f(x^*)$$

x^* is a candidate for minimizing $q(c_k, x^*)$, and so:

$$\min_{x\in\mathbb{R}^n} q(c_k, x) \le q(c_k, x^*) = f(x^*)$$

From the assumption in the problem, the sequence $\{x_k\}$ *satisfies:*

$$q(c_k, x_k) \leq \min_{x \in \mathbb{R}^n} q(c_k, x) + \epsilon$$

Combining the two results:

$$q(c_k, x_k) \leq f(x^*) + \epsilon$$

Expanding the definition of q*, we get the desired inequality:*

$$f(x_k) + c_k P(x_k) \leq f(x^*) + \epsilon, \quad \forall k.$$

the terms above rearrange to isolate the penalty:

$$c_k P(x_k) \leq f(x^*) + \epsilon - f(x_k)$$

Then, the both sides of the inequality above is divided by c_k*:*

$$P(x_k) \leq \frac{f(x^*) + \epsilon - f(x_k)}{c_k}$$

Given the satisfactory condition ($c_k \to \infty$*), meaning that the right-hand side converges to 0, assuming* $f(x_k) \geq \inf_x f(x)$ *is bounded below, so it would be:*

$$\frac{f(x^*) + \epsilon - f(x_k)}{c_k} \to 0 \Rightarrow P(x_k) \to 0$$

Hence, the penalty term would be:

$$\lim_{k \to \infty} P(x_k) = 0.$$

(Or, suppose $\forall \delta > 0$*, and for any* $\delta > 0$*, since* $c_k \to \infty$*, there exists* K *such that for all* $k > K$*,*

$$\frac{f(x^*) + \epsilon - f(x_k)}{c_k} < \delta \Rightarrow P(x_k) < \delta \Rightarrow \lim_{k \to \infty} P(x_k) = 0$$

)

This means that as the penalty parameter increases, the violation of the constraint vanishes.

As described in the hint, $\bar{x} \in \mathbb{R}^n$ *is any limit point of* $\{x_k\}$*, and there exists a subsequence* $\{x_k\}_{k \in \mathcal{K}} \subseteq \{x_k\}$ *such that:*

$$x_k \to \bar{x}, \quad \text{as } k \in \mathcal{K} \to \infty$$

From the problem assumptions, the penalty function $P(x)$ is continuous. The limit of penalty values is:

$$\lim_{k\to\infty} P(x_k) = 0 \Rightarrow \lim_{k\in\mathcal{K}} P(x_k) = 0$$

Since $x_k \to \bar{x}$ along the subsequence and P is continuous:

$$P(\bar{x}) = \lim_{k\in\mathcal{K}} P(x_k) = 0.$$

Because $P(x) = 0$ if and only if $x \in S$, we can say:

$$\bar{x} \in S$$

This means $\bar{x}$ is feasible.

From the inequality:

$$f(x_k) + c_k P(x_k) \leq f(x^*) + \epsilon$$

Since $P(x_k) \geq 0$, the penalty can be dropped to get a looser but still valid inequality:

$$f(x_k) \leq f(x^*) + \epsilon$$

This inequality holds for all k, so it also holds along any subsequence $k \in \mathcal{K}$. Hence:

$$\limsup_{k\in\mathcal{K}} f(x_k) \leq f(x^*) + \epsilon.$$

Since $x_k \to \bar{x}$ along $\mathcal{K}$ and f is continuous:

$$\lim_{k\in\mathcal{K}} f(x_k) = f(\bar{x})$$

So the $\limsup$ would be:

$$f(\bar{x}) = \limsup_{k\in\mathcal{K}} f(x_k)$$

Thus, we conclude:

$$f(\bar{x}) \leq f(x^*) + \epsilon.$$

3. (50 points) **Penalty Method.** Consider the constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^4} \quad & \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + b^T\mathbf{x} \\ \text{subject to} \quad & \sum_{i=1}^{4} x_i^2 = 1 \\ & x_i \geq 0, \quad i = 1,2,3,4 \end{aligned} \tag{7}$$

$$\mathbf{Q} = \begin{bmatrix} 2 & 1 & 0 & 10 \\ 1 & 4 & 3 & 0.5 \\ 0 & 3 & -5 & 6 \\ 10 & 0.5 & 6 & -7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ -2 \\ 3 \end{bmatrix} \tag{8}$$

(a) Write the constrained optimization problem (4) in the standard form by properly defining $f, h, g_1, ..., g_4$:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^4} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h(\mathbf{x}) = 0 \\ & g_i(\mathbf{x}) \leq 0, i = 1,2,3,4; \end{aligned}$$

The standard form of the objective function f in the given problem is:

$$f(\mathbf{x}) = \frac{1}{2}x^T Qx + b^T x$$

The equality constraint can be written as:

$$h(\mathbf{x}) := \sum_{i=1}^{4} x_i^2 - 1 = 0$$

The condition $x_i \geq 0$ in the inequality constraint are written as:

$$g_i(\mathbf{x}) := -x_i \leq 0 \quad \text{for } i = 1,2,3,4$$

Thus, the standard form of this problem is:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^4} \quad & f(\mathbf{x}) := \frac{1}{2}x^T Qx + b^T x \\ \text{subject to} \quad & h(\mathbf{x}) := \sum_{i=1}^{4} x_i^2 - 1 = 0 \\ & g_i(\mathbf{x}) := -x_i \leq 0, \quad i = 1,2,3,4; \end{aligned}$$

(b) Implement the penalty method with the *quadratic penalty function* to find a local minimum point of this constrained optimization problem. Provide a pseudocode of your implementation that explains the details of main steps clearly and the stopping criteria.

The following parameters were used in the implementation:

- *Initial point:* $x_0 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^\top$
- *Normalized to:* $x_0/\|x_0\| = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}^\top$
- *Initial penalty parameter:* $c_0 = 1$
- *Penalty increase factor:* $\beta = 2$
- *Stopping tolerance:* $\epsilon = 10^{-6}$
- *Maximum iterations:* $max\_iter = 100$

Algorithm 1 The Penalty Method

Input: Objective function $f(x) = \frac{1}{2}x^\top Qx + b^\top x$
    Equality constraint: $h(x) = \sum_{i=1}^{4} x_i^2 - 1 = 0$
    Inequality constraints: $g_i(x) = -x_i \leq 0, \quad i = 1, \ldots, 4$
    Initial guess $x_{init} \in \mathbb{R}^4$
    Initial penalty parameter $c_0 > 0$
    Penalty update factor $\beta > 1$
    Tolerance $\epsilon > 0$, maximum iterations max_iter

Normalize initial point: $x \leftarrow x_0/\|x_0\|$
Initialize penalty: $c \leftarrow c_0$
for $k = 1$ to max_iter **do**
    Minimize the penalized objective by using gradient descent with backtracking line search:

$$q(c_k, x) = f(x) + c_k \cdot P(x), \text{ where } P(x) = \left(\sum_{i=1}^{4} x_i^2 - 1 \right)^2 + \sum_{i=1}^{4} \max(0, -x_i)^2$$

    Compute penalty value: $P_{\text{val}} \leftarrow P(x)$
    if $c \cdot P_{\text{val}} < \epsilon$ **then**
        break
    end if
    Update penalty: $c \leftarrow \beta \cdot c$
end for
Return final solution x

(c) Let $\mathbf{x}_k$ denote the solution at the end of the k-th iteration of the penalty method. And, let $c_k P(\mathbf{x}_k)$ denote the penalty term value at the end of the k-th iteration. Plot (i) $f(\mathbf{x}_k)$ versus k, (ii) $c_k P(\mathbf{x}_k)$ versus k, and (iii) $P(\mathbf{x}_k)$ versus k. And, interpret the plots.

According to Figure 1. The function value $f(x_k)$ increases monotonically and then stabilizes around iteration $k \approx 15$. Since the original objective is quadratic and not necessarily convex, the function value starting from a low value and increasing is expected as the iterates become feasible. The final plateau indicates convergence to a local minimum of the original constrained problem, after constraint satisfaction becomes strict.

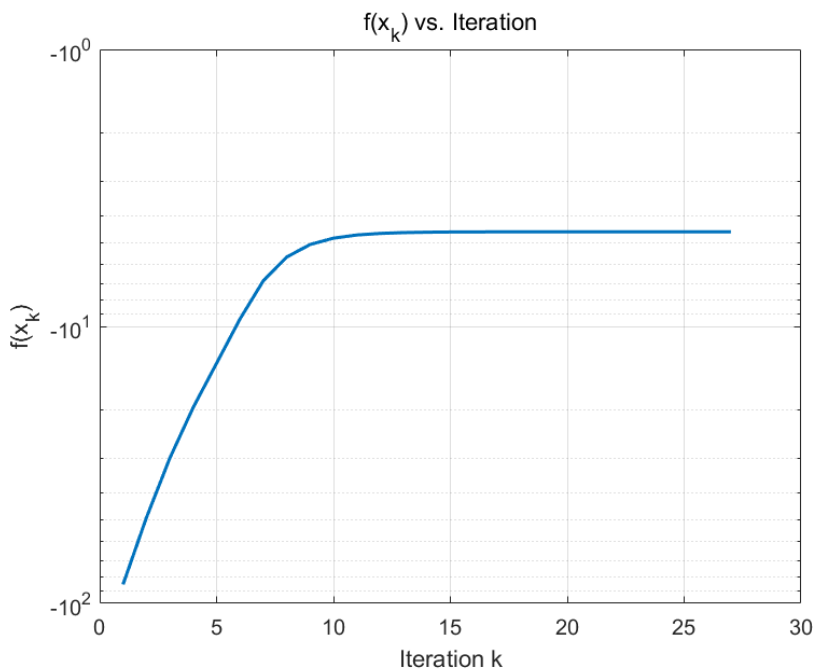

Figure 1: $f(x_k)$ values vs. Iteration k

Figure 2 shows the values of the penalty term in the penalized objective $q(c_k, \mathbf{x}) = f(\mathbf{x}) + c_k \cdot P(\mathbf{x})$. The linear decay on the log-scale indicates that although c_k increases (since penalty weight increases), $P(x_k)$ decreases faster, leading to convergence. The curve reaches below 10^{-6}, which matches your stopping tolerance $\epsilon = 10^{-6}$, explaining why the algorithm stops at around $k = 27$.

$P(x_k)$ measures constraint violation (equality and inequality). As can be seen in Figure 3, the plot shows a steady exponential decrease, approaching 10^{-15} by around $k = 27$. This indicates that the sequence x_k converges to a feasible point satisfying both the equality constraint $\sum x_i^2 = 1$ and inequality constraints $x_i \geq 0$.

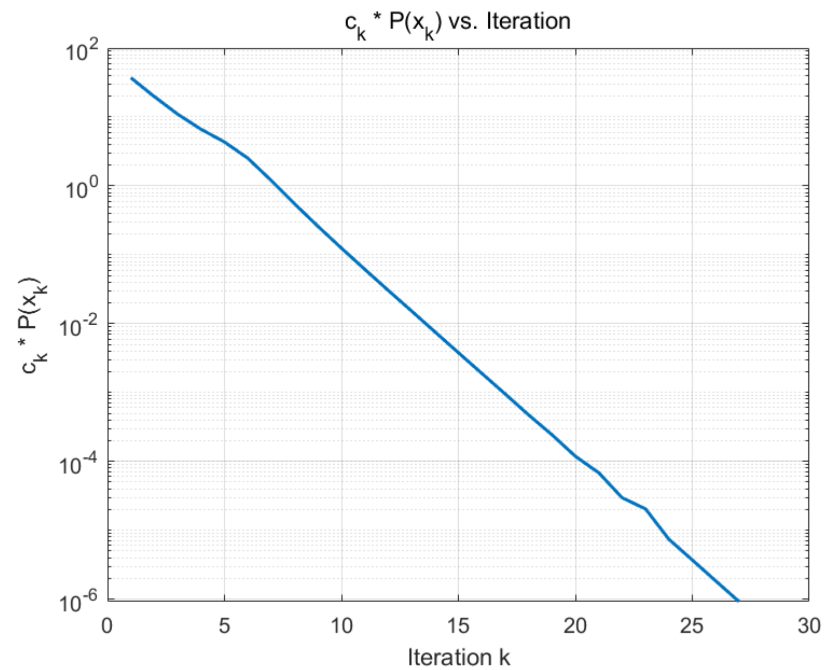

Figure 2: Penalty term-$c_k \cdot P(x_k)$ values vs. Iteration k

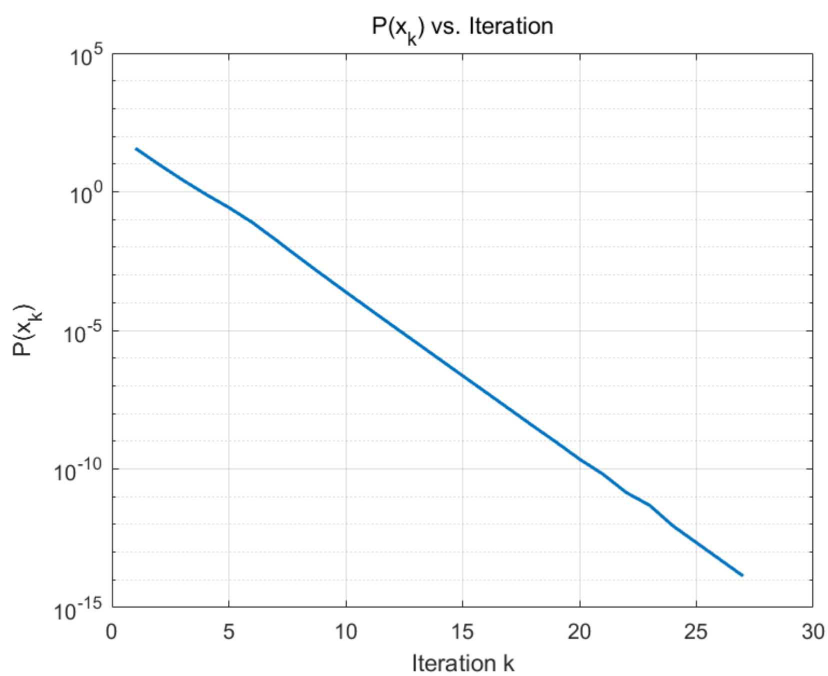

Figure 3: $P(x_k)$ values vs. Iteration k

(d) Let $\mathbf{x}_{final}$ denote the solution found by your implementation of the penalty method. Check whether the KKT conditions hold at $\mathbf{x}_{final}$ (you can use the results of Question 1). Check whether the second order necessary condition holds at $\mathbf{x}_{final}$. Also check whether the second order sufficient condition holds at $\mathbf{x}_{final}$.

- *KKT conditions at x_{final}*
  From the result of (b), x_{final} is:

$$x_{final} = \begin{bmatrix} 0.1120 \\ -0.0000 \\ 0.9937 \\ -0.0000 \end{bmatrix}$$

The elements of x_{final} are examined to check whether the KKT conditions hold at $\mathbf{x}_{final}$. Active constraints at x_{final} are:

$$x_2 = -0.0000 \approx 0 \rightarrow active$$

$$x_4 = -0.0000 \approx 0 \rightarrow active$$

$$x_1, x_3 > 0 \rightarrow inactive$$

So, the active set is:

$$\mathcal{W} = \{2, 4\}$$

The gradients of constraints:

$$\nabla h(x) = 2x = \begin{bmatrix} 0.2240 \\ 0 \\ 1.9874 \\ 0 \end{bmatrix}$$

$$\nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g_4(x) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$$

By using Jacobian matrix A in Question 1-(c), the gradients above are in A:

$$A = \begin{bmatrix} 0.2240 & 0 & 1.9874 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 4}$$

The gradient of objective $\nabla f(x) = Qx + b$, and Q, b are given in the problem. So, the computation would be:

$$Qx + b = \begin{bmatrix} 2 \cdot 0.112 + 1 \cdot 0 + 0 + 10 \cdot 0 \\ 1 \cdot 0.112 + 4 \cdot 0 + 3 \cdot 0.9937 + 0.5 \cdot 0 \\ 0 + 3 \cdot 0 + (-5) \cdot 0.9937 + 6 \cdot 0 \\ 10 \cdot 0.112 + 0.5 \cdot 0 + 6 \cdot 0.9937 + (-7) \cdot 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ -2 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.224 - 1 \\ 0.112 + 2.981 + 0 \\ -4.968 - 2 \\ 1.12 + 5.962 + 3 \end{bmatrix} = \begin{bmatrix} -0.776 \\ 3.093 \\ -6.968 \\ 10.082 \end{bmatrix}$$

$$\nabla f(x) = \begin{bmatrix} -0.776 \\ 3.093 \\ -6.968 \\ 10.082 \end{bmatrix}$$

Then, to solve for multipliers, KKT matrix equation is used:

$$A^T \begin{bmatrix} \lambda \\ \mu_2 \\ \mu_4 \end{bmatrix} = -\nabla f(x)$$

$$\begin{bmatrix} 0.2240 & 0 & 0 \\ 0 & -1 & 0 \\ 1.9874 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \lambda \\ \mu_2 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} 0.224 \cdot \lambda \\ -\mu_2 \\ 1.9874 \cdot \lambda \\ -\mu_4 \end{bmatrix} = -\nabla f(x) = \begin{bmatrix} 0.776 \\ -3.093 \\ 6.968 \\ -10.082 \end{bmatrix}$$

From second and fourth rows:

$$\mu_2 = 3.093 \geq 0, \quad \mu_4 = 10.082 \geq 0$$

From first and third:

$$0.224\lambda = 0.776 \Rightarrow \lambda \approx 3.464$$

$$1.9874\lambda = 6.968 \Rightarrow \lambda \approx 3.506$$

So, KKT stationary holds with:

$$\lambda \approx 3.48, \quad \mu_2 \approx 3.093, \quad \mu_4 \approx 10.082$$

All $\mu_i \geq 0$, so complementary slackness and dual feasibility also hold. Thus, KKT conditions hold.

- *The second order necessary condition at x_{final}*

  According to Theorem in the lecture slides, If x^* is a relative minimum point for the above problem, then there is a $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$ with $\mu \geq 0$ such that the KKT conditions hold and such that:

  $$\mathbf{L}(x^*) := \mathbf{F}(x^*) + \lambda^T \mathbf{H}(x^*) + \mu^T \mathbf{G}(x^*)$$

  is positive semidefinite on the tangent subspace $\mathcal{M}$ of the active constraints at $\mathbf{x}^*$.
  Since each $g_i(x) = -x_i$, the second derivatives $\nabla^2 g_i$ are zero matrices. So, the Hessian of the Lagrangian at x_{final}:

  $$\nabla^2 \mathbf{L}(x_{final}) = Q + 2\lambda I = Q + 6.96I$$

  $\mathcal{M}$ is the tangent subspace at x_{final} given by:

  $$\mathcal{M} = \{y \in \mathbb{R}^4 : y^T \nabla h_i(x_{final}) = 0, i = 1, 2, 3, 4, \quad y^T \nabla g_j(x_{final}) = 0, j \in J\}$$

  , where $J = \{j : g_j(x_{final}) = 0\}$ This means:

  $$\mathcal{M} = \{y \in \mathbb{R}^4 | \begin{bmatrix} 0.2240 & 0 & 1.9874 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} y = 0\}$$

  It defines a 1-dimensional subspace since 3 independent linear equations in $\mathbb{R}^4$ leave a 1D null space.
  Then, $v_1 \in \mathbb{R}^4$ is the orthonormal basis vector of the null space $\mathcal{M}$, which we computed numerically. The second order necessary condition requires:

  $$v_1^T (Q + 2\lambda I) v_1 > 0 \Rightarrow 8.872 > 0$$

  This shows that the Hessian of the Lagrangian is positive definite and hence positive semidefinite on the tangent subspace $\mathcal{M}$.
  Therefore, the second order necessary condition holds at x_{final}.

- *The second order sufficient condition at x_{final}*

  x_{final} is a feasible point that satisfies the KKT conditions. Then, second-order sufficient condition holds at x_{final} if the Hessian of the Lagrangian:

  $$\nabla^2 L(x_{final}) = Q + 2\lambda I$$

  is positive definite on the tangent subspace:

  $$\{y \in \mathbb{R}^4 : y^T \nabla h_i(x_{final}) = 0, i = 1, 2, 3, 4, \quad y^T \nabla g_j(x_{final}) = 0, j \in J\}$$

where $J = \{j : g_j(x_{final}) = 0, \mu_j > 0\}$.
The tangent subspace $\mathcal{M}$ *is:*

$$\mathcal{M} = \{y \in \mathbb{R}^4 | \begin{bmatrix} 0.2240 & 0 & 1.9874 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} y = 0\}$$

It is a one-dimensional subspace of $\mathbb{R}^4$*. This means that there is only one independent direction* y *in which we can move while remaining tangent to the feasible set at* x^**.*
The Hessian of the Lagrangian is:

$$\nabla^2 L(x_{final}) = Q + 2\lambda I = Q + 6.96I$$

Then, $V \in \mathbb{R}^{4\times1}$ *is the orthonormal basis matrix of the tangent space* $\mathcal{M}$*. The restriction of the Hessian to* $\mathcal{M}$ *is:*

$$H = V^T(Q + 2\lambda I)V$$

This is a scalar since $\mathcal{M}$ *is 1-dimensional, so:*

$$y^T\nabla^2 L(x_{final})y = v_1^T(Q + 6.96I)v_1$$

where $v_1 \in \mathbb{R}^4$ *is the orthonormal basis vector of* $\mathcal{M}$*. From the numerical computation, we get:*

$$v_1^T(Q + 6.96I)v_1 > 0 \Rightarrow 8.872 > 0$$

Thus, we conclude that the second-order sufficient condition holds at x_{final}*.*