Q1.2: Given two functions \( f_0(x) = x^2 - 4x + 4 \) and \( f_1(x) = x + 3 \), where \( x \in \mathbb{R} \). Solve \( \min_{x \in \mathbb{R}} f_0(x) \) subject to \( f_1(x) \leq 0 \).

Concepts Emphasized:
- Kuhn Tucker (KT) Conditions:
  - Let \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) be a twice-continuously differentiable function, then...
    - (Necessary Condition): if \( x^* \) is a local minimizer, then \( \nabla f(x^*) = 0 \) and \( \nabla^2 f(x^*) \) is positive semidefinite (i.e. \( x^T \nabla^2 f(x^*) x \geq 0 \) for all \( x \)).
    - (Sufficient Condition): if \( \nabla f(x^*) = 0 \) and \( \nabla^2 f(x^*) \) is positive definite (i.e. \( x^T \nabla^2 f(x^*) x > 0 \) for all \( x \neq 0 \)), then \( x^* \) is a local minimizer.
  - In the single-variable case, \( \nabla^2 f(x^*) \) being positive (semi-)definite simply corresponds to having \( f''(x^*) \geq 0 \).
- Lagrangian & Duality:
  - Given the constrained optimization problem:
    \[
    \begin{align*}
    \text{minimize} & \quad f_0(x) \\
    \text{subject to} & \quad f_1(x) \leq 0 \\
    & \quad \vdots \\
    & \quad f_m(x) \leq 0, \\
    & \quad h_1(x) = 0 \\
    & \quad \vdots \\
    & \quad h_p(x) = 0
    \end{align*}
    \]  
  - The Lagrangian function associated with (1) is given by
    \[
    L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)
    \]
  - Rewriting (1) in terms of the Lagrangian, we can obtain an equivalent primal formulation given by \( \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} \min_{\mu \in \mathbb{R}^p} L(x, \lambda, \mu) \). If we swap the order of the min/max operations, we obtain the associated dual problem, given by \( \max_{\lambda \in \mathbb{R}^m} \min_{\mu \in \mathbb{R}^p} \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \).
  - The solution to the dual problem is guaranteed to be a lower bound to the primal solution (i.e. \( \max_{\lambda \in \mathbb{R}^m} \min_{\mu \in \mathbb{R}^p} L(x, \lambda, \mu) \leq \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \)). This phenomenon is known as weak duality.
• In certain special cases, we can guarantee that the primal/dual formulations will have the same solution (i.e. \( \max_{\lambda \in \mathbb{R}^2_0} \min_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda,b) = \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^2_0} \mathcal{L}(x,\lambda,b) \)), in which case we have strong duality.

• One such case is Slater's condition, which states that if \( f_0, \ldots, f_m \) are all convex and \( h_1, \ldots, h_p \) are all linear, then the existence of a point \( \bar{x} \in \mathbb{R}^n \) such that \( h_1(\bar{x}) = h_2(\bar{x}) = \ldots = h_p(\bar{x}) = 0 \) and \( \begin{cases} f_i(\bar{x}) < 0 \\ f_k(\bar{x}) < 0 \\ s_m(\bar{x}) < 0 \end{cases} \) is sufficient to guarantee strong duality.

Suggestions for Solving:
- The Lagrangian for our constrained minimization problem is given by \( \mathcal{L}(x,\lambda) = (x^2 - 4x + 4) + \lambda(x + 3) \), so our primal formulation is given by \( \min_{x \in \mathbb{R}} \max_{\lambda \in \mathbb{R}_0^2} (x^2 - 4x + 4) + \lambda(x + 3) \). The dual formulation is given by \( \max_{\lambda \in \mathbb{R}_0^2} \min_{x \in \mathbb{R}} (x^2 - 4x + 4) + \lambda(x + 3) \) and will produce an equivalent solution by Slater's condition.

- To solve the dual formulation, we first solve \( \min_{x \in \mathbb{R}} (x^2 - 4x + 4) + \lambda(x + 3) \).

By the KT conditions, we know an optimal \( x^* \) must satisfy \( \frac{\partial}{\partial x} \mathcal{L}(x^*) = 0 \), so \( \frac{\partial}{\partial x} = 2x - 4 + \lambda = 0 \), suggesting \( x^* = \frac{4 - \lambda}{2} = 2 - 0.5\lambda \). To confirm this, note that \( \frac{\partial^2}{\partial x^2} = 2 > 0 \), so \( x^* = 2 - 0.5\lambda \) is a local minimizer.

- With the optimal \( x^* \) for any given \( \lambda \) now determined, our dual problem \( \max_{\lambda \in \mathbb{R}_0^2} \min_{x \in \mathbb{R}} (x^2 - 4x + 4) + \lambda(x + 3) \) now reduces to \( \max_{\lambda \in \mathbb{R}_0^2} \left[ (2 - 0.5\lambda)^2 - 4(2 - 0.5\lambda) + 4 \right] + \lambda(2 - 0.5\lambda + 3) \)

\[ = \max_{\lambda \in \mathbb{R}_0^2} 5\lambda - 0.25\lambda^2 \]. Using KT conditions, we find \( \frac{\partial}{\partial \lambda} (5\lambda - 0.25\lambda^2) = 5 - 0.5\lambda = 0 \Rightarrow \lambda^* = 10 \), which we confirm is optimal by computing \( \frac{\partial^2}{\partial \lambda^2} (5\lambda - 0.25\lambda^2) = -0.5 < 0 \). The optimal point for our original problem is given by \( x^* = 2 - 0.5\lambda = 2 - 0.5(10) = -3 \), corresponding to the value \( f_0(-3) = (-3)^2 - 4(-3) + 4 = 25 \).
Note that if you were to try solving the primal (i.e., \( \min_{x \in \mathbb{R}} \max_{\lambda \geq 0} (x^2 - 4x + 4) + \lambda (x+3) \)) directly, you would first have to solve \( \max_{\lambda \in \mathbb{R}^2} (x^2 - 4x + 4) + \lambda (x+3) \), but this problem doesn't actually have a solution in most cases, as shown by the fact that \( \frac{\partial L}{\partial x} = x + 3 \) does not have any terms containing \( \lambda \). While we can satisfy \( \frac{\partial L}{\partial x} = x + 3 = 0 \) when \( x = -3 \), which happens to correspond to the solution in this case, we cannot expect this to occur in general.
Q 3.2: Write the gradient and Hessian matrix of the quadratic function \( f(\bar{x}) = \bar{x}^T A \bar{x} + 2 \bar{b}^T \bar{x} + c \), where \( \bar{x} \in \mathbb{R}^n \), matrix \( A \in \mathbb{R}^{n \times n} \), vector \( \bar{b} \in \mathbb{R}^n \), and \( c \in \mathbb{R} \).

Concepts Emphasized:
- Gradient and Hessian:
  - For a (twice-differentiable) function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \), the gradient is given by \( \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \) and the Hessian is given by \( \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \)

Suggestions for Solving:
- Letting \( \bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \), \( A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \), and \( \bar{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \), we can expand

\[
\begin{align*}
  f(\bar{x}) &= \bar{x}^T A \bar{x} + 2 \bar{b}^T \bar{x} + c \\
  &= \bar{x}^T \left( x_1 \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix} + x_2 \begin{bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2n} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{n1} \\ a_{n2} \\ \vdots \\ a_{nn} \end{bmatrix} \right) + 2 \bar{b}^T \bar{x} + c \\
  &= x_1 \left( \sum_{i=1}^{n} a_{1i} x_i \right) + x_2 \left( \sum_{i=1}^{n} a_{2i} x_i \right) + \cdots + x_n \left( \sum_{i=1}^{n} a_{ni} x_i \right) + 2 \bar{b}^T \bar{x} + c \\
  &= x_1 \sum_{i=1}^{n} a_{1i} x_i + x_2 \sum_{i=1}^{n} a_{2i} x_i + \cdots + x_n \sum_{i=1}^{n} a_{ni} x_i + 2 \sum_{i=1}^{n} b_i x_i + c \\
  &= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i + 2 \sum_{i=1}^{n} b_i x_i + c \\
  &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i + 2 \sum_{i=1}^{n} b_i x_i + c
\end{align*}
\]

To compute the gradient, note that for any general \( 1 \leq k \leq n \), the partial derivative with respect to \( x_k \) is given by
\[ \frac{d^2 f}{dx_k} = \frac{2}{dx_k} \left( \sum_{i=1}^n \sum_{j=1}^m x_i a_{ij} x_j \right) + 2 \frac{d^2}{dx_k} \left( \sum_{i=1}^n b_i x_i \right) + \frac{d^2}{dx_k}(c) \]

\[ = \sum_{i=1}^n a_{ik} x_i + \sum_{j=1}^m a_{kj} x_j + 2 b_k \text{ (only terms containing } x_k \text{ are preserved).} \]

Repeating this approach for each variable, we get that

\[ \nabla^2 f(x) = \begin{bmatrix} \frac{d^2}{dx_1} & \cdots & \frac{d^2}{dx_n} \\ \frac{d^2}{dx_2} & \ddots & \vdots \\ \frac{d^2}{dx_n} & \cdots & \frac{d^2}{dx_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n a_{i1} x_i + \sum_{j=1}^m a_{ij} x_j + 2 b_i \\ \sum_{i=1}^n a_{i2} x_i + \sum_{j=1}^m a_{ij} x_j + 2 b_i \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n a_{in} x_i + \sum_{j=1}^m a_{ij} x_j + 2 b_i \end{bmatrix} \]

\[ = \begin{bmatrix} \frac{d^2}{dx_1} \\ \frac{d^2}{dx_2} \\ \vdots \\ \frac{d^2}{dx_n} \end{bmatrix} A^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n a_{i1} x_i \\ \sum_{i=1}^n a_{i2} x_i \\ \vdots \\ \sum_{i=1}^n a_{in} x_i \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 2 b_1 \\ 2 b_2 \\ \vdots \\ 2 b_n \end{bmatrix} = (A + A^T) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + 2 \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \]

- Try applying these principles to compute the Hessian in your HW ☺️
Q 3.6: Assume that the matrix $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is not singular. Derive the analytical form of the derivative of $f$ over matrix $A$ (i.e., $[u_{i,j}] = \nabla_A f$, where $u_{i,j} = \frac{\partial f}{\partial a_{i,j}}$) for the function $f = \text{tr}(A^{-1})$.

**Concepts Emphasized:**

- **Matrix Derivative:**
  
  Given a function $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ and a matrix of its independent variables $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$, the derivative with respect to the matrix is defined as
  
  \[ \frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} \]

**Suggestions for Solving:**

- “The Matrix Cookbook” has a section on matrix derivative properties that may be helpful.