Assignment 6

Quill Healey
CX 4803, Computational Foundations of Machine Learning

April 4, 2022

Problem 1

Consider the following constrained optimization problem:

$$\min_{x_1,x_2} \; 2x_1 + 3x_2 - x_1^3 - 2x_2^2$$

subject to

$$\begin{aligned} x_1 + 3x_2 &\le 6 \\ 5x_1 + 2x_2 &\le 10 \\ x_1 &\ge 0 \\ x_2 &\ge 0 \end{aligned}$$

What are all the points that satisfy the Karush-Kuhn-Tucker (KKT) conditions?
Firstly, recall the set of KKT conditions:

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{1a}$$
$$c_i(x^*) = 0, \quad i \in \mathcal{E} \tag{1b}$$
$$c_i(x^*) \ge 0, \quad i \in \mathcal{I} \tag{1c}$$
$$\lambda_i^* \ge 0, \quad i \in \mathcal{I} \tag{1d}$$
$$\lambda_i^* c_i(x^*) = 0, \quad i \in \mathcal{E} \cup \mathcal{I} \tag{1e}$$

where x^* is a locally optimal solution to the optimization problem, the functions c refer to both the equality and inequality constraints, and the function $\mathcal{L}(x^*, \lambda^*)$ is the Lagrangian function given by the following equation

$$L(x, \lambda) \equiv f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x).$$

Thus we refer to the lambdas as the *Lagrange multipliers*, and λ^* are the multipliers which give us the local optimal solutions, x^*, of our objective function $f(x)$.
With these underlying expressions in our tool belt we can begin to solve our desired problem. We will start by identifying the objective function and constraints of the optimization problem, and then use these to construct and solve the Lagrangian function.

$$\begin{aligned} &f(x) = 2x_1 + 3x_2 - x_1^3 - 2x_2^2, \\ &c_1(x) \equiv x_1 + 3x_2 \le 6 \xrightarrow{standardize} 6 - x_1 - 3x_2 \ge 0, \\ &c_2(x) \equiv 5x_1 + 2x_2 \le 10 \xrightarrow{standardize} 10 - 5x_1 - 2x_2 \ge 0, \\ &c_3(x) \equiv x_1 \ge 0, \\ &c_4(x) \equiv x_2 \ge 0 \end{aligned}$$

Plugging into the Lagrangian:

$$
\begin{aligned}
L(x,\lambda) &= 2x_1 + 3x_2 - x_1^3 - 2x_2^2 - \sum_{i=1}^{4} \lambda_i c_i(x) \\
L(x,\lambda) &= 2x_1 + 3x_2 - x_1^3 - 2x_2^2 - (\lambda_1(6 - x_1 - 3x_2) + \lambda_2(10 - 5x_1 - 2x_2) + \lambda_3 x_1 + \lambda_4 x_2) \\
L(x,\lambda) &= 2x_1 + 3x_2 - x_1^3 - 2x_2^2 - \lambda_1(6 - x_1 - 3x_2) - \lambda_2(10 - 5x_1 - 2x_2) - \lambda_3 x_1 - \lambda_4 x_2
\end{aligned}
$$

With this function defined let's find its gradient.

$$
\begin{aligned}
\nabla_{x_1} L(x,\lambda) &= \frac{\partial}{\partial x_1}\left(2x_1 + 3x_2 - x_1^3 - 2x_2^2 - \lambda_1(6 - x_1 - 3x_2) - \lambda_2(10 - 5x_1 - 2x_2) - \lambda_3 x_1 - \lambda_4 x_2\right) \\
&= 2 - 3x_1^2 + \lambda_1 + 5\lambda_2 - \lambda_3
\end{aligned} \tag{2a}
$$

$$
\begin{aligned}
\nabla_{x_2} L(x,\lambda) &= \frac{\partial}{\partial x_1}\left(2x_1 + 3x_2 - x_1^3 - 2x_2^2 - \lambda_1(6 - x_1 - 3x_2) - \lambda_2(10 - 5x_1 - 2x_2) - \lambda_3 x_1 - \lambda_4 x_2\right) \\
&= 3 - 4x_2 + 3\lambda_1 + 2\lambda_2 - \lambda_4
\end{aligned} \tag{2b}
$$

Using (2a) and (2b) we can now write out our problem-specific KKT conditions:

$$
\begin{aligned}
-3x_1^2 + \lambda_1 + 5\lambda_2 - \lambda_3 + 2 &= 0, \\
-4x_2 + 3\lambda_1 + 2\lambda_2 - \lambda_4 + 3 &= 0
\end{aligned} \tag{3a}
$$

$$6 - x_1 - 3x_2 \geq 0, \tag{3b}$$

$$10 - 5x_1 - 2x_2 \geq 0, \tag{3c}$$

$$x_1 \geq 0, \tag{3d}$$

$$x_2 \geq 0, \tag{3e}$$

$$\lambda_i \geq 0, i = 1, \ldots, 4 \tag{3f}$$

$$\lambda_1(6 - x_1 - 3x_2) = 0, \tag{3g}$$

$$\lambda_2(10 - 5x_1 - 2x_2) = 0, \tag{3h}$$

$$\lambda_3 x_1 = 0, \tag{3i}$$

$$\lambda_4 x_2 = 0 \tag{3j}$$

Recall that each inequality constraint can be either active or inactive. Consequently, as we have four inequality constraints, there are 2^4 ways to "guess and check" which points match our KKT conditions. As this would be exhausting to manually calculate, recall that the points satisfying the KKT conditions are locally optimal points. Additionally, keep in mind that since $\nabla f(x)^T = [2 - 3x_1^2,\ 3 - 4x_2]$, increasing x_1 and x_2 will (in the long term) decrease our objective function. Utilizing this knowledge, we can eliminate some particular combinations of constraint equations.

Beginning our search by comparing two edge points $(0,0)$ and $(2,0)$, we find that $f(0,0) = 0$ and $f(2,0) = -4$. Consequently, we can eliminate all combinations of equations where both the (3d) and (3e) constraints are active. Continuing with this meta solution technique, we now compare the edge points $(2,0)$ and $(0,2)$. Finding that $f(0,2) = -2 > -4 = f(2,0)$, we can rule out any combination of constraints where (3d) is active and (this assumption can be deduced as $\nabla_{x_2} f(x)$ implies that in order to decrease $f(x)$ we should increase x_2 over time, however, the feasible region does not allow $x_2 > 2$ and any $0 < x_2 < 2$ where $x_1 = 0$ will be even smaller). This leaves us with the options where both (3d) and (3e) are inactive, or solely (3d) is inactive. Before further exploring the former of the two, let's perform some more reasoning. Clearly the objective function will desire that x_1 and x_2 increase infinitely. It is only ($3b$) and ($3c$) which enforce that x_1 and x_2 stop their positive growth. Consequently, if we desire both (3d) and (3e) to be inactive then both (3b) and (3c) must be active. We can thus set up the equations (notice that we set $\lambda_3 = \lambda_4 = 0$):

$$
\begin{aligned}
-3x_1^2 + \lambda_1 + 5\lambda_2 + 2 &= 0 \\
-4x_2 + 3\lambda_1 + 2\lambda_2 + 3 &= 0 \\
6 - x_1 - 3x_2 &= 0 \\
10 - 5x_1 - 2x_2 &= 0
\end{aligned}
$$

Solving for this system gives $x_1 = \frac{18}{13}$ and $x_2 = \frac{20}{13}$ with optimal solution of $f(\frac{18}{13}, \frac{20}{13}) = -0.0036413$, which is clearly greater than $f(2, 0) = -4$.

We have finally reached the conclusion that (3b) and (3d) are inactive while (3c) and (3e) are active. Keeping in mind $\nabla_{x_1} f(x)$ indicates that we should increase x_1 in the long term, and that $f(1, 0) = 2 - 1 = 1 > -4 = f(2, 0)$, we can conclude that $(2, 0)$ is the sole point that satisfies the KKT conditions.

Problem 2

To find the solution to the following equality-constrained optimization problem:

$$\min_x \frac{1}{2}x^T Hx + h^T x$$

subject to $Ax = b$, we will first compute the matrix algebra necessary to formulate the dual.
We now recall that the Lagrangian dual of a QP is also a QP. Firstly, we write the Lagrangian function as:

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T Hx + h^T x - \lambda^T (Ax - b). \tag{4}$$

Now, defining the (Lagrangian) dual function $q(\lambda)$ as $q(\lambda) = \inf_x L(x, \lambda)$, we find the infimum of $\mathcal{L}$ using $\nabla_x \mathcal{L}(x, \lambda) = 0$. We recall from class that this is

$$x^* = H^{-1}(A^T \lambda - h).$$

We now plug back into (4) to offically define the dual function as:

$$q(\lambda) = -\frac{1}{2}(A^T \lambda - h)H^{-1}(A^T \lambda - h) + \lambda^t b$$

Thus the dual problem is:

$$\begin{aligned} max \quad & q(\lambda) = -\frac{1}{2}(A^T \lambda - h)H^{-1}(A^T \lambda - h) + \lambda^T b \\ s.t. \quad & \lambda \geq 0 \end{aligned} \tag{5}$$

However, solving for the given matrix and vector values (given in the problem), we find the dual can be simplified to

$$\begin{aligned} \max_\lambda \quad & q(\lambda) = -9\lambda^2 + 17\lambda - 4.75 \\ s.t. \quad & \lambda \geq 0 \end{aligned} \tag{6}$$

From here we can implement basic differential calculus to obtain:

$$\frac{dq(\lambda)}{d\lambda} = -18\lambda + 17 \to \lambda = \frac{17}{18},$$

which satisfies the non-negativity constraint. Using this value we find that

$$x^* = \begin{bmatrix} 0.3889 \\ 1.2222 \\ 1.1667 \end{bmatrix},$$

which does fulfill our equality constraint $Ax = b = 4$.

Problem 3

Consider the problem

$$\min_{x_1,x_2} \; x_1^2 + 2x_1x_2$$

subject to $x_1^2x_2 \geq 10$.

(a)

We will use the same method of Lagrange multipliers that we have previously implemented. Recognize that $f(\vec{x}) = x_1^2 + 2x_1x_2$ and $g(\vec{x}) = x_1^2x_2 - 10$. Thus the Lagrangian is $\mathcal{L}(\vec{x}, \lambda) = x_1^2 + 2x_1x_2 - \lambda(x_1^2x_2 - 10)$. To solve for our optimal values and multipliers we firstly recall (1a) from the KKT conditions:

$$\nabla_{x_1}\mathcal{L}(\vec{x}, \lambda) = 2x_1 + 2x_2 - 2\lambda x_1x_2 = 0 \tag{7a}$$
$$\nabla_{x_2}\mathcal{L}(\vec{x}, \lambda) = 2x_1 - \lambda x_1^2 = 0 \tag{7b}$$

Clearly we cannot solve this system of equations as our number of unknowns is greater than our number of equations. However, we can now recall KKT constraint (1e), separating it into two cases: when the problem constraint: $g(\vec{x}) = x_1^2x_2 - 10$ is active and when it is not. We will start with the latter.

Case 1: $\mathbf{g(\tilde{x})}$ is not active ($\lambda = \mathbf{0}$)

Plugging $\lambda = 0$ into (3a) and (3b) we achieve the system:

$$\begin{bmatrix} 2 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which when solved gives $\vec{x}' = [0\ 0]$. Plugging these values into our original problem we see that $x_1^2x_2 = 0 * 0 \ngeq 10$. Consequently we reject our previous claim that $\vec{x}' = [0\ 0]$ is a solution.

Case 2: $\mathbf{g(\tilde{x})}$ is active: Collecting (3a), (3b), and KKT constraint (1e) we find the following system of equations:

$$\begin{aligned} 2x_1 + 2x_2 - 2\lambda x_1x_2 &= 0 \\ 2x_1 - \lambda x_1^2 &= 0 \\ x_1^2x_2 &= 10 \end{aligned}$$

When solved we find $\vec{x}^* = \begin{bmatrix} \sqrt[3]{10} \\ \sqrt[3]{10} \end{bmatrix}$, with corresponding $\lambda^* = \frac{2^{2/3}}{\sqrt[3]{5}}$. This leads to a locally optimal solution of ≈ 13.9248.

(b)

Recall that the dual problem is given by

$$\begin{aligned} max \quad & q(\lambda) \\ s.t. \quad & \lambda \;\geq\; 0 \end{aligned}$$

where $q(\lambda)$ is known as the *dual function*, and is given by the infimum of the primal's Lagrangian: $q(\lambda) = \inf_{x1,x_2} = \mathcal{L}(\vec{x}, \lambda) = x_1^2 + 2x_1x_2 - \lambda(x_1^2x_2 - 10)$. We can find this function by solving for $\nabla_x\mathcal{L}(\vec{x}, \lambda)$.

$$2x_1 + 2x_2 - 2\lambda x_1x_2 = 0$$
$$2x_1 - \lambda x_1^2 = 0 \rightarrow \mathbf{x_1} = \frac{\mathbf{2}}{\lambda}$$
$$\Rightarrow \text{ Plugging into eq.1: } 2 * \frac{2}{\lambda} + 2x_2 - 4x_2 = 0$$
$$\rightarrow -2x_2 = -\frac{4}{\lambda}$$
$$\rightarrow \mathbf{x_2} = \frac{\mathbf{2}}{\lambda}$$

Plugging $_1 = x_2 = \frac{2}{\lambda}$ back into our dual function:

$$q(\lambda) = \frac{4}{\lambda^2} + 2 * \frac{2}{\lambda} * \frac{2}{\lambda} - \lambda(\frac{4}{\lambda^2} * \frac{2}{\lambda} - 10)$$
$$q(\lambda) = \frac{4}{\lambda^2} + 10\lambda,$$

which we can use to formulate our dual problem:

$$\begin{aligned} max \quad & q(\lambda) = \frac{4}{\lambda^2} + 10\lambda \\ s.t. \quad & \lambda \geq 0 \end{aligned}$$

(c) and Correction

Observing the "dual Problem" above, we can quickly realize that this problem is unbounded. Recall that strong duality does state that it is possible for either the primal or dual to be unbounded. **However**, the necessary condition for this is that the other problem is infeasible. As we see in part (a), the primal does have a local optimal solution. Consequently, strong duality dictates that the dual should also have a solution and that the objective functions should be equal.

With this being said, I must wave the white flag. I have a midterm tomorrow and consequently cannot attempt a "buzzer beater", so to speak, on this problem. However, I will state that when solving the dual function we should see that $\lambda^* = \frac{2^{2/3}}{\sqrt[3]{5}}$.
(Other thoughts for solving this revolve around slack variables or somehow re-incorporating x^*. However, typically one problem has both λ^* and x^*, while the other only has the former.)