Name: **Present group members:**

Worksheet 9-1: Q1

Consider the problem

$$\min_{\mathbf{x}=(x_1,x_2)} f(\mathbf{x}) = -x_1 x_2, \quad \text{s.t. } \mathbf{e}^\top \mathbf{x} = 1, \text{ where } \mathbf{e}^T = \begin{bmatrix} 1 & 1 \end{bmatrix},$$

The feasible set for this problem is $U = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{e}^\top \mathbf{x} = 1\} = \{\mathbf{x} \in \mathbb{R}^n : x_1 + x_2 = 1\}$.

1. Take a look at this desmos plot. What is the blue surface? What is the red line? What is the green line? Move the slider for a around, what is the grey point?

   - *The blue surface is the function $f(\mathbf{x})$.*
   - *The red line is the constraint set, U*
   - *The green line is the function restricted to U.*
   - *The grey point is a point on the function. So the height of this thing is what we're trying to minimize.*

2. Based on the plot, does this appear to be a convex problem?

   The constraint set is a line $x_1 + x_2 = 1$, so it is convex. While the function $f(x_1, x_2) = -x_1 x_2$ is a saddle, and so is not convex, the restriction to the line $x_1 + x_2 = 1$ is indeed convex. That means this is actually a convex problem.

3. Let's find the solution without using the stationarity condition first. For a point $\mathbf{x} = (x_1, x_2) \in U$, write down $\mathbf{x}$ in terms of just x_1. Then write down $f(\mathbf{x})$ restricted to U in terms of just x_1.

   - *Since $\mathbf{x} \in U$, $x_1 + x_2 = 1$, so $x_2 = 1 - x_1$.*
   - *So $\mathbf{x} = (x_1, 1 - x_1)$.*
   - *This means that in U, $f(x_1, x_2) = f(x_1, 1 - x_1) = -x_1(1 - x_1) = -x_1 + x_1^2$*

4. Great, this is a function in one variable! Find the minimum. Use this to determine the minimum for the problem. Move the grey point in the desmos plot to check your answer.

   - $\frac{d}{dx_1}(-x_1(1 - x_1)) = -1 + 2x_1$
   - *The derivative above is 0 if $-1 + 2x_1 = 0$, so the minimum is at $x_1 = \frac{1}{2}$*
   - *Going back to the original problem, this means the minimum occurs at $\left(\frac{1}{2}, 1 - \frac{1}{2}\right) = \left(\frac{1}{2}, \frac{1}{2}\right)$*

5. Let's go back and understand the stationarity condition for this problem. First, what is $\nabla f(x_1, x_2)$?

   $\nabla f(x_1, x_2) = \begin{bmatrix} -x_2 \\ -x_1 \end{bmatrix}$

6. We'll start with a point that *isn't* a stationary point and show that the stationarity condition *doesn't* hold. For the point $\mathbf{x}^* = (0, 1)$ and some other point $\mathbf{x} = (x_1, x_2) \in U$, write down the stationarity condition we would check. Put it in terms of only x_1.

   $$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0$$
   $$(-1, 0)^\top (x_1 - 0, (1 - x_1) - 1) \geq 0$$
   $$(-1, 0)^\top (x_1, -x_1) \geq 0$$
   $$-x_1 \geq 0$$

7. To show that $\mathbf{x}^* = (0, 1)$ is *not* a stationary point, use your calculation above to find a point $\mathbf{x} \in U$ that does not satisfy the stationarity condition.

   Any point $(x_1, 1 - x_1)$ with $x_1 > 0$ will work. So, for example, choose $(7, -6)$. This point is in the set U since the sum of the coordinates is 1. However, when we plug it into the stationaryity condition,

   $$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) = -x_1 = -7$$

   so it is not ≥ 0. This means $\mathbf{x}^* = (0, 1)$ is not a stationary point.

8. Now, we'll do this for $\mathbf{x}^*$ which gives the minimum that you found on the first page, which should have been $\mathbf{x}^* = \left(\frac{1}{2}, \frac{1}{2}\right)$. For $\mathbf{x}^*$ equal to that point, what is $\nabla f(\mathbf{x}^*)$?

$$\nabla f(x_1, x_2) = \begin{bmatrix} -x_2 \\ -x_1 \end{bmatrix}, \qquad \nabla f(1/2, 1/2) = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix},$$

9. Say we have some point $\mathbf{x} = (x_1, x_2) \in U$. Write the stationarity condition for this problem we would check for $\mathbf{x}^*$ found above in terms of only x_1. Is there any possible $\mathbf{x} \in U$ that does not satisfy the stationarity condition?

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0$$
$$(-\tfrac{1}{2}, -\tfrac{1}{2})^\top (x_1 - \tfrac{1}{2}, (1 - x_1) - \tfrac{1}{2}) \geq 0 \; f$$
$$(-\tfrac{1}{2}, -\tfrac{1}{2})^\top (x_1 - \tfrac{1}{2}, \tfrac{1}{2} - x_1) \geq 0$$
$$-\tfrac{1}{2} \cdot (x_1 - \tfrac{1}{2}) + \tfrac{-1}{2}(\tfrac{1}{2} - x_1) \geq 0$$

This turns into $0 \geq 0$ *which is always true. So no matter what* $\mathbf{x}$ *is chosen* 0 *is always* ≥ 0*, so it trivially satisfies the stationarity condition. That means the point* $\mathbf{x}^* = \left(-\frac{1}{2}, -\frac{1}{2}\right)$ *is a stationary point.*

Of course we knew it was going to be a stationary point because it's a minimum of a convex problem.

Worksheet 9-1: Q2

Let's extend the example above to the more general case. Consider the optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad \text{s.t. } \mathbf{e}^\top \mathbf{x} = 1, \text{ where } \mathbf{e}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix},$$

where f is a continuously differentiable function over $\mathbb{R}^n$. The feasible set for the problem is

$$U = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^\top \mathbf{x} = 1\} = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1\}.$$

We will show that the stationarity condition here, namely

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \text{ for all } \mathbf{x} \text{ satisfying } \mathbf{e}^\top \mathbf{x} = 1 \tag{1}$$

is satisfied when

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \frac{\partial f}{\partial x_2}(\mathbf{x}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*) \tag{2}$$

1. First, go back to the previous problem. Check that the solution you found for $\mathbf{x}^*$ satisfies the second condition above (Eqn. 2).

   In the example above, $\nabla f(1/2, 1/2) = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$, and the condition above is just that each entry is the same. Here, they're all $-\frac{1}{2}$ so this satisfies the condition.

2. Now we will check that if Eqn. 2 is true, then Eqn. 1 is true. Say that every entry in $\nabla f(\mathbf{x}^*)$ is a (so this is all the things in Eqn. 2). Simplify $\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x})^*$ as much as possible.

$$\begin{aligned} \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}^*)(x_i - x_i^*) \\ &= \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \left(\sum_{i=1}^n x_i - \sum_{i=1}^n x_i^* \right) \\ &= \frac{\partial f}{\partial x_1}(\mathbf{x}^*)(1 - 1) = 0, \end{aligned}$$

3. Why does the result above imply that $\mathbf{x}^*$ is a stationary point?

   To be a stationary point, we need $\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0$ for all $\mathbf{x}$, but since the left side is always 0, this is always true.

Worksheet 9-1: Q3

Consider the convex optimization problem

$$\min_{x,y,z} 2x^2 + 3y^2 + 4z^2 + 2xy - 2xz - 8x - 4y - 2z, \quad \text{s.t. } x, y, z \geq 0.$$

(a) What is the gradient of $f(\mathbf{x}) = 2x^2 + 3y^2 + 4z^2 + 2xy - 2xz - 8x - 4y - 2z$?

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 4x + 2y - 2z - 8 \\ 6y + 2x - 4 \\ 8z - 2x - 2 \end{bmatrix}$$

(b) Fix $\mathbf{x}^* = (\frac{17}{7}, 0, \frac{6}{7})$. What is $\nabla f(\mathbf{x}^*)$? Is $\mathbf{x}^*$ a stationary point of the function f?

$$\nabla f(\mathbf{x}^*) = \begin{bmatrix} 4 \cdot \frac{17}{7} + 2 \cdot 0 - 2 \cdot \frac{6}{7} - 8 \\ 6 \cdot 0 + 2 \cdot \frac{17}{7} - 4 \\ 8 \cdot \frac{6}{7} - 2 \cdot \frac{17}{7} - 2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{6}{7} \\ 0 \end{bmatrix}$$

Since $\nabla f(\mathbf{x}^*)$ is not 0, this point is not a stationary point of the function.

(c) Show that the vector $(\frac{17}{7}, 0, \frac{6}{7})$ is a stationary point of the problem.

To show that it is a stationary point of the problem, we need to check that for any $\mathbf{x}$ in the constraint set,

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0.$$

The point $\mathbf{x} = (x, y, z)$ is in the constraint set if $x, y, z \geq 0$.
We calculate that

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) = \begin{bmatrix} 0 & \frac{6}{7} & 0 \end{bmatrix} \begin{bmatrix} x - \frac{17}{7} \\ y - 0 \\ z - \frac{6}{7} \end{bmatrix} = 0 \cdot \left(x - \tfrac{17}{7}\right) + \tfrac{6}{7} \cdot (y) + 0 \cdot \left(z - \tfrac{6}{7}\right) = \tfrac{6}{7} y.$$

Since to be in the set, $y \geq 0$, this means that $\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0$, which is the definition of being a stationary point of the problem.

(d) Find the first iteration of the gradient projection method starting with $\mathbf{x}_0 = (1, 1, 1)$, and using a constant step size 0.5.

- *The equation for the gradient projection algorithm update step is*

$$\boldsymbol{x}_{k+1} = P_C(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k)).$$

- *For $k = 0$, this is $\boldsymbol{x}_1 = P_C(\mathbf{x}_0 - t_0 \nabla f(\mathbf{x}_0))$.*
- *We have constant step size, so $t_0 = 0.5$.*
- *We have $\mathbf{x}_0 = (1, 1, 1)$. From above, we know ∇f so we can calculate:*

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 4x + 2y - 2z - 8 \\ 6y + 2x - 4 \\ 8z - 2x - 2 \end{bmatrix}, \qquad \nabla f(1, 1, 1) = \begin{bmatrix} 4 + 2 - 2 - 8 \\ 6 + 2 - 4 \\ 8 - 2 - 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \\ 4 \end{bmatrix}$$

- *So*

$$\mathbf{x}_0 - t_0 \nabla f(\mathbf{x}_0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 0.5 \cdot \begin{bmatrix} -4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 - (-2) \\ 1 - 2 \\ 1 - 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}$$

- *In this problem, the constraint set $C = \{(x, y, z) \mid x, y, z \geq 0\} = \mathbb{R}^3_+$. From the projection class earlier, we know that $P_{\mathbb{R}^3_+}(\mathbf{x}) = [\mathbf{x}]_+$.*
- *So,*

$$\boldsymbol{x}_1 = P_C(\mathbf{x}_0 - t_0 \nabla f(\mathbf{x}_0)) = \left[\begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix} \right]_+ = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$