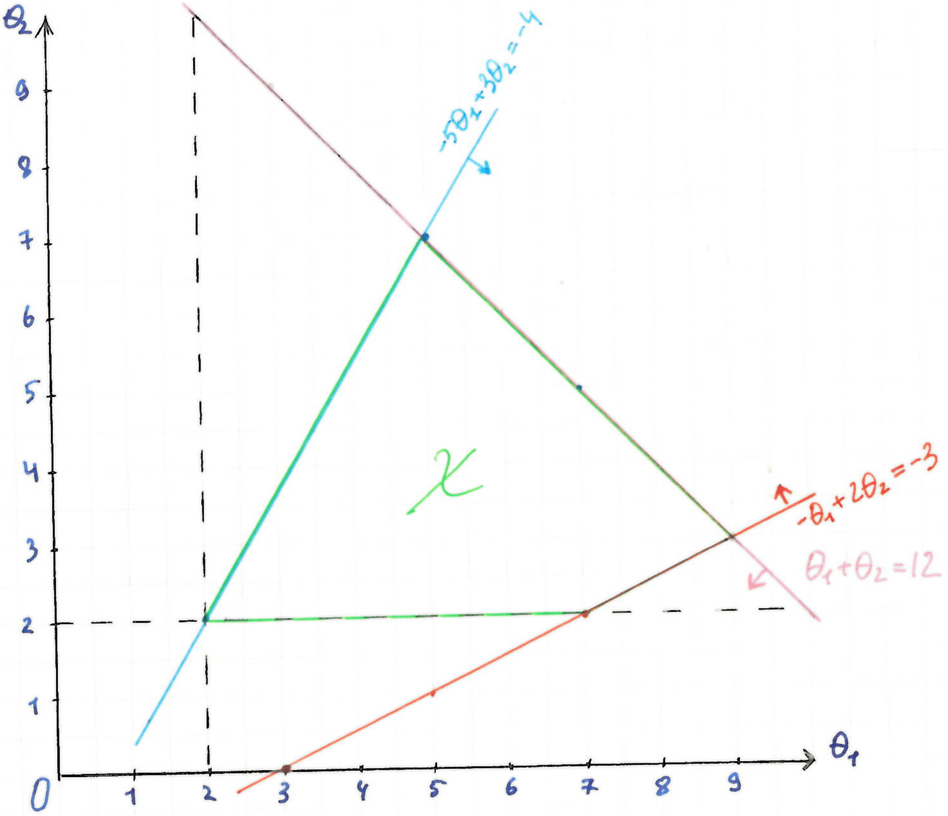
Iuliia Skobleva 03723809
Hardik Arora 03694982

Constrained Optimization.

Constrained Optimization / Projected GD

Problem 1.

$\mathcal{X} = \{\bar{\theta} \in \mathbb{R}^2 : \theta_1 + \theta_2 \leq 12, -\theta_1 + 2\theta_2 \geq -3, -5\theta_1 + 3\theta_2 \leq -4, \theta_1 \geq 2, \theta_2 \geq 2\}$



to find the minimizer and the maximizer of $f(\bar{\theta}) = 2\theta_1 - 3\theta_2$ we only need to consider the vertices of the convex set $\mathcal{X}$. the function $f(\bar{\theta})$ is linear in $\bar{\theta}$ and is therefore convex/concave on a convex set. its extrema can be found on the vertices

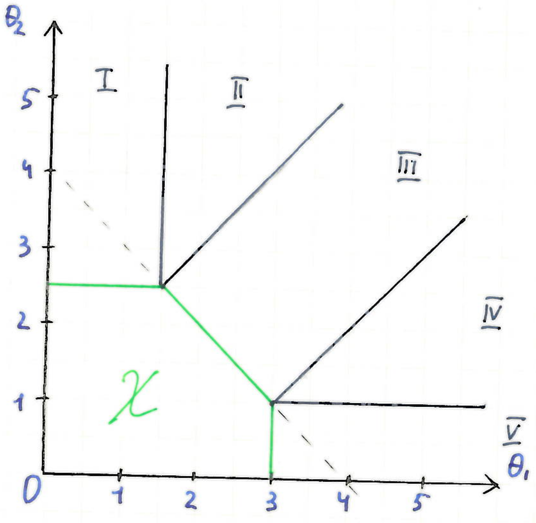
- $f(2,2) = -2$
- $f(7,2) = 8$
- $f(9,3) = 9$ (max)
- $f(5,7) = -11$ (min)

$\hookrightarrow \quad \bar{\theta}_{min} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$

$\bar{\theta}_{max} = \begin{pmatrix} 9 \\ 3 \end{pmatrix}$

Problem 2.

$\mathcal{X} = \{\bar{\theta} \in \mathbb{R}^2 : \theta_1 + \theta_2 \leq 4,\ 0 \leq \theta_1 \leq 3,\ 0 \leq \theta_2 \leq 2.5\}$



the convex set $\mathcal{X}$ can be seen on the left in green.

to find the projection $\pi_{\mathcal{X}}(\bar{p}) = \underset{\bar{\theta} \in \mathcal{X}}{\operatorname{argmin}} \|\bar{\theta} - \bar{p}\|_2^2$ we consider 5 regions, shown on the left.

we use the formula:

$$\pi_{\mathcal{X}_{a,b}}(\bar{p}) = a + \frac{(\bar{p}-\bar{a})^T(\bar{b}-\bar{a})}{\|\bar{b}-\bar{a}\|_2^2}(\bar{b}-\bar{a}),$$

where $\bar{a} \neq \bar{b}$ lie on a hyperplane $\mathcal{X}_{a,b}$

I: let $\bar{a} = \begin{pmatrix} 0 \\ 2.5 \end{pmatrix}$, $\bar{b} = \begin{pmatrix} 1 \\ 2.5 \end{pmatrix}$

$$\pi^{I}_{\mathcal{X}_{a,b}}(\bar{p}) = \begin{pmatrix} 0 \\ 2.5 \end{pmatrix} + \left(\bar{p} - \begin{pmatrix} 0 \\ 2.5 \end{pmatrix}\right)^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

II: any point from II will be projected onto $\begin{pmatrix} 1.5 \\ 2.5 \end{pmatrix}$, so:

$$\pi^{II}_{\mathcal{X}_{a,b}}(\bar{p}) = \begin{pmatrix} 1.5 \\ 2.5 \end{pmatrix}$$

III: let $\bar{a} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, $\bar{b} = \begin{pmatrix} 1.5 \\ 2.5 \end{pmatrix}$

$$\pi^{III}_{\mathcal{X}_{a,b}}(\bar{p}) = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 2 \cdot \left(\bar{p} - \begin{pmatrix} 2 \\ 2 \end{pmatrix}\right)^T \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \cdot \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix}$$

IV: $\pi^{IV}_{\mathcal{X}_{a,b}}(\bar{p}) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

V: let $\bar{a} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $\bar{b} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

$$\pi^{V}_{\mathcal{X}_{a,b}}(\bar{p}) = \begin{pmatrix} 3 \\ 0 \end{pmatrix} + \left(\bar{p} - \begin{pmatrix} 3 \\ 0 \end{pmatrix}\right)^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Next, we perform projected gradient descent starting from $\bar{\theta}^{(0)} = \begin{pmatrix} 2.5 \\ 1 \end{pmatrix}$ with $\tau = 0.05$ for the following problem:

$$\underset{\bar{\theta}}{\text{minimize}} \quad (\theta_1 - 2)^2 - (2\theta_2 - 7)^2$$
$$\text{subject to} \quad \bar{\theta} \in \mathcal{X}$$

First, $\nabla f(\theta_1, \theta_2) = \begin{pmatrix} 2\theta_1 - 4 \\ 8\theta_2 - 28 \end{pmatrix}$

Step 1.

$$\tilde{\theta}^{(1)} = \theta^{(0)} - \tau \cdot \nabla f(\theta^{(0)})$$

$$\tilde{\theta}^{(1)} = \begin{pmatrix} 2.5 \\ 1 \end{pmatrix} - 0.05 \cdot \begin{pmatrix} 1 \\ -20 \end{pmatrix} = \begin{pmatrix} 2.45 \\ 2 \end{pmatrix}$$ ... this point falls into $\underline{\text{III}}$

$$\theta^{(1)} = \pi^{\text{III}}_{\chi_{a,b}} \begin{pmatrix} 2.45 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0.45 \\ 0 \end{pmatrix}^T \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \cdot \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} =$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - 2 \cdot 0.225 \cdot \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 2.225 \\ 1.775 \end{pmatrix}$$

Step 2.

$$\tilde{\theta}^{(2)} = \theta^{(1)} - \tau \cdot \nabla f(\theta^{(1)})$$

$$\tilde{\theta}^{(2)} = \begin{pmatrix} 2.225 \\ 1.775 \end{pmatrix} - 0.05 \cdot \begin{pmatrix} 0.45 \\ -13.8 \end{pmatrix} = \begin{pmatrix} 2.2025 \\ 2.465 \end{pmatrix}$$ ... this point falls into $\underline{\text{III}}$

$$\theta^{(2)} = \pi^{\text{III}}_{\chi_{a,b}} \begin{pmatrix} 2.2025 \\ 2.465 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0.2025 \\ 0.465 \end{pmatrix}^T \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \cdot \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} =$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 0.2625 \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 1.86875 \\ 2.13125 \end{pmatrix}$$

Lagrangian / Duality

Problem 3.

minimize $\theta_1 - \sqrt{3}\theta_2$

subject to $\theta_1^2 + \theta_2^2 - 4 \leq 0$

1. $\mathcal{L}(\bar{\theta}, \alpha) = \theta_1 - \sqrt{3}\,\theta_2 + \alpha(\theta_1^2 + \theta_2^2 - 4)$
2. obtain $g(\alpha) = \min_{\theta} \mathcal{L}(\theta, \alpha)$

$$\frac{d}{d\theta_1}\mathcal{L}(\bar{\theta}, \alpha) = 1 + 2\alpha\theta_1 \overset{!}{=} 0 \quad \Rightarrow \quad \theta_1 = -\frac{1}{2\alpha}$$

$$\frac{d}{d\theta_2}\mathcal{L}(\bar{\theta}, \alpha) = -\sqrt{3} + 2\alpha\theta_2 \overset{!}{=} 0 \quad \Rightarrow \quad \theta_2 = \frac{\sqrt{3}}{2\alpha}$$

$$g(\alpha) = -\frac{1}{2\alpha} - \frac{3}{2\alpha} + \alpha\left(\frac{1}{4\alpha^2} + \frac{3}{4\alpha^2} - 4\right) = -\frac{2}{\alpha} + \frac{1}{\alpha} - 4\alpha = -\frac{1}{\alpha} - 4\alpha$$

3. $\underset{\alpha}{\text{maximize}} \; g(\alpha)$

subject to $\alpha \geq 0$

$$\frac{d}{d\alpha} g(\alpha) = \frac{1}{\alpha^2} - 4 \overset{!}{=} 0$$

$\Rightarrow \quad \alpha = \pm 1/2$

$\Rightarrow \quad \alpha = 1/2 \qquad (\text{since } \alpha \geq 0)$

4. minimum of $f_0(\theta_1, \theta_2) = \theta_1 - \sqrt{3}\theta_2$ is obtained at:

$$\theta_1^* = - \frac{1}{2 \cdot 1/2} = -1$$

$$\theta_2^* = \frac{\sqrt{3}}{2 \cdot 1/2} = \sqrt{3}$$

Problem 4

$$\underset{\bar{w}, b}{\text{minimize}} \quad \frac{1}{2}\bar{w}^T\bar{w}$$

$$\text{subject to} \quad y_i(\bar{w}^T\bar{x}_i + b) - 1 \geq 0 \quad \text{for } i = 1 \dots N$$

we show that strong duality holds for this problem. from the lechure we know, that strong duality holds if f_0 is convex and there exists a feasible $\bar{x}_i \in \mathbb{R}^d$ such that for $i = 1 \dots N$ the constraint is affine.

1. $\nabla_{\bar{w}} \frac{1}{2}\bar{w}^T\bar{w} = \frac{1}{2} \cdot 2\bar{w} = \bar{w}$

   $\Delta_w \frac{1}{2}\bar{w}^T\bar{w} = 1$

   the second derivative of $f_0(\bar{w}) = \frac{1}{2}\bar{w}^T\bar{w}$ is positive, which means that $f_0(\bar{w})$ is convex.

2. the constraint must be of form $f_i(\bar{\theta}) \leq 0$, so that we can apply the above rule (Slater's constraint qualification). in our case:

   $1 - y_i(\bar{w}^T\bar{x}_i + b) \leq 0$

   for $y_i = \{-1, 1\}$:

(1) $1 + \bar{w}^T\bar{x}_i + b \leq 0$

(2) $1 - \bar{w}^T\bar{x}_i - b \leq 0$

the constraints (1)-(2) are affine.

Therefore, the duality gap for our constrained optimization problem is zero.