

# Constrained Optimization I

---

Nipun Batra

IIT Gandhinagar

August 30, 2025

# Lagrangian and Duality

---

Nipun Batra

June 28, 2020

IIT Gandhinagar

Lectures heavily inspired by the Maths for Machine learning book

# Minimax Inequality

- Minimax inequality

$$\max_{\mathbf{y}} \min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{x}} \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$$

# Minimax Inequality

- Minimax inequality  
states:  $\max_{\mathbf{y}} \min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{x}} \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$
- We first prove For all  $\mathbf{x}, \mathbf{y}$   $\min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$

# Minimax Inequality

- Let us choose  $q(x, y) = xy$

# Minimax Inequality

- Let us choose  $q(x, y) = xy$
- Let us first find  $\max_y q(\mathbf{x}, \mathbf{y})$

# Minimax Inequality

- Let us choose  $q(x, y) = xy$
- Let us first find  $\max_y q(\mathbf{x}, \mathbf{y})$

# Minimax Inequality

- Let us choose  $q(x, y) = xy$
- Let us first find  $\max_y q(\mathbf{x}, \mathbf{y})$

$x$	1	1	2	3	4
$\Downarrow$	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			



# Minimax Inequality

- For each value of  $x$ , we find  $y$  that maximizes  $q(x, y)$

$x$	1	1	2	3	4
$\Downarrow$	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- For each value of  $x$ , we find  $y$  that maximizes  $q(x, y)$
- $y = 4$  maximizes  $q(x, y) \forall x$

$x$	1	1	2	3	4
$\Downarrow$	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- For each value of  $y$ , we find  $x$  that minimizes  $q(x, y)$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- For each value of  $y$ , we find  $x$  that minimizes  $q(x, y)$
- $x = 1$  minimizes  $q(x, y) \forall y$

$x$	1	1	2	3	4
$\Downarrow$	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y$	$\Rightarrow$		

# Minimax Inequality

- We just showed For all  $\mathbf{x}, \mathbf{y}$   $\min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- We just showed For all  $\mathbf{x}, \mathbf{y}$   $\min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leq \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$
- The equality occurs at  $x = 1, y = 4$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- Let us now find  $\max_y \min_x q(x, y)$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

# Minimax Inequality

- Similarly, let us now find  $\min_x \max_y q(x, y)$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			



# Minimax Inequality

- Similarly, let us now find  $\min_x \max_y q(\mathbf{x}, \mathbf{y})$
- We can thus see our Minimax inequality  
 $\max_y \min_x q(\mathbf{x}, \mathbf{y}) \leq \min_x \max_y q(\mathbf{x}, \mathbf{y})$

$x$ $\Downarrow$	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		$y \Rightarrow$			

## Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

## Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \mathbf{1}(g_i(\mathbf{x}))$$

## Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \mathbf{1}(g_i(\mathbf{x}))$$

where  $\mathbf{1}(z)$  is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

## Revisiting the Lagrange multipliers

Our problem is of the form

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \mathbf{1}(g_i(\mathbf{x}))$$

where  $\mathbf{1}(z)$  is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

This would give infinite penalty if constraint is not satisfied. But, this formulation is hard to solve too.

## Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ( $\lambda_i \geq 0$ ) to “approximate”  $J(\mathbf{x})$

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

## Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ( $\lambda_i \geq 0$ ) to “approximate”  $J(\mathbf{x})$

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

What is the relationship between  $\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda})$  and  $J(\mathbf{x})$  given  $\lambda_i \geq 0$ ?

## Revisiting the Lagrange multipliers

Idea: Introduce Lagrange multipliers ( $\lambda_i \geq 0$ ) to “approximate”  $J(\mathbf{x})$

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

What is the relationship between  $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$  and  $J(\mathbf{x})$  given  $\lambda_i \geq 0$ ?

When  $\lambda \geq 0$ , the Lagrangian  $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$  is a lower bound of  $J(\mathbf{x})$ .

Hence, the maximum of  $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$  with respect to  $\boldsymbol{\lambda}$  is

$$J(\mathbf{x}) = \max_{\boldsymbol{\lambda} \geq 0} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$$



## Revisiting the Lagrange multipliers

$$J(\mathbf{x}) = \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

## Revisiting the Lagrange multipliers

$$J(\mathbf{x}) = \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

But, our original problem was minimizing  $J(\mathbf{x})$ , which is equivalent to:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

## Revisiting the Lagrange multipliers

$$J(\mathbf{x}) = \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

But, our original problem was minimizing  $J(\mathbf{x})$ , which is equivalent to:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

Using the Minimax inequality, we can write:

## Revisiting the Lagrange multipliers

$$J(\mathbf{x}) = \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

But, our original problem was minimizing  $J(\mathbf{x})$ , which is equivalent to:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

Using the Minimax inequality, we can write:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda) \geq \max_{\lambda \geq 0} \min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \lambda)$$

## Revisiting the Lagrange multipliers

$$J(\mathbf{x}) = \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

But, our original problem was minimizing  $J(\mathbf{x})$ , which is equivalent to:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda)$$

Using the Minimax inequality, we can write:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \max_{\lambda \geq 0} \mathcal{L}(\mathbf{x}, \lambda) \geq \max_{\lambda \geq 0} \min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \lambda)$$

We can write the dual objective as a function of  $\lambda$  as

$$\mathcal{D}(\lambda) = \min_{\mathbf{x} \in \mathbb{R}^d} \mathcal{L}(\mathbf{x}, \lambda)$$

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

- Or, primal objective =  $J(\mathbf{x}) \geq \max_{\lambda} \mathcal{D}(\lambda)$

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

- Or, primal objective =  $J(\mathbf{x}) \geq \max_{\lambda} \mathcal{D}(\lambda)$
- Or, primal objective (in terms of  $\mathbf{x}$ )  $\geq$  dual objective (in terms of  $\lambda$ )



## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

- Or, primal objective =  $J(\mathbf{x}) \geq \max_{\lambda} \mathcal{D}(\lambda)$
- Or, primal objective (in terms of  $\mathbf{x}$ )  $\geq$  dual objective (in terms of  $\lambda$ )
- For SVM like formulations, primal objective is the same as dual objective (strong duality)

## Revisiting the Lagrange multipliers

- Primal objective:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0 \quad \text{for all } i = 1, \dots, m \end{array}$$

- Or, primal objective =  $J(\mathbf{x}) \geq \max_{\lambda} \mathcal{D}(\lambda)$
- Or, primal objective (in terms of  $\mathbf{x}$ )  $\geq$  dual objective (in terms of  $\lambda$ )
- For SVM like formulations, primal objective is the same as dual objective (strong duality)
- For some problems, there is a “duality-gap” between the two objectives