Lecture 12: Duality Examples
Saddle Point Theory

Nov 1, 2024

Happy Halloween - Part Two!

Typos c/o ChatGPT

Recall (Convex) Duality Framework

$$\begin{aligned} \text{minimize}_{x \in X} \; & f_0(x) \\ \text{subject to} \; & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ & h_j(x) = 0, \quad j = 1, \ldots, s. \end{aligned}$$

With λ_i, ν_j denoting Lagrange multipliers for g_i and $h_j(x) = 0$, respectively,

Lagrangian is:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{s} \nu_j h_j(x),$$

With $g(\lambda, \nu) := \inf_{x \in X} \mathcal{L}(x, \lambda, \nu)$, the dual problem becomes:

$$\begin{aligned} \text{maximize} \; & g(\lambda, \nu) \\ \text{subject to} \; & \lambda \geq 0. \end{aligned}$$

No sign constraints on ν!

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\min \frac{1}{2}x^\intercal P x + c^\intercal x$$
$$A_1 x = b_1$$
$$A_2 x \leq b_2$$

where $P = P^\intercal$.

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P x + c^\intercal x \\ & A_1 x = b_1 \\ & A_2 x \leq b_2 \end{aligned}$$

where $P = P^\intercal$.

Quadratically Constrained Quadratic Programs

A **Quadratically Constrainted Quadratic Program (QCQP)** is a problem:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P_0 x + c^\intercal x \\ & x^\intercal P_i x + q_i^\intercal x + b_i \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where $Q_i, i = 0, \ldots, m$ are **symmetric** matrices.

Convex if $P \succeq 0, P_i \succeq 0$. Gurobi can now handle **non-convex** QCQPs!

Two Problems to Warm Up

QP with Inequality Constraint

$$\underset{x}{\text{minimize}} \ \frac{1}{2}x^\intercal Qx + c^\intercal x$$
$$Ax \leq b$$

where $Q \succ 0$ is a **positive definite** matrix.

QCQP

$$\begin{aligned} \text{minimize} \ & \frac{1}{2}x^\intercal P_0 x + q_0^\intercal x + r_0 \\ \text{subject to} \ & \frac{1}{2}x^\intercal P_i x + q_i^\intercal x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $P_0 \succ 0$ and $P_i \succeq 0$

- **What is the Lagrangian? What is the dual? Does Slater Condition hold?**

Convex QP With Inequality Constraints

QP with Inequality Constraint

$$\underset{x}{\text{minimize}}\ \frac{1}{2}x^\intercal Qx + c^\intercal x$$
$$Ax \leq b$$

where $Q \succ 0$ is a **positive definite** matrix.

- *What is the Lagrangian? What is the dual? Does Slater Condition hold?*
- The Langragian function is:

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^\intercal Qx + c^\intercal x + \lambda^\intercal (Ax - b)$$

- To compute $g(\lambda)$, take the gradient. Infimum achieved at $x = -Q^{-1}(c + A^\intercal \lambda)$
- The dual function becomes:

$$g(\lambda) = -\frac{1}{2}\lambda^\intercal AQ^{-1}A^\intercal \lambda - \lambda^\intercal (b + AQ^{-1}c) - \frac{1}{2}c^\intercal Q^{-1}c.$$

- Assuming $Ax \leq b$ feasible, Slater condition holds
- Dual easier to solve? (maximize a concave quadratic function with constraints $\lambda \geq 0$)

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^\intercal P_0 x + q_0^\intercal x + r_0 \\ \text{subject to} \quad & \frac{1}{2}x^\intercal P_i x + q_i^\intercal x + r_i \leq 0, \quad i = 1, \dots, m, \end{aligned}$$

where $P_0 \succ 0$ and $P_i \succeq 0$

- The Lagrangian is:
$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^\intercal P(\lambda) x + q(\lambda)^\intercal x + r(\lambda),$$
where $P(\lambda) = P_0 + \sum_{i=1}^m \lambda_i P_i, \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$
- Because $\lambda \geq 0$, we have $P(\lambda) \succ 0$ and therefore:
$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{2}q(\lambda)^\intercal P(\lambda)^{-1} q(\lambda) + r(\lambda).$$
- We can express the dual problem as:
$$\max_{\lambda \geq 0} -\frac{1}{2}q(\lambda)^\intercal P(\lambda)^{-1} q(\lambda) + r(\lambda)$$
- Slater condition holds if there exists an x with
$$\frac{1}{2}x^\intercal P_i x + q_i^\intercal x + r_i < 0, \quad i = 1, \dots, m.$$

A Non-Convex QCQP

A Special Non-Convex QCQP

For $A = A^\intercal$ and $A \not\succeq 0$, consider:

$$\text{minimize } x^\intercal A x + 2b^\intercal x$$
$$x^\intercal x \le 1$$

- Lagrangian is:

$$\mathcal{L}(x, \lambda) = x^\intercal A x + 2b^\intercal x + \lambda(x^\intercal x - 1) = x^\intercal (A + \lambda I) x + 2b^\intercal x - \lambda,$$

$$g(\lambda) = \begin{cases} -b^\intercal (A + \lambda I)^\dagger b - \lambda & A + \lambda I \succeq 0, \; b \in \mathcal{R}(A + \lambda I), \\ -\infty & \text{otherwise}, \end{cases}$$

where $M^\dagger$ is the (Moore-Penrose) pseudo-inverse of M

- The dual problem is

$$\begin{aligned} \text{maximize}_{\lambda \ge 0} \quad & -b^\intercal (A + \lambda I)^\dagger b - \lambda \\ \text{subject to} \quad & A + \lambda I \succeq 0, \; b \in \mathcal{R}(A + \lambda I) \end{aligned}$$

- Readily solved because it can be expressed as

$$\text{maximize} \left\{ -\sum_{i=1}^{n} \frac{(q_i^\intercal b)^2}{\lambda_i + \lambda} - \lambda \; : \; \lambda \ge -\lambda_{\min}(A) \right\}$$

where λ_i, q_i are eigen-decomposition of A and $(q_i^\intercal b)^2/0 := 0$ if $q_i^\intercal b = 0$ and ∞ otherwise.

A Non-Convex QCQP

A Special Non-Convex QCQP

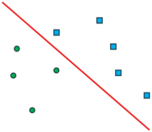
For $A = A^\intercal$ and $A \not\succeq 0$, consider:

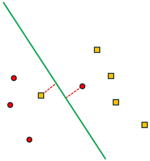
$$\begin{aligned} \text{minimize } & x^\intercal A x + 2b^\intercal x \\ & x^\intercal x \leq 1 \end{aligned}$$

- Slater condition trivially satisfied!
- We actually have **zero duality gap**, $p^\star = d^\star$!
- A more general result: strong duality for any quadratic optimization problem with two constraints $\ell \leq x^\intercal P x \leq u$ if P and A are simultaneously diagonalizable

Regularized Support Vector Machines (SVM)

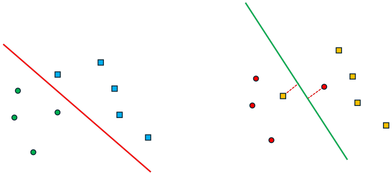
- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.





Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

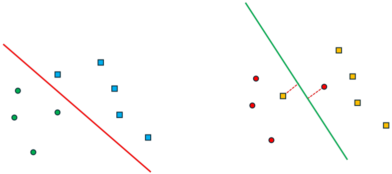


- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b = 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



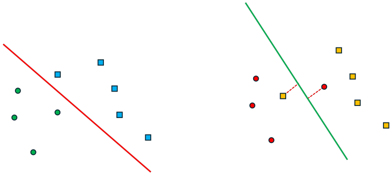
- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b = 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \dots, m.$$

- *How to solve this problem?*

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b = 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

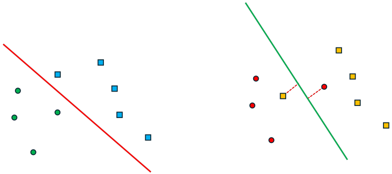
- *How to solve this problem?* This is **an LP**!

- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \text{ where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b = 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

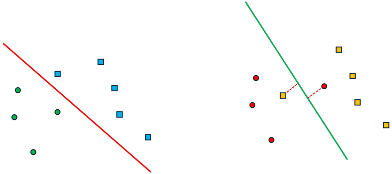
- *How to solve this problem?* This is **an LP**!

- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \text{ where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

- **Hard (MIP) problem!**

Regularized Support Vector Machines (SVM)



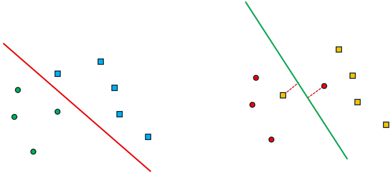
- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

  - Separable if and only if $y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m$.

  - Minimize $\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**

  - Replace $\psi(t)$ with upper bound $h(t) = (1 - t)_+ = \max(0, 1 - t)$ (**hinge function**)

Regularized Support Vector Machines (SVM)



- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

  - Separable if and only if $y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m$.

  - Minimize $\sum_{i=1}^m \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**

  - Replace $\psi(t)$ with upper bound $h(t) = (1-t)_+ = \max(0, 1-t)$ (**hinge function**)

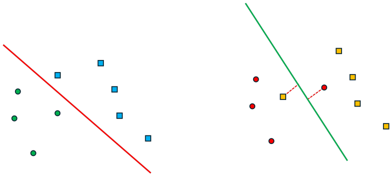
  - Solve **regularized** version:

$$\min_{w,b} \; C \cdot \sum_{i=1}^m (1 - y_i(w^\intercal x_i + b))_+ + \frac{1}{2}\|w\|_2^2,$$

  where parameter $C > 0$ controls trade-off between robustness and performance

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Solve $\min_{w,b} \; C \cdot \sum_{i=1}^m (1 - y_i(w^\intercal x_i + b))_+ + \frac{1}{2}\|w\|_2^2$
- Can be written as a QP by introducing slack variables:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^m v_i \quad : \quad v \geq 0, \; y_i(w^\intercal x_i + b) \geq 1 - v_i, \; i = 1, \ldots, m,$$

or more compactly:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C \cdot 1^\intercal v \quad : \quad v \geq 0, \; v + Z^\intercal w + by \geq 1,$$

where $Z^\intercal \in \mathbb{R}^{m \times n}$ is the matrix with rows given by $y_i \cdot x_i^\intercal$

- **What is the Lagrangian? What is the dual? Does Slater Condition hold?**

Regularized Support Vector Machines (SVM)

- Solve
$$\min_{w,b,v} \frac{1}{2}\|w\|_2^2 + C \cdot 1^\intercal v \quad : \quad v \geq 0,\ v + Z^\intercal w + by \geq 1,$$
where $Z^\intercal \in \mathbb{R}^{m \times n}$ is the matrix with rows given by $y_i \cdot x_i^\intercal$
- $\mathcal{L}(w, b, \lambda, \mu) = \frac{1}{2}\|w\|_2^2 + C \cdot v^\intercal 1 + \lambda^\intercal(1 - v - Z^\intercal w - by) - \mu^\intercal v$
- $g(\lambda, \mu) = \min_{w,b} \mathcal{L}(w, b, \lambda, \mu)$
- Taking gradients : $w(\lambda, \mu) = Z\lambda,\ C \cdot 1 = \lambda + \mu,\ \lambda^\intercal y = 0$
- We obtain
$$g(\lambda, \mu) = \begin{cases} \lambda^\intercal 1 - \frac{1}{2}\|Z\lambda\|_2^2 & \text{if } \lambda^\intercal y = 0,\ \lambda + \mu = C \cdot 1, \\ +\infty & \text{otherwise.} \end{cases}$$
- Dual problem
$$d^\star = \max_\lambda \Big\{ \lambda^\intercal 1 - \frac{1}{2}\lambda^\intercal Z^\intercal Z \lambda \quad : \quad 0 \leq \lambda \leq C \cdot 1,\ \lambda^\intercal y = 0 \Big\}.$$
- Strong duality holds, because the primal problem is a QP
- Dual objective depends only on the **kernel matrix** $K = Z^\intercal Z \in S_+^m$, and dual problem involves only m variables and $m + 1$ constraints
- Only dependence on the number of dimensions (features) n is through Z; this requires all products $x_i^\intercal x_j$, $1 \leq i \leq j \leq m$ but still more memory-efficient than solving the primal!

Saddle Point Theory

Primal Problem

$$
\begin{array}{lll} (\mathscr{P}) \text{ minimize}_x & f_0(x) & \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m & (1) \\ & x \in X. & \end{array}
$$

- There is a very insightful way to make the primal and dual look more "symmetric"
- Recall: Lagrangian $\mathcal{L}(x, \lambda)$ and dual objective $g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda)$.
- **Claim:**

$$
\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) =
$$

Saddle Point Theory

Primal Problem

$$(\mathscr{P}) \quad \text{minimize}_x \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m \qquad x \in X. \tag{1}$$

- There is a very insightful way to make the primal and dual look more "symmetric"
- Recall: Lagrangian $\mathcal{L}(x, \lambda)$ and dual objective $g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda)$.
- **Claim:**

$$\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x) & \text{if } f_i(x) \leq 0, \; i = 1, \ldots, m, \\ \infty & \text{otherwise.} \end{cases}$$

- So we can express **the optimal values of the primal and dual** as:

$$p^\star = \inf_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda) \qquad\qquad d^\star = \sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda)$$

Saddle Point Theory

Alternative Formulation of Primal and Dual Problems

We can express **the optimal values of the primal and dual** as:

$$p^\star = \inf_{x \in X} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda) \qquad d^\star = \sup_{\lambda \geq 0} \inf_{x \in X} \mathcal{L}(x, \lambda)$$

- How to restate **weak duality** and **strong duality** in terms of the problems above?
- **Weak duality**:

$$\sup_{\lambda \geq 0} \inf_{x \in X} \mathcal{L}(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda)$$

- **Strong duality**:

$$\sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda) = \inf_{x \in X} \sup_{\lambda \geq 0} \mathcal{L}(x, \lambda).$$

- Strong duality holds exactly when we can interchange the order of **min** and **max**

Min-Max and Max-Min Problems

Min-Max and Max-Min

Consider more broadly the pair of problems:

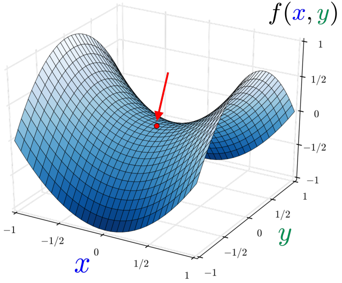
$$\sup_{y \in Y} \inf_{x \in X} f(x,y) \qquad \inf_{x \in X} \sup_{y \in Y} f(x,y)$$

- For any f, Z, W, the **max-min inequality** (i.e., "weak duality") holds:

$$\sup_{y \in Y} \inf_{x \in X} f(x,y) \leq \inf_{x \in X} \sup_{y \in Y} f(x,y)$$

- f, Z, W satisfy the **saddle-point property** if equality holds:

$$\sup_{y \in Y} \inf_{x \in X} f(x,y) = \inf_{x \in X} \sup_{y \in Y} f(x,y).$$



Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x, z)$
  - x wants to **minimize** the amount, z wants to **maximize** it

Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x, z)$
  - x wants to **minimize** the amount, z wants to **maximize** it
- Min-max inequality: the player who moves **second** has an advantage!
  - x moves first and y moves second $\Rightarrow$ larger payment
  - y moves first and x moves second $\Rightarrow$ smaller payment

Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x, z)$
  - x wants to **minimize** the amount, z wants to **maximize** it
- Min-max inequality: the player who moves **second** has an advantage!
  - x moves first and y moves second $\Rightarrow$ larger payment
  - y moves first and x moves second $\Rightarrow$ smaller payment
- Player moving second has information about first player's move and can use a **strategy**, i.e., make a choice that depends on the first player's choice

$$\text{Left problem: } \inf_{x \in X} f(x, y) \text{ for any given } y \;\Rightarrow x^\star(y)$$

$$\text{Right problem: } \sup_{y \in Y} f(x, y) \text{ for any given } x \;\Rightarrow y^\star(x)$$

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Sion-Kakutani Theorem

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be convex and compact subsets and let $f : X \times Y \to \mathbb{R}$ be a continuous function that is convex in $x \in X$ for any fixed $y \in Y$ and is concave in $y \in Y$ for any fixed $x \in X$. Then,

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Sion-Kakutani Theorem

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be convex and compact subsets and let $f : X \times Y \to \mathbb{R}$ be a continuous function that is convex in $x \in X$ for any fixed $y \in Y$ and is concave in $y \in Y$ for any fixed $x \in X$. Then,

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

Generalizations: Y only needs to be convex (not compact); $f(\cdot, y)$ must be quasi-convex on X and with closed lower level sets (for any $y \in Y$); and $f(x, \cdot)$ must be quasi-concave on Y and with closed upper level sets (for any $x \in X$)

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \le 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^\star \in X$. Then:

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P}) \ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \le 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^\star \in X$. Then:

(i) A **sufficient condition** for $x^\star$ to be optimal is the existence of $\lambda^\star \ge 0$ such that $(x^\star, \lambda^\star)$ is a saddle point of the Lagrange function $\mathcal{L}(x, \lambda)$:

$$\mathcal{L}(x, \lambda^\star) \ge \mathcal{L}(x^\star, \lambda^\star) \ge \mathcal{L}(x^\star, \lambda) \quad \forall x \in X, \ \lambda \ge 0.$$

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^\star \in X$. Then:

(i) A **sufficient condition** for $x^\star$ to be optimal is the existence of $\lambda^\star \geq 0$ such that $(x^\star, \lambda^\star)$ is a saddle point of the Lagrange function $\mathcal{L}(x, \lambda)$:

$$\mathcal{L}(x, \lambda^\star) \geq \mathcal{L}(x^\star, \lambda^\star) \geq \mathcal{L}(x^\star, \lambda) \quad \forall x \in X,\ \lambda \geq 0.$$

(ii) If $(\mathscr{P})$ is a convex optimization problem and satisfies the Slater condition, then the above condition is also **necessary** for the optimality of $x^\star$.