COMS 4771 Lecture 26

1. Lagrangian duality

Lagrangian duality

Lagrangian duality

Uses of Lagrangian duality:

- Analysis: get insight about convex optimization problems.
- Algorithms: simple algorithms for constrained convex optimization.

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n : \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n : \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \le 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

- You MUST find $\boldsymbol{x}$ satisfying $f_i(\boldsymbol{x}) \leq 0$ for each $i = 1, 2, \ldots, n$ (i.e., a *feasible* $\boldsymbol{x}$).

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x} \in \mathbb{R}^d} & f_0(\boldsymbol{x}) \\ \text{s.t.} & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{array}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

- You MUST find $\boldsymbol{x}$ satisfying $f_i(\boldsymbol{x}) \leq 0$ for each $i = 1, 2, \ldots, n$ (i.e., a *feasible* $\boldsymbol{x}$).
- Therefore, **you are playing the following game**:

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n : \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

- You MUST find $\boldsymbol{x}$ satisfying $f_i(\boldsymbol{x}) \leq 0$ for each $i = 1, 2, \ldots, n$ (i.e., a *feasible* $\boldsymbol{x}$).
- Therefore, **you are playing the following game**:
  - If you pick a feasible $\boldsymbol{x}$, then you are charged $f_0(\boldsymbol{x})$.

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

- You MUST find $\boldsymbol{x}$ satisfying $f_i(\boldsymbol{x}) \leq 0$ for each $i = 1, 2, \ldots, n$ (i.e., a *feasible* $\boldsymbol{x}$).
- Therefore, **you are playing the following game**:
  - If you pick a feasible $\boldsymbol{x}$, then you are charged $f_0(\boldsymbol{x})$.
  - If you pick an infeasible $\boldsymbol{x}$, then you are charged $f_0(\boldsymbol{x})$, plus $\lambda_i \cdot f_i(\boldsymbol{x})$ for each constraint $i \in [n]$ that you violate.

Constrained convex optimization problem

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n. \end{aligned}$$

Always think of optimization as a game against an adversary who is trying to prevent you from solving the problem.

- You MUST find $\boldsymbol{x}$ satisfying $f_i(\boldsymbol{x}) \leq 0$ for each $i = 1, 2, \ldots, n$ (i.e., a *feasible* $\boldsymbol{x}$).
- Therefore, **you are playing the following game**:
  - If you pick a feasible $\boldsymbol{x}$, then you are charged $f_0(\boldsymbol{x})$.
  - If you pick an infeasible $\boldsymbol{x}$, then you are charged $f_0(\boldsymbol{x})$, plus $\lambda_i \cdot f_i(\boldsymbol{x})$ for each constraint $i \in [n]$ that you violate.
  - The adversary gets to pick any non-negative $\lambda_i \geq 0$, and he gets to pick them **after you pick your** $\boldsymbol{x}$!!!!

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

- **Observation**: If $f_i(\boldsymbol{x}) \leq 0$, adversary might as well pick $\lambda_i = 0$.

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

- **Observation**: If $f_i(\boldsymbol{x}) \leq 0$, adversary might as well pick $\lambda_i = 0$.

Equivalent game: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \ := \ f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x}).$$

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

- **Observation**: If $f_i(\boldsymbol{x}) \leq 0$, adversary might as well pick $\lambda_i = 0$.

Equivalent game: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \ := \ f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x}).$$

- If you manage to pick a feasible $\boldsymbol{x}$, you incur cost $f_0(\boldsymbol{x})$.

  (Adversary would not want to set $\lambda_i > 0$ for any i in this case.)

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

- **Observation**: If $f_i(\boldsymbol{x}) \leq 0$, adversary might as well pick $\lambda_i = 0$.

Equivalent game: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \ := \ f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x}).$$

- If you manage to pick a feasible $\boldsymbol{x}$, you incur cost $f_0(\boldsymbol{x})$.

  (Adversary would not want to set $\lambda_i > 0$ for any i in this case.)

- If you pick an infeasible $\boldsymbol{x}$, adversary punishes you by sending $\lambda_i \to +\infty$ for the constraint $i \in [n]$ you violate.

The game against the adversary

The game you are playing: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \mathbb{1}\{f_i(\boldsymbol{x}) > 0\} \cdot \lambda_i \cdot f_i(\boldsymbol{x}).$$

- **Observation**: If $f_i(\boldsymbol{x}) \leq 0$, adversary might as well pick $\lambda_i = 0$.

Equivalent game: you pick $\boldsymbol{x} \in \mathbb{R}^d$, then adversary picks $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \geq \mathbf{0}$, then you incur cost

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;:=\; f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x}).$$

- If you manage to pick a feasible $\boldsymbol{x}$, you incur cost $f_0(\boldsymbol{x})$.

  (Adversary would not want to set $\lambda_i > 0$ for any i in this case.)

- If you pick an infeasible $\boldsymbol{x}$, adversary punishes you by sending $\lambda_i \to +\infty$ for the constraint $i \in [n]$ you violate.

  In this case, you incur $+\infty$ cost.

The best you can do against the adversary

Best cost for you:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\lambda_1,\lambda_2,\ldots,\lambda_n\geq 0} \underbrace{f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x})}_{\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})}.$$

The best you can do against the adversary

Best cost for you:

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\lambda_1,\lambda_2,\ldots,\lambda_n\geq 0} \underbrace{f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x})}_{\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})}.$$

$\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})$ is called the **Lagrangian function**.

The best you can do against the adversary

Best cost for you:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0} \underbrace{f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x})}_{\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})}.$$

$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is called the **Lagrangian function**.

Summary of game:

- You pick $\boldsymbol{x} \in \mathbb{R}^d$.
- Adversary picks $\boldsymbol{\lambda} \geq \mathbf{0}$.
- You incur cost $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$.

The best you can do against the adversary

Best cost for you:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0} \underbrace{f_0(\boldsymbol{x}) + \sum_{i=1}^{n} \lambda_i \cdot f_i(\boldsymbol{x})}_{\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})}.$$

$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is called the **Lagrangian function**.

Summary of game:

- You pick $\boldsymbol{x} \in \mathbb{R}^d$.
- Adversary picks $\boldsymbol{\lambda} \geq \mathbf{0}$.
- You incur cost $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$.

Only way for you not to incur $+\infty$ cost is to pick a feasible $\boldsymbol{x} \in \mathbb{R}^d$.

Weak duality

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

Weak duality

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

- **LHS**: best you can do in game where you must pick $\boldsymbol{x}$ first, and then adversary gets to pick $\boldsymbol{\lambda}$ after seeing what you picked.

Weak duality

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

- **LHS**: best you can do in game where you must pick $\boldsymbol{x}$ first, and then adversary gets to pick $\boldsymbol{\lambda}$ after seeing what you picked.
- **RHS**: worst you can do in game where adversary must pick $\boldsymbol{\lambda}$ first, and then you get to pick $\boldsymbol{x}$ after seeing what adversary picked.

Weak duality

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

- **LHS**: best you can do in game where you must pick $\boldsymbol{x}$ first, and then adversary gets to pick $\boldsymbol{\lambda}$ after seeing what you picked.
- **RHS**: worst you can do in game where adversary must pick $\boldsymbol{\lambda}$ first, and then you get to pick $\boldsymbol{x}$ after seeing what adversary picked.
- RHS is better for you in terms of guaranteeing smaller cost.

Weak duality

lower bound, the right part always convex, have solution
as the lower bound for the original problem

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

- **LHS**: best you can do in game where you must pick $\boldsymbol{x}$ first, and then adversary gets to pick $\boldsymbol{\lambda}$ after seeing what you picked.
- **RHS**: worst you can do in game where adversary must pick $\boldsymbol{\lambda}$ first, and then you get to pick $\boldsymbol{x}$ after seeing what adversary picked.
- RHS is better for you in terms of guaranteeing smaller cost.

In fact, for all $\tilde{\boldsymbol{x}}\in\mathbb{R}^d$ and all $\tilde{\boldsymbol{\lambda}}\geq\mathbf{0}$,

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

STRONG DUALITY

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Doesn't matter if you go first or if adversary goes first.

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Doesn't matter if you go first or if adversary goes first.

- **LHS**: the primal (i.e., original) problem.

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Doesn't matter if you go first or if adversary goes first.

- **LHS**: the primal (i.e., original) problem.
- **RHS**:

  $$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} g(\boldsymbol{\lambda})$$

  where

  $$g(\boldsymbol{\lambda}) := \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}).$$

  This maximization problem is called the **Lagrange dual problem**.

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Doesn't matter if you go first or if adversary goes first.

- **LHS**: the primal (i.e., original) problem.
- **RHS**:

  $$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} g(\boldsymbol{\lambda})$$

  where

  $$g(\boldsymbol{\lambda}) := \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}).$$

  This maximization problem is called the **Lagrange dual problem**.

  - $\boldsymbol{\lambda} \geq \mathbf{0}$: **Lagrange multipliers** (dual variables).

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Doesn't matter if you go first or if adversary goes first.

- **LHS**: the primal (i.e., original) problem.
- **RHS**:

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} g(\boldsymbol{\lambda})$$

where

$$g(\boldsymbol{\lambda}) := \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}).$$

This maximization problem is called the **Lagrange dual problem**.

  - $\boldsymbol{\lambda} \geq \mathbf{0}$: **Lagrange multipliers** (dual variables).
  - $g \colon \mathbb{R}^n \to \mathbb{R}$: **Lagrange dual function**, which is always **concave**.

L(x, \lambda) is convave

min {concave functions} is also a concave function

Equality constraints: shortcut

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{k \times d}$, and $\boldsymbol{b} \in \mathbb{R}^k$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n; \\ & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Equality constraints: shortcut

For convex functions $f_0, f_1, f_2, \ldots, f_n \colon \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{k \times d}$, and $\boldsymbol{b} \in \mathbb{R}^k$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \ldots, n; \\ & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Equality constraint $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}$ is same as conjunction of two inequality constraints:

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}.$$

Equality constraints: shortcut

For convex functions $f_0, f_1, f_2, \dots, f_n \colon \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{k \times d}$, and $\boldsymbol{b} \in \mathbb{R}^k$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, 2, \dots, n; \\ & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Equality constraint $\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}$ is same as conjunction of two inequality constraints:

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{0}.$$

Shortcut: Lagrange multipliers associated with equality constraints do not have to be non-negative. (Easy to derive by reasoning about "game against adversary".)

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

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

Strong duality

Under some conditions (e.g., some $\boldsymbol{x} \in \mathbb{R}^d$ strictly satisfies all constraints),

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\boldsymbol{\lambda}\geq\mathbf{0},\boldsymbol{\nu}\in\mathbb{R}^k} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}) \;=\; \max_{\boldsymbol{\lambda}\geq\mathbf{0},\boldsymbol{\nu}\in\mathbb{R}^k} \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}).$$

- **LHS**: the primal (i.e., original) problem.
- **RHS**:

  $$\max_{\boldsymbol{\lambda}\geq\mathbf{0},\boldsymbol{\nu}\in\mathbb{R}^k} g(\boldsymbol{\lambda},\boldsymbol{\nu})$$

  where

  $$g(\boldsymbol{\lambda},\boldsymbol{\nu}) \;:=\; \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu}).$$

Maximum entropy

Example: maximum entropy

The maximum entropy optimization problem (over domain $[d]$):

$$
\begin{aligned}
\min_{\boldsymbol{w}\in\mathbb{R}^d} \quad & \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} \\
\text{s.t.} \quad & \mu_j - \sum_{i=1}^{d} t_{i,j} w_i = 0 \quad j = 1, 2, \ldots, k; \\
& 1 - \sum_{i=1}^{d} w_i = 0.
\end{aligned}
$$

(We drop the constraint $\boldsymbol{w} \geq \boldsymbol{0}$, since it'll be satisfied w/o explicitly imposing it.)

Example: maximum entropy

The maximum entropy optimization problem (over domain $[d]$):

$$\begin{aligned} \min_{\boldsymbol{w} \in \mathbb{R}^d} \quad & \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} \\ \text{s.t.} \quad & \mu_j - \sum_{i=1}^{d} t_{i,j} w_i = 0 \quad j = 1, 2, \dots, k; \\ & 1 - \sum_{i=1}^{d} w_i = 0. \end{aligned}$$

(We drop the constraint $\boldsymbol{w} \geq \mathbf{0}$, since it'll be satisfied w/o explicitly imposing it.)

Lagrange multipliers: $\boldsymbol{\eta} \in \mathbb{R}^k$, $\nu \in \mathbb{R}$.

Example: maximum entropy

The maximum entropy optimization problem (over domain $[d]$):

$$\begin{aligned} \min_{\boldsymbol{w} \in \mathbb{R}^d} \quad & \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} \\ \text{s.t.} \quad & \mu_j - \sum_{i=1}^{d} t_{i,j} w_i = 0 \quad j = 1, 2, \dots, k; \\ & 1 - \sum_{i=1}^{d} w_i = 0. \end{aligned}$$

(We drop the constraint $\boldsymbol{w} \geq \mathbf{0}$, since it'll be satisfied w/o explicitly imposing it.)

Lagrange multipliers: $\boldsymbol{\eta} \in \mathbb{R}^k$, $\nu \in \mathbb{R}$.

Lagrangian:

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\eta}, \nu) \;=\; \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} + \sum_{j=1}^{k} \eta_j \left(\mu_j - \sum_{i=1}^{d} t_{i,j} w_i \right) + \nu \left(1 - \sum_{i=1}^{d} w_i \right).$$

Example: maximum entropy

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\eta}, \nu) \;=\; \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} + \sum_{j=1}^{k} \eta_j \left(\mu_j - \sum_{i=1}^{d} t_{i,j} w_i \right) + \nu \left(1 - \sum_{i=1}^{d} w_i \right).$$

We want to minimize this w.r.t. $\boldsymbol{w}$.

Example: maximum entropy

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\eta}, \nu) \;=\; \sum_{i=1}^{d} w_i \ln \frac{w_i}{\pi_i} + \sum_{j=1}^{k} \eta_j \left(\mu_j - \sum_{i=1}^{d} t_{i,j} w_i \right) + \nu \left(1 - \sum_{i=1}^{d} w_i \right).$$

We want to minimize this w.r.t. $\boldsymbol{w}$.

Partial derivative w.r.t. w_i is

$$1 + \ln(w_i) - \ln(\pi_i) - \sum_{j=1}^{k} \eta_j t_{i,j} - \nu,$$

which is zero when

$$w_i = \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} - (1 - \nu) \right).$$

(This is always non-negative!)

Example: maximum entropy

Plugging-in the choice of w_i for each $i = 1, 2, \ldots, n$ that minimizes $\mathcal{L}(\boldsymbol{w}, \boldsymbol{\eta}, \boldsymbol{\nu})$ yields the dual problem

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^k, \nu \in \mathbb{R}} \qquad g(\boldsymbol{\eta}, \nu),$$

where

$$g(\boldsymbol{\eta}, \nu) \;=\; -\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} - (1 - \nu) \right) + \sum_{j=1}^{k} \eta_j \mu_j + \nu.$$

Example: maximum entropy

Maximum entropy dual optimization problem:

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^k, \nu \in \mathbb{R}} \qquad -\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} - (1-\nu) \right) + \sum_{j=1}^{k} \eta_j \mu_j + \nu.$$

Example: maximum entropy

Maximum entropy dual optimization problem:

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^k, \nu \in \mathbb{R}} \qquad -\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} - (1 - \nu) \right) + \sum_{j=1}^{k} \eta_j \mu_j + \nu.$$

This is concave in $\boldsymbol{\eta}$ and ν.

Example: maximum entropy

Maximum entropy dual optimization problem:

$$\max_{\boldsymbol{\eta}\in\mathbb{R}^k,\nu\in\mathbb{R}} \qquad -\sum_{i=1}^{d}\pi_i \exp\left(\sum_{j=1}^{k}\eta_j t_{i,j} - (1-\nu)\right) + \sum_{j=1}^{k}\eta_j\mu_j + \nu.$$

This is concave in $\boldsymbol{\eta}$ and ν.

Derivative w.r.t. ν is

$$-\sum_{i=1}^{d}\pi_i \exp\left(\sum_{j=1}^{k}\eta_j t_{i,j} - (1-\nu)\right) + 1,$$

which is zero when

$$1-\nu = \ln\left\{\sum_{i=1}^{d}\pi_i \exp\left(\sum_{j=1}^{k}\eta_j t_{i,j}\right)\right\}.$$

Example: maximum entropy

Plugging-in maximizing choice of ν leaves us with

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^k} \qquad -\ln\left\{ \sum_{i=1}^{d} \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} \right) \right\} + \sum_{j=1}^{k} \eta_j \mu_j .$$

Example: maximum entropy

Plugging-in maximizing choice of ν leaves us with

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^k} \qquad -\ln \left\{ \sum_{i=1}^{d} \pi_i \exp\left(\sum_{j=1}^{k} \eta_j t_{i,j} \right) \right\} + \sum_{j=1}^{k} \eta_j \mu_j .$$

Derivative w.r.t. η_j is

$$-\frac{\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j'=1}^{k} \eta_{j'} t_{i,j'}\right) t_{i,j}}{\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j'=1}^{k} \eta_{j'} t_{i,j'}\right)} + \mu_j ,$$

Te

which is zero when

$$\sum_{i=1}^{d} \hat{w}_i t_{i,j} \;=\; \mu_j$$

where

$$\hat{w}_i \;=\; \frac{\pi_i \exp\left(\sum_{j'=1}^{k} \eta_{j'} t_{i,j'}\right)}{\sum_{i=1}^{d} \pi_i \exp\left(\sum_{j'=1}^{k} \eta_{j'} t_{i,j'}\right)} \quad i = 1, 2, \ldots, n.$$

Lagrangian relaxation method

Lagrangian relaxation method

Equality constrained convex optimization problem for $\boldsymbol{A} \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^{k}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Lagrangian relaxation method

Equality constrained convex optimization problem for $\boldsymbol{A} \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Lagrangian: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;=\; f_0(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \rangle$.

Lagrangian relaxation method

Equality constrained convex optimization problem for $\boldsymbol{A} \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^d} \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0}. \end{aligned}$$

Lagrangian: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \; = \; f_0(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \rangle$.

Lagrangian relaxation algorithm
Initialize $\boldsymbol{\lambda}^{(1)} \in \mathbb{R}^k$.
For $t = 1, 2, \ldots$:

- Let $\boldsymbol{x}^{(t)}$ be minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)})$ w.r.t. $\boldsymbol{x}$ (e.g., via gradient descent).
- If $\boldsymbol{A}\boldsymbol{x}^{(t)} = \boldsymbol{b}$, then done; return current $\boldsymbol{x}^{(t)}$.
- If $\boldsymbol{A}\boldsymbol{x}^{(t)} \neq \boldsymbol{b}$, then update $\boldsymbol{\lambda}^{(t)}$ (via gradient *ascent* step of size $\eta_t > 0$):

$$\boldsymbol{\lambda}^{(t+1)} \; := \; \boldsymbol{\lambda}^{(t)} + \eta_t \Big(\boldsymbol{A}\boldsymbol{x}^{(t)} - \boldsymbol{b} \Big).$$

graident of \lambada

Since $\boldsymbol{x}^{(t)}$ is the minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)})$,

$$\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)}) \;=\; \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)}) \;\leq\; \max_{\boldsymbol{\lambda}\in\mathbb{R}^k} \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;\leq\; \min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\boldsymbol{\lambda}\in\mathbb{R}^k} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

where last step uses weak duality.

Since $\boldsymbol{x}^{(t)}$ is the minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)})$,

$$\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)}) \;=\; \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)}) \;\leq\; \max_{\boldsymbol{\lambda} \in \mathbb{R}^k} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;\leq\; \min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \in \mathbb{R}^k} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

where last step uses weak duality.

- $\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)})$ is always a lower-bound on optimal value.

Since $\boldsymbol{x}^{(t)}$ is the minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)})$,

$$\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)}) \;=\; \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)}) \;\leq\; \max_{\boldsymbol{\lambda}\in\mathbb{R}^k} \min_{\boldsymbol{x}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;\leq\; \min_{\boldsymbol{x}\in\mathbb{R}^d} \max_{\boldsymbol{\lambda}\in\mathbb{R}^k} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

where last step uses weak duality.

- $\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)})$ is always a lower-bound on optimal value.
- If $\boldsymbol{A}\boldsymbol{x}^{(t)} = \boldsymbol{b}$, then $\boldsymbol{x}^{(t)}$ achieves optimal value.

Lagrangian relaxation method

Since $\boldsymbol{x}^{(t)}$ is the minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)})$,

$$\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)}) \;=\; \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^{(t)}) \;\leq\; \max_{\boldsymbol{\lambda} \in \mathbb{R}^k} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \;\leq\; \min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \in \mathbb{R}^k} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

where last step uses weak duality.

- $\mathcal{L}(\boldsymbol{x}^{(t)}, \boldsymbol{\lambda}^{(t)})$ is always a lower-bound on optimal value.
- If $\boldsymbol{A}\boldsymbol{x}^{(t)} = \boldsymbol{b}$, then $\boldsymbol{x}^{(t)}$ achieves optimal value.

Can also apply to inequality constrained problems via slack variables.

Recap

- **Lagrangian duality**: every convex optimization problem has an associated dual problem.
- **Lagrange dual problem**: maximization of concave objective function subject to convex constraints.

  Actually, we can apply this approach to **non-convex** optimization problems: the dual problem **always** has a concave objective function and convex constraints!
- In MaxEnt, dual problem reveals form of solution.
- **Lagrangian relaxation method**: exploit Lagrangian duality to get a simple algorithm for constrained optimization.