# Lecture 12: Applications of LP Duality

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

October 21, 2021

# Overview

- Game Theory - Minimax Theorems
- Learning Theory - Boosting
- Combinatorics - Bipartite Matching
- Conclusion
- Acknowledgements

# Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$

# Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively

# Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
  - If Alice plays $i$ and Bob plays $j$, then
  - Alice gets $A_{ij}$
  - Bob gets $B_{ij}$

# Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \dots, m\}$ and $S_B = \{1, \dots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
  - If Alice plays $i$ and Bob plays $j$, then
  - Alice gets $A_{ij}$
  - Bob gets $B_{ij}$
- Example: *battle of the sexes* game

# Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
  - If Alice plays $i$ and Bob plays $j$, then
  - Alice gets $A_{ij}$
  - Bob gets $B_{ij}$
- Example: *battle of the sexes* game

$$\begin{array}{c|cc} & F & O \\ \hline F & 1 & 0 \\ O & 0 & 2 \end{array} = B$$

Bob

$$A = \begin{array}{|c|c|} \hline 2 & 0 \\ \hline 0 & 1 \\ \hline \end{array}$$

Alice

| | Football | Opera |
|---|---|---|
| Football | (2,1) | (0,0) |
| Opera | (0,0) | (1,2) |

Table: Battle of the sexes payoff matrices

$m = n = 2$

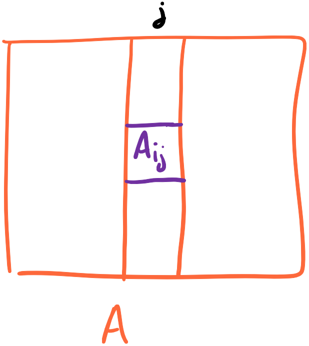
# Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

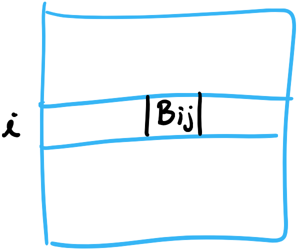
**Definition (Nash Equilibrium)**

A strategy profile $(i, j)$ is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$



Aij largest entry in the column



Bij largest in ith row of B

# Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

## Definition (Nash Equilibrium)

A strategy profile $(i, j)$ is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$

| | Football | Opera |
|---|---|---|
| Football | (2,1) | (0,0) |
| Opera | (0,0) | (1,2) |

Nash equilibria

Table: Battle of the sexes payoff matrices

# Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

## Definition (Nash Equilibrium)

A strategy profile $(i, j)$ is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$

| | Football | Opera |
|---|---|---|
| Football | (2,1) | (0,0) |
| Opera | (0,0) | (1,2) |

Table: Battle of the sexes payoff matrices

B

| | Silent | Snitch |
|---|---|---|
| Silent | (-1,-1) | (-10,0) |
| Snitch | (0,-10) | (-5,-5) |

A

Nash eq.

Table: Prisoner's dilemma

# Mixed Strategies

**Definition (Mixed Strategy)**

A mixed strategy is a probability distribution over a set of pure strategies $S$. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \underbrace{\{x \in \mathbb{R}^n}_{\text{encodes probability distribution}} \mid \underline{x \geq 0} \text{ and } \underline{\|x\|_1 = 1}\}$$

$$x_i = \Pr[A \text{ plays } i]$$

# Mixed Strategies

**Definition (Mixed Strategy)**

A mixed strategy is a probability distribution over a set of pure strategies $S$. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"

# Mixed Strategies

**Definition (Mixed Strategy)**

A mixed strategy is a probability distribution over a set of pure strategies $S$. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"
- Payoffs for each player defined as *expected gain*. That is, $(x, y)$ is the profile of mixed strategies used by Alice and Bob, we have:

# Mixed Strategies

**Definition (Mixed Strategy)**

A mixed strategy is a probability distribution over a set of pure strategies $S$. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"
- Payoffs for each player defined as *expected gain*. That is, $(x, y)$ is the profile of mixed strategies used by Alice and Bob, we have:

Pr[A plays i and B plays j]

$= x_i y_j$

$$v_A(x, y) = \sum_{(i,j) \in S_A \times S_B} A_{ij} \underbrace{x_i y_j}_{(i,j)} = \boxed{x^T A y}$$

$$v_B(x, y) = \sum_{(i,j) \in S_A \times S_B} B_{ij} x_i y_j = \boxed{x^T B y}$$

# Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

### Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$ ← Alice no incentive to deviate
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$ ← Bob " " " "

Conditioned on the fact that other player is playing x or y respectively

# Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

### Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

Bob

Alice

|  | Jump left | Jump right |
|---|---|---|
| kick left | (-1,1) | (1,-1) |
| kick right | (1,-1) | (-1,1) |

Table: Penalty Kick

No pure Nash equilibrium!

# Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

## Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

| | Jump left | Jump right |
|---|---|---|
| kick left | (-1,1) | (1,-1) |
| kick right | (1,-1) | (-1,1) |

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$ $(A+B=0)$

# Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

**Definition ((Mixed) Nash Equilibrium)**

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

| | Jump left | Jump right |
|---|---|---|
| kick left | (-1,1) | (1,-1) |
| kick right | (1,-1) | (-1,1) |

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$
- No pure Nash Equilibrium!

# Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

**Definition ((Mixed) Nash Equilibrium)**

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T A y \geq z^T A y$ for all $z \in \Delta_A$
2. $x^T B y \geq x^T B w$ for all $w \in \Delta_B$

| | Jump left | Jump right |
|---|---|---|
| kick left | (-1,1) | (1,-1) |
| kick right | (1,-1) | (-1,1) |

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$
- No pure Nash Equilibrium!
- One mixed Nash equilibrium: $x = y = (1/2, 1/2)$

# Von Neumann's Minimax Theorem

**Theorem**

*In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:*

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

Alice tries maximize gain A

Bob tries to minimize his loss A (= -B)

expected gain for Alice is $\boxed{x^T A y}$

# Von Neumann's Minimax Theorem

**Theorem**

*In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:*

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

For given $x \in \Delta_A$: fixed strategy for Alice

$$\underbrace{\min_{y \in \Delta_B} x^T A y}_{\text{Bob's best mixed strategy}} \overset{\geq}{\underset{\leq}{=}} \underbrace{\min_{j \in S_B} (x^T A)_j}_{\text{Bob's best pure strategy}}$$

$$x^T A y = \sum_{j=1}^{n} \underbrace{y_j}_{\geq 0} \cdot \underbrace{(x^T A)_j}_{\geq \min_{j \in S_B} (x^T A)_j} \geq \min_{j \in S_B} (x^T A)_j \cdot \underbrace{\sum_{j=1}^{n} y_j}_{= 1}$$

# Von Neumann's Minimax Theorem

**Theorem**

*In a zero-sum game, for any payoff matrix* $A \in \mathbb{R}^{m \times n}$*:*

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

For given $x \in \Delta_A$:

$$\min_{y \in \Delta_B} x^T A y = \min_{j \in S_B} (x^T A)_j$$

$\therefore$ Alice need only think about Bob's pure strategies

*Left hand side* can be written as

think of $s$ as $\min x^T A y$

$$\begin{array}{ll} \max & s \\ \text{s.t.} & s \leq (x^T A)_j \quad \text{for } j \in S_B \Big\} \\ & \left.\begin{array}{l} \sum_{i \in S_A} x_i = 1 \\ x \geq 0 \end{array}\right\} \end{array}$$

once Alice plays $x$ Bob will play his best response.

prob. dist. $\iff x \in \Delta_A$

# Von Neumann's Minimax Theorem

**Theorem**

*In a* *zero-sum game*, *for any payoff matrix* $A \in \mathbb{R}^{m\times n}$:

$$\max_{x\in\Delta_A}\ \min_{y\in\Delta_B}\ x^TAy = \boxed{\min_{y\in\Delta_B}\ \max_{x\in\Delta_A}\ x^TAy}$$

Bob plays first

For given $x \in \Delta_A$:

$$\min_{y\in\Delta_B} x^TAy = \min_{j\in S_B}(x^TA)_j$$

For given $y \in \Delta_B$:

$$\max_{x\in\Delta_A} x^TAy = \max_{i\in S_A}(Ay)_i$$

*Left hand side* can be written as

$$\begin{aligned} \max \quad & s \\ \text{s.t.} \quad & s \le (x^TA)_j \quad \text{for } j \in S_B \\ & \sum_{i\in S_A} x_i = 1 \\ & x \ge 0 \end{aligned}$$

# Von Neumann's Minimax Theorem

**Theorem**

*In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m\times n}$:*

$$\boxed{\max_{x\in\Delta_A}\min_{y\in\Delta_B} x^T A y} = \boxed{\min_{y\in\Delta_B}\max_{x\in\Delta_A} x^T A y}$$

For given $x \in \Delta_A$:

$$\min_{y\in\Delta_B} x^T A y = \min_{j\in S_B} (x^T A)_j$$

*Left hand side* can be written as

$$\begin{aligned} \max \quad & s \\ \text{s.t.} \quad & s \le (x^T A)_j \quad \text{for } j \in S_B \\ & \sum_{i\in S_A} x_i = 1 \\ & x \ge 0 \end{aligned}$$

For given $y \in \Delta_B$:

$$\max_{x\in\Delta_A} x^T A y = \max_{i\in S_A} (Ay)_i$$

*Right hand side* can be written as

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t \ge (Ay)_i \quad \text{for } i \in S_A \\ & \sum_{j\in S_B} y_j = 1 \\ & y \ge 0 \end{aligned}$$

# Von Neumann's Minimax Theorem

**Theorem**

*In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:*

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

These programs are dual to each other!

*Left hand side* can be written as

$$\begin{array}{ll} \max & s \\ \text{s.t.} & s \le (x^T A)_j \quad \text{for } j \in S_B \\ & \sum_{i \in S_A} x_i = 1 \\ & x \ge 0 \end{array}$$

*Right hand side* can be written as

$$\begin{array}{ll} \min & t \\ \text{s.t.} & t \ge (Ay)_i \quad \text{for } i \in S_A \\ & \sum_{j \in S_B} y_j = 1 \\ & y \ge 0 \end{array}$$

# Proof of Duality

$(e_1 - e_2)^T \cdot z$

$$\min \quad t_+ - t_-$$

$$\text{s.t.} \quad -(Ay)_i + \overset{(t_+ - t_-)}{t} - w_i = 0$$

$$\sum y_j = 1$$

$$t_+, t_-, y \geq 0 \quad w \geq 0$$

$$z = \begin{pmatrix} t_+ \\ t_- \\ y \\ w \end{pmatrix}$$

$$B = \left(\begin{array}{cc|c|c} 0 & 0 & 1 \; 1 \cdots 1 & 0 \; 0 \cdots 0 \\ \hline +1 & -1 & & \\ +1 & -1 & -A & -I \\ \vdots & \vdots & & \\ +1 & -1 & & \end{array}\right) z = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = b_1$$

# Proof of Duality

dual program:

$$u = \begin{pmatrix} s \\ \hline \vec{v} \end{pmatrix} \qquad \vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

$$\max \quad u^T f_1$$

$$\text{s.t.} \quad u^T B \leq \underbrace{(e_1 - e_2)^T}_{(1,-1,0,0,\dots,0)}$$

$$\Longleftrightarrow \quad \max \quad s$$

$$\text{s.t.} \quad \left.\begin{aligned} -\sum v_i &\leq 1 \\ \sum v_i &\leq -1 \\ s + (v^T A)_j &\leq 0 \\ -v^T I &\leq 0 \end{aligned}\right\} \quad \begin{aligned} &\sum v_i = 1 \\ &s \leq (v^T A)_j \\ &v_i \geq 0 \end{aligned}$$

- Game Theory - Minimax Theorems
- Learning Theory - Boosting
- Combinatorics - Bipartite Matching
- Conclusion
- Acknowledgements

# Learning Theory

Consider classification problem over $\mathcal{X}$:

$\uparrow$ set of inputs

$\mathcal{X} = \{x_1, x_2, \dots, x_m\}$

# Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$

# Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0, 1\}$

# Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0, 1\}$
- Data is sampled from unknown distribution $q \in \Delta_{\mathcal{X}}$

# Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0, 1\}$
- Data is sampled from unknown distribution $q \in \Delta_{\mathcal{X}}$
- *Weak learning assumption*:

  For any distribution $q \in \Delta_{\mathcal{X}}$, there is a hypothesis $h \in \mathcal{H}$ which is wrong less than half the time.

$$\exists \gamma > 0,\ \forall \underline{q \in \Delta_{\mathcal{X}}},\ \underline{\exists h \in \mathcal{H}},\quad \Pr_{x \sim q}[h(x) \neq c(x)] \leq \underbrace{\frac{1-\gamma}{2}}_{\gamma\text{ - away from } 1/2}$$

# Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0,1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0,1\}$
- Data is sampled from unknown distribution $q \in \Delta_{\mathcal{X}}$
- *Weak learning assumption*:

  For any distribution $q \in \Delta_{\mathcal{X}}$, there is a hypothesis $h \in \mathcal{H}$ which is wrong less than half the time.

  $$\exists \gamma > 0,\ \forall q \in \Delta_{\mathcal{X}},\ \exists h \in \mathcal{H}, \quad \Pr_{x \sim q}[h(x) \neq c(x)] \leq \frac{1-\gamma}{2}$$

- Surprisingly, weak learning assumption implies something much stronger: it is possible to *combine* classifiers in $\mathcal{H}$ to construct a *classifier* that is *always right* (known as *strong learning*).

# Boosting

**Theorem**

*Let $\mathcal{H}$ be a set of hypotheses satisfying weak learning assumption. Then there is distribution $\underline{p \in \Delta}_{\mathcal{H}}$ such that the weighed majority classifier*

$$c_p(x) := \begin{cases} 1, & \textit{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \textit{otherwise} \end{cases}$$

*is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$*

# Boosting

**Theorem**

*Let $\mathcal{H}$ be a set of hypotheses satisfying weak learning assumption. Then there is distribution $p \in \Delta_{\mathcal{H}}$ such that the weighed majority classifier*

$$c_p(x) := \begin{cases} 1, & \text{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

*is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$*

- Let $M \in \{-1, 1\}^{m \times n}$, where $\underline{m = |\mathcal{X}|}$ and $\underline{n = |\mathcal{H}|}$.

$$M_{ij} = \begin{cases} +1, & \text{if classifier } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

zero-sum game $\mathcal{X}$ vs $\mathcal{H}$

# Boosting

**Theorem**

*Let $\mathcal{H}$ be a set of hypotheses satisfying weak learning assumption. Then there is distribution $p \in \Delta_{\mathcal{H}}$ such that the weighed majority classifier*

$$c_p(x) := \begin{cases} 1, & \textit{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \textit{otherwise} \end{cases}$$

*is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$*

- Let $M \in \{-1, 1\}^{m \times n}$, where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if classifier } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

- Weak learning:

$$\sum_{1 \leq i \leq n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

# Boosting - Proof

Let $M \in \{-1,1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\underbrace{\sum_{1 \le i \le n} q_i \cdot \delta_{h_j(x_i) \neq c(x_i)}}_{\mathbb{E}[h_j \text{ being wrong}]} \le \frac{1-\gamma}{2}$$

$\exists \gamma > 0$ s.t. $\forall q \in \Delta_{\mathcal{X}}$ $\exists h \in \mathcal{H}$ $(h_j)$ s.t.

$h_j$ is wrong $\boxed{< \frac{1-\gamma}{2} \text{ of the time}}$

$$\delta_{ij} := \delta_{h_j(x_i) \neq c(x_i)} = \begin{cases} 1 & \text{if } h_j(x_i) \neq c(x_i) \quad (\text{mistake}) \\ 0 & \text{o.w.} \quad (\text{correct}) \end{cases}$$

predicted by $h_j$

correct value

$$\mathbb{E}[h_j \text{ being wrong}] = \sum_i \underbrace{\Pr[\text{pick } x_i]}_{q_i} \delta_{ij}$$

# Boosting - Proof

Let $M \in \{-1,1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \le i \le n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \le \frac{1-\gamma}{2}$$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

min-max argument

payoff when $\mathcal{X}$ plays q, $\mathcal{H}$ plays $h_j$ →

$$\boxed{q^T M e_j \le -\gamma} \Rightarrow \boxed{q^T M p \le -\gamma}$$

payoff when both play mixed strategies

for any $p \in \Delta_{\mathcal{H}}$.

if $h_j$ wrong on $x_i$ then $\delta_{ij} = 1 \Rightarrow M_{ij} = 1$

right $\delta_{ij} = 0 \Rightarrow M_{ij} = -1$

$$q^T M e_j = \sum_{i=1}^{m} M_{ij} \cdot q_i = \sum_{i=1}^{m} q_i \cdot (2\delta_{ij} - 1) = 2 \cdot \frac{1-\gamma}{2} - 1 = -\gamma$$

# Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \leq i \leq n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

$$q^T M e_j \leq -\gamma \quad \Rightarrow \quad q^T M p \leq \underline{-\gamma}$$

for any $p \in \Delta_{\mathcal{H}}$.

- By minimax, we have:

$$\underbrace{\boxed{\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p}}_{\substack{\leq -\gamma \\ \text{previous bullet}}} = \underbrace{\boxed{\min_{p \in \Delta_{\mathcal{H}}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p}}_{\leq -\gamma} \leq -\gamma$$

# Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \leq i \leq n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

  $$q^T M e_j \leq -\gamma \quad \Rightarrow \quad q^T M p \leq -\gamma$$

  for any $p \in \Delta_{\mathcal{H}}$.
- By minimax, we have:

  $$\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p = \min_{p \in \Delta_{\mathcal{H}}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p \leq -\gamma$$

- In particular, right hand side implies weighted classifier given by optimum solution $p^*$ *always* correct.

# Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \leq i \leq n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

$(p^*, q^*)$ optimum solution $\therefore$ $q^T M p^* \leq -\gamma \quad \forall \ q \in \Delta_{\mathcal{X}}$

$$\Rightarrow \underline{e_i^T M p^* \leq -\gamma}$$

- By minimax, we have:

$$\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p = \min_{p \in \Delta_{\mathcal{H}}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p \leq -\gamma$$

- In particular, right hand side implies weighted classifier given by optimum solution $p^*$ *always* correct.

# Proof of Correctness of Classifier

$$C_{p^*}(x_i) = \begin{cases} 1 & \text{if } \underline{\sum_{j=1}^{n} p_j^* h_j(x_i) \geq \frac{1}{2}} \quad (*) \\ 0 & \text{otherwise} \end{cases}$$

$$\underbrace{2h_j(x_i) - 1}_{-M_{ij}} = \begin{cases} 1 & \text{if } h_j(x_i) = 1 \quad \overset{h_j}{\text{correct}} \\ -1 & \text{otherwise } (\text{not } \overset{h_j}{\text{correct}}) \end{cases} \quad \boxed{\text{if } c(x_i) = 1}$$

$$C_{p^*}(x_i) \text{ correct} \iff -\sum_{j=1}^{n} M_{ij} p_j^* \geq \underbrace{1}_{(*)} - \underbrace{1}_{\Sigma p_i^*} = 0$$

$$C_{p^*}(x_i) \text{ correct} \iff \underbrace{\sum_{j=1}^{n} M_{ij} p_j^*}_{e_i^T M p^*} < 0 \qquad \leq -\gamma < 0 \text{ by previous page}$$

# Proof of Correctness of Classifier

- Game Theory - Minimax Theorems
- Learning Theory - Boosting
- Combinatorics - Bipartite Matching
- Conclusion
- Acknowledgements

# Bipartite Matching

- Given a bipartite graph $G(L \sqcup R, E)$, does it have a perfect matching?

# Bipartite Matching

- Given a bipartite graph $G(L \sqcup R, E)$, does it have a perfect matching?
- We saw in lecture 7 that we can randomly isolate a perfect matching, if one exists

# Bipartite Matching

- Given a bipartite graph $G(L \sqcup R, E)$, does it have a perfect matching?
- We saw in lecture 7 that we can randomly isolate a perfect matching, if one exists
- Can we remove the randomness in that process? This would lead to a fast parallel algorithm for matching.

# Bipartite Matching

- Given a bipartite graph $G(L \sqcup R, E)$, does it have a perfect matching?
- We saw in lecture 7 that we can randomly isolate a perfect matching, if one exists
- Can we remove the randomness in that process? This would lead to a fast parallel algorithm for matching.
- Breakthrough result of [Fenner, Gurjar and Thierauf 2019]

# Bipartite Matching

- Given a bipartite graph $G(L \sqcup R, E)$, does it have a perfect matching?
- We saw in lecture 7 that we can randomly isolate a perfect matching, if one exists
- Can we remove the randomness in that process? This would lead to a fast parallel algorithm for matching.
- Breakthrough result of [Fenner, Gurjar and Thierauf 2019]
- We will see just a piece of the proof

# Bipartite Matching & Circulation

- Given an even cycle $C = (e_1, e_2, \ldots, e_{2k})$, we say that the *circulation* of $C$ is given by

$$circ(C) = |w(e_1) - w(e_2) + \ldots + w(e_{2k-1}) - w(e_{2k})|$$

# Bipartite Matching & Circulation

- Given an even cycle $C = (e_1, e_2, \ldots, e_{2k})$, we say that the *circulation* of $C$ is given by

$$circ(C) = |w(e_1) - w(e_2) + \ldots + w(e_{2k-1}) - w(e_{2k})|$$

- Lemma: if we assign weights $w(e_i)$ such that $circ(C) \neq 0$ for each cycle $C$ of the bipartite graph $G$, then we get that the minimum weight PM is unique!

# Bipartite Matching & Circulation

- Given an even cycle $C = (e_1, e_2, \ldots, e_{2k})$, we say that the *circulation* of $C$ is given by

$$circ(C) = |w(e_1) - w(e_2) + \ldots + w(e_{2k-1}) - w(e_{2k})|$$

- Lemma: if we assign weights $w(e_i)$ such that $circ(C) \neq 0$ for each cycle $C$ of the bipartite graph $G$, then we get that the minimum weight PM is unique!
- The approach of [Fenner, Gurjar and Thierauf 2019] is to construct a set of weights which make all circulations non-zero!
  - To do that, they iteratively construct a weight assignment that kills small cycles (i.e., make their circulation non-zero)
  - Once we have a bipartite graph with no cycles of length $2k$, then in next iteration we kill cycles of length up to $4k$
  - One can show that no cycles of length $2k$ then *few cycles* of length $4k$ – similar to Karger's min cut lemma!

# Bipartite Matching

- Suppose we have a weight assignment $w$. Let $G_w$ be the subgraph of $G$ given by the union of *all min $w$-weight perfect matchings in $G$*.

# Bipartite Matching

- Suppose we have a weight assignment $w$. Let $G_w$ be the subgraph of $G$ given by the union of *all min $w$-weight perfect matchings in $G$*.
- Claim: circulation of each (even) cycle in $G_w$ is zero

# Bipartite Matching

- Suppose we have a weight assignment $w$. Let $G_w$ be the subgraph of $G$ given by the union of *all min $w$-weight perfect matchings in $G$*.
- Claim: circulation of each (even) cycle in $G_w$ is zero
- Proof: LP duality! (complementary slackness)
  - Linear programs:

*Primal*

$$\begin{array}{ll} \text{min} & \sum_{e \in E} w_e x_e \\ \text{s.t.} & x \geq 0 \\ & \sum_{e \in \delta(u)} x_e = 1 \\ & \text{for } u \in L \sqcup R \end{array}$$

# Bipartite Matching

- Suppose we have a weight assignment $w$. Let $G_w$ be the subgraph of $G$ given by the union of *all min $w$-weight perfect matchings in $G$*.
- Claim: circulation of each (even) cycle in $G_w$ is zero
- Proof: LP duality! (complementary slackness)
  - Linear programs:

*Primal*

$$\begin{aligned} \min \quad & \sum_{e\in E} w_e x_e \\ \text{s.t.} \quad & x \geq 0 \\ & \sum_{e\in\delta(u)} x_e = 1 \\ & \text{for } u \in L \sqcup R \end{aligned}$$

*Dual*

$$\begin{aligned} \max \quad & \sum_{u\in L\sqcup R} y_u \\ \text{s.t.} \quad & y_u + y_v \leq w_e \\ & \text{for } e = \{u, v\} \in E \end{aligned}$$

  - Complementary slackness says $x_e \neq 0$ in primal, where $e = \{u, v\}$ $\Rightarrow y_u + y_v = w_e$ in dual optimal.

# Bipartite Matching - Dual

# Bipartite Matching - Circulation

# Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life

# Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)

# Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

# Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

- Linear Programming and Duality - fundamental concepts, lots of applications!

# Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

- Linear Programming and Duality - fundamental concepts, lots of applications!
  - Applications in Combinatorial Optimization (a lot of it happened here at UW!)
  - Applications in Game Theory (minimax theorem)
  - Applications in Learning Theory (boosting)
  - Applications in Parallel Computation/Derandomization (Perfect Matching)
  - many more

# Acknowledgement

- Lecture based largely on:
  - Lectures 3-6 of Yarom Singer's Advanced Optimization class
  - [Schrijver 1986, Chapter 7]
  - Personal Communication with Rohit
- See Yarom's notes at `https://people.seas.harvard.edu/~yaron/AM221-S16/schedule.html`

# References I

Schirjver, Alexander (1986)
*Theory of Linear and Integer Programming*

Fourier, J. B. 1826
Analyse des travaux de l'Académie Royale des Sciences pendant l'année 1823.
*Partie mathématique (1826)*

Fourier, J. B. 1827
Analyse des travaux de l'Académie Royale des Sciences pendant l'année 1824.
*Partie mathématique (1827)*

Fenner, Stephen, and Gurjar, Rohit, and Thierauf, Thomas (2019)
Bipartite perfect matching is in quasi-NC
SIAM Journal on Computing, 2019