Online learning
in
repeated matrix games

Yoav Freund

January 31, 2011

Outline

Repeated Games

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Outline

Repeated games

- Game between two players.

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices ,**C**

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices ,**C**
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Players Observe each other's action.

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Players Observe each other's action.
- Row player gains $(i, j) \in [0, 1]$

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Players Observe each other's action.
- Row player gains $(i, j) \in [0, 1]$
- Column player gains $\mathbf{C}(i, j) \in [0, 1]$

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Players Observe each other's action.
- Row player gains $(i, j) \in [0, 1]$
- Column player gains $\mathbf{C}(i, j) \in [0, 1]$
- Game repeated many times.

Repeated games

- Game between two players.
- Defined by two $n \times m$ matrices $,\mathbf{C}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Players Observe each other's action.
- Row player gains $(i, j) \in [0, 1]$
- Column player gains $\mathbf{C}(i, j) \in [0, 1]$
- Game repeated many times.
- Player choices can depend on the past.

Notions of Convergence

- 1 The players have diminising regret (External or internal)

Notions of Convergence

- 1 The players have diminising regret (External or internal)
- 2 The empirical distributions over the actions converge to a set of distributions over action k-tuples $< i_1, i_2, i_3, \ldots, i_k >$

Notions of Convergence

- 1 The players have diminising regret (External or internal)
- 2 The empirical distributions over the actions converge to a set of distributions over action k-tuples $< i_1, i_2, i_3, \ldots, i_k >$
- The set of Nash Equilibria.

Notions of Convergence

- 1 The players have diminising regret (External or internal)
- 2 The empirical distributions over the actions converge to a set of distributions over action k-tuples $< i_1, i_2, i_3, \ldots, i_k >$
- The set of Nash Equilibria.
- The set of Correlated equilibria.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as “follow the leader”.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as "follow the leader".
- If both sides use fictitious play and the game is two-player zero sum game, then empirical distribution converges to the set of Nash Equilibria.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as “follow the leader”.
- If both sides use fictitious play and the game is two-player zero sum game, then empirical distribution converges to the set of Nash Equilibria.
- If each player has only two actions, then cnvergence to Nash set also guaranteed.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as “follow the leader”.
- If both sides use fictitious play and the game is two-player zero sum game, then empirical distribution converges to the set of Nash Equilibria.
- If each player has only two actions, then cnvergence to Nash set also guaranteed.
- Convergence to Nash set is not guaranteed in general.

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as “follow the leader”.
- If both sides use fictitious play and the game is two-player zero sum game, then empirical distribution converges to the set of Nash Equilibria.
- If each player has only two actions, then cnvergence to Nash set also guaranteed.
- Convergence to Nash set is not guaranteed in general.
- What if the other side does not follow fictitious play?

Fictitious play

- **Strategy:** Use best response to the empirical distribution of the actions of the opponent so far.
- Same as “follow the leader”.
- If both sides use fictitious play and the game is two-player zero sum game, then empirical distribution converges to the set of Nash Equilibria.
- If each player has only two actions, then cnvergence to Nash set also guaranteed.
- Convergence to Nash set is not guaranteed in general.
- What if the other side does not follow fictitious play?
- Conforming player can suffer non-diminishing regret.

Hannan's Algorithm

- *Approximation to Bayes risk in reapeated play*, Contributions to the theory of Games, 1957.

Hannan's Algorithm

- *Approximation to Bayes risk in reapeated play*, Contributions to the theory of Games, 1957.
- 1957: IBM announces it will no longer be using vacuum tubes and releases its first computer that had 2000 transistors.

Hannan's Algorithm

- *Approximation to Bayes risk in reapeated play*, Contributions to the theory of Games, 1957.
- 1957: IBM announces it will no longer be using vacuum tubes and releases its first computer that had 2000 transistors.
- Instead of using "follow the leader" use "follow the perturbed leader", i.e. add a small amount of noise to the cumulative utility of each action, *then* pick the leader.

Hannan's Algorithm

- *Approximation to Bayes risk in reapeated play*, Contributions to the theory of Games, 1957.
- 1957: IBM announces it will no longer be using vacuum tubes and releases its first computer that had 2000 transistors.
- Instead of using "follow the leader" use "follow the perturbed leader", i.e. add a small amount of noise to the cumulative utility of each action, *then* pick the leader.
- **Hannan consistency:** Cumulative regret / Cumulative utility $\rightarrow 0$.

Hannan's Set

- Diminishing regret relative to playing a fixed pure strategy.

Hannan's Set

- Diminishing regret relative to playing a fixed pure strategy.
- Hannan's set contains all joint distributions over player's action where all players have no external regret.

Hannan's Set

- Diminishing regret relative to playing a fixed pure strategy.
- Hannan's set contains all joint distributions over player's action where all players have no external regret.
- Hannan's set contains the set of correlated equilibrium which contains the set of Nash Equilibria.

Reaching correlated equilibrium

- By all players minimizing internal regret.

Reaching correlated equilibrium

- By all players minimizing internal regret.
- By making a calibrated predictions of the opponent's next move and playing best response.

Zero sum games in matrix form

- Game between two players.

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Row player gains $\mathbf{M}(i, j) \in [0, 1]$

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Row player gains $\mathbf{M}(i, j) \in [0, 1]$
- Column player looses $\mathbf{M}(i, j)$

Zero sum games in matrix form

- Game between two players.
- Defined by $n \times m$ matrix $\mathbf{M}$
- Row player chooses $i \in \{1, \ldots, n\}$
- Column player chooses $j \in \{1, \ldots, m\}$
- Row player gains $\mathbf{M}(i, j) \in [0, 1]$
- Column player looses $\mathbf{M}(i, j)$
- Game repeated many times.

Pure vs. mixed strategies

- Choosing a single action = pure strategy.

Pure vs. mixed strategies

- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.

Pure vs. mixed strategies

- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.
- Row player chooses dist. over rows **P**

Pure vs. mixed strategies

- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.
- Row player chooses dist. over rows **P**
- Column player chooses dist. over columns **Q**

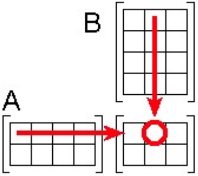
Pure vs. mixed strategies

- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.
- Row player chooses dist. over rows $\mathbf{P}$
- Column player chooses dist. over columns $\mathbf{Q}$
- Row player gains $\mathbf{M}(\mathbf{P}, \mathbf{Q})$.

Pure vs. mixed strategies

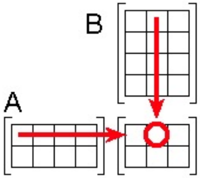
- Choosing a single action = pure strategy.
- Choosing a Distribution over actions = mixed strategy.
- Row player chooses dist. over rows $\mathbf{P}$
- Column player chooses dist. over columns $\mathbf{Q}$
- Row player gains $\mathbf{M}(\mathbf{P}, \mathbf{Q})$.
- Column player looses $\mathbf{M}(\mathbf{P}, \mathbf{Q})$.

Mixed strategies in matrix notation



$$(A \times B)_{12} = \sum_{r=1}^{4} a_{1r} b_{r2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} + a_{14}b_{42}$$

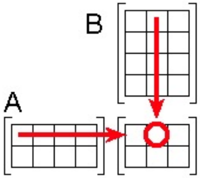
Mixed strategies in matrix notation



$$(A \times B)_{12} = \sum_{r=1}^{4} a_{1r} b_{r2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} + a_{14}b_{42}$$

Q is a column vector. $\mathbf{P}^T$ is a row vector.

Mixed strategies in matrix notation



$$(A \times B)_{12} = \sum_{r=1}^{4} a_{1r} b_{r2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} + a_{14}b_{42}$$

Q is a column vector. $\mathbf{P}^T$ is a row vector.

$$\mathbf{M}(\mathbf{P}, \mathbf{Q}) = \mathbf{P}^T\mathbf{M}\mathbf{Q} = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{P}(i)\mathbf{M}(i,j)\mathbf{Q}(j)$$

The basic algorithm

- Choose an initial distribution $\mathbf{P}_1$

The basic algorithm

- Choose an initial distribution $\mathbf{P}_1$
- $$\mathbf{P}_{t+1}(i) = \mathbf{P}_t(i) \frac{e^{-\eta \mathbf{M}(i, \mathbf{Q}_t)}}{Z_t}$$

The basic algorithm

- Choose an initial distribution $\mathbf{P}_1$
- $$\mathbf{P}_{t+1}(i) = \mathbf{P}_t(i)\frac{e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}}{Z_t}$$
- Where $Z_t = \sum_{i=1}^{n} \mathbf{P}_t(i) e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}$

The basic algorithm

- Choose an initial distribution $\mathbf{P}_1$
- $$\mathbf{P}_{t+1}(i) = \mathbf{P}_t(i)\frac{e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}}{Z_t}$$
- Where $Z_t = \sum_{i=1}^{n} \mathbf{P}_t(i) e^{-\eta \mathbf{M}(i,\mathbf{Q}_t)}$
- $\eta > 0$ is the learning rate.

Main Theorem

- For any game matrix $\mathbf{M}$.

Main Theorem

- For any game matrix $\mathbf{M}$.
- Any sequence of mixed strat. $\mathbf{Q}_1, \ldots, \mathbf{Q}_T$

Main Theorem

- For any game matrix **M**.
- Any sequence of mixed strat. $\mathbf{Q}_1, \ldots, \mathbf{Q}_T$
- The sequence $\mathbf{P}_1, \ldots, \mathbf{P}_T$ produced by basic alg using $\eta > 0$ satisfies

$$\sum_{t=1}^{T} \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \left(\frac{1}{1 - e^{-\eta}} \right) \min_{\mathbf{P}} \left[\eta \sum_{t=1}^{T} \mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + \mathrm{RE}\left(\mathbf{P} \;\|\; \mathbf{P}_1\right) \right]$$

Corollary

- Setting $\eta = \ln\left(1 + \sqrt{\frac{2 \ln n}{T}}\right)$

Corollary

- Setting $\eta = \ln\left(1 + \sqrt{\frac{2\ln n}{T}}\right)$
- the average per-trial loss is

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + \Delta_{T,n}$$

Corollary

- Setting $\eta = \ln\left(1 + \sqrt{\frac{2\ln n}{T}}\right)$
- the average per-trial loss is

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{M}(\mathbf{P}, \mathbf{Q}_t) + \Delta_{T,n}$$

- Where

$$\Delta_{T,n} = \sqrt{\frac{2\ln n}{T}} + \frac{\ln n}{T} = O\left(\sqrt{\frac{\ln n}{T}}\right).$$

Main Lemma

On any iteration t

Main Lemma

On any iteration t

For any mixed strategy $\tilde{\mathbf{P}}$

Main Lemma

On any iteration t

For any mixed strategy $\tilde{\mathbf{P}}$

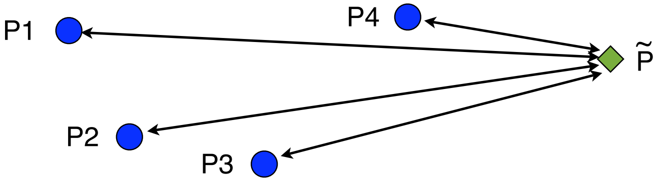
$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$$

Visual intuition

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_t\right) \leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1 - e^{-\eta}) \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$$

Visual intuition

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \parallel \mathbf{P}_t\right) \le \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) - (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)$$



Proof of Lemma (1)

$$\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right)$$

Proof of Lemma (1)

$$
\begin{aligned}
&\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t}\right) \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t+1}(i)} - \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t}(i)}
\end{aligned}
$$

Proof of Lemma (1)

$$
\begin{aligned}
& \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t}\right) \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t+1}(i)} - \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t}(i)} \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\mathbf{P}_{t}(i)}{\mathbf{P}_{t+1}(i)}
\end{aligned}
$$

Proof of Lemma (1)

$$
\begin{aligned}
&\mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_{t+1}\right) - \mathrm{RE}\left(\tilde{\mathbf{P}} \;\|\; \mathbf{P}_t\right) \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_{t+1}(i)} - \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\tilde{\mathbf{P}}(i)}{\mathbf{P}_t(i)} \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{\mathbf{P}_t(i)}{\mathbf{P}_{t+1}(i)} \\
&= \sum_{i=1}^{n} \tilde{\mathbf{P}}(i) \ln \frac{Z_t}{e^{\eta \mathbf{M}(i,\mathbf{Q}_t)}}
\end{aligned}
$$

Proof of Lemma (2)

$$= \quad \eta \sum_{i=1}^{n} \tilde{\mathbf{P}}(i)\mathbf{M}(i, \mathbf{Q}_t) + \ln Z_t$$

Proof of Lemma (2)

$$
\begin{aligned}
&= \eta \sum_{i=1}^{n} \tilde{\mathbf{P}}(i)\mathbf{M}(i, \mathbf{Q}_t) + \ln Z_t \\
&\leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln \left[\sum_{i=1}^{n} \mathbf{P}_t(i) \left(1 - (1 - e^{-\eta})\mathbf{M}(i, \mathbf{Q}_t)\right) \right]
\end{aligned}
$$

Proof of Lemma (2)

$$
\begin{aligned}
&= \eta \sum_{i=1}^{n} \tilde{\mathbf{P}}(i)\mathbf{M}(i, \mathbf{Q}_t) + \ln Z_t \\
&\leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln \left[\sum_{i=1}^{n} \mathbf{P}_t(i) \left(1 - (1 - e^{-\eta})\mathbf{M}(i, \mathbf{Q}_t)\right) \right] \\
&= \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln \left(1 - (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right)
\end{aligned}
$$

Proof of Lemma (2)

$$
\begin{aligned}
&= \eta \sum_{i=1}^{n} \tilde{\mathbf{P}}(i)\mathbf{M}(i, \mathbf{Q}_t) + \ln Z_t \\
&\leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln \left[\sum_{i=1}^{n} \mathbf{P}_t(i) \left(1 - (1 - e^{-\eta})\mathbf{M}(i, \mathbf{Q}_t)\right) \right] \\
&= \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + \ln \left(1 - (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)\right) \\
&\leq \eta \mathbf{M}(\tilde{\mathbf{P}}, \mathbf{Q}_t) + (1 - e^{-\eta})\mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t)
\end{aligned}
$$

The minmax Theorem

John von Neumann, 1928.

The minmax Theorem

John von Neumann, 1928.

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) \leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

The minmax Theorem

John von Neumann, 1928.

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}, \mathbf{Q}) \leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

In words: for mixed strategies, choosing second gives no advantage.

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$
Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$ and $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$
Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_t$ and $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T}\mathbf{Q}_t$

$$\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \quad \leq \quad \max_{\mathbf{Q}} \overline{\mathbf{P}}^{\mathrm{T}}\mathbf{M}\mathbf{Q}$$

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$
Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$ and $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$

$$
\begin{aligned}
\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} &\leq \max_{\mathbf{Q}} \overline{\mathbf{P}}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \\
&= \max_{\mathbf{Q}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \quad \text{by definition of } \overline{\mathbf{P}}
\end{aligned}
$$

Proving minmax Theorem using online learning (1)

Row player chooses $\mathbf{P}_t$ using learning alg.
Column player chooses $\mathbf{Q}_t$ after row player so that
$\mathbf{Q}_t = \arg\max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$
Let $\overline{\mathbf{P}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t$ and $\overline{\mathbf{Q}} \doteq \frac{1}{T}\sum_{t=1}^{T} \mathbf{Q}_t$

$$\begin{aligned}
\min_{\mathbf{P}} \max_{\mathbf{Q}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} &\leq \max_{\mathbf{Q}} \overline{\mathbf{P}}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \\
&= \max_{\mathbf{Q}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q} \quad \text{by definition of } \overline{\mathbf{P}} \\
&\leq \frac{1}{T}\sum_{t=1}^{T} \max_{\mathbf{Q}} \mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}
\end{aligned}$$

Proving minmax Theorem using online learning (2)

$$= \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t \qquad \text{by definition of } \mathbf{Q}_t$$

Proving minmax Theorem using online learning (2)

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t{}^{\mathrm{T}} \mathbf{M} \mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}^{\mathrm{T}} \mathbf{M} \mathbf{Q}_t + \Delta_{T,n} && \text{by the Corollary}
\end{aligned}
$$

Proving minmax Theorem using online learning (2)

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_t^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\le \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t + \Delta_{T,n} && \text{by the Corollary} \\
&= \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\overline{\mathbf{Q}} + \Delta_{T,n} && \text{by definition of } \overline{\mathbf{Q}}
\end{aligned}
$$

Proving minmax Theorem using online learning (2)

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}_t{}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q}_t + \Delta_{T,n} && \text{by the Corollary} \\
&= \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\overline{\mathbf{Q}} + \Delta_{T,n} && \text{by definition of } \overline{\mathbf{Q}} \\
&\leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{Q} + \Delta_{T,n}.
\end{aligned}
$$

Proving minmax Theorem using online learning (2)

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}_t^{\mathrm{T}} \mathbf{M} \mathbf{Q}_t && \text{by definition of } \mathbf{Q}_t \\
&\leq \min_{\mathbf{P}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{P}^{\mathrm{T}} \mathbf{M} \mathbf{Q}_t + \Delta_{T,n} && \text{by the Corollary} \\
&= \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}} \mathbf{M} \overline{\mathbf{Q}} + \Delta_{T,n} && \text{by definition of } \overline{\mathbf{Q}} \\
&\leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{P}^{\mathrm{T}} \mathbf{M} \mathbf{Q} + \Delta_{T,n}.
\end{aligned}
$$

but $\Delta_{T,n}$ can be set arbitrarily small.

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.
- If all sides use learning, then game will converge to minmax equilibrium.

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.
- If all sides use learning, then game will converge to minmax equilibrium.
- If opponent is not optimally adversarial (limited by knowledge, computationa power...) then learning gives better performance than min-max.

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.
- If all sides use learning, then game will converge to minmax equilibrium.
- If opponent is not optimally adversarial (limited by knowledge, computationa power...) then learning gives better performance than min-max.
- Is it realistic to assume that markets are at equilibrium?

Minmax is weaker than diminishing regret

- The minmax theorem proves the existence of an Equilibrium.
- Learning guarantees no regret with respect to the past.
- If all sides use learning, then game will converge to minmax equilibrium.
- If opponent is not optimally adversarial (limited by knowledge, computationa power...) then learning gives better performance than min-max.
- Is it realistic to assume that markets are at equilibrium?
- If game is not zero sum (allows incentives to collaborate) and all players use learning then game converges to correlated equilibrium.