

Online learning in repeated matrix games

Yoav Freund

January 18, 2018

Outline

Repeated Matrix Games

Fictitious play

Strategy using Hedge

The basic analysis

Proof of minmax theorem

Approximately solving games

Fixed Learning rate

Variable learning rate

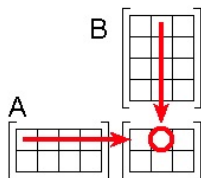
Zero sum games in matrix form

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

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**
- ▶ **Row** player gains **$M(P, Q)$** .
- ▶ **Column** player loses **$M(P, Q)$** .

Mixed strategies in matrix notation



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

Q is a **column** vector. **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 minmax Theorem

When using pure strategies, second player has an advantage.

John von Neumann, 1928.

$$\min_P \max_Q \mathbf{M}(\mathbf{P}, \mathbf{Q}) = \max_Q \min_P \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

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

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, computational power...) then learning gives **better** performance than min-max.
- ▶ Our goal is to minimize regret.

Fictitious play

- ▶ Choose the best action with respect to the sum of past loss vectors.
- ▶ Might not converge to optimal mixed strategy.
- ▶ Consider playing the matching coins game against an adversary that alternates HTTHHTTTHHTTHH....

Randomized Fictitious play

- ▶ Choose the best action with respect to the sum of past loss vectors **plus noise**.
- ▶ Adding noise allows us to choose responses that are slightly worse than best response.
- ▶ **Hannan 1957** Randomized ficticonverge to regret minimizing strategy.

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.

Generalized regret bound

- ▶ Regret relative to the best *pure strategy* i

$$\sum_{t=1}^T \mathbf{M}(\mathbf{P}_t, \mathbf{Q}_t) \leq \left(\frac{1}{1 - e^{-\eta}} \right) \min_i \left[\eta \sum_{t=1}^T \mathbf{M}(i, \mathbf{Q}_t) - \ln \mathbf{P}_1(i) \right]$$

- ▶ regret with respect the the best *mixed strategy* \mathbf{P} :

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

- ▶ Where

$$\text{RE}(\mathbf{P} \parallel \mathbf{Q}) \doteq \sum_{i=1}^n \mathbf{P}(i) \ln \frac{\mathbf{P}(i)}{\mathbf{Q}(i)}$$

Main Theorem

- ▶ For **any** game matrix **M**.
- ▶ Any sequence of mixed strat. **Q**₁, ..., **Q**_T
- ▶ The sequence **P**₁, ..., **P**_T produced by basic alg using **η** > 0 satisfies

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

Corollary

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

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

- ▶ Where

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

Main Lemma

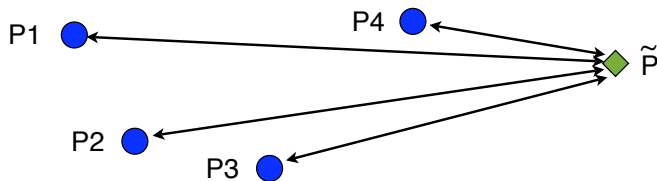
On any iteration t

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

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

Visual intuition

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



Proof of Lemma (1)

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

Proof of Lemma (2)

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

The minmax Theorem

John von Neumann, 1928.

$$\min_P \max_Q \mathbf{M}(\mathbf{P}, \mathbf{Q}) = \max_Q \min_P \mathbf{M}(\mathbf{P}, \mathbf{Q})$$

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.

Column player chooses \mathbf{Q}_t after row player so that

$$\mathbf{Q}_t = \arg \max_{\mathbf{Q}} \mathbf{M}(\mathbf{P}_t, \mathbf{Q})$$

$$\text{Let } \bar{\mathbf{P}} \doteq \frac{1}{T} \sum_{t=1}^T \mathbf{P}_t \text{ and } \bar{\mathbf{Q}} \doteq \frac{1}{T} \sum_{t=1}^T \mathbf{Q}_t$$

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

Proving minmax Theorem using online learning (2)

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

$$\leq \min_{\mathbf{P}} \frac{1}{T} \sum_{t=1}^T \mathbf{P}^T \mathbf{M} \mathbf{Q}_t + \Delta_{T,n} \quad \text{by the Corollary}$$

$$= \min_{\mathbf{P}} \mathbf{P}^T \mathbf{M} \overline{\mathbf{Q}} + \Delta_{T,n} \quad \text{by definition of } \overline{\mathbf{Q}}$$

$$\leq \max_{\mathbf{Q}} \min_{\mathbf{P}} \mathbf{P}^T \mathbf{M} \mathbf{Q} + \Delta_{T,n}.$$

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

Solving a game

- ▶ to **solve** a game is to find the min-max mixed strategies **P, Q**
- ▶ Suppose that **Hedge**(η) is playing **P**₁, **P**₂, against an adversary that plays **Q**₁, **Q**₂, ... such that

- └ Approximately solving games
- └ Fixed Learning rate

Using average row distribution

- Using the

- └ Approximately solving games
- └ Variable learning rate

Using the final row distribution

► XXX