Game-Theoretic Learning:
Regret Minimization vs. Utility Maximization

Amy Greenwald
with David Gondek, Amir Jafari, and Casey Marks
Brown University

University of Pennsylvania
November 17, 2004
Background

No-external-regret learning converges to the set of minimax equilibria, in zero-sum games. [e.g., Freund and Schapire 1996]

No-internal-regret learning converges to the set of correlated equilibria, in general-sum games. [e.g., Foster and Vohra 1997]
1. Definitions
   - A continuum of no-regret properties, called no-Φ-regret.
   - A continuum of game-theoretic equilibria, called Φ-equilibria.

2. Existence Theorem
   - Constructive proof: No-Φ-regret learning algorithms exist, ∀Φ.

3. Convergence Theorem
   - No-Φ-regret learning converges to the set of Φ-equilibria, ∀Φ.

4. Surprising Result
   - No-internal-regret is the strongest form of no-Φ-regret learning.
   - Therefore, no no-Φ-regret algorithm learns Nash equilibria.
Outline

- Game Theory
- Single Agent Learning Model
- Multiagent Learning & Game-Theoretic Equilibria
Game Theory: A Crash Course

1. General-Sum Games
   - Nash Equilibrium
   - Correlated Equilibrium

2. Zero-Sum Games
   - Minimax Equilibrium
An Example

Prisoners’ Dilemma

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>4,4</td>
<td>0,5</td>
</tr>
<tr>
<td>$D$</td>
<td>5,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

$C$: Cooperate
$D$: Defect
One-Shot Games

A one-shot game is a 3-tuple $\Gamma = (I, (A_i, r_i)_{i \in I})$, where

- $I$ is a set of players

- for all players $i \in I$,
  - a set of pure actions $A_i$ with $a_i \in A_i$
  - a reward function $r_i : A \to \mathbb{R}$, where $A = \prod_{i \in I} A_i$ with $a \in A$
One-Shot Games

A one-shot game is a 3-tuple $\Gamma = (I, (A_i, r_i)_{i \in I})$, where

- $I$ is a set of players
- for all players $i \in I$,
  - a set of pure actions $A_i$ with $a_i \in A_i$
  - a reward function $r_i : A \to \mathbb{R}$, where $A = \prod_{i \in I} A_i$ with $a \in A$

The players can employ randomized or mixed actions:

- for all players $i \in I$,
  - a set of mixed actions $Q_i = \{q_i \in \mathbb{R}^{A_i} | \sum_j q_{ij} = 1 & q_{ij} \geq 0, \forall j\}$, with $q_i \in Q_i$
  - an expected reward function $r_i : Q \to \mathbb{R}$, where $Q = \prod_{i \in I} Q_i$ with $q \in Q$, s.t. $r_i(q) = \sum_{a \in A} q(a)r_i(a)$
Nash Equilibrium

Notation
Write \( a = (a_i, a_{-i}) \in A \) for \( a_i \in A_i \) and \( a_{-i} \in A_{-i} = \prod_{j \neq i} A_j \).
Write \( q = (q_i, q_{-i}) \in Q \) for \( q_i \in Q_i \) and \( q_{-i} \in Q_{-i} = \prod_{j \neq i} Q_j \).

Definition
A Nash equilibrium is a mixed action profile \( q^* \) s.t. \( r_i(q^*) \geq r_i(q_i, q_{-i}^*) \), for all players \( i \) and for all mixed actions \( q_i \in Q_i \).

Theorem [Nash 51]
Every finite strategic form game has a mixed strategy Nash equilibrium.
Correlated Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>6,6</td>
<td>2,7</td>
</tr>
<tr>
<td>$B$</td>
<td>7,2</td>
<td>0,0</td>
</tr>
</tbody>
</table>

$max \, 12\pi_{TL} + 9\pi_{TR} + 9\pi_{BL} + 0\pi_{BR}$

subject to

\[ \pi_{TL} + \pi_{TR} + \pi_{BL} + \pi_{BR} = 1 \]
\[ \pi_{TL}, \pi_{TR}, \pi_{BL}, \pi_{BR} \geq 0 \]

\[ 6\pi_{L|T} + 2\pi_{R|T} \geq 7\pi_{L|T} + 0\pi_{R|T} \]
\[ 7\pi_{L|B} + 0\pi_{R|B} \geq 6\pi_{L|B} + 2\pi_{R|B} \]
\[ 6\pi_{T|L} + 2\pi_{B|L} \geq 7\pi_{T|L} + 0\pi_{B|L} \]
\[ 7\pi_{T|R} + 0\pi_{B|R} \geq 6\pi_{T|R} + 2\pi_{B|R} \]
Correlated Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T</em></td>
<td>6,6</td>
<td>2,7</td>
</tr>
<tr>
<td><em>B</em></td>
<td>7,2</td>
<td>0,0</td>
</tr>
</tbody>
</table>

max $12\pi_{TL} + 9\pi_{TR} + 9\pi_{BL} + 0\pi_{BR}$

subject to

$$\pi_{TL} + \pi_{TR} + \pi_{BL} + \pi_{BR} = 1$$

$$\pi_{TL}, \pi_{TR}, \pi_{BL}, \pi_{BR} \geq 0$$

$$6\pi_{TL} + 2\pi_{TR} \geq 7\pi_{TL} + 0\pi_{TR}$$

$$7\pi_{BL} + 0\pi_{BR} \geq 6\pi_{BL} + 2\pi_{BR}$$

$$6\pi_{TL} + 2\pi_{BL} \geq 7\pi_{TL} + 0\pi_{BL}$$

$$7\pi_{TR} + 0\pi_{BR} \geq 6\pi_{TR} + 2\pi_{BR}$$
Correlated Equilibrium

Definition
A mixed action profile $q^* \in Q$ is a correlated equilibrium iff for all pure actions $j, k \in A_i$,

\[ \sum_{a_{-i} \in A_{-i}} q(j, a_{-i}) (r_i(j, a_{-i}) - r_i(k, a_{-i})) \geq 0 \]  

Observe
Every Nash equilibrium is a correlated equilibrium \( \Rightarrow \)
Every finite normal form game has a correlated equilibrium.
Zero-Sum Games

Matching Pennies

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>−1,1</td>
<td>1,−1</td>
</tr>
<tr>
<td>T</td>
<td>1,−1</td>
<td>−1,1</td>
</tr>
</tbody>
</table>

Rock-Paper-Scissors

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0,0</td>
<td>−1,1</td>
<td>1,−1</td>
</tr>
<tr>
<td>P</td>
<td>1,−1</td>
<td>0,0</td>
<td>−1,1</td>
</tr>
<tr>
<td>S</td>
<td>−1,1</td>
<td>1,−1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

\[ \sum_{i \in I} r_i(a) = 0, \text{ for all } a \in A \]
\[ \sum_{i \in I} r_i(a) = c, \text{ for all } a \in A, \text{ for some } c \in \mathbb{R} \]
Minimax Equilibrium

Example

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Definition

A mixed action profile \((q_1^*, q_2^*) \in Q\) is a minimax equilibrium in a two-player, zero-sum game iff

\[
\circ \quad r_1(q_1^*, q_2^*) \geq r_1(j, q_2^*), \quad \forall j \in A_1
\]

\[
\circ \quad l_2(q_1^*, q_2^*) \leq l_2(q_1^*, k), \quad \forall k \in A_2
\]
Single Agent Learning Model

- set of actions $N = \{1, \ldots, n\}$

- for all times $t$,
  - mixed action vector $q^t \in Q = \{q \in \mathbb{R}^n | \sum_i q_i = 1 \& q_i \geq 0, \forall i\}$
  - pure action vector $a^t = e_i$ for some pure action $i$
  - reward vector $r^t = (r_1, \ldots, r_n) \in [0, 1]^n$

A learning algorithm $\mathcal{A}$ is a sequence of functions $q^t : \text{History}^{t-1} \rightarrow Q$, where a History is a sequence of action-reward pairs $(a^1, r^1), (a^2, r^2), \ldots$. 
Transformations

Mixed Transformations

$\Phi_{\text{LINEAR}} = \{ \phi : Q \to Q \}$

$\Phi_{\text{LINEAR}}$ = the set of all linear transformations
$\Phi_{\text{LINEAR}}$ = the set of all row stochastic matrices

$\Phi_{\text{SWAP}} = \{ \phi : Q \to Q \mid \phi \text{ deterministic} \} \subset \Phi_{\text{LINEAR}}$

Pure Transformations

$\mathcal{F}_{\text{SWAP}} = \{ F : N \to N \}$

$\mathcal{F}_{\text{SWAP}}$ = the set of all pure transformations
Isomorphism

The operation of elements of $\mathcal{F}_{\text{SWAP}}$ on $N \cong N$ is the operation of elements of $\Phi_{\text{SWAP}}$ on $Q$

\begin{align*}
\phi_{ij} &= \delta_{F(i)=j} \\
\forall k \quad e_k \phi &= e_{F(k)}
\end{align*} \hspace{1cm} (2) \hspace{1cm} (3)

**Example**  If $n = 4$ and $F = \{1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 4, 4 \mapsto 1\}$, then

$$
\phi = \begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0
\end{bmatrix}
$$

Thus, $\langle q_1, q_2, q_3, q_4 \rangle \phi = \langle q_4, q_1, q_2, q_3 \rangle$, for all $\langle q_1, q_2, q_3, q_4 \rangle \in Q$. 

16
External Regret Matrices

\[ \mathcal{F}_{\text{EXT}} = \{ F^j \in \mathcal{F}_{\text{SWAP}} | j \in N \}, \text{ where } F^j(k) = j \]

\[ \Phi_{\text{EXT}} = \{ \phi^j \in \Phi_{\text{SWAP}} | j \in N \}, \text{ where } e_k \phi^j = e_j \]

**Example**  If \( n = 4 \), then

\[ \phi^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \]

Thus, \( \langle q_1, q_2, q_3, q_4 \rangle \phi = \langle 0, 1, 0, 0 \rangle \), for all \( \langle q_1, q_2, q_3, q_4 \rangle \in Q \).
Internal Regret Matrices

\[ \mathcal{F}_{\text{INT}} = \{ F^{ij} \in \mathcal{F}_{\text{SWAP}} | ij \in N \} \text{, where } F^{ij}(k) = \begin{cases} j & \text{if } k = i \\ k & \text{otherwise} \end{cases} \]

\[ \Phi_{\text{INT}} = \{ \phi^{ij} \in \Phi_{\text{SWAP}} | ij \in N \} \text{, where } e_k \phi^{ij} = \begin{cases} e_j & \text{if } k = i \\ e_k & \text{otherwise} \end{cases} \]

Example  If \( n = 4 \), then

\[ \phi^{23} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

Thus, \( \langle q_1, q_2, q_3, q_4 \rangle \phi = \langle q_1, 0, q_2 + q_3, q_4 \rangle \), for all \( \langle q_1, q_2, q_3, q_4 \rangle \in Q \).
Regret Vector $\rho \in \mathbb{R}^\Phi$

<table>
<thead>
<tr>
<th>Regret Type</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Regret Vector</td>
<td>$\tilde{\rho}_\phi(r, a) = r \cdot a\phi - r \cdot a$</td>
</tr>
<tr>
<td>Expected Regret Vector</td>
<td>$\hat{\rho}<em>\phi(r, q) = \mathbb{E}[\rho</em>\phi(r, a) \mid a \sim q]$</td>
</tr>
<tr>
<td></td>
<td>$= \rho_\phi(r, \mathbb{E}[a \mid a \sim q])$</td>
</tr>
<tr>
<td></td>
<td>$= r \cdot q\phi - r \cdot q$</td>
</tr>
</tbody>
</table>

No Observed $\Phi$-Regret

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} \tilde{\rho}_\phi(r^\tau, a^\tau) \leq 0, \text{ for all } \phi \in \Phi, \text{ a.s.}$$

No Expected $\Phi$-Regret

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} \hat{\rho}_\phi(r^\tau, q^\tau) \leq 0, \text{ for all } \phi \in \Phi$$
Approachability

$U \subseteq V$ is said to be approachable iff there exists learning algorithm $A = q^1, q^2, \ldots$ s.t. for any sequence of rewards $r^1, r^2, \ldots$,

$$\lim_{t \to \infty} d(U, \bar{\rho}^t) = \lim_{t \to \infty} \inf_{u \in U} d(u, \bar{\rho}^t) = 0$$

a.s., where $\bar{\rho}^t$ denotes the average value of $\rho$ through time $t$.

A $\Phi$-no-regret learning algorithm is one whose observed regret approaches the negative orthant $\mathbb{R}_\Phi$. 
Blackwell’s Theorem

The negative orthant $\mathbb{R}_\Phi$ is approachable iff there exists a learning algorithm $\mathcal{A} = q^1, q^2, \ldots$ s.t. for any sequence of rewards $r^1, r^2, \ldots$,

$$\rho(r^{t+1}, q^{t+1}) \cdot (\bar{\rho}^t)^+ \leq 0$$

(4)

for all times $t$, where $x^+ = \max\{x, 0\}$.

Moreover, this procedure can be used to approach the negative orthant $\mathbb{R}_\Phi$:

- if $\bar{\rho}^t \in \mathbb{R}_\Phi$, play arbitrarily;

- if $\bar{\rho}^t \in V \setminus \mathbb{R}_\Phi$, play according to $\mathcal{A}$. 

21
Regret Matching Algorithm

Given $\Phi$
Given $Y \in \mathbb{R}^\Phi$

If $\sum_{\phi \in \Phi} Y_\phi = 0$, play arbitrarily
If $\sum_{\phi \in \Phi} Y_\phi > 0$, define stochastic matrix

$$A \equiv A(\Phi, Y) = \frac{\sum_{\phi \in \Phi} \phi Y_\phi}{\sum_{\phi \in \Phi} Y_\phi}$$  \hspace{1cm} (5)$$

play mixed strategy $q = qA$

Regret Matching Theorem

Regret matching satisfies the generalized Blackwell condition:

$$\rho(r, q) \cdot Y = 0$$
Proof

\[ \rho(r, q) \cdot Y = \sum_{\phi \in \Phi} \rho\phi(r, q)Y_{\phi} \]  
\[ = \sum_{\phi \in \Phi} (r \cdot q\phi - r \cdot q)Y_{\phi} \]  
\[ = \sum_{\phi \in \Phi} r \cdot (qY_{\phi} - q\phi) \]  
\[ = r \cdot \left( q \sum_{\phi \in \Phi} \phi Y_{\phi} - q \sum_{\phi \in \Phi} Y_{\phi} \right) \]  
\[ = \left( \sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot \left( q \frac{\sum_{\phi \in \Phi} \phi Y_{\phi}}{\sum_{\phi \in \Phi} Y_{\phi}} - q \right) \]  
\[ = \left( \sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot (qA - q) \]  
\[ = \left( \sum_{\phi \in \Phi} Y_{\phi} \right) r \cdot (q - q) \]  
\[ = 0 \]
Generic Regret Matching Algorithm \((\Phi, g)\)

for \(t = 1, \ldots, T\)

1. play mixed strategy \(q^t\)
2. realize pure action \(i\)
3. observe rewards \(r^t\)
4. for all \(\phi \in \Phi\)
   - compute instantaneous regret
     * observed \(\rho^t_{\phi} \equiv \rho_{\phi}(r^t, e_i) = r^t \cdot e_i \phi - r^t \cdot e_i\)
     * expected \(\rho^t_{\phi} \equiv \rho_{\phi}(r^t, q^t) = r^t \cdot q^t \phi - r^t \cdot q^t\)
   - update cumulative regret vector \(X^t_{\phi} = X^{t-1}_{\phi} + \rho^t_{\phi}\)
5. compute \(Y = g(X^t)\)
6. compute \(A = \frac{\sum_{\phi \in \Phi} \phi Y_{\phi}}{\sum_{\phi \in \Phi} Y_{\phi}}\)
7. solve for the fixed point \(q^{t+1} = q^{t+1} A\)
Special Cases of Regret Matching

Foster and Vohra 97 ($\Phi_{\text{INT}}$)
Hart and Mas-Colell 00 ($\Phi_{\text{EXT}}$)

Choose $G(X) = \frac{1}{2} \sum_k (X_k^+)^2$ so that $g_k(X) = X_k^+$

Freund and Schapire 95 ($\Phi_{\text{EXT}}$)
Cesa-Bianchi and Lugosi 03 ($\Phi_{\text{INT}}$)

Choose $G(X) = \frac{1}{\eta} \ln \left( \sum_k e^{\eta X_k} \right)$ so that $g_k(X) = e^{\eta X_k} / \sum_k e^{\eta X_k}$
Multiagent Model

- a set of players $I$ ($i \in I$)

- for all players $i$,
  - a set of pure actions $A_i$ with $a_i \in A_i$
  - a set of mixed actions $Q_i$ with $q_i \in Q_i$

- a reward function $r_i : A \rightarrow [0, 1]$, where $A = \prod_i A_i$ with $a \in A$
- an expected reward function $r_i : Q \rightarrow [0, 1]$, where $Q = \prod_i Q_i$ with $q \in Q$ s.t. $r_i(q) = \sum_{a \in A} q(a) r_i(a)$

- a set $\Phi_i$ ($\phi_i \in \Phi_i$)
Φ-Equilibrium

An mixed action profile \( q \in Q \) is a Φ-equilibrium iff
\[ r_i(\phi_i(q)) \leq r_i(q), \text{ for all players } i \text{ and for all } \phi_i \in \Phi_i. \]

Examples
Correlated Equilibrium: \( \Phi_i = \Phi_{\text{INT}} \), for all players \( i \)
Generalized Minimax Equilibrium: \( \Phi_i = \Phi_{\text{EXT}} \), for all players \( i \)
Convergence Theorem

If all players $i$ play via some no-$\Phi_i$-regret algorithm, then the joint empirical distribution of play converges to the set of $\Phi$-equilibria, almost surely.

Proof

For all players $i$, for all $\phi_i \in \Phi_i$,

$$
\limsup_{t \to \infty} r_i(\phi_i(z^t)) - r_i(z^t) = \limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} r_i(\phi_i(a_i^\tau), a_{-i}^\tau) - \frac{1}{t} \sum_{\tau=1}^{t} r_i(a_i^\tau, a_{-i}^\tau) \leq 0
$$

almost surely.
Zero-Sum Games

Matching Pennies

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>−1, 1</td>
<td>1, −1</td>
</tr>
<tr>
<td>T</td>
<td>1, −1</td>
<td>−1, 1</td>
</tr>
</tbody>
</table>

Rock-Paper-Scissors

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0, 0</td>
<td>−1, 1</td>
<td>1, −1</td>
</tr>
<tr>
<td>P</td>
<td>1, −1</td>
<td>0, 0</td>
<td>−1, 1</td>
</tr>
<tr>
<td>S</td>
<td>−1, 1</td>
<td>1, −1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>
Matching Pennies

Weighs

NER exponential $[\eta = 0.037233]$: Matching Pennies

Frequencies

NER exponential $[\eta = 0.037233]$: Matching Pennies
Rock-Paper-Scissors

Weights

Frequencies

NER polynomial \([p = 2]\): Rock, Paper, Scissors

Player 1: R
Player 1: P
Player 1: S
Player 2: R
Player 2: P
Player 2: S

Empirical Frequency

Time

Weight

Time
General-Sum Games

Shapley Game

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$C$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>0,0</td>
<td>1,0</td>
<td>0,1</td>
</tr>
<tr>
<td>$M$</td>
<td>0,1</td>
<td>0,0</td>
<td>1,0</td>
</tr>
<tr>
<td>$B$</td>
<td>1,0</td>
<td>0,1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Correlated Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$C$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>0</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>$M$</td>
<td>1/6</td>
<td>0</td>
<td>1/6</td>
</tr>
<tr>
<td>$B$</td>
<td>1/6</td>
<td>1/6</td>
<td>0</td>
</tr>
</tbody>
</table>
Shapley Game: No Internal Regret Learning

Frequencies

NIR polynomial \([p = 2]\) : Shapley Game

NIR exponential \([\eta = 0.014823]\) : Shapley Game
Shapley Game: No Internal Regret Learning

Joint Frequencies

NIR polynomial \([p = 2]\): Shapley Game

NIR exponential \([\eta = 0.014823]\): Shapley Game
Shapley Game: No External Regret Learning

Frequencies

NER polynomial \([p = 2]\): Shapley Game

NER exponential \([\eta = 0.014823]\): Shapley Game
Summary

- No-external- and no-internal-regret can be defined along one continuum, no-Φ-regret.

- No-Φ-regret learning algorithms exist, ∀Φ.

- No-Φ-regret learning converges to the set of Φ-equilibria, ∀Φ.

- No-internal-regret learning is the strongest form of no-Φ-regret learning. Therefore, Nash equilibrium cannot be learned via no-Φ-regret learning.
“A little rationality goes a long way” [Hart 03]

Regret Minimization vs. Utility Maximization

- RM is easy to implement.
- RM justifies randomness in actions.
- Can RM be used to explain human behavior?