# CSCE 631 — Algorithmic Game Theory

## Lecture 8: Regret Minimization II

Alan Kuhnle

Summer 2026

Texas A&M University

# Today's Plan

Regret Matching

Internal Regret and Swap Regret

From Regret to Equilibria

Playing Games Repeatedly

Summary

# Where We Left Off: From L7 to L8

**Recap of Lecture 7**

- A single learner facing adversarial losses can guarantee **no external regret**: MWU achieves $R_T \le 2\sqrt{T \ln n}$, so $R_T/T \to 0$.
- This controls each player's *marginal* time-average $\bar{x}_i = \frac{1}{T}\sum_t x_i^t$ — but not yet the *joint* play across players.

**Where we are headed today:**

- **Regret matching** — a parameter-free no-regret algorithm, the canonical choice for learning in games and the basis of CFR.
- **Stronger deviation classes** (internal / swap regret) and the payoff: no-external-regret $\Rightarrow$ CCE and no-swap-regret $\Rightarrow$ CE for the *joint* empirical distribution of play.

# Regret Matching

# Regret Matching Algorithm

**Regret Matching (Hart & Mas-Colell, 2000)**

1. Initialize: play uniformly $x^1 = (1/n, \ldots, 1/n)$.
2. After round $t$, compute cumulative regret for each action $i$:

$$R_i^t = \sum_{s=1}^{t} [\ell_{i_s}^s - \ell_i^s] = \sum_{s=1}^{t} [\langle x^s, \ell^s \rangle - \ell_i^s]$$

   (how much better action $i$ would have been).
3. At round $t+1$, play:

$$x_i^{t+1} = \frac{[R_i^t]^+}{\sum_j [R_j^t]^+} \qquad \text{where } [z]^+ = \max(z, 0)$$

   If all regrets are non-positive, play uniformly.

# Regret Matching: Properties

- Proportional to **positive regret**: actions that would have been better get more weight.
- Simpler than MWU: no learning rate parameter.
- **Regret bound:**

$$R_T \leq O\left(\sqrt{T\,n}\right)$$

- Slightly worse dependence on $n$ compared to MWU ($\sqrt{n}$ vs. $\sqrt{\ln n}$).
- But: very natural, easy to implement, widely used in practice (especially in poker AI).

# Regret Matching: Intuition

**Why does it work?**

- Actions with high cumulative regret get played more $\Rightarrow$ their future regret decreases.
- Actions with negative cumulative regret get played less (weight $= 0$).
- Self-correcting: the algorithm naturally balances exploration toward under-explored good actions.

**Connection to Fixed Points**

At a steady state where $R_i^t/t \to r_i$, the strategy $x_i \propto [r_i]^+$ is a **correlated equilibrium** strategy (Hart & Mas-Colell).

# Internal Regret and Swap Regret

# Beyond External Regret

**External regret** compares to the best *fixed* action.

**Stronger notions:**

- **Internal regret:** compares the performance of action $i$ (when played) to what would have happened if action $j$ was played *instead of $i$ every time $i$ was chosen*.
- **Swap regret:** the most general version — allows a different "swap" for each action simultaneously.

Hierarchy: swap regret $\geq$ internal regret $\geq$ external regret.

Each leads to convergence to a different equilibrium concept.

# Internal Regret: Definition

**Definition (Internal Regret)**

The **internal regret** for swapping action $i$ to action $j$ is:

$$R_{i\to j}^{T} = \sum_{t=1}^{T} x_i^t \left(\ell_i^t - \ell_j^t\right)$$

The **internal regret** is:

$$R_T^{\text{int}} = \max_{i,j\in[n]} R_{i\to j}^{T}$$

**Interpretation:** Had I replaced action $i$ with action $j$ *every time I played* $i$, how much would I have saved?

An algorithm has **no internal regret** if $R_T^{\text{int}}/T \to 0$.

# Swap Regret: Definition

**Definition (Swap Regret)**

A **modification function** is $\sigma : [n] \to [n]$. The **swap regret** is:

$$R_T^{\text{swap}} = \max_{\sigma:[n]\to[n]} \sum_{t=1}^{T} \sum_{i=1}^{n} x_i^t \left(\ell_i^t - \ell_{\sigma(i)}^t\right)$$

- $\sigma$ specifies: whenever action $i$ is recommended, play $\sigma(i)$ instead.
- External regret: $\sigma$ is a *constant* function ($\sigma(i) = j$ for all $i$).
- Internal regret: $\sigma$ differs from identity in at most one coordinate.
- Swap regret: $\sigma$ is *arbitrary*.

No swap regret $\Rightarrow$ no internal regret $\Rightarrow$ no external regret.

# Achieving No-Swap-Regret

**Theorem (Blum & Mansour, 2007)**

Given any no-external-regret algorithm $\mathcal{A}$ with regret bound $R_T^{\text{ext}}$, there exists a no-swap-regret algorithm with:

$$R_T^{\text{swap}} \leq n \cdot R_T^{\text{ext}}$$

**Construction:**

1. Run $n$ copies of $\mathcal{A}$: one for each action $i \in [n]$.
2. Copy $i$ outputs a distribution $q^{t,i} \in \Delta_n$ suggesting what to swap action $i$ to.
3. Combine: $x^t$ is the stationary distribution of the Markov chain with transition matrix $Q^t$ where row $i$ is $q^{t,i}$.
4. Feed appropriate losses to each copy.

# No-Swap-Regret: Regret Bound

Using MWU as the base algorithm with regret $O(\sqrt{T \ln n})$:

$$R_T^{\text{swap}} \leq n \cdot O(\sqrt{T \ln n}) = O(n\sqrt{T \ln n})$$

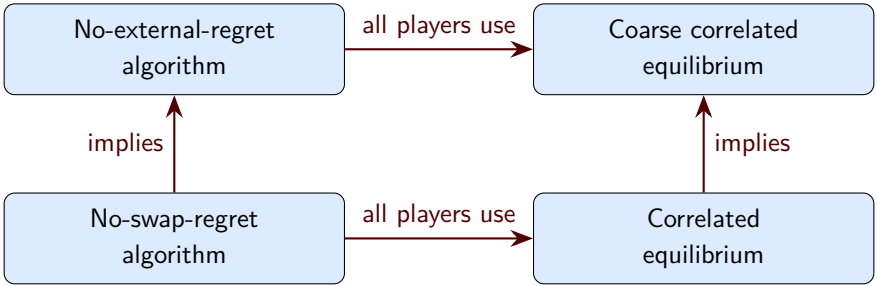
Per-round swap regret:

$$\frac{R_T^{\text{swap}}}{T} \leq O\left(\frac{n\sqrt{\ln n}}{\sqrt{T}}\right) \to 0$$

- Convergence rate is $O(1/\sqrt{T})$, but with a factor of $n$.
- For $n$ actions: need $T = \Omega(n^2 \ln n/\varepsilon^2)$ rounds for $\varepsilon$-swap-regret.
- Slower than external regret, but polynomial.

# From Regret to Equilibria

# The Key Connection



# Coarse Correlated Equilibrium (Recall)

**Definition (Coarse Correlated Equilibrium)**

A distribution $p$ over action profiles is a CCE if for every player $i$ and every action $a_i'$:

$$\mathbb{E}_{a \sim p}\big[u_i(a)\big] \;\geq\; \mathbb{E}_{a \sim p}\big[u_i(a_i', a_{-i})\big]$$

- No player benefits from *unconditionally* switching to a fixed action (before seeing the recommendation).
- Weaker than CE: in CE, you can condition on your recommendation.
- Every CE is a CCE; every NE is a CE.

# Theorem: No-External-Regret $\Rightarrow$ CCE

**Theorem**

In a repeated game, if every player $i$ uses a no-external-regret algorithm, then the empirical distribution of play converges to the set of coarse correlated equilibria.

**Proof sketch.** Let $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_i^t$ be player $i$'s empirical strategy.

For any fixed action $a_i'$ of player $i$:

$$\frac{1}{T}\sum_{t=1}^{T} u_i(x^t) - u_i(a_i', x_{-i}^t) \geq -\frac{R_T^i}{T} \to 0$$

This is exactly the CCE condition for the empirical distribution. $\square$

# Proof Details: Making It Precise

**Formal statement.** Define the empirical distribution over action profiles:

$$p^T(a) = \frac{1}{T} \sum_{t=1}^{T} \prod_{i=1}^{n} x_{i,a_i}^t$$

*Link to L7:* the marginals of $p^T$ are the L7 averages $\bar{x}_i = \frac{1}{T} \sum_t x_i^t$; $p^T$ also records the temporal correlation that CCE/CE exploit.

For player $i$ and deviation $a_i'$:

$$\mathbb{E}_{a \sim p^T}[u_i(a)] - \mathbb{E}_{a \sim p^T}[u_i(a_i', a_{-i})] = \frac{1}{T} \sum_{t=1}^{T} \left[ \langle x_i^t, u_i(\cdot, x_{-i}^t) \rangle - u_i(a_i', x_{-i}^t) \right]$$

$$\geq -\frac{R_T^i}{T}$$

by the definition of external regret.

As $T \to \infty$, the RHS $\to 0$, so $p^T$ approaches a CCE.

# Theorem: No-Swap-Regret $\Rightarrow$ CE

**Theorem**

In a repeated game, if every player $i$ uses a no-swap-regret algorithm, then the empirical distribution of play converges to the set of correlated equilibria.

**Proof sketch.** CE requires: no profitable deviation *conditional on each recommendation*.

For player $i$, the CE constraint for swapping $a_i \to a_i'$:

$$\sum_{a_{-i}} p(a_i, a_{-i})\big[u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})\big] \geq 0$$

This corresponds exactly to the *internal regret* term $R^T_{a_i \to a_i'}$.

No swap regret $\Rightarrow$ all such terms vanish $\Rightarrow$ CE. $\square$

# Convergence Rates

| Regret Type | Equilibrium | Rate |
|---|---|---|
| External regret $\leq \varepsilon T$ | CCE | $T = O(\ln n/\varepsilon^2)$ |
| Swap regret $\leq \varepsilon T$ | CE | $T = O(n^2 \ln n/\varepsilon^2)$ |

- CCE convergence is fast: $O(\log n)$ dependence on number of actions.
- CE convergence is slower but still polynomial.
- Both are **uncoupled** dynamics: each player uses only their own payoff information.
- Contrast: computing CE directly via LP requires full game knowledge.

# Playing Games Repeatedly

# The Repeated Game Model

**Setup**

- A fixed game $G$ with players $1, \ldots, n$ and action sets $A_1, \ldots, A_n$.
- At each round $t$: all players simultaneously choose actions.
- Each player observes their own utility (or the full action profile).
- Each player runs an online learning algorithm.

**Important:** The game is fixed, but the "losses" each player faces are determined by *other players' actions* — which change over time.

The adversarial regret bound still applies because it holds for *any* loss sequence.

# Why Adversarial Bounds Suffice

**Concern:** Other players are also adapting. Is the loss sequence really "adversarial"?

**Answer:** The regret bound for MWU (and similar algorithms) holds for *any* sequence of losses, including:

- Losses generated by an adaptive adversary.
- Losses generated by other learning agents.
- Losses generated by nature.

The bound is *oblivious to the loss-generating mechanism*.

This is the power of worst-case regret analysis: no assumptions needed.

# Formal Convergence to CCE

**Theorem**

Let $G$ be an $n$-player game. If each player $i$ uses MWU with $\eta = \sqrt{\ln |A_i| / T}$, then the empirical distribution of play $\bar{p}^T$ satisfies:

$$\max_i \max_{a_i' \in A_i} \left[ \mathbb{E}_{a \sim \bar{p}^T}[u_i(a_i', a_{-i})] - \mathbb{E}_{a \sim \bar{p}^T}[u_i(a)] \right] \leq 2\sqrt{\frac{\ln(\max_i |A_i|)}{T}}$$

After $T = O(\ln n / \varepsilon^2)$ rounds, the empirical play is an $\varepsilon$-CCE.

# Formal Convergence to CE

**Theorem**

Let $G$ be an $n$-player game. If each player $i$ uses the Blum–Mansour no-swap-regret algorithm, then the empirical distribution $\bar{p}^T$ is an $\varepsilon$-CE after

$$T = O\left(\frac{(\max_i |A_i|)^2 \cdot \ln(\max_i |A_i|)}{\varepsilon^2}\right) \text{ rounds.}$$

- Quadratic dependence on the number of actions (vs. logarithmic for CCE).
- Still **polynomial-time** and **uncoupled**.
- Each player only needs to know their own utility function.

## Implications for Equilibrium Computation

1. **Behavioral justification:** CE/CCE arise naturally from simple, myopic learning — not from coordinated reasoning.
2. **Computational:** running no-regret dynamics for $T = O(\text{poly}(n)/\varepsilon^2)$ rounds yields an $\varepsilon$-CE. Compare: direct LP has $\prod_i |A_i|$ variables.
3. **Decentralized:** each player acts independently using only local information.
4. **Robustness:** works even if some players are adversarial (remaining no-regret players still have bounded regret).

**Caveat**

These dynamics do **not** converge to Nash equilibria in general games. For NE convergence, we need zero-sum structure (next lecture).

# Non-Convergence to NE: Example

**Shapley's game** (3 actions each):

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

- Unique NE: $x^* = y^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
- Under many learning dynamics (including fictitious play): strategies **cycle** and do not converge to NE.
- The *empirical frequency* converges to CCE / CE, but *last-iterate* behavior cycles.

General games: last-iterate convergence to NE is **not guaranteed** by no-regret dynamics.

# Summary

## Summary

1. **Regret matching:** play proportional to positive cumulative regret; simple, practical, $O(\sqrt{Tn})$ regret.
2. **Internal regret:** regret for swapping one action to another whenever it was played.
3. **Swap regret:** regret for applying any modification function; strongest notion; achievable via $n$ copies of an external-regret algorithm.
4. **No-external-regret $\Rightarrow$ CCE:** if all players minimize external regret, empirical play converges to coarse correlated equilibrium.
5. **No-swap-regret $\Rightarrow$ CE:** if all players minimize swap regret, empirical play converges to correlated equilibrium.
6. **NE not guaranteed:** general games can cycle under learning dynamics.
7. **Next time:** zero-sum games, where regret minimization *does* yield Nash equilibria.