Algorithmic Game Theory | Summer 2015, Week 3

# No-Regret Dynamics

ETH Zürich | Paul Dütting

In settings where Nash equilibria are hard to compute, it is at least questionable that players will reach them. Fortunately, there are other, weaker equilibrium concepts that generalize Nash equilibria but are easy to compute. We will introduce a hierarchy of equilibrium concepts; and study natural learning dynamics that are required to meet the less stringent requirements of two of these equilibrium concepts.

## 1 Hierarchy of Equilibrium Concepts

We have already seen pure and mixed Nash equilibria in cost-minimization games, and observed that every pure Nash equilibrium is also a mixed Nash equilibrium.

**Definition 3.1.** *An* $\epsilon$-approximate correlated equilibrium *(or* $\epsilon$-correlated equilibrium*) of a cost-minimization game is a probability distribution* $p$ *on the set of strategy profiles* $S = \Pi_{i \in \mathcal{N}} S_i$ *such that for every* $i \in \mathcal{N}$*, every strategy* $s_i \in S_i$*, and every deviation* $s_i' \in S_i$ *we have*

$$\mathbf{E}_{s \sim p} \left[ C_i(s) \mid s_i \right] \leq \mathbf{E}_{s \sim p} \left[ C_i(s_i', s_{-i}) \mid s_i \right] + \epsilon \ .$$

*The case of* $\epsilon = 0$ *is called* correlated equilibrium*.*

Importantly, the distribution $p$ in the above definition need not be a product distribution. A correlated equilibrium protects agains conditional deviations of the form "whenever a player played $s_i$, he now plays $s_i'$."

**Definition 3.2.** *An* $\epsilon$-approximate coarse correlated equilibrium *(or* $\epsilon$-coarse correlated equilibrium*) of a cost-minimization game is a probability distribution* $p$ *on the set of strategy profiles* $S = \Pi_{i \in \mathcal{N}} S_i$ *such that for every* $i \in \mathcal{N}$ *and every deviation* $s_i' \in S_i$ *we have*

$$\mathbf{E}_{s \sim p} \left[ C_i(s) \right] \leq \mathbf{E}_{s \sim p} \left[ C_i(s_i', s_{-i}) \right] + \epsilon \ .$$

*The case of* $\epsilon = 0$ *is called* coarse correlated equilibrium*.*

A coarse correlated equilibrium differs from a correlated equilibrium in that it only protects against unconditional unilateral deviations.

One can show that every mixed Nash equilibrium is also a correlated equilibrium, and every correlated equilibrium is also a coarse correlated equilibrium. This leaves us with the following hierarchy of equilibrium concepts:

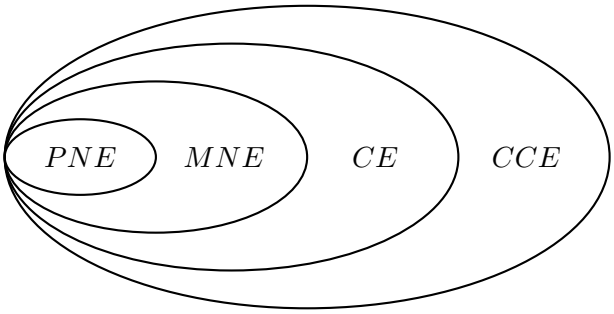



**Figure 1:** Venn diagram of equilibrium concepts.

We will first study the dynamics needed to reach a coarse correlated equilibrium. Afterwards we will give a black-box reduction for correlated equilibria.

## 2 Minimizing External Regret

Consider the following problem. There is a single player playing $T$ rounds against an adversary, trying to minimize his cost. In each round, the player chooses a probability distribution over $N$ strategies (also termed actions here). After the player has commited to a probability distribution, or mixed strategy as we will say, the adversary picks a cost vector fixing the cost for each of the $N$ strategies.

In round $t = 1, \ldots, T$, the following happens:

- The player picks a probability distribution $p^t = (p_1^t, \ldots, p_N^t)$ over his strategies.
- The adversary picks a cost vector $\ell^t = (\ell_1^t, \ldots, \ell_N^t)$, where $\ell_i^t \in [0, 1]$ for all $i$.
- A strategy $a^t$ is chosen according to the probability distribution $p^t$. The player incurs this strategy's cost and gets to know the entire cost vector.

What is the right benchmark for an algorithm in this setting? The *best action sequence in hindsight* achieves a cost of $\sum_{t=1}^{T} \min_{i \in [N]} \ell_i^t$. However, getting close to this number is generally hopeless as the following example shows.

**Example 3.3.** *Suppose $N = 2$ and consider an adversary that chooses $\ell^t = (1, 0)$ if $p_1^t \geq 1/2$ and $\ell^t = (0, 1)$ otherwise. Then the expected cost of the player is at least $T/2$, while the best action sequence in hindsight has cost* 0.

Instead, we will swap the sum and the minimum, and compare to $L_{\min}^T = \min_{i \in [N]} \sum_{t=1}^{T} \ell_i^t$. That is, instead of comparing to the best action sequence in hindsight, we compare to the *best fixed action in hindsight*.

The expected cost of some algorithm $\mathcal{A}$ that uses probability distributions $p^1, \ldots, p^T$ against cost vectors $\ell^1, \ldots, \ell_T$ is given as $L_{\mathcal{A}}^T = \sum_{t=1}^{T} \sum_{i=1}^{N} p_i^t \ell_i^t$. The difference of this cost and the cost of the best single strategy in hindsight is called *external regret*.

**Definition 3.4.** *The* external regret *of algorithm $\mathcal{A}$ is defined as $R_{\mathcal{A}}^T = L_{\mathcal{A}}^T - L_{min}^T$.*

**Definition 3.5.** *An algorithm is called* no-external-regret algorithm *if for any adversary and all $T$ we have $R_{\mathcal{A}}^T = o(T)$.*

This means that the *average* cost per round of a no-external-regret algorithm approaches the one of the best fixed strategy in hindsight or even beats it.

The next two examples show that there can be no deterministic no-external-regret algorithm, and provide a lower bound on the speed of convergence.

**Example 3.6** (Randomization is necessary)**.** *Suppose there are $N \geq 2$ actions. In each round $t$ the algorithm commits to a strategy $i$. The adversary can set $\ell_i^t = 1$ and $\ell_j^t = 0$ for $j \neq i$. The total cost of the algorithm will be $T$, while the cost of the best fixed action in hindsight is at most $T/N$.*

**Example 3.7** (Lower bound on speed of convergence)**.** *Consider the case where $N = 2$ and an adversary that, independently in each round $t$, chooses uniformly at random between $\ell^t = (1, 0)$ and $\ell^t = (0, 1)$. The expected cost of any algorithm will be exactly $T/2$. Due to a standard deviation of $\Theta(\sqrt{T})$, however, the best fixed strategy in hindsight has expected cost $T/2 - \Theta(\sqrt{T})$. So the external regret will be $\Theta(\sqrt{T})$.*

A similar argument shows that for $N \geq 2$ the external regret achievable by any algorithm will be at least $\Theta(\sqrt{T \ln N})$.

## 2.1 The Multiplicative-Weights Algorithm

By the definition it is not even clear that there are no-external-regret algorithms. Fortunately, there are. In this section, we will get to know the *multiplicative-weights algorithm* (also known as randomized weighted majority or hedge).

The algorithm maintains weights $w_i^t$, which are proportional to the probability that strategy $i$ will be used in round $t$. After each round, the weights are updated by a multiplicative factor, which depends on the cost in the current round.

Let $\eta \in (0, \frac{1}{2}]$; we will choose $\eta$ later.

- Initially, set $w_i^1 = 1$, for every $i \in [N]$.
- At every time $t$,
  - Let $W^t = \sum_{i=1}^N w_i^t$;
  - Choose strategy $i$ with probability $p_i^t = w_i^t / W^t$;
  - Set $w_i^{t+1} = w_i^t \cdot (1-\eta)^{\ell_i^t}$.

Let's build up some intuition for what this algorithm does. First suppose $\ell_i^t \in \{0,1\}$. Strategies with cost 0 maintain their weight, while the weight of strategies with cost 1 is multiplied by $(1-\eta)$. So the weight decays exponentially quickly in the number of 1's. Next consider the impact of $\eta$. Setting $\eta$ to zero means that we pick a strategy uniformly at random and continue to do so, on the other hand the higher $\eta$ the more we punish strategies which incurred a high cost. So we can think of $\eta$ as controlling the tradeoff between exploration (small $\eta$) and exploitation (large $\eta$).

**Theorem 3.8** (Littlestone and Warmuth, 1994)**.** *The multiplicative-weights algorithm, for any sequence of cost vectors from* $[0,1]$*, guarantees*

$$L_{MW}^T \leq (1+\eta) L_{min}^T + \frac{\ln N}{\eta} \ .$$

Setting $\eta = \sqrt{\frac{\ln N}{T}}$ yields

$$L_{\text{MW}}^T \leq L_{\text{min}}^T + 2\sqrt{T \ln N} \ .$$

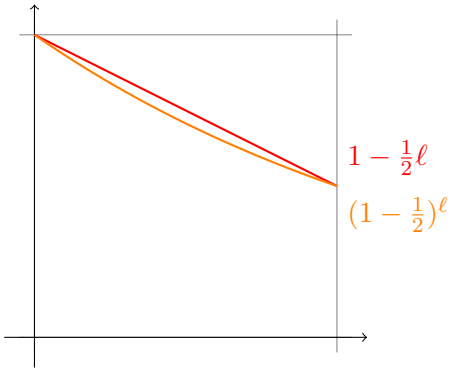
**Corollary 3.9.** *The multiplicative-weights algorithm with* $\eta = \sqrt{\frac{\ln N}{T}}$ *has external regret at most* $2\sqrt{T \ln N} = o(T)$ *and hence is a no-external-regret algorithm.*

Notice that this matches the above lower bound.

*Proof.* Let us analyze how the sum of weights $W^t$ decreases over time. It holds

$$W^{t+1} \ = \ \sum_{i=1}^N w_i^{t+1} \ = \ \sum_{i=1}^N w_i^t (1-\eta)^{\ell_i^t} \ .$$

Observe that $(1-\eta)^\ell \ = \ (1-\ell\eta)$, for both $\ell = 0$ and $\ell = 1$. Furthermore, $(1-\eta)^\ell$ is a convex function in $\ell$. For $\ell \in [0,1]$ this implies $(1-\eta)^\ell \ \leq \ (1-\ell\eta)$.




This gives us

$$W^{t+1} \;\leq\; \sum_{i=1}^{N} w_i^t(1 - \ell_i^t\eta) \;=\; W^t - \eta\sum_{i=1}^{N} w_i^t\ell_i^t \;.$$

Let $\ell^t$ denote the expected cost of MW in step $t$. It holds $\ell^t = \sum_{i=1}^{N} \ell_i^t w_i^t/W^t$. Substituting this into the bound for $W^{t+1}$ gives

$$W^{t+1} \;\leq\; W^t - \eta\ell^t W^t \;=\; W^t(1 - \eta\ell^t) \;.$$

As a consequence,

$$W^{T+1} \;\leq\; W^1\prod_{t=1}^{T}(1 - \eta\ell^t) \;=\; N\prod_{t=1}^{T}(1 - \eta\ell^t) \;.$$

The sum of weights after step $T$ can be upper bounded in terms of the expected costs of MW. On the other hand, the sum of weights after step $T$ can be lower bounded in terms of the costs of the best strategy as follows:

$$W^{T+1} \;\geq\; \max_{1\leq i\leq N}(w_i^{T+1}) \;=\; \max_{1\leq i\leq N}\left(w_i^1\prod_{t=1}^{T}(1-\eta)^{\ell_i^t}\right) \;=\; \max_{1\leq i\leq N}\left((1-\eta)^{\sum_{t=1}^{T}\ell_i^t}\right) \;=\; (1-\eta)^{L_{\min}^T} \;.$$

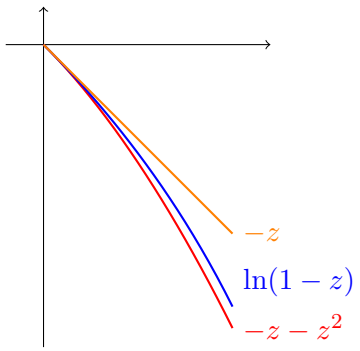
Combining the bounds and taking the logarithm on both sides gives us

$$L_{\min}^T\ln(1-\eta) \;\leq\; (\ln N) + \sum_{t=1}^{T}\ln(1-\eta\ell^t) \;.$$

In order to simplify, we will now use the following estimation

$$-z - z^2 \;\leq\; \ln(1-z) \;\leq\; -z \;,$$

which holds for every $z \in [0, \frac{1}{2}]$.



This gives us

$$\begin{aligned} L^T_{\min}(-\eta - \eta^2) &\leq (\ln N) + \sum_{t=1}^{T}(-\eta \ell^t) \\ &= (\ln N) - \eta L^T_{\mathrm{MW}} \ . \end{aligned}$$

Finally, solving for $L^T_{\mathrm{MW}}$ gives

$$L^T_{\mathrm{MW}} \leq (1+\eta) L^T_{\min} + \frac{\ln N}{\eta} \ . \quad \square$$

# 3 Connection to Coarse Correlated Equilibria

Let us now connect this back to cost-minimization games. For this fix a cost-minimization game. Without loss of generality, assume that all costs are in $[0, 1]$. We consider *no-external-regret dynamics* defined as follows.

In each time step $t = 1, \ldots, T$:

1. Each player $i$ simultaneously and independently chooses a mixed strategy $\sigma_i^t$ using a no-external-regret algorithm.
2. Each player $i$ receives a cost vector $c_i^t$, where $c_i^t(s_i)$ is the expected cost of strategy $s_i$ when the other players play their chosen mixed strategies. That is, $c_i^t(s_i) = \mathbf{E}_{s_{-i} \sim \sigma_{-i}}[C_i(s_i, s_{-i})]$.

Do such dynamics converge to Nash equilibria? Not necessarily. However, "on average" the players play according to an approximate coarse correlated equilibrium.

**Proposition 3.10.** *Let $\sigma^1, \ldots, \sigma^T$ be generated by no-external-regret dynamics such that each player's external regret is at most $\epsilon T$. Let $p$ be the probability distribution that first selects a single $t \in [T]$ uniformly at random and then chooses for every $i \in \mathcal{N}$ one $s_i$ according to $\sigma_i^t$. Then $p$ is an $\epsilon$-coarse correlated equilibrium.*

*Proof.* By definition, for each player $i$,

$$\mathbf{E}_{s \sim \sigma}[C_i(s)] - \mathbf{E}_{s \sim \sigma}[C_i(s'_i, s_{-i})] = \frac{1}{T} \sum_{t=1}^{T} \left( \mathbf{E}_{s \sim \sigma^t}[C_i(s)] - \mathbf{E}_{s \sim \sigma^t}[C_i(s'_i, s_{-i})] \right) \leq \epsilon.$$

where the inequality follows by observing that the first term in the summation is the expected cost achieved by the regret-minimization algorithm and the second term is bounded by the cost achieved by the best fixed cost in hindsight. $\square$

Notice that a player that uses the multiplicative-weights algorithm needs only $O(\frac{\ln N}{\epsilon^2})$ iterations to achieve the required bound on the external regret.

# 4 Swap Regret and Correlated Equilibria

Let's be even more ambitious and ask whether there is an analogue of the theory that we just developed for coarse correlated equilibria that applies to correlated equilibria.

For this consider the same dynamics as in the case of external regret, but a different notion of regret. As before denote the expected cost of some algorithm $\mathcal{A}$ that uses probability distributions $p^1, \ldots, p^T$ against cost vectors $\ell^1, \ldots, \ell^t$ by $L^T_{\mathcal{A}} = \sum_{t=1}^{T} \sum_{i=1}^{N} p_i^t \ell_i^t$.

A *switching function* is a function $\delta : [N] \to [N]$. The expected cost under a fixed switching function $\delta$ is $L^T_\delta = \sum_{t=1}^{T} \sum_{i=1}^{N} p_i^t \ell^t_{\delta(i)}$. Let $L^T_{swap} = \min_\delta L^T_\delta$.

**Definition 3.11.** *The* swap regret *of an algorithm $\mathcal{A}$ is defined as $R^T_{swap} = L^T_{\mathcal{A}} - L^T_{swap}$.*

**Definition 3.12.** *An algorithm is called* no-swap-regret algorithm *if for any adversary and all $T$ we have $R^T_{swap} = o(T)$.*

It is not difficult to show a result analogous to Proposition 3.10 for swap regret and correlated equilibria; so all that we are left with is to show that no-swap-regret algorithms do exist.

## 4.1 A Black-Box Reduction

It turns out that there is a surprisingly clean and simple reduction from the problem of finding a no-swap-regret algorithm to that of finding a no-external-regret algorithm.

**Theorem 3.13** (Blum and Mansour, 2007)**.** *If there is a no-external-regret algorithm, then there is a no-swap-regret algorithm.*

*Proof.* We construct the algorithm for swap regret from $N$ algorithms for external regret. The basic idea is to have one algorithm $\mathcal{A}^i$ for each action $a_i$, which is responsible for protecting against profitable deviations from action $a_i$ to some other action.

At time $t = 1, \dots, N$:

1. Receive distributions $q^{1,t}, \dots, q^{N,t}$ over actions from algorithms $\mathcal{A}^1, \dots, \mathcal{A}^N$.
2. Compute and output a *consensus distribution* $p^t$.
3. Receive a cost vector $\ell^t$ from the adversary.
4. Give algorithm $\mathcal{A}^i$ the cost vector $p^t_i \ell^t$.

We will leave the definition of the consensus distribution and how to compute it open for now; instead we will re-engineer how this distribution needs to look like.

Let's first take the perspective of the no-swap-regret algorithm. The expected cost of the no-swap-regret algorithm is

$$\sum_{t=1}^{T} \sum_{j=1}^{N} p^t_j \ell^t_j. \tag{1}$$

The expected cost under a fixed switching function $\delta$ is

$$\sum_{t=1}^{T} \sum_{j=1}^{N} p^t_j \ell^t_{\delta(j)}. \tag{2}$$

Our goal is to show that for every switching function "(1)" $\leq$ "(2)" $+ o(T)$.

Let's switch to the perspective of the no-external regret algorithm $\mathcal{A}^i$. For any fixed strategy $k$ we know that

$$\sum_{t=1}^{T} \sum_{j=1}^{N} q^{i,t}_j (p^t_i \ell^t_j) \leq \sum_{t=1}^{T} p^t_i \ell^t_k + o(T). \tag{3}$$

The left-hand side is the expected cost of algorithm $\mathcal{A}^i$, and it is at most the right-hand side because we assumed $\mathcal{A}^i$ to be a no-external-regret algorithm.

Now fix a switching function $\delta$ and sum inequality (3) over all $i$ with $k$ instantiated as $\delta(i)$. Then,

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N} q^{i,t}_j (p^t_i \ell^t_j) \leq \sum_{t=1}^{T} \sum_{i=1}^{N} p^t_i \ell^t_{\delta(i)} + o(T). \tag{4}$$

Note that the right-hand side is identical to equality (2), up to the additive error term $o(T)$. So it remains to link the left-hand side to equality (1). Specifically, we would like to argue that

$$p_j^t = \sum_{i=1}^{N} q_j^{i,t} p_i^t, \tag{5}$$

in which case the two would be indentical and we would be done.

In fact, equation (5) looks familiar. It looks like an equation defining a stationary distribution of a Markov chain. This is the key insight, and the basic idea behind the definition of the consensus distribution.

Namely, from algorithms $\mathcal{A}^1, \ldots, \mathcal{A}^N$ at time $t$ we will construct the following Markov chain:

- The set of states is $A = \{1, \ldots, N\}$.
- For every $i, j \in A$ the transition probability from $i$ to $j$ is $q_j^{i,t}$.

Then $p^t$ satisfies (5) if and only if it is the stationary distribution of this Markov chain. At least one such distribution exists and can be computed via an eigenvector computation. This completes the proof. □

## Recommended Literature


- A. Blum and Y. Mansour. Learning, Regret Minimization, and Equilibria. In: Algorithmic Game Theory, N. Nisan et al., pages 79–101, 2007. (General reference)
- Tim Roughgarden's lecture notes, Chapters 17 and 18, `http://theory.stanford.edu/~tim/f13/f13.pdf` (General reference)
- N. Littlestone, M. Warmuth. The Weighted Majority Algorithm. Information and Computation, 108(2):212–261, 1994. (The external regret result)
- A. Blum and Y. Mansour. From external to internal regret. Journal of Machine Learning Research, 8:13071324, 2007. (Reduction from swap to external regret)
- G. H. Golub and C. F. van Loan. Matrix Computations, 4th edition. Johns Hopkins University Press, 2012. (Existence and poly-time computation of a stationary distribution of a Markov chain)