15-888 Computational Game Solving (Fall 2021) Thu, September 16th 2021

Lecture 4

Blackwell approachability and regret minimization on simplex domains

Instructor: Gabriele Farina*

In Lecture 3 we have seen how no-external-regret dynamics converge to several solution concepts of interest, including Nash equilibria in two-player zero-sum games, normal-form coarse-correlated equilibria in multiplayer general-sum games, and more generally convex-concave saddle point problems. Furthermore, we have seen that no-external-regret dynamics are a fundamental building block for constructing no-Φ-regret dynamics for general sets Φ of deviation functions.

In this lecture, we start exploring how no-external-regret dynamics have been historically constructed for action spaces of one-shot decision making problems (that is, probability simplexes). The historical construction we will present today is not only very elegant, but also it leads to an algorithm that is still extremely popular in practice today, called *regret matching*.

We will see other, more modern, techniques to construct no-external-regret dynamics for general convex and compact domains later on in this course.

1 No-external-regret dynamics for probability simplexes

In this lecture we focus on the task of constructing no-external-regret dynamics for one-shot decision making problems, that is, from minimizing regret on a simplex domain

$$\Delta^n = \big\{(x_1, \dots, x_n) \in \mathbb{R}^n_{\geq 0} : x_1 + \cdots + x_n = 1\big\}.$$

As we will see in the next lecture, it will be later possible to "upgrade" any regret minimizer for one-shot decision making to a regret minimizer for tree-form decision making.

Remember that constructing a regret minimizer for Δ^n means that we need to build a mathematical object that supports two operations:

- NextElement has the effect that the regret minimizer will output an element $\boldsymbol{x}^t \in \Delta^n$;
- ObserveUtility($\boldsymbol{\ell}^t$) provides the environment's feedback to the regret minimizer, in the form of a linear utility vector $\boldsymbol{\ell}^t \in \mathbb{R}^n$ that evaluates how good the last-output point.

Our goal will be to make sure that the (external) *regret*

$$R^T \coloneqq \max_{\hat{\boldsymbol{x}} \in \Delta^n} \left\{ \sum_{t=1}^T \Big(\langle \boldsymbol{\ell}^t, \hat{\boldsymbol{x}} \rangle - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \Big) \right\}.$$

grows sublinearly in T no matter the utility vectors $\boldsymbol{\ell}^t$ chosen by the environment.

We will construct a regret minimizer for Δ^n by means of an ancient (and very elegant!) construction, called *Blackwell approachability*. Blackwell approachability is a precursor of the theory of regret minimization, and played a fundamental role in the historical development of several efficient online optimization methods.

*Computer Science Department, Carnegie Mellon University. ✉ gfarina@cs.cmu.edu.

In particular, as we will show in a minute, the problem of minimizing regret on a simplex can be rewritten as a Blackwell approachability game. The solution of the Blackwell approachability game is a regret-minimizing dynamic called *regret matching* (RM). As of today, regret matching and its variants are still often some of the most practical algorithms for learning in games.

1.1 Blackwell approachability game

Blackwell approachability generalizes the problem of playing a repeated two-player game to games whose utilites are vectors instead of scalars.

Definition 1.1. A Blackwell approachability game is a tuple $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}, S)$, where $\mathcal{X}, \mathcal{Y}$ are closed convex sets, $\boldsymbol{u} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ is a biaffine function, and $S \subseteq \mathbb{R}^d$ is a closed and convex *target set*. A Blackwell approachability game represents a vector-valued repeated game between two players. At each time t, the two payers interact in this order:

- first, Player 1 selects an action $\boldsymbol{x}^t \in \mathcal{X}$;
- then, Player 2 selects an action $\boldsymbol{y}^t \in \mathcal{Y}$, which can depend adversarially on all the $\boldsymbol{x}^t$ output so far;
- finally, Player 1 incurs the vector-valued payoff $\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \mathbb{R}^d$, where $\boldsymbol{u}$ is a biaffine function.

Player 1's objective is to guarantee that the average payoff converges to the target set S. Formally, given target set $S \subseteq \mathbb{R}^d$, Player 1's goal is to pick actions $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots \in \mathcal{X}$ such that no matter the actions $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots \in \mathcal{Y}$ played by Player 2,

$$\min_{\hat{\boldsymbol{s}} \in S} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \right\|_2 \to 0 \quad \text{as} \quad T \to \infty. \tag{1}$$

1.2 External regret minimization on the simplex via Blackwell approachability

Hart and Mas-Colell [2000] noted that the construction of a regret minimizer for a simplex domain Δ^n can be reduced to constructing an algorithm for a particular Blackwell approachability game $\Gamma \coloneqq (\Delta^n, \mathbb{R}^n, \boldsymbol{u}, \mathbb{R}^n_{\leq 0})$ which we now describe. For all $i \in \{1, \ldots, n\}$, the i-th component of the vector-valued payoff function $\boldsymbol{u}$ measures the change in regret incurred at time t, compared to always playing the i-th vertex $\boldsymbol{e}_i$ of the simplex. Formally, $\boldsymbol{u} : \Delta^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) = \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}, \tag{2}$$

where $\mathbf{1}$ is the n-dimensional vector whose components are all 1.

The following lemma establishes an important link between Blackwell approachability on Γ and external regret minimization on the simplex Δ^n.

Lemma 1.1. The regret $R^T = \max_{\hat{\boldsymbol{x}} \in \Delta^n} \frac{1}{T} \sum_{t=1}^{T} \langle \boldsymbol{\ell}^t, \hat{\boldsymbol{x}} - \boldsymbol{x}^t \rangle$ cumulated up to any time T by any sequence of decisions $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^T \in \Delta^n$ is related to the distance of the average Blackwell payoff from the target cone $\mathbb{R}^n_{\leq 0}$ as

$$\frac{R^T}{T} \leq \min_{\hat{\boldsymbol{s}} \in \mathbb{R}^n_{\leq 0}} \left\| \hat{\boldsymbol{s}} - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \right\|_2. \tag{3}$$

So, a strategy for the Blackwell approachability game Γ is a regret-minimizing strategy for the simplex

domain Δ^n.

Proof. For any $\hat{\boldsymbol{x}} \in \Delta^n$, the regret cumulated compared to always playing $\hat{\boldsymbol{x}}$ satisfies

$$\begin{aligned}\frac{1}{T}R^T(\hat{\boldsymbol{x}}) &\coloneqq \frac{1}{T}\sum_{t=1}^{T}\Big(\langle \boldsymbol{\ell}^t, \hat{\boldsymbol{x}}\rangle - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t\rangle\Big) = \frac{1}{T}\sum_{t=1}^{T}\Big(\langle \boldsymbol{\ell}^t, \hat{\boldsymbol{x}}\rangle - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t\rangle\langle \mathbf{1}, \hat{\boldsymbol{x}}\rangle\Big) \\ &= \left\langle \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t\rangle \mathbf{1}, \hat{\boldsymbol{x}}\right\rangle = \left\langle \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t), \hat{\boldsymbol{x}}\right\rangle \\ &= \min_{\hat{\boldsymbol{s}}\in\mathbb{R}^n_{\le 0}}\left\langle -\hat{\boldsymbol{s}} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t), \hat{\boldsymbol{x}}\right\rangle, \end{aligned} \tag{4}$$

where we used the fact that $\hat{\boldsymbol{x}} \in \Delta^n$ in the second equality, and that $\min_{\hat{\boldsymbol{s}}\in\mathbb{R}^n_{\le 0}}\langle -\hat{\boldsymbol{s}}, \hat{\boldsymbol{x}}\rangle = 0$ since $\hat{\boldsymbol{x}} \ge \mathbf{0}$. Applying the Cauchy-Schwarz inequality to the right-hand side of (4), we obtain

$$\frac{1}{T}R^T(\hat{\boldsymbol{x}}) \le \min_{\hat{\boldsymbol{s}}\in\mathbb{R}^n_{\le 0}}\left\| -\hat{\boldsymbol{s}} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\right\|_2 \|\hat{\boldsymbol{x}}\|_2.$$

So, using the fact that $\|\hat{\boldsymbol{x}}\|_2 \le 1$ for any $\hat{\boldsymbol{x}} \in \Delta^n$,

$$\frac{1}{T}R^T(\hat{\boldsymbol{x}}) \le \min_{\hat{\boldsymbol{s}}\in\mathbb{R}^n_{\le 0}}\left\| -\hat{\boldsymbol{s}} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\right\|_2.$$

Taking a max over $\hat{\boldsymbol{x}} \in \Delta^n$ yields the statement. □

1.3 Solving Blackwell games: Blackwell's algorithm

A central concept in the theory of Blackwell approachability is the following.

Definition 1.2 (Forceable halfspace)**.** Let $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}, S)$ be a Blackwell approachability game and let $\mathcal{H} \subseteq \mathbb{R}^d$ be a halfspace, that is, a set of the form $\mathcal{H} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{a}^\top \boldsymbol{x} \le b\}$ for some $\boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}$. The halfspace $\mathcal{H}$ is said to be *forceable* if there exists a strategy of Player 1 that guarantees that the payoff is in $\mathcal{H}$ no matter the actions played by Player 2, that is, if there exists $\boldsymbol{x}^* \in \mathcal{X}$ such that

$$\boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{y}) \in \mathcal{H} \qquad \forall\, \boldsymbol{y} \in \mathcal{Y}.$$

When that is the case, we call action $\boldsymbol{x}^*$ a *forcing action* for $\mathcal{H}$.

Blackwell's *approachability theorem* [Blackwell, 1956] states the following.

Theorem 1.1 (Blackwell's theorem)**.** Goal (1) can be attained if and only if every halfspace $H \supseteq S$ is forceable.

We constructively prove the direction that shows how forceability translates into a sequence of strategies that guarantees that goal (1) is attained. Let $(\mathcal{X}, \mathcal{Y}, \boldsymbol{u}, S)$ be the Blackwell game. The method is pretty simple: at each time step $t = 1, 2, \dots$ operate the following:

1. Compute the average payoff received so far, that is, $\boldsymbol{\phi}^t = \frac{1}{t}\sum_{\tau=1}^{t-1}\boldsymbol{u}(\boldsymbol{x}^\tau, \boldsymbol{y}^\tau)$.

2. Compute the Euclidean projection $\boldsymbol{\psi}^t$ of $\boldsymbol{\phi}^t$ onto the target set S.

3. If $\boldsymbol{\phi}^t \in S$ (that is, goal (1) has already been met), pick and play any $\boldsymbol{x}^t \in \mathcal{X}$, observe the opponent's action $\boldsymbol{y}^t$, and return.

4. Else, consider the halfspace $\mathcal{H}^t$ tangent to S at the projection point $\boldsymbol{\psi}^t$, that contains S. In symbols,

$$\mathcal{H}^t \coloneqq \{\boldsymbol{z} \in \mathbb{R}^d : (\boldsymbol{\phi}^t - \boldsymbol{\psi}^t)^\top \boldsymbol{z} \le (\boldsymbol{\phi}^t - \boldsymbol{\psi}^t)^\top \boldsymbol{\psi}^t\}.$$

5. By hypothesis, $\mathcal{H}^t$ is forceable. Pick $\boldsymbol{x}^t$ to be a forcing action for $\mathcal{H}^t$, observe the opponent's action $\boldsymbol{y}^t$, and return.

The above method is summarized in Figure 1.

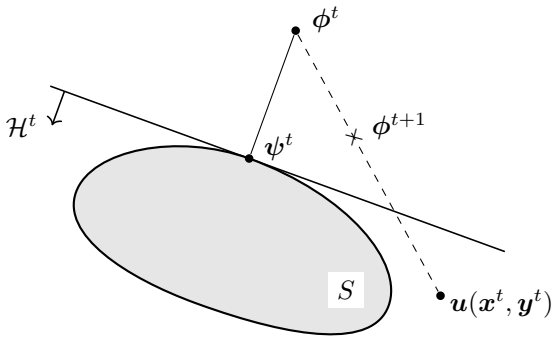

Figure 1: Construction of the approachability strategy described in Section 1.3.

Let's see how the average payoff $\boldsymbol{\phi}^t$ changes when we play as described above. Clearly,

$$\boldsymbol{\phi}^{t+1} = \frac{1}{t}\sum_{\tau=1}^{t} \boldsymbol{u}(\boldsymbol{x}^\tau, \boldsymbol{y}^\tau) = \frac{t-1}{t}\boldsymbol{\phi}^t + \frac{1}{t}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t).$$

Hence, denoting with ρ^{t+1} the squared Euclidean distance between $\boldsymbol{\phi}^{t+1}$ and the target set, that is,

$$\rho^t \coloneqq \min_{\hat{\boldsymbol{s}} \in S} \left\|\hat{\boldsymbol{s}} - \boldsymbol{\phi}^t\right\|_2^2,$$

we have

$$\begin{aligned}
\rho^{t+1} &\le \left\|\psi^t - \boldsymbol{\phi}^{t+1}\right\|_2^2 = \left\|\boldsymbol{\psi}^t - \frac{t-1}{t}\boldsymbol{\phi}^t - \frac{1}{t}\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\right\|_2^2 \\
&= \left\|\frac{t-1}{t}\left(\boldsymbol{\psi}^t - \boldsymbol{\phi}^t\right) + \frac{1}{t}\left(\boldsymbol{\psi}^t - \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\right)\right\|_2^2 \\
&= \frac{(t-1)^2}{t^2}\rho^t + \frac{1}{t^2}\left\|\boldsymbol{\psi}^t - \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\right\|_2^2 + \frac{2(t-1)}{t^2}\left\langle \boldsymbol{\psi}^t - \boldsymbol{\phi}^t, \boldsymbol{\psi}^t - \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\right\rangle. \qquad (5)
\end{aligned}$$

The proof so far does not use any particular assumption about how $\boldsymbol{x}^t$ is picked. Here is where that enters the picture. If $\boldsymbol{\phi}^t \in S$, then $\boldsymbol{\psi}^t = \boldsymbol{\phi}^t$ and therefore the last inner product is equal to 0. Otherwise, we have that $\boldsymbol{\psi}^t - \boldsymbol{\phi}^t \neq 0$. In that case, $\boldsymbol{x}^t$ is constructed by forcing the halfspace $\mathcal{H}^t$, and therefore, no matter how $\boldsymbol{y}^t$ is picked by the opponent we have

$$\begin{aligned}
\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \in \mathcal{H}^t \quad &\iff \quad (\boldsymbol{\phi}^t - \boldsymbol{\psi}^t)^\top \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \ge (\boldsymbol{\phi}^t - \boldsymbol{\psi}^t)^\top \boldsymbol{\psi}^t \\
&\iff \quad \left\langle \boldsymbol{\psi}^t - \boldsymbol{\phi}^t, \boldsymbol{\psi}^t - \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\right\rangle \le 0.
\end{aligned}$$

Plugging in the last inequality into (5) and bounding $\|\boldsymbol{\psi}^t - \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t)\|_2^2 \le \Omega^2$ where Ω^2 is a diameter parameter of the game (which only depends on $\boldsymbol{u}$ and S), we obtain

$$\rho^{t+1} \le \frac{(t-1)^2}{t^2}\rho^t + \frac{\Omega^2}{t^2} \quad \implies \quad t^2\rho^{t+1} - (t-1)^2\rho^t \le \Omega^2 \qquad \forall\, t = 1, 2, \dots.$$

Summing the inequality above for $t = 0, \dots, T-1$ and removing the telescoping terms, we obtain

$$T^2\rho^{T+1} \le T\Omega^2 \quad \implies \quad \rho^{T+1} \le \frac{\Omega^2}{T} \quad \implies \quad \min_{\hat{\boldsymbol{s}} \in S} \left\| \hat{\boldsymbol{s}} - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{y}^t) \right\|_2 \le \frac{\Omega}{\sqrt{T}}, \tag{6}$$

which implies that the average payoff in the Blackwell game converges to S at a rate of $O(1/\sqrt{T})$.

1.4 The regret matching (RM) algorithm

At this point we have all the ingredients necessary to construct the first regret minimizer for simplex domains of this course: the regret matching (RM) algorithm.

First, recall from Section 1.2 that the external regret minimization on the simplex can be solved via the Blackwell game $\Gamma \coloneqq (\Delta^n, \mathbb{R}^n, \boldsymbol{u}, \mathbb{R}^n_{\le 0})$ where $\boldsymbol{u} : \Delta^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) = \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}, \tag{7}$$

where $\mathbf{1}$ is the n-dimensional vector whose components are all 1. We will solve this Blackwell approachability game using the strategy explained in Section 1.3.

Computation of $\boldsymbol{\psi}^t$ (Step 2). Let's start from looking at how to compute the projection $\boldsymbol{\psi}^t$ of $\boldsymbol{\phi}^t$ onto $S = \mathbb{R}_{\le 0}$. Projection onto the nonpositive orthant amounts to a component-wise minimum with 0, that is, $\boldsymbol{\psi}^t = [\boldsymbol{\phi}^t]^-$. Hence,

$$\boldsymbol{\phi}^t - \boldsymbol{\psi}^t = [\boldsymbol{\phi}^t]^+ \quad \implies \quad (\boldsymbol{\phi}^t - \boldsymbol{\psi}^t)^\top \boldsymbol{\psi}^t = 0.$$

Halfspace to be forced (Step 4). Following on with Blackwell's algorithm, when $[\boldsymbol{\phi}^t]^+ \neq \mathbf{0}$, the halfspace to be forced at each time t is

$$\mathcal{H}^t \coloneqq \{\boldsymbol{z} \in \mathbb{R}^n : \langle [\boldsymbol{\phi}^t]^+, \boldsymbol{z} \rangle \le 0\}.$$

Forcing action for $\mathcal{H}^t$ (Step 5). We now show that a forcing action for $\mathcal{H}^t$ indeed exists. Remember that by definition, that is an action $\boldsymbol{x}^* \in \Delta^n$ such that no matter the $\boldsymbol{\ell} \in \mathbb{R}^n$, $\boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{\ell}) \in \mathcal{H}^t$. Expanding the definition of $\mathcal{H}^t$ and $\boldsymbol{u}$, we are looking for a $\boldsymbol{x}^* \in \Delta^n$ such that

$$\begin{aligned}
\langle [\boldsymbol{\phi}^t], \boldsymbol{\ell} - \langle \boldsymbol{\ell}, \boldsymbol{x}^* \rangle \mathbf{1} \rangle \le 0 \quad \forall\, \boldsymbol{\ell} \in \mathbb{R}^n \quad &\iff \quad \langle [\boldsymbol{\phi}^t], \boldsymbol{\ell} \rangle - \langle \boldsymbol{\ell}, \boldsymbol{x}^* \rangle \langle [\boldsymbol{\phi}^t]^+, \mathbf{1} \rangle \le 0 \qquad \forall\, \boldsymbol{\ell} \in \mathbb{R}^n \\
&\iff \quad \langle [\boldsymbol{\phi}^t], \boldsymbol{\ell} \rangle - \langle \boldsymbol{\ell}, \boldsymbol{x}^* \rangle \|[\boldsymbol{\phi}^t]^+\|_1 \le 0 \qquad \forall\, \boldsymbol{\ell} \in \mathbb{R}^n \\
&\iff \quad \left\langle \boldsymbol{\ell}, \frac{[\boldsymbol{\phi}^t]}{\|[\boldsymbol{\phi}^t]^+\|_1} \right\rangle - \langle \boldsymbol{\ell}, \boldsymbol{x}^* \rangle \le 0 \qquad \forall\, \boldsymbol{\ell} \in \mathbb{R}^n \\
&\iff \quad \left\langle \boldsymbol{\ell}, \frac{[\boldsymbol{\phi}^t]}{\|[\boldsymbol{\phi}^t]^+\|_1} - \boldsymbol{x}^* \right\rangle \le 0 \qquad \forall\, \boldsymbol{\ell} \in \mathbb{R}^n.
\end{aligned}$$

Note that we are lucky: $[\boldsymbol{\phi}^t]^+ / \|[\boldsymbol{\phi}^t]\|_1$ is a nonnegative vector whose entries sum to 1. So, the above inequality can be satisfied with equality for the choice

$$\boldsymbol{x}^* = \frac{[\boldsymbol{\phi}^t]^+}{\|[\boldsymbol{\phi}^t]^+\|_1} \in \Delta^n.$$

In other words, we have that Blackwell's algorithm in this case picks

$$\boldsymbol{x}^{t+1} = \frac{[\boldsymbol{\phi}^t]^+}{\|[\boldsymbol{\phi}^t]^+\|_1} \in \Delta^n \iff \boldsymbol{x}^{t+1} \propto [\boldsymbol{\phi}^t]^+ \propto \left[\boldsymbol{r}^t\right]^+, \text{where } \boldsymbol{r}^t \coloneqq \sum_{\tau=1}^{t} \boldsymbol{\ell}^\tau - \langle \boldsymbol{\ell}^\tau, \boldsymbol{x}^\tau \rangle \mathbf{1}.$$

Remark 1.1. The vector $\boldsymbol{r}^t$ contains the regret cumulated compared to always playing each of the available actions with probability 1. Since $\boldsymbol{x}^{t+1}$ is proportional to $\boldsymbol{r}^t$, effectively the algorithm picks a distribution over actions that is proportional to the regret cumulated compared each of the actions. This justifies why this algorithm is known under the name "*regret matching*".

We also remark the following useful property of regret matching.

Remark 1.2. One very appealing property of the regret matching algorithm is its *lack of hyperparameters*. It just works "out of the box".

The resulting algorithm is summarized in Algorithm 1. By combining the analysis of the Blackwell approach-

Algorithm 1: Regret matching

$\boldsymbol{r}^0 \leftarrow \mathbf{0} \in \mathbb{R}^n, \quad \boldsymbol{x}^0 \leftarrow \mathbf{1}/n \in \Delta^n$

function NEXTSTRATEGY()
  $\boldsymbol{\theta}^t \leftarrow [\boldsymbol{r}^{t-1}]^+$
  if $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \;/\; \|\boldsymbol{\theta}^t\|_1$
  else **return** $\boldsymbol{x}^t \leftarrow$ any point in Δ^n

function OBSERVEUTILITY($\boldsymbol{\ell}^t$)
  $\boldsymbol{r}^t \leftarrow \boldsymbol{r}^{t-1} + \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}$

Algorithm 2: Regret matching$^+$

$\boldsymbol{z}^0 \leftarrow \mathbf{0} \in \mathbb{R}^n, \quad \boldsymbol{x}^0 \leftarrow \mathbf{1}/n \in \Delta^n$

function NEXTSTRATEGY()
  $\boldsymbol{\theta}^t \leftarrow [\boldsymbol{z}^{t-1}]^+$
  if $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \;/\; \|\boldsymbol{\theta}^t\|_1$
  else **return** $\boldsymbol{x}^t \leftarrow$ any point in Δ^n

function OBSERVEUTILITY($\boldsymbol{\ell}^t$)
  $\boldsymbol{z}^t \leftarrow [\boldsymbol{z}^{t-1} + \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}]^+$

ability strategy (6) together with Lemma 1.1, we obtain the following result.

Theorem 1.2. The regret matching algorithm (Algorithm 1) is an external regret minimizer, and satisfies the regret bound $R^T \leq \Omega\sqrt{T}$, where Ω is the maximum norm $\|\boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}\|_2$ up to time T.

1.5 The regret matching$^+$ (RM$^+$) algorithm

The regret matching$^+$ algorithm was introduced by Tammelin [2014], Tammelin et al. [2015], and is given in Algorithm 2. It differs from RM only on the last line, where a further thresholding is added. That small change has the effect that actions with negative cumulated regret (that is, "bad" actions) are treated as actions with 0 regret. Hence, intuitively, if a bad action were to become good over time, it would take less time for RM$^+$ to notice and act on that change. Because of that, regret matching$^+$ has stronger practical performance and is often preferred over regret matching in the game solving literature.

With a simple modification to the analysis in Section 1.3, the same bound as RM can be proven.

Theorem 1.3. The regret matching$^+$ algorithm (Algorithm 2) is an external regret minimizer, and satisfies the regret bound $R^T \leq \Omega\sqrt{T}$, where Ω is the maximum norm $\|\boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}\|_2$ up to time T.

We will not get into further details about RM$^+$ in this lecture, but we might touch back on it again later (probably around Lecture 8), when we will discuss a surprising connection between RM$^+$ and online mirror descent, probably the most well-studied algorithm in the field of online learning.

References

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

David Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6: 1–8, 1956.

Oskari Tammelin. Solving large imperfect information games using CFR+, 2014.

Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit Texas hold'em. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.