15-888 Computational Game Solving (Fall 2021) Tue, Oct 5th 2021

# Lecture 9

## Predictive regret matching and regret matching$^+$

Instructor: Gabriele Farina*

# 1 Predictive Blackwell approachability for simplex domains

In Lecture 8 we saw that a Blackwell approachability game with a conic target set can be solved by means of an external regret minimization algorithm whose domain is the polar of the cone, using a construction by Abernethy et al. [2011].

In this lecture, we will specialize that algorithm in the particular Blackwell game $\Gamma = (\Delta^n, \mathbb{R}^n, \boldsymbol{u}, S \coloneqq \mathbb{R}^n_{\le 0})$ we introduced in Lecture 4, where the bilinear Blackwell utility of the game was defined as

$$\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\ell}) \coloneqq \boldsymbol{\ell} - \langle \boldsymbol{\ell}, \boldsymbol{x} \rangle \, \mathbf{1} \in \mathbb{R}^n.$$

As we showed back then, any solution to $\Gamma$—that is, any algorithm that picks strategies $\boldsymbol{x}^t \in \Delta^n$ so that the average Blackwell payoff is close to the target set $S = \mathbb{R}^n_{\le 0}$—is a regret minimizer for $\Delta^n$. In particular, recall that the external regret

$$R^T \coloneqq \max_{\hat{\boldsymbol{x}} \in \Delta^n} \sum_{t=1}^T (\boldsymbol{\ell}^t)^\top \hat{\boldsymbol{x}} - \sum_{t=1}^T (\boldsymbol{\ell}^t)^\top \boldsymbol{x}^t$$

cumulated by strategies $\boldsymbol{x}^t$ with respect to any sequence of utilities $\boldsymbol{\ell}^t$ satisfies the inquality.

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

As we already mentioned, we are interested in solving the Blackwell game $\Gamma$ by means of the general framework introduced in Lecture 8, which for the particular case of target set $\mathbb{R}^n_{\le 0}$ boils down to Algorithm 1.

**Algorithm 1:** (Predictive) Blackwell approachability for simplex domain

**Data:** $\mathcal{R}_S$ (predictive) regret minimizer for $\mathbb{R}^n_{\ge 0}$

**function** NextStrategy($\boldsymbol{v}^t$)
  [▷ Set $\boldsymbol{v}^t = \mathbf{0}$ for the non-predictive version)]
  $\boldsymbol{\theta}^t \leftarrow \mathcal{R}_S$.NextStrategy($\boldsymbol{v}^t$)
  **if** $\boldsymbol{\theta}^t \ne \mathbf{0}$ **then return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t / \|\boldsymbol{\theta}^t\|_1 \in \Delta^n$
  **else return** an arbitrary point $\boldsymbol{x}^t \in \Delta^n$

**function** ReceivePayoff($\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)$)
  $\mathcal{R}_S$.ObserveLoss($\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)$)

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

Algorithm 1 gives a way to solve $\Gamma$ starting from *any* regret minimizer $\mathcal{R}_S$ for the nonpositive orthant $\mathbb{R}^n_{\geq 0}$. In the rest of the lecture we will explore what happens when $\mathcal{R}_S$ is set to FTRL and OMD, as well as their predictive variants.

## 2 Recovering regret matching (RM) and regret matching plus (RM$^+$)

In this section we show that when the Blackwell game $\Gamma$ is solved by means of Algorithm 1 instantiated with $\mathcal{R}_S$ set to FTRL, the regret matching (RM) algorithm is recovered [Farina et al., 2021]. Even more surprising, when $\mathcal{R}_S$ is set to OMD the regret matching plus (RM$^+$) algorithm is recovered instead. We will use these connections for two purposes:

- The fact that RM$^+$ can be recovered from Algorithm 1 (which was proven sound for every choice of regret minimizer $\mathcal{R}_S$ in Lecture 8) immediately implies correctness of RM$^+$.Even better, the connections between FTRL, OMD and RM, RM$^+$ will enable us to quickly give a regret bound for RM and RM$^+$ starting from the known regret bounds for FTRL and OMD seen in Lecture 7. We do so in Section 2.3.
- The connections suggest that if we started instead from the *predictive* versions of FTRL and OMD, we could hope to arrive to predictive versions of RM and RM$^+$, respectively. We will show that that is indeed the case in Section 3.

**Algorithm 2:** Regret matching

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

2 **function** NextStrategy()

3     **if** $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \ / \ \|\boldsymbol{\theta}^t\|_1$

4     **else**     **return** $\boldsymbol{x}^t \leftarrow$ any point in $\Delta^n$

5 **function** ObserveUtility($\boldsymbol{\ell}^t$)

6     $\boldsymbol{\theta}^{t+1} \leftarrow \boldsymbol{\theta}^t + \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}$

**Algorithm 3:** Regret matching$^+$

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

2 **function** NextStrategy()

3     **if** $\boldsymbol{\theta}^t \neq \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \ / \ \|\boldsymbol{\theta}^t\|_1$

4     **else**     **return** $\boldsymbol{x}^t \leftarrow$ any point in $\Delta^n$

5 **function** ObserveUtility($\boldsymbol{\ell}^t$)

6     $\boldsymbol{\theta}^{t+1} \leftarrow [\boldsymbol{\theta}^t + \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}]^+$

### 2.1 FTRL leads to regret matching (RM)

The regret minimizer $\mathcal{R}_S$ is used in Algorithm 1 to pick the next vector $\boldsymbol{\theta}^t \in \mathbb{R}^n_{\geq 0}$ to force observes utilities $\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) = \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}$. Consider now $\mathcal{R}_S$ to be the FTRL algorithm with regularizer $\varphi = \frac{1}{2}\| \cdot \|_2^2$ and step size $\eta > 0$ (recalled in Algorithm 4). In that case, the vector $\boldsymbol{\theta}^t$ (Line 2 in Algorithm 1) has the closed-form solution

$$\boldsymbol{\theta}^t = \arg\max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0}} \left\{ \left( \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \right)^{\top} \hat{\boldsymbol{\theta}} - \frac{\|\hat{\boldsymbol{\theta}}\|_2^2}{2\eta} \right\} = \eta \left[ \sum_{t=1}^{T} \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \right]^+ = \eta \left[ \sum_{t=1}^{T} \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} \right]^+ .$$

Since the forcing action $\boldsymbol{\theta^t}/\|\boldsymbol{\theta^t}\|_1$ is invariant to positive constants, we see that the action $\boldsymbol{x}^t$ picked by **??** (Line 3) is the same for all values of $\eta > 0$ and is computed as

$$\boldsymbol{x}^t = \frac{\left[ \sum_{t=1}^{T} \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} \right]^+}{\left\| \left[ \sum_{t=1}^{T} \boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} \right]^+ \right\|_1}. \tag{2}$$

provided $\boldsymbol{\theta}^t \neq \mathbf{0}$, and is an arbitrary vector $\boldsymbol{x}^t \in \Delta^n$ otherwise. These iterates coincide at all times $t$ with the iterates produced by the regret matching algorithm seen in Lecture 4 and recalled in Algorithm 2.

**Algorithm 4:** (Predictive) FTRL

**Data:** $\mathcal{X} \subseteq \mathbb{R}^n$ convex and compact set
$\varphi : \mathcal{X} \to \mathbb{R}_{\geq 0}$ strongly convex regularizer
$\eta > 0$ step-size parameter

$\boldsymbol{L}^0 \leftarrow \boldsymbol{0} \in \mathbb{R}^n$

**function** NextStrategy($\boldsymbol{m}^t$)
[$\triangleright$ Set $\boldsymbol{m}^t = \boldsymbol{0}$ for non-predictive version]
**return** $\displaystyle \arg\max_{\hat{\boldsymbol{x}} \in \mathcal{X}} \left\{ (\boldsymbol{L}^{t-1} + \boldsymbol{m}^t)^\top \hat{\boldsymbol{x}} - \frac{1}{\eta}\varphi(\hat{\boldsymbol{x}}) \right\}$

**function** ObserveUtility($\boldsymbol{\ell}^t$)
$\boldsymbol{L}^t \leftarrow \boldsymbol{L}^{t-1} + \boldsymbol{\ell}^t$

**Algorithm 5:** (Predictive) OMD

**Data:** $\mathcal{X} \subseteq \mathbb{R}^n$ convex and compact set
$\varphi : \mathcal{X} \to \mathbb{R}_{\geq 0}$ strongly convex regularizer
$\eta > 0$ step-size parameter

$\boldsymbol{z}^0 \in \mathcal{X}$ such that $\nabla\varphi(\boldsymbol{z}^0) = \boldsymbol{0}$

**function** NextStrategy($\boldsymbol{m}^t$)
[$\triangleright$ Set $\boldsymbol{m}^t = \boldsymbol{0}$ for non-predictive version]
**return** $\displaystyle \arg\max_{\hat{\boldsymbol{x}} \in \mathcal{X}} \left\{ (\boldsymbol{m}^t)^\top \hat{\boldsymbol{x}} - \frac{1}{\eta} D_\varphi(\hat{\boldsymbol{x}} \,\|\, \boldsymbol{z}^{t-1}) \right\}$

**function** ObserveUtility($\boldsymbol{\ell}^t$)
$\displaystyle \boldsymbol{z}^t \leftarrow \arg\max_{\hat{\boldsymbol{z}} \in \mathcal{X}} \left\{ (\boldsymbol{\ell}^t)^\top \hat{\boldsymbol{z}} - \frac{1}{\eta} D_\varphi(\hat{\boldsymbol{z}} \,\|\, \boldsymbol{z}^{t-1}) \right\}$

## 2.2 OMD corresponds to regret matching plus (RM$^+$)

Consider now $\mathcal{R}_S$ to be the OMD algorithm with the regularizer $\varphi = \frac{1}{2}\|\cdot\|_2^2$ and step size $\eta > 0$ (recalled in Algorithm 5). In that case, the vector $\boldsymbol{\theta}^t$ (Line 2 in Algorithm 1) has the closed-form solution

$$\boldsymbol{\theta}^t = \arg\max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0}} \left\{ \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)^\top \hat{\boldsymbol{\theta}} - \frac{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{t-1}\|_2^2}{2\eta} \right\} = \left[\boldsymbol{\theta}^{t-1} + \eta\, \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\right]^+. \tag{3}$$

Since (3) is homogeneous in $\eta > 0$ (that is, the only effect of $\eta$ is to rescale all $\boldsymbol{\theta}^t$ by the same constant) and the forcing action $\boldsymbol{\theta}^t/\|\boldsymbol{\theta}^t\|_1$ is invariant to positive rescaling of $\boldsymbol{\theta}^t$, we see that Algorithm 1 outputs the same iterates no matter the choice of stepsize parameter $\eta > 0$. In particular, we can assume without loss of generality that $\eta = 1$. In that case, Equation (3) corresponds exactly to Line 6 in RM$^+$ (Algorithm 3).

## 2.3 Regret Analysis

The connectsion between regret matching (RM), regret matching plus (RM$^+$) and FTRL, OMD we uncovered in Sections 2.1 and 2.2 can help us establish regret bounds for RM and RM$^+$ starting from the regret bounds for FTRL and OMD. To do so, let's start from recalling the relationship—seen in Lecture 8—between the regret of $\mathcal{R}_S$ and the distance of the average Blackwell payoff to the target set, that is,

$$\min_{\hat{\boldsymbol{s}} \in \mathbb{R}^n_{\leq 0}} \left\| -\hat{\boldsymbol{s}} + \frac{1}{T}\sum_{t=1}^T \boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \right\|_2 \leq \frac{1}{T} \max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}^n_2} R_S^T(\hat{\boldsymbol{\theta}}). \tag{4}$$

Combining (4) with (1), we obtain that the regret cumulated by the sequence of strategies $\boldsymbol{x}^t$ produced by Algorithm 1 with respect to any sequence of utilities $\boldsymbol{\ell}^t$ satisfies

$$\frac{1}{T} R^T \leq \frac{1}{T} \max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}^n_2} R_S^T(\hat{\boldsymbol{\theta}}) \quad \Longrightarrow \quad R^T \leq \max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}^n_2} R_S^T(\hat{\boldsymbol{\theta}}), \tag{5}$$

where $R_S^T$ is the regret cumulated by the regret minimizer $\mathcal{R}_S$ oracle used in Algorithm 1. As we know from Lecture 7, both FTRL and OMD with regularizer $\varphi = \frac{1}{2}\|\cdot\|_2^2$ and step size $\eta > 0$ guarantee that

$$R_S^T(\hat{\boldsymbol{\theta}}) \leq \frac{\|\hat{\boldsymbol{\theta}}\|_2^2}{2\eta} + \eta \sum_{t=1}^T \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\|_2^2 \quad \Longrightarrow \quad \max_{\hat{\boldsymbol{\theta}} \in \mathbb{R}^n_{\geq 0} \cap \mathbb{B}^n_2} R_S^T(\hat{\boldsymbol{\theta}}) \leq \frac{1}{2\eta} + \eta \sum_{t=1}^T \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\|_2^2, \tag{6}$$

where we used the fact that $\hat{\boldsymbol{\theta}} \in \mathbb{B}_2^n$ on the right side of the implication. So, plugging (6) into (5), we have

$$R^T \le \frac{1}{2\eta} + \eta \sum_{t=1}^{T} \|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\|_2^2.$$

Since we have shown above that the iterates produced by regret matching (Section 2.1) and regret matching plus (Section 2.2) are independent of $\eta > 0$, we can minimize the right-hand side over $\eta > 0$, obtaining the bound

$$R^T \le \sqrt{2 \sum_{t=1}^{T} \left\|\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t)\right\|_2^2}.$$

Finally, expanding the definition of $\boldsymbol{u}(\boldsymbol{x}^t, \boldsymbol{\ell}^t) \coloneqq \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1} - \boldsymbol{\ell}^t$, we obtain the following statement.

**Theorem 2.1.** At every time $T$, the regret cumulated by the regret matching (Algorithm 2) and regret matching plus algorithms (Algorithm 3) satisfy the regret bound

$$R^T \le \sqrt{2 \sum_{t=1}^{T} \left\|\boldsymbol{\ell}^t - \langle \boldsymbol{\ell}^t, \boldsymbol{x}^t \rangle \mathbf{1}\right\|_2^2}.$$

## 3 Predictive regret matching and regret matching plus

We can repeat the same analysis we did in Section 2.1 (which used FTRL) and Section 2.2 (which used OMD) using the *predictive* versions of FTRL and OMD. The resulting algorithms are again independent on the stepsize parameter, and are given in Algorithm 6 and Algorithm 7.

**Algorithm 6:** (Predictive) 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($\boldsymbol{m}^t$)
- [$\triangleright$ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version
- $\boldsymbol{\theta}^t \leftarrow [\boldsymbol{r}^{t-1} + \langle \boldsymbol{m}^t, \boldsymbol{x}^{t-1} \rangle \mathbf{1} - \boldsymbol{m}^t]^+$
- **if** $\boldsymbol{\theta}^t \ne \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \,/\, \|\boldsymbol{\theta}^t\|_1$
- **else** **return** $\boldsymbol{x}^t \leftarrow$ arbitrary point in $\Delta^n$

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

**Algorithm 7:** (Predictive) 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($\boldsymbol{m}^t$)
- [$\triangleright$ Set $\boldsymbol{m}^t = \mathbf{0}$ for non-predictive version
- $\boldsymbol{\theta}^t \leftarrow [\boldsymbol{z}^{t-1} + \langle \boldsymbol{m}^t, \boldsymbol{x}^{t-1} \rangle \mathbf{1} - \boldsymbol{m}^t]^+$
- **if** $\boldsymbol{\theta}^t \ne \mathbf{0}$ **return** $\boldsymbol{x}^t \leftarrow \boldsymbol{\theta}^t \,/\, \|\boldsymbol{\theta}^t\|_1$
- **else** **return** $\boldsymbol{x}^t \leftarrow$ arbitrary point in $\Delta^n$

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

The same regret analysis of Section 2.3 holds verbatim. In particular, we have the following.

**Theorem 3.1.** At every time $T$, the regret cumulated by the predictive regret matching (Algorithm 6) and predictive regret matching plus algorithms (Algorithm 7) satisfy the regret bound

$$R^T \le \sqrt{2 \sum_{t=1}^{T} \left\|(\boldsymbol{\ell}^t - \boldsymbol{m}^t) - \langle \boldsymbol{\ell}^t - \boldsymbol{m}^t, \boldsymbol{x}^t \rangle \mathbf{1}\right\|_2^2}.$$

## References

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 27–46, 2011.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2021.