IEOR8100: Economics, AI, and Optimization
Lecture Note 4: Blackwell Approachability and Regret Matching

Christian Kroer*

February 16, 2020

1 Recap

The last few times we have learned about online convex optimization, and we saw algorithms for minimizing regret such as Hedge and Online Mirror Descent. In this lecture we are going to introduce a new type of online-learning problem concerned with *vector-valued games*. This framework will eventually be shown to lead to one of the fastest algorithms for game solving in practice.

2 Blackwell Approachability

In two-player zero-sum games we saw that there exists a value for the game v such that the row player can choose a strategy x assuring that the payoff will be in the set $(-\infty, v]$ no matter what the column player does, and vice versa the column player can assure that the payoff lies in the set $[v, \infty)$, no matter what the row player does.

In Blackwell approachability we ask whether there is a way to generalize the notion of forcing the payoffs to lie in a particular set to *vector-valued games*.

We consider the following setup:

- The row and column players choose strategies from compact convex sets X and Y respectively.
- There is a bilinear vector-valued payoff function $f(x, y) \in \mathbb{R}^m$.
- There is a closed convex *target set* C.
- We will assume that $f(x, y) \in B(0, 1), C \subseteq B(0, 1)$, where $B(0, 1) = \{g : \|g\|_2 \leq 1\}$.

The goal for the row player is to get payoffs $f(x, y)$ to lie inside C. The case of a single-shot game is trivially analyzed: it is generally only possible to do this if there exists x such that $f(x, y) \in C, \ \forall y \in Y$. So in general this won't be possible. However, it turns out that much more interesting things can be said about a variant where the two players are playing a repeated game. In particular, the players choose actions x_t, y_t at each timestep t. The goal for the row player is to have the average payoff vector $\bar{f}_t = \frac{1}{t} \sum_{i=1}^{t} f(x_t, y_t)$ approach C, while the goal of the column player is to keep $\bar{f}_t$ from approaching C. We will measure the distance as $d(\bar{f}_t, C) = \min_{z \in C} \|z - \bar{f}_t\|_2$

We will say that

*Department of Industrial Engineering and Operations Research, Columbia University. Email: christian.kroer@columbia.edu.

Definition 1. *A target set C is* approachable *if there exists an algorithm for picking x_t based on $x_1, \ldots, x_{t-1}, y_1, \ldots, y_{t-1}$ such that $d(\bar{f}_t, C) \to 0$ as t goes to infinity.*

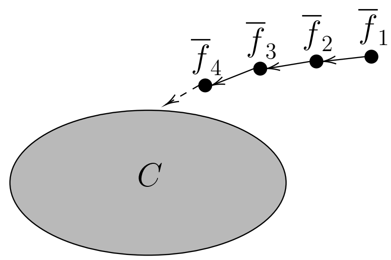

Figure 1: Blackwell approachability requires that the sequence $\{\bar{f}_t\}_{t=1}$ approaches C no matter the choices of the y player.

A stronger notion is that

Definition 2. *A target set C is* forceable *if there exists x such that $f(x, y) \in C$ for all $y \in Y$.*

2.1 Scalar Approachability

In the special case where $m = 1$ we get a scalar approachability game. As discussed at the beginning of this section, this can be analyzed via minimax theorems. In particular, for the scalar case target sets are intervals, and we may analyze only intervals of the form $(-\infty, \lambda]$ without loss of generality. Clearly, an interval $(-\infty, \lambda]$ is approachable if $\lambda \geq v$, where v is the value of the game associated to the bilinear function f in Sion's minimax theorem. This follows because if the row player plays any strategy x such that they are guaranteed at least v, then $f(x, y_t) \in (-\infty, \lambda]$ for all t no matter the y_t. Conversely, if $\lambda < v$, then by Sion's theorem the column player may play a strategy y such that no matter the x_t, $f(x_t, y) \geq v > \lambda$. We thus have the lemma

Lemma 1. *In scalar approachability games, the following three statements are equivalent:*

- *A target set $(-\infty, \lambda]$ is approachable.*
- *A target set $(-\infty, \lambda]$ is forceable.*
- *$\lambda \geq v$, where v is the value of the game associated to f, X, Y in Sion's minimax theorem.*

2.2 Halfspace Approachability

We first analyze the special case where the target set is a halfspace $H = \{h : \langle h, a \rangle \leq b\}$. Halfspaces turn out to have the nice property that forceability is equivalent to approachability:

Lemma 2. *A halfspace H is approachable if and only if it is forceable.*

Proof. The proof consists in reducing halfspace approachability to a scalar approachability game. To do that, let $\hat{f}(x, y) = \langle a, f(x, y) \rangle$. Now we clearly have that forcing H wrt. f is equivalent to forcing $(-\infty, b]$ wrt. $\hat{f}$. Say x^* forces $(-\infty, b]$, then

$$b \geq \hat{f}(x^*, y) = \langle a, f(x^*, y) \rangle, \ \forall y \in Y,$$

and so x^* also forces H, and vice versa.

For approachability we have that the distance from $\bar{f}_t$ to H satisfies

$$d(\bar{f}_t, H) = d\left(\langle a, \bar{f}_t \rangle, (-\infty, b]\right) = d\left(\frac{1}{t}\sum_{i=1}^{t} \langle a, f_i \rangle, (-\infty, b]\right).$$

Thus approachability of H is equivalent to approachability of $(-\infty, b]$.

From Lemma 1 we have that approachability and forceability are equivalent for $(-\infty, b]$, so they must be equivalent for H. □

2.3 Blackwell's Approachability Theorem

Now we are ready to analyze the general case of when a convex compact set C is approachable. Blackwell proved the following:

Theorem 1. *A convex compact set C is approachable if and only if every halfspace $H \supseteq C$ is forceable. If every halfspace is forceable then C can be approached at a rate of $\frac{2}{\sqrt{T}}$.*

Blackwell's proof is constructive. It is based on the following algorithm for approaching C when all halfspaces containing C are forceable: At every timestep t, do the following:

- If $\bar{f}_t \in C$, play any x_t.
- Else consider the projection ϕ_t of $\bar{f}_t$ onto C. We construct a halfspace H with normal vector $a_t = \phi_t - \bar{f}_t$, and constant $b_t = \langle a_t, \phi_t \rangle$. Play any x_t forcing H.

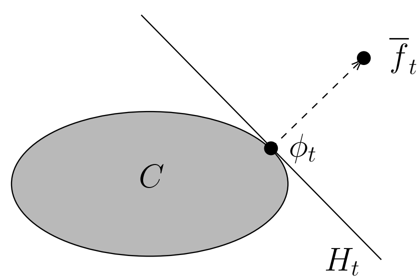

Figure 2: The tangent halfspace forced in Blackwell's theorem.

The algorithm repeatedly takes the halfspace tangent to the projection of $\bar{f}_t$, and forces it. We now prove Blackwell's theorem.

Proof. Say that C is approachable. Then we may play any algorithm guaranteed to approach C, and we will then be guaranteed to approach every $H \supseteq C$.

Now assume that all $H \supseteq C$ are approachable, and play Blackwell's algorithm. First note that since ϕ_t is the projection of $\bar{f}_t$ onto a convex set H (this follows from how we constructed H) we have from first-order optimality:

$$\langle \phi_t - \bar{f}_t, z - \phi_t \rangle \geq 0, \ \forall z \in H. \tag{1}$$

Let $f_{t+1} = f(x_{t+1}, y_{t+1})$. We have

$$
\begin{aligned}
d(\bar{f}_{t+1}, C)^2 =& \min_{z \in C} \|\bar{f}_{t+1} - z\|_2^2 \\
\leq& \|\bar{f}_{t+1} - \phi_t\|_2^2 \\
=& \left\| \frac{t}{t+1} \bar{f}_t + \frac{1}{t+1} f_{t+1} - \phi_t \right\|_2^2 ; \quad \text{by definition of } \bar{f}_{t+1} \\
=& \left\| \frac{t}{t+1} (\bar{f}_t - \phi_t) + \frac{1}{t+1} (f_{t+1} - \phi_t) \right\|_2^2 \\
=& \frac{1}{(t+1)^2} \left(t^2 \|(\bar{f}_t - \phi_t)\|_2^2 + \|(f_{t+1} - \phi_t)\|_2^2 + 2t\langle \bar{f}_t - \phi_t, f_{t+1} - \phi_t \rangle \right) \\
\leq& \frac{1}{(t+1)^2} \left(t^2 \|(\bar{f}_t - \phi_t)\|_2^2 + \|(f_{t+1} - \phi_t)\|_2^2 \right) ; \quad \text{by (1)} \\
=& \frac{1}{(t+1)^2} \left(t^2 d(\bar{f}_t, C)^2 + \|(f_{t+1} - \phi_t)\|_2^2 \right)
\end{aligned}
$$

Telescoping this inequality we have

$$
d(\bar{f}_{t+1}, C)^2 \leq \frac{1}{(t+1)^2} \sum_{i=1}^{t} \|f_{t+1} - \phi_t\|_2^2 \leq \frac{4t}{(t+1)^2} \leq \frac{4}{t+1},
$$

where the second inequality is from the fact that we assumed payoffs to lie in the norm-ball $B(0, 1)$. Taking the square root of both sides gives the theorem. □

3 Regret Matching

Blackwell's constructive result can easily be converted to a regret minimization algorithm for linear losses over a simplex Δ^n. For each pure action i we say that $r_{t,i} = \langle g_t, x_t \rangle - g_{t,i}$ is the regret from not playing action i rather than x_t, and we let r_t be the vector of all n regrets. We will use $\frac{r}{\sqrt{n}}$ as our vector-valued payoff. Note that the regret is now $R_T = \max_i \sum_{t=1}^{T} r_{t,i}$, and having regret grow sublinearly is equivalent to $\bar{r}_t = \frac{1}{t} \sum_{k=1}^{t} r_k$ approaching the non-positive orthant as t tends to infinity. Thus our target set is $C = \mathbb{R}^n_-$.

By Blackwell's theorem having $\bar{r}_t$ approach $\mathbb{R}^n_-$ can be done by repeatedly forcing tangent halfspaces. To do so, let ϕ_t be the projection of $\bar{r}_t$ onto $\mathbb{R}^n_-$. Note that the normal vector $a_t = \bar{r}_t - \phi_t$ simply thresholds $\bar{r}_t$ at zero, setting all negative entries to zero. Now, we will force r_{t+1} to be in the halfspace with normal vector a_t by ensuring $\langle a_t, r_{t+1} \rangle = 0$. To do so, first consider the square matrix of pairwise regrets B, where B_{ij} is the regret from playing i rather than j under g_{t+1}. We have that $B_{ij} = -B_{ij}$, so B is skew-symmetric, which means that $\langle q, Bq \rangle = 0$ for all q. We can choose $x_{t+1} = \frac{a_t}{\|a_t\|_1}$, in which case we get that the next regret is $r_{t+1} = Bx_{t+1} = B\frac{a_t}{\|a_t\|_1}$, and now it satisfies $\langle a_t, r_{t+1} \rangle = 0$, and thus we forced the desired halfspace.

Summarizing what we did in terms of our standard regret minimization framework, we have an algorithm that works as follows:

- Play arbitrary x_1
- Keep a sum $\hat{r}_t = \sum_{k=1}^{t} r_k$ of regret vectors
- At time $t+1$ set $x_{t+1,i} = \frac{[\hat{r}_{t,i}]^+}{\sum_{k=1}^{t} [\hat{r}_k]^+}$ ($[\cdot]^+$ denotes thresholding at 0)

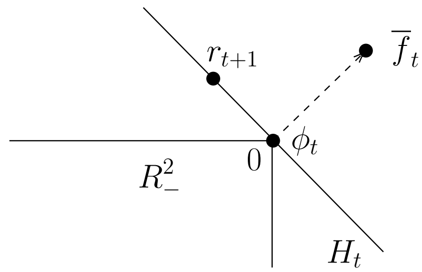

Figure 3: The next regret vector r_{t+1} lies in the halfspace forced in Blackwell's theorem.

- If no regrets are positive, play uniform strategy

This algorithm is called *regret matching*, and by Blackwell's theorem regret matching has regret that grows on the order of $O\left(\sqrt{T}\right)$, assuming $g_t \in B(0,1)$ for all t (if this does not hold we may simply normalize the payoffs).

4 Regret Matching$^+$

Finally, we present a variation on regret matching, which turns out to be immensely useful in practice. In regret matching, remember that we took the sum of the regret vectors and thresholded it at zero when generating x_{t+1}. In *regret matching*$^+$ (RM$^+$), we only keep track of positive regrets. Formally, we have the following algorithm:

- initialize $Q_1 = 0$ and play x_1 arbitrarily
- After seeing g_t, set $Q_t = \left[\frac{t-1}{t}Q_t + \frac{1}{t}g_t\right]^+$
- At time $t+1$, play $x_{t+1,i} = \frac{Q_{t,i}}{\|Q_t\|_1}$

The important observation for RM$^+$ is that we are constructing a sequence that upper-bounds regret, i.e. $Q_t \geq \bar{r}_t$. This is easy to see, as we are only dropping negative terms in the summation that makes up $\bar{r}_t$.

Visually, we may think of it as moving along a face of $\mathbb{R}^n_-$, while maintaining the same distance d to $\mathbb{R}^n_-$ while moving towards 0. See Figure 4

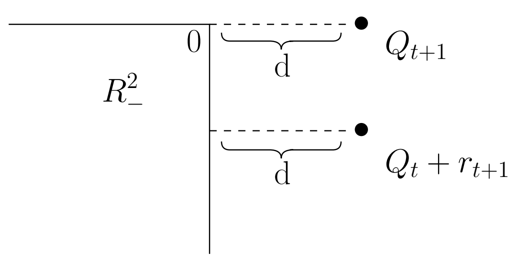

Figure 4: The thresholding used in constructing Q_{t+1}.

Theorem 2. RM$^+$ *approaches* $C = \mathbb{R}^n_-$ *at a rate of* $\frac{2}{\sqrt{T+1}}$

Proof. Let Q_t^* be the projection of Q_t onto C. Let H be the halfspace $\{q : \langle Q_t, q\rangle \leq 0\}$ corresponding to forcing in Blackwell's theorem (since $Q_t^* = 0$). We have

$$\begin{aligned}
d(Q_{t+1}, C)^2 =& \min_{z\in C} \|Q_{t+1} - z\|^2 \\
\leq& \|Q_{t+1} - Q_t^*\|^2 \\
=& \|Q_{t+1}\|^2; \text{ since } Q_t^* = 0 \\
=& \|\frac{t}{t+1}Q_t + \frac{1}{t+1}r_{t+1}\|^2; \text{ since thresholding is distance preserving} \\
=& \frac{1}{(t+1)^2}\left(t^2\|Q_t\|^2 + \|r_{t+1}\|^2 + 2t\langle Q_t, r_{t+1}\rangle\right) \\
=& \frac{1}{(t+1)^2}\left(t^2\|Q_t\|^2 + \|r_{t+1}\|^2\right); \text{ by forcing } r_{t+1} \in H.
\end{aligned}$$

By telescoping we now get

$$\begin{aligned}
d(Q_{t+1}, C)^2 \leq& \frac{1}{(t+1)^2}\left(t^2 d(Q_t, C) + \|r_{t+1}\|^2\right) \\
\leq& \frac{1}{(t+1)^2}\sum_{k=1}^{t}\|r_{k+1}\|^2 \\
\leq& \frac{1}{(t+1)^2}4t \\
\leq& \frac{4}{(t+1)}
\end{aligned}$$

Taking square roots concludes the theorem. □

5 Overview of Regret Minimizers

At this point we have covered quite a few regret minimizers. In the coming lectures we will start to look at how they can be used to solve zero-sum games, both matrix games and extensive-form games. For now, let us quickly recap and compare our options. Say that we want to minimize linear losses from $[0, 1]^n$ over a simplex Δ^n (note that this covers convex losses with bounded dual norm of the gradients). In that case we have covered 5 algorithms with two different types of regret bounds:

- Regret bound: $O\left(\sqrt{T \log n}\right)$: **Hedge** and **OMD (entropy)**
- Regret bound: $O\left(\sqrt{nT}\right)$: **OMD (Euclidean)**, **Regret Matching**, and **Regret Matching$^+$**.

6 Historical Notes and Further Reading

Blackwell approachability was introduced in [1]. Regret matching was introduced by [4]. The RM$^+$ algorithm was introduced in [6] and proven correct by [7]. The proof of RM$^+$ via modified Blackwell approachability is, I believe, new. It was developed together with Gabriele Farina when working on the papers Farina et al. [2, 3].

There aren't many places to find coverage of Blackwell approachability, and furthermore all the sources I know of cover it in quite different ways and levels of generality. Lectures notes 13 and 14 of Ramesh Johari [5] cover the finite-action space case as well as regret matching and the relationship to calibration. Another nice presentation for that case if that of Young [8]. The more general proof of Blackwell's theorem given here largely follows the one given in a blog post by Farina at `http://www.cs.cmu.edu/~gfarina/2016/approachability/`.

References

[1] David Blackwell et al. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[2] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Regret minimization in behaviorally-constrained zero-sum games. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1107–1116. JMLR. org, 2017.

[3] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Online convex optimization for sequential decision processes and extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1917–1925, 2019.

[4] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[5] Ramesh Johari. MS&E 336: dynamics and learning in games, 2007.

[6] Oskari Tammelin. Solving large imperfect information games using cfr+. *arXiv preprint arXiv:1407.5042*, 2014.

[7] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit texas hold'em. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[8] H Peyton Young. *Strategic learning and its limits*. OUP Oxford, 2004.