4. Blackwell approachability and regret minimization on simplex domains

Link

---

1 - Blackwell Approachability and Regret Minimization on Simplex Domains

- This lecture focuses on constructing no-external-regret dynamics for one-shot decision making problems.

- Strategies output from the regret minimizers must be in the simplex domain, $x^t \in \Delta^n$.

  - *Note:* a strategy assigns a probability to each action and the sum of all probabilities must be equal to 1.

- External cumulative regret for the simplex domain:

$$R^T := \max_{\hat{x} \in \Delta^n} \left\{ \sum_{t=1}^{T} (< \ell^t, \hat{x} > - < \ell^t, x^t >) \right\}$$

1.1 - Blackwell Approachability Game

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

  - Let,
    - $\mathcal{X}$ and $\mathcal{Y}$ be closed convex strategy sets for Player 1 and 2.
    - $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ be a biaffine utility function
    - $S \subseteq \mathbb{R}^d$ be a closed and convex *target set*.

  - Game order:
    - Player 1 selects an action $x^t \in \mathcal{X}$
    - Player 2 selects an action $y^t \in \mathcal{Y}$ which depends adversarially on all the x^t output.
    - Player 1 incurs the vector-valued payoff $u(x^t, y^t) \in \mathbb{R}^d$

> Player 1's objective is to guarantee the average payoff converges to the target set S.
>
> $$\min_{\hat{s} \in S} \left\| \hat{s} - \frac{1}{T} \sum_{t=1}^{T} u(x^t, y^t) \right\|_2 \to 0 \quad \text{as} \quad T \to \infty$$

1.2- External Regret Minimization on the Simplex Via Blackwell Approachability

- A regret minimizer for a simplex domain can be reduced to constructing an algorithm for a particular Blackwell approachability game: [2000-Hart]

  - $\Gamma := \left(\Delta^n,\, \mathbb{R}^n,\, u,\, \mathbb{R}^n_{\leq 0}\right)$

  - The utility function is a vector-valued payoff function that measures the change in regret incurred at time t, $u : \Delta^n \times \mathbb{R}^n \to \mathbb{R}^n$ defined as

  $u(x^t,\, \ell^t) = \ell^t - < \ell^t,\, x^t > 1$

  Note: Each component of the utility function is the regret we have for playing our mixed strategy x^t rather than purely playing that action. Strongly negative utility means that action performs much worse than our mixed strategy. Logically, we want to improve our mixed strategy until it performs the same or better than purely playing any single action. Hence, our target space for the utility is $S = \mathbb{R}_{\leq 0}$.

- **Lemma** - The regret R^T cumulated up to any time T by any sequence of decisions $x^1, \ldots, x^T \in \Delta^n$ is related to the distance of the average Blackwell payoff from the target cone $\mathbb{R}^n_{\leq 0}$ as

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

  So the Blackwell approachability game Γ is a regret-minimizing strategy for the simplex.
  (See notes for proof.)

> By choosing a clever utility function and defining our target set to be the non-positive reals, the objective function for the Blackwell approachability game reduces to external regret minimization.

1.3 - Solving Blackwell Games: Blackwell's Algorithm

- *Forceable halfspace* - Let $(\mathcal{X},\, \mathcal{Y},\, u,\, S)$ be a Blackwell approachability game and let $\mathcal{H} \subseteq \mathbb{R}^d$ be a halfspace, that is, a set of the form $\mathcal{H} = \{x \in \mathbb{R}^d : a^T x \leq b\}$ for some $a \in \mathbb{R}^d$, $b \in \mathbb{R}$.

  Note: It's a "half" space because $a^T x \leq b$ forms a hyperplane that splits $\mathbb{R}^d$ into two halves.

  The halfspace is *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, there exists $x^* \in \mathcal{X}$ such that

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

Note: Looking back at the objective function (defined in the Lemma), we want to be able to force the utility (by selecting a strategy x) toward a target space S. So we can select a halfspace (hyperplane which defines a region) which contains S. If $\mathcal{H}$ is forceable, then we can force u into $\mathcal{H}$ by selecting x^*. We can do this progressively, each time moving the utility closer to the target, until the utility is contained in the target set.

- Blackwell approachability theorem - Player 1's objective in the Blackwell approachability game can be attained if and only if every halfspace $H \supseteq S$ is forceable.
- Proof by construction, at each time step $t = \, 1,\, 2,\, \ldots$ operate the following:
  - Compute the average payoff received so far, $\phi^t = \frac{1}{t}\sum_{\tau=1}^{t-1} u(x^\tau, y^\tau)$.
  - Compute the Euclidean projection, ψ^t, of ϕ^t onto the target set S.
  - If $\phi^t \in S$, meaning the objective has already been met, then pick and play any $x^t \in \mathcal{X}$, observe the opponent's action y^t, and return.
  - Else, the objective hasn't been met. Consider the halfspace $\mathcal{H}^t$ tangent to S at the projection point ψ^t, that contains S.

    $\mathcal{H}^t := \{z \in \mathbb{R}^d : (\phi^t - \psi^t)^T z \leq (\phi^t - \psi^t)^T \psi^t\}$
  - By hypothesis, $\mathcal{H}^t$ is forceable. Pick x^t to be a forcing action for $\mathcal{H}^t$, observe the opponent's action y^t, and return.

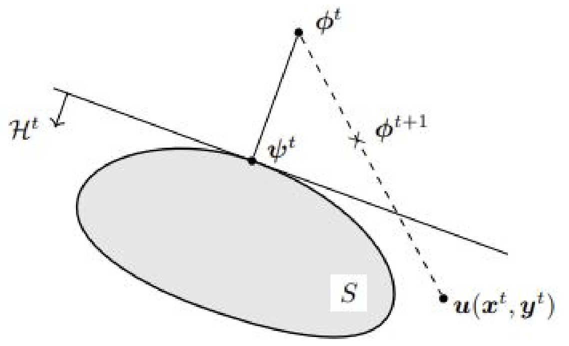


Figure 1: Construction of the approachability strategy described in Section 1.3.

  - Average payoff when playing the strategy above:

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

    Note: This is just a matter of moving terms in and out of the summation.
  - Euclidean distance between ϕ^{t+1} and the target set:

    $\rho^t := \min_{\hat{s} \in S} \|\hat{s} - \phi^t\|_2^2$
  - ρ^{t+1} is the minimum distance between ϕ^{t+1} and S. So by definition, ψ^t can only be as close to the closest point in S to ϕ^{t+1}, or farther away.

    $\rho^{t+1} \leq \|\psi^t - \phi^{t+1}\|_2^2$

    $\|\psi^t - \phi^{t+1}\|_2^2 = \frac{(t-1)^2}{t^2}\rho^t \;+\; \frac{1}{t^2}\|\psi^t - u(x^t, y^t)\|_2^2 \;+\; \frac{2(t-1)}{t^2}\langle\, \psi^t - \phi^t,\; \psi^t - u(x^t, y^t)\,\rangle$
  - We choose an x^t that forces the halfspace $\mathcal{H}^t$, therefore, no matter how y^t is picked by the opponent we have:

$$u(x^t, y^t) \Leftrightarrow (\phi^t - \psi^t)^T u(x^t, y^t) \geq (\phi^t - \psi^t)^T \psi^t$$
$$u(x^t, y^t) \Leftrightarrow \;\langle\, \psi^t - \phi^t,\; \psi^t - u(x^t, y^t)\,\rangle\; \leq 0$$

  - Plugging this back into the equation before and bounding $\|\psi^t - u(x^t, y^t)\|_2^2 \;\leq \Omega^2$ where Ω is a diameter parameter:

$$\rho^{t+1} \leq \frac{(t-1)^2}{t^2}\rho^t \;+\; \frac{\Omega^2}{t^2} \;\implies\; t^2\rho^{t+1} - (t-1)^2\rho^t \leq \Omega^2 \quad \forall t = 1, 2, \ldots$$

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

$$\min_{\hat{s} \in S} \left\| \,\hat{s} - \frac{1}{T}\sum_{t=1}^{T} u(x^t, y^t) \right\|_2 \leq \frac{\Omega}{\sqrt{T}}$$

  - This implies 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

- We will use the steps described in last section and apply them to Γ, the regret minimizer for the simplex domain:
  - **Computation of** ψ^t - the projection onto the nonpositive orthant (projecting ϕ^t onto $S = \mathbb{R}_{\leq 0}$) amounts to a component-wise minimum with 0, denoted $\psi^t = [\phi^t]^-$.

    Note: ψ^t is all the negative components of ϕ^t and zero for all positive components of ϕ^t.

    Hence,

    $\phi^t - \psi^t = [\phi^t]^+ \implies (\phi^t - \psi^t)^T \psi^t = 0$

  - **Halfspace to be forced** - when $[\phi^t]^+ \neq 0$ (otherwise we will have met the objective), the halfspace to be forced is

    $\mathcal{H}^t := \{z \in \mathbb{R}^n \,:\, \langle\, [\phi^t]^+, z \,\rangle \leq 0\}$

  - **Forcing action for** $\mathcal{H}^t$ - Recall, a forcing action is an action $x^* \in \Delta^n$ such that no matter the $\ell \in \mathbb{R}^d$, $u(x^*, \ell) \in \mathcal{H}^t$.

    Expanding the definition of $\mathcal{H}^t$ and u , we are looking for a $x^* \in \Delta^n$ such that:

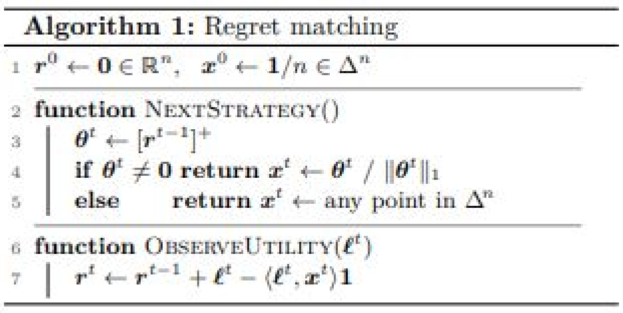
    $$\begin{gathered}
    \langle\, [\phi^t]^+,\, \ell - \langle\, \ell,\, x^* \,\rangle \,\rangle \leq 0 \quad \forall \ell \in \mathbb{R}^n \\
    \langle\, [\phi^t]^+,\, \ell \,\rangle \,-\, \langle\, \ell,\, x^* \,\rangle \langle\, [\phi^t]^+, 1 \,\rangle \,\leq\, 0 \quad \forall \ell \in \mathbb{R}^n \\
    \langle\, [\phi^t]^+,\, \ell \,\rangle \,-\, \langle\, \ell,\, x^* \rangle \, \|[\phi^t]^+\|_1 \,\leq\, 0 \quad \forall \ell \in \mathbb{R}^n \\
    \left\langle\, \ell,\, \frac{[\phi^t]^+}{\|[\phi^t]^+\|_1} \,\right\rangle \,-\, \langle \ell,\, x^* \rangle \,\leq 0 \quad \forall \ell \in \mathbb{R}^n \\
    \left\langle\, \ell,\, \frac{[\phi^t]^+}{\|[\phi^t]^+\|_1} - x^* \,\right\rangle \,\leq\, 0 \quad \forall \ell \in \mathbb{R}^n
    \end{gathered}$$

    This is solved for by:

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

    > This amounts to playing the actions more frequently which outperform the current mixed strategy. And if an action does not outperform the mixed strategy (i.e. its cumulated regret is negative) then never play it.

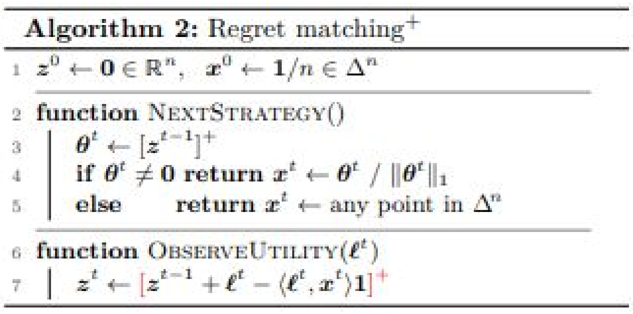
- Let r^t be the cumulated regret, $r^t := \sum_{\tau=1}^{\tau} \ell^\tau \,-\, \langle\, \ell^\tau,\, x^\tau \,\rangle 1$. Then, $x^{t+1} \propto [\phi^t]^+ \propto [r^t]^+$
- A nice property of regret matching is the lack of hyperparameters, it just works.
- Resulting algorithm:



```
Algorithm 1: Regret matching
$r^0 \leftarrow 0 \in \mathbb{R}^n, \quad x^0 \leftarrow 1/n \in \Delta^n$
function NextStrategy()
    $\theta^t \leftarrow [r^{t-1}]^+$
    if $\theta^t \neq 0$ return $x^t \leftarrow \theta^t \,/\, \|\theta^t\|_1$
    else      return $x^t \leftarrow$ any point in $\Delta^n$
function ObserveUtility($\ell^t$)
    $r^t \leftarrow r^{t-1} + \ell^t - \langle \ell^t, x^t \rangle 1$
```

1.5 - The Regret Matching$^+$ (RM$^+$) Algorithm

- RM$^+$ algorithm assigns zero regret to actions with negative cumulated regret:



```
Algorithm 2: Regret matching$^+$
$z^0 \leftarrow 0 \in \mathbb{R}^n, \quad x^0 \leftarrow 1/n \in \Delta^n$
function NextStrategy()
    $\theta^t \leftarrow [z^{t-1}]^+$
    if $\theta^t \neq 0$ return $x^t \leftarrow \theta^t \,/\, \|\theta^t\|_1$
    else      return $x^t \leftarrow$ any point in $\Delta^n$
function ObserveUtility($\ell^t$)
    $z^t \leftarrow [z^{t-1} + \ell^t - \langle \ell^t, x^t \rangle 1]^+$
```

- This has the practical advantage of adjusting more quicklym to when a bad action becomes good (i.e. an action initially with negative regret becomes positive over time).

- First introduced in [2014-Tammelin] and [2015-Tammelin].