5. Regret circuits and counterfactual regret minimization (CFR)

Link

---

1 - Regret Circuits

- Last lecture, we studied online optimization for simplex domains. In this section, we show how we can construct an efficient regret minimizer that outputs sequence-form strategies, by combining many local regret minimizers at each decision point.

- Recall from Lecture 2 that we can construct a sequence-form strategy space Q from the bottom up using convex hulls and Cartesian products.

- *Regret circuit* - local regret minimizer for a node in a tree-form decision problem.
- The resulting family of algorithms is known as counterfactual regret minimization (CFR).

1.1 - Regret Circuit for Cartesian Product

- Let $\mathcal{X} \subseteq \mathbb{R}^m$ and $\mathcal{Y} \subseteq \mathbb{R}^n$ be two sets, and let $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$ be regret minimizers for $\mathcal{X}$ and $\mathcal{Y}$ respectively.

- We can combine $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$ to form a regret minimizer for the Cartesian product $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^{m+n}$ by doing the following:

  - Output next strategy by asking $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$ for their strategies, x^t and y^t , and return the pair $(x^t, y^t) \in \mathcal{X} \times \mathcal{Y}$.

  - Decompose any utility vector $\ell \in \mathbb{R}^{m+n}$ into two constituent parts, $\ell = (\ell_\mathcal{X}, \ell_\mathcal{Y}) \in \mathbb{R}^m \times \mathbb{R}^n$.

    At time t, we forward the utility vector we receive $\ell^t = (\ell^t_\mathcal{X}, \ell^t_\mathcal{Y})$ onto $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$, respectively.

- Algorithm:

Algorithm 1: Regret minimizer for $\mathcal{X} \times \mathcal{Y}$

```
function NextStrategy()
  x^t ← R_X.NextStrategy()
  y^t ← R_Y.NextStrategy()
  return (x^t, y^t) ∈ X × Y
function ObserveUtility(ℓ^t = (ℓ^t_X, ℓ^t_Y) ∈ R^m × R^n)
  R_X.ObserveUtility(ℓ^t_X)
  R_Y.ObserveUtility(ℓ^t_Y)
```

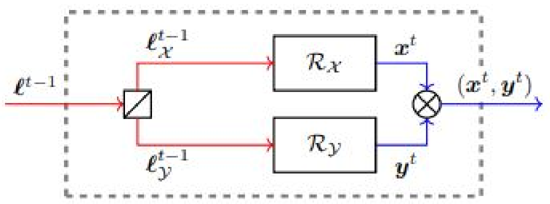

Figure 1: Regret circuit for the Cartesian product $\mathcal{X} \times \mathcal{Y}$.

- Cumulative regret analysis for $\mathcal{R}_{\mathcal{X}\times\mathcal{Y}}$:

  - Cumulative regret for $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$ up to any time T :

$$R^T_\mathcal{X} = \max_{\hat{x}\in\mathcal{X}} \left\{ \sum_{t=1}^{T} (\ell^t_\mathcal{X})^T \hat{x} \;-\; (\ell^t_\mathcal{X})^T x^t \right\}$$

$$R^T_\mathcal{Y} = \max_{\hat{y}\in\mathcal{Y}} \left\{ \sum_{t=1}^{T} (\ell^t_\mathcal{Y})^T \hat{y} \;-\; (\ell^t_\mathcal{Y})^T y^t \right\}$$

  - Using our definitions above for the cartesian product strategy and utility sets, the cumulative regret for $\mathcal{R}_{\mathcal{X}\times\mathcal{Y}}$ up to any time T :

$$R^T_{\mathcal{X}\times\mathcal{Y}} = \max_{(\hat{x},\hat{y})\in\mathcal{X}\times\mathcal{Y}} \left\{ \sum_{t=1}^{T} (\ell^t_\mathcal{X}, \ell^t_\mathcal{Y})^T\, (\hat{x}, \hat{y}) \;-\; (\ell^t_\mathcal{X}, \ell^t_\mathcal{Y})^T\, (\hat{x}, \hat{y}) \right\}$$

$$R^T_{\mathcal{X}\times\mathcal{Y}} = \max_{\hat{x}\in\mathcal{X}} \left\{ \sum_{t=1}^{T} (\ell^t_\mathcal{X})^T \hat{x} \;-\; (\ell^t_\mathcal{X})^T x^t \right\} \;+\; \max_{\hat{y}\in\mathcal{Y}} \left\{ \sum_{t=1}^{T} (\ell^t_\mathcal{Y})^T \hat{y} \;-\; (\ell^t_\mathcal{Y})^T y^t \right\}$$

$$R^T_{\mathcal{X}\times\mathcal{Y}} = R^T_\mathcal{X} \;+\; R^T_\mathcal{Y}$$

  - If $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$ guarantee sublinear regret, then so does $\mathcal{R}_{\mathcal{X}\times\mathcal{Y}}$.

    > In other words, we can minimize regret on $\mathcal{X} \times \mathcal{Y}$ by minimizing it on $\mathcal{X}$ and $\mathcal{Y}$ independently, then combining the decisions.

- Extensions to cartesian product of more than two sets is direct.

1.2 - Regret Circuit for Convex Hull

- In order to construct a regret minimizer for the convex hull $\text{co}\{\mathcal{X}, \mathcal{Y}\}$, we must use a third regret minimizer $\mathcal{R}_\Delta$ for the 2-simplex Δ^2 to decide how to "mix" x^t and y^t.

  where $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathcal{R}^n$.

- Constructing NextStrategy and ObserveUtility :

- Output next strategy by getting x^t and y^t from $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$.
  Then, we ask $\mathcal{R}_\Delta$ for its next strategy $\lambda^t = (\lambda_1^t, \lambda_2^t) \in \Delta^2$.
  And, return the convex combination $\lambda_1^t x^t + \lambda_2^t y^t \in \mathrm{co}\{\mathcal{X},\, \mathcal{Y}\}$.

- At time t, we receive the utility vector $\ell^t \in \mathbb{R}^n$ and forward it to $\mathcal{R}_\mathcal{X}$ and $\mathcal{R}_\mathcal{Y}$. Then, we forward the utility vector $\ell^t_\Delta := ((\ell^t)^T x^t,\, (\ell^t)^T y^t)$ to $\mathcal{R}_\Delta$.

- Algorithm:

Algorithm 2: Regret minimzer for $\mathrm{co}\{\mathcal{X}, \mathcal{Y}\}$

function NEXTSTRATEGY()
  $\boldsymbol{x}^t \leftarrow \mathcal{R}_\mathcal{X}$.NEXTSTRATEGY()
  $\boldsymbol{y}^t \leftarrow \mathcal{R}_\mathcal{Y}$.NEXTSTRATEGY()
  $\boldsymbol{\lambda}^t = (\lambda_1^t, \lambda_2^t) \in \Delta^2 \leftarrow \mathcal{R}_\Delta$.NEXTSTRATEGY()
  return $\lambda_1^t \boldsymbol{x}^t + \lambda_2^t \boldsymbol{y}^t \in \mathrm{co}\{\mathcal{X}, \mathcal{Y}\}$

function OBSERVEUTILITY($\boldsymbol{\ell}^t$)
  $\mathcal{R}_\mathcal{X}$.OBSERVEUTILITY($\boldsymbol{\ell}^t$)
  $\mathcal{R}_\mathcal{Y}$.OBSERVEUTILITY($\boldsymbol{\ell}^t$)
  $\mathcal{R}_\Delta$.OBSERVEUTILITY$\left(\boldsymbol{\ell}^t_\Delta := \begin{pmatrix} (\boldsymbol{\ell}^t)^\top \boldsymbol{x}^t \\ (\boldsymbol{\ell}^t)^\top \boldsymbol{y}^t \end{pmatrix}\right)$

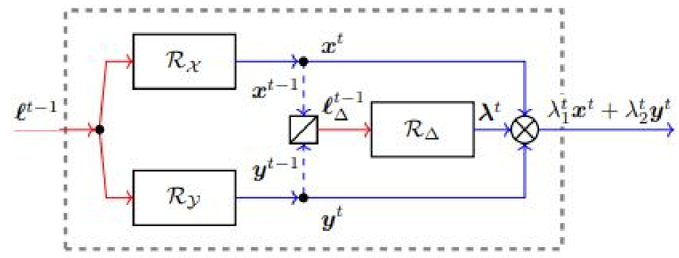

Figure 2: Regret circuit for the convex hull $\mathrm{co}\{\mathcal{X}, \mathcal{Y}\}$. The utility vector ℓ^t_λ is defined in Equation (1).

- Cumulative regret analysis for $\mathcal{R}_{\mathrm{co}\{\mathcal{X},\mathcal{Y}\}}$:

  - Regret minimizer $\mathcal{R}_\Delta$ outputs the strategy $\lambda^t = (\lambda_1^t, \lambda_2^t)$ and observes the utility ℓ^t_Δ.

    Regret for any time up to T :

$$R^T_\Delta = \max_{\hat{\lambda} \in \Delta^2} \left\{ \left(\sum_{t=1}^{T} \hat{\lambda}_1 \, (\ell^t)^T \, x^t + \hat{\lambda}_2 \, (\ell^t)^T \, y^t \right) \right\} - \left(\sum_{t=1}^{T} \lambda_1^t \, (\ell^t)^T \, x^t + \lambda_2^t \, (\ell^t)^T \, y^t \right)$$

  - Regret of $\mathcal{R}_{\mathrm{co}\{\mathcal{X},\mathcal{Y}\}}$, for any time up to T :

$$R^T = \max_{\lambda \in \hat{\Delta}^2,\, \hat{x} \in \mathcal{X},\, \hat{y} \in \mathcal{Y}} \left\{ \sum_{t=1}^{T} (\ell^t)^T \, (\hat{\lambda}_1 \, \hat{x} + \hat{\lambda}_2 \, \hat{y}) \; - \; (\ell^t)^T \, (\lambda_1^t \, x^t + \lambda_2^t \, y^t) \right\}$$

$$R^T = \max_{\lambda \in \hat{\Delta}^2,\, \hat{x} \in \mathcal{X},\, \hat{y} \in \mathcal{Y}} \left\{ \lambda_1 \sum_{t=1}^{T} (\ell^t)^T \, \hat{x} \; + \lambda_2 \sum_{t=1}^{T} (\lambda^t)^T \, \hat{y} \right\} - \left(\sum_{t=1}^{T} \lambda_1^t \, (\ell^t)^T \, x^t + \lambda_2^t \, (\ell^t)^T \, y^t \right)$$

  - Observe

$$\max_{\lambda \in \hat{\Delta}^2,\, \hat{x} \in \mathcal{X},\, \hat{y} \in \mathcal{Y}} \left\{ \lambda_1 \sum_{t=1}^{T} (\ell^t)^T \, \hat{x} \; + \lambda_2 \sum_{t=1}^{T} (\lambda^t)^T \, \hat{y} \right\} = \max_{\hat{\lambda} \in \Delta^2} \left\{ \lambda_1 \max_{\hat{x} \in \mathcal{X}} \left\{ \sum_{t=1}^{T} (\ell^t)^T \, \hat{x} \right\} + \lambda_2 \max_{\hat{y} \in \mathcal{Y}} \left\{ \sum_{t=1}^{T} (\lambda^t)^T \, \hat{y} \right\} \right\}$$

    and

$$\max_{\hat{x} \in \mathcal{X}} \left\{ \sum_{t=1}^{T} (\ell^t)^T \, \hat{x} \right\} = R^T_\mathcal{X} + \sum_{t=1}^{T} (\ell^t)^T \, x^t$$

  - Inserting these equations into the regret and doing some algebra:

$$R^T = \max_{\hat{\lambda} \in \Delta^2} \left\{ \left(\sum_{t=1}^{T} \hat{\lambda}_1 \, (\ell^t)^T \, x^t + \hat{\lambda}_2 \, (\ell^t)^T \, y^t \right) + \left(\hat{\lambda}_1 \, R^T_\mathcal{X} + \hat{\lambda}_2 \, R^T_\mathcal{Y} \right) \right\} - \left(\sum_{t=1}^{T} \lambda_1^t \, (\ell^t)^T \, x^t + \lambda_2^t \, (\ell^t)^T \, y^t \right)$$

    Note:

$$\hat{\lambda}_1 \, R^T_\mathcal{X} + \hat{\lambda}_2 \, R^T_\mathcal{Y} \le \max\{R^T_\mathcal{X} \, , \, R^T_\mathcal{Y}\}$$

  - Inserting and rearranging terms:

$$R^T \le R^T_\Delta + \max\{R^T_\mathcal{X} \, , \, R^T_\mathcal{Y}\}$$

    This guarantees R^T grows sublinearly if the set $\{R^T_\Delta,\, R^T_\mathcal{X} \, , \, R^T_\mathcal{Y}\}$ grows sublinearly.

- Extension to multiple sets:

$$R^T \le R^T_\Delta + \max\{R^T_{\mathcal{X}_1} \, , \, \dots \, R^T_{\mathcal{X}_2}\}$$

2 - Counterfactual Regret Minimization

- *Counterfactual regret (CFR)* - regret minimizer for a sequence-form strategy space, constructed recursively using regret circuits.

- ρ - transition function of the process, selecting action a at decision point j results in the transition $\rho(j,\ a) \in \mathcal{J} \cup \mathcal{K} \cup \{\perp\}$, or for observation points $\rho(k,\ s) \in \mathcal{J} \cup \mathcal{K} \cup \{\perp\}$

- Algorithm:

Algorithm 3: CFR regret minimizer

Data: $\mathcal{R}_j$, one regret minimizer for $\Delta^{|A_j|}$; one for each decision point $j \in \mathcal{J}$ of the TFSDP.

function NextStrategy()
- [▷ Step 1: we ask each of the $\mathcal{R}_j$ for their next strategy local at each decision point]
- **for each** decision point $j \in \mathcal{J}$ **do**
  - $\boldsymbol{b}_j^t \in \Delta^{|A_j|} \leftarrow \mathcal{R}_j$.NextStrategy()
- [▷ Step 2: we construct the sequence-form representation of the strategy that plays according to the distribution b_j^t at each decision point $j \in \mathcal{J}$]
- $\boldsymbol{x}^t = \mathbf{0} \in \mathbb{R}^{|\Sigma|}$
- **for each** decision point $j \in \mathcal{J}$ in *top-down traversal* order in the TFSDP **do**
  - **for each** action $a \in A_j$ **do**
    - **if** $p_j = \varnothing$ **then**
      - $\boldsymbol{x}^t[ja] \leftarrow \boldsymbol{b}_j^t[a]$
    - **else**
      - $\boldsymbol{x}^t[ja] \leftarrow \boldsymbol{x}^t[p_j] \cdot \boldsymbol{b}_j^t[a]$
- [▷ You should convince yourself that the vector x^t we just filled in above is a valid sequence-form strategy, that is, it satisfies the required consistency constraints we saw in Lecture 2. In symbols, $x^t \in Q$]
- **return** $\boldsymbol{x}^t$

function ObserveUtility($\boldsymbol{\ell}^t \in \mathbb{R}^{|\Sigma|}$)
- [▷ Step 1: we compute the expected utility for each subtree rooted at each node $v \in \mathcal{J} \cup \mathcal{K}$]
- $V^t \leftarrow$ empty dictionary [▷ eventually, it will map keys $\mathcal{J} \cup \mathcal{K} \cup \{\perp\}$ to real numbers]
- $V^t[\perp] \leftarrow 0$
- **for each** node in the tree $v \in \mathcal{J} \cup \mathcal{K}$ in *bottom-up traversal* order in the TFSDP **do**
  - **if** $v \in \mathcal{J}$ **then**
    - Let $j = v$
    - $V^t[j] \leftarrow \sum_{a \in A_j} \boldsymbol{b}_j^t[a] \cdot \left(\boldsymbol{\ell}^t[ja] + V^t[\rho(j,a)]\right)$
  - **else**
    - Let $k = v$
    - $V^t[k] \leftarrow \sum_{s \in S_k} V^t[\rho(k,s)]$
- [▷ Step 2: at each decision point $j \in \mathcal{J}$, we now construct a local utility vector ℓ_j^t called *counterfactual utility*]
- **for each** decision point $j \in \mathcal{J}$ **do**
  - $\boldsymbol{\ell}_j^t \leftarrow \mathbf{0} \in \mathbb{R}^{|A_j|}$
  - **for each** action $a \in A_j$ **do**
    - $\boldsymbol{\ell}_j^t[a] \leftarrow \boldsymbol{\ell}^t[ja] + V^t[\rho(j,a)]$
  - $\mathcal{R}_j$.ObserveUtility($\boldsymbol{\ell}_j^t$)