Solving Normal-Form Games

Brian Zhang

Recap: Normal-Form Games

			
0.2 	0	-1	+1
0.5 	+1	0	-1
0.3 	-1	+1	0

- SIMULTANEOUS

(No turns)

- Strategy for a player is just a probability distribution over actions

Two-Player Zero-Sum Normal-Form Games

- NE doesn’t have problems as in general-sum or multiplayer games
- In a sense, NE is optimal in that no opponent can exploit you
  - If I were to play any other strategy than ⅓, ⅓, ⅓ in rock paper scissors, you could exploit me
- NE can leave utility on the table against imperfect opponents
  - If you always play Rock, NE will still just play ⅓, ⅓, ⅓
- But this is a price usually worth paying when playing experts or other AI programs

	R	P	S
R	0	-1	1
P	1	0	-1
S	-1	1	0

Computing NE in Two-Player Zero-Sum Normal-Form Games (This Lecture)

1. LP for small games
2. Iterative Approaches
   - Best-Response Dynamics (doesn’t converge)
   - Fictitious Play aka Follow the Leader (FTL)
3. No-Regret Algorithms
   - Follow the Regularized Leader (FTRL)
   - Regret Matching
4. Optimistic regret minimization

Running example:
Weighted RPS

	R	P	S
R	0	-2	1
P	2	0	-1
S	-1	1	0

LP Approach

$$\max_{\boldsymbol{x} \in \Delta^m} \min_{\boldsymbol{y} \in \Delta^n} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

			$\boldsymbol{y}$		
			1/4	1/4	1/2
			R	P	S
$\boldsymbol{x}$	1/4	R	0	-2	1
	1/4	P	2	0	-1
	1/2	S	-1	1	0

LP Approach

$$\max_{\boldsymbol{x}\in\mathbb{R}^m} \left\{ \begin{array}{l} \min_{\boldsymbol{y}\in\mathbb{R}^n} \ \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y} \\ \text{s.t. } \boldsymbol{1}^\top \boldsymbol{y} = \boldsymbol{1}, \\ \qquad \boldsymbol{y} \geq \boldsymbol{0} \end{array} \right.$$

$$\text{s.t. } \boldsymbol{1}^\top \boldsymbol{x} = \boldsymbol{1}, \quad \boldsymbol{x} \geq \boldsymbol{0}$$

LP duality

$$\max_{v\in\mathbb{R}} \ v$$

$$\text{s.t. } \boldsymbol{A}^\top \boldsymbol{x} \geq \boldsymbol{1} v$$

find the largest value v s.t.

every strategy of the opponent gives us expected value at least $\boldsymbol{v}$

$\boldsymbol{x}$ \ $\boldsymbol{y}$		1/4	1/4	1/2
		R	P	S
1/4	R	0	-2	1
1/4	P	2	0	-1
1/2	S	-1	1	0

LP Approach

$$\max_{\substack{\boldsymbol{x}\in\mathbb{R}^m \\ v\in\mathbb{R}}} v$$

find the largest value v s.t. for some $\boldsymbol{x}$

$$\text{s.t.}\ \mathbf{1}^\top \boldsymbol{x} = 1,\quad \boldsymbol{x} \geq \mathbf{0}$$

$\boldsymbol{x}$ is a valid mixed strategy

$$\boldsymbol{A}^\top \boldsymbol{x} \geq \mathbf{1}v$$

every strategy of the opponent gives us expected value at least $\boldsymbol{v}$

find the largest value v s.t.

every strategy of the opponent gives us expected value at least $\boldsymbol{v}$

$$\max_{v\in\mathbb{R}} v$$

$$\text{s.t.}\quad \boldsymbol{A}^\top \boldsymbol{x} \geq \mathbf{1}v$$

				$\boldsymbol{y}$	
			1/4	1/4	1/2
			R	P	S
	1/4	R	0	-2	1
$\boldsymbol{x}$	1/4	P	2	0	-1
	1/2	S	-1	1	0

LP Approach

- Solving our game results in the following
- We maximize the value that the opponent can get against us
- Any deviation would allow the opponent to exploit us more

		R	P	S
1/4	R	0	-2	1
1/4	P	2	0	-1
1/2	S	-1	1	0

	R	P	S
EV	0	0	0

For **P2's** strategy: take dual values of constraint $\boldsymbol{A}^\top \boldsymbol{x} \geq \mathbf{1}v$, or solve

$$\min_{\boldsymbol{y} \in \Delta^n} \max_{\boldsymbol{x} \in \Delta^m} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

Iterative Approaches

We'll make this precise soon

- Only relatively small games can be solved via LP
- For larger games we need iterative approaches
- Most iterative approaches *approach* a NE *on average*
  - Can be stopped any time
- What we'll cover
  - Best Response Dynamics (doesn't converge to NE)
  - Fictitious Play aka Follow the Leader (isn't no-regret)
  - Follow the Regularized Leader (*e.g.,* gradient descent, multiplicative weights)
  - Regret Matching
  - Regret Matching Plus
  - Optimistic regret minimization

Best Reponse Dynamics

$$x_i^{t+1} = \arg\max_{x_i} \; u_i(x_i, x_{-i}^t)$$

*Best respond to the opponent's **last** strategy*

Question: Does

$$\frac{1}{T}\sum_{t=1}^{T} x_i^t \xrightarrow{T\to\infty} \text{NE?}$$

No!

$$\frac{1}{T}\sum_{t=1}^{T} x_i^t \to \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \neq \text{NE} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/2 \end{bmatrix}$$

	R	P	S
R	0	-2	1
P	2	0	-1
S	-1	1	0

Fictitious Play (Follow the Leader)

$$x_i^{t+1} = \arg\max_{x_i} \frac{1}{t}\sum_{\tau=1}^{t} u_i(x_i, x_{-i}^{\tau})$$

Best respond to the opponent's ***average*** *strategy*

Question: Does

$$\frac{1}{T}\sum_{t=1}^{T} x_i^t \xrightarrow{T\to\infty} \text{NE?}$$

Yes! (for zero-sum games)
[Robinson 1951]

...but possibly with very slow rate $T^{-1/n}$ if the tiebreaking is done adversarially [Daskalakis & Pan 2014]

Open question [“Karlin’s weak conjecture”, Karlin 1959]: Does FP *with non-adversarial tiebreaking* converge with rate $O_n(T^{-1/2})$ in all zero-sum normal-form games?

No-Regret Algorithms

- What if I’m playing a repeated game against someone who knows I am playing fictitious play?
- Then they would know exactly what my next move will be and could choose a best response every time
- Can we find iterative algorithms that will not be *too bad* even when the opponent knows the algorithm?
- No-regret algorithms do exactly this
    - And achieve faster convergence than FP as well!

Regret Minimization

for $t = 1, \ldots, T$:

- Agent chooses an *action distribution* $\boldsymbol{x}^t \in X \coloneqq \Delta^n$
- Environment chooses a *utility vector* $\boldsymbol{u}^t \in [0, 1]^n$
- Agent observes $\boldsymbol{u}^t$ and gets utility $\langle \boldsymbol{u}^t, \boldsymbol{x}^t \rangle$

Δ^n = set of distributions on n things $= \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq 0, \sum \boldsymbol{x}_i = 1\}$

Agent goal: Minimize *regret*.

"How well do we do against best, fixed strategy in hindsight?"

$$R^T \coloneqq \max_{\hat{\boldsymbol{x}} \in X} \left\{ \sum_{t=1}^{T} \langle \boldsymbol{u}^t, \hat{\boldsymbol{x}} \rangle \right\} - \sum_{t=1}^{T} \langle \boldsymbol{u}^t, \boldsymbol{x}^t \rangle$$

Maximum utility that was achievable by the **best fixed** action in hindsight

Utility that was actually accumulated

Goal: have R^T grow sublinearly with respect to time T, e.g., $R^T = O_n(\sqrt{T})$

No assumption on utilities!
Must be able to handle adversarial environments

What does regret minimization have to do with zero-sum games?

Nash equilibrium in a 2-player 0-sum normal-form game with payoff matrix A:

$$\max_{x \in \Delta^m} \min_{y \in \Delta^n} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

IDEA: Self-play. Make two regret minimizers play each other

for $t = 1, \ldots, T$:

- $\boldsymbol{x}^t \leftarrow$ request strategy from P1's regret minimizer
- $\boldsymbol{y}^t \leftarrow$ request strategy from P2's regret minimizer
- Pass utility $\boldsymbol{A}\boldsymbol{y}^t$ to P1's regret minimizer
- Pass utility $-\boldsymbol{A}^\top \boldsymbol{x}^t$ to P2's regret minimizer

$$R_1^T \coloneqq \max_{\hat{x} \in \Delta^m} \left\{ \sum_{t=1}^{T} \langle \boldsymbol{A}\boldsymbol{y}^t, \hat{\boldsymbol{x}} \rangle \right\} - \sum_{t=1}^{T} \langle \boldsymbol{A}\boldsymbol{y}^t, \boldsymbol{x}^t \rangle \le O_m(\sqrt{T})$$

$$R_2^T \coloneqq \max_{\hat{y} \in \Delta^n} \left\{ \sum_{t=1}^{T} \langle -\boldsymbol{A}^\top \boldsymbol{x}^t, \hat{\boldsymbol{y}} \rangle \right\} - \sum_{t=1}^{T} \langle -\boldsymbol{A}^\top \boldsymbol{x}^t, \boldsymbol{y}^t \rangle \le O_n(\sqrt{T})$$

Add these two lines and divide by T to get the average

$$\max_{\hat{x} \in \Delta^m} \{\hat{\boldsymbol{x}}^\top \boldsymbol{A} \bar{\boldsymbol{y}}\} - \min_{\hat{y} \in \Delta^n} \{\bar{\boldsymbol{x}}^\top \boldsymbol{A} \hat{\boldsymbol{y}}\} \le O_{m,n}\left(\frac{1}{\sqrt{T}}\right)$$

where $\bar{\boldsymbol{x}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}^t$ and $\bar{\boldsymbol{y}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{y}^t$

TAKEAWAY

The average strategies converge to a Nash equilibrium!

Regret Minimization: Follow the Leader (Fictitious Play)

First attempt: Follow the leader. That is, play the best action in hindsight so far:

$$\boldsymbol{x}^{t+1} = \arg\max_{\boldsymbol{x}\in X} \sum_{\tau\le t} \langle \boldsymbol{u}^\tau, \boldsymbol{x} \rangle$$

This does not work!

Counterexample: $n = 2$ actions,

$$\boldsymbol{u}^t = \begin{cases} [{}^1/_2, 0] & t = 1 \\ [0, 1] & t > 1, \text{even} \\ [1, 0] & t > 1, \text{odd} \end{cases}$$

Best action in hindsight has utility $\approx T/2$

Follow-the-leader always plays the wrong action and therefore gets utility ≈ 0

More generally: No algorithm outputting only pure actions can have no regret

Follow the *Regularized* Leader

Idea: Add a strictly convex *regularizer* $R : X \to \mathbb{R}$

$$\boldsymbol{x}^{t+1} = \arg\max_{\boldsymbol{x} \in X} \sum_{\tau \le t} \langle \boldsymbol{u}^{\tau}, \boldsymbol{x} \rangle - \frac{1}{\eta} R(\boldsymbol{x})$$

- This prevents each iterate from being deterministic
- The resulting algorithm **is** no-regret (for $\eta \propto 1/\sqrt{T}$)
- Intuitively, **updates toward high-regret actions, but not too much**

Follow the leader will always play deterministic actions

Follow the regularized leader will mix

Follow the *Regularized* Leader

Idea: Add a strictly convex *regularizer* $R : X \to \mathbb{R}$

$$\boldsymbol{x}^{t+1} = \arg\max_{\boldsymbol{x}\in X} \sum_{\tau\le t} \langle \boldsymbol{u}^{\tau}, \boldsymbol{x}\rangle - \frac{1}{\eta} R(\boldsymbol{x})$$

Example 1: *quadratic*

$$R(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x}\|_2^2$$

Closed-form optimization:

Π_X = projection onto X

$$\boldsymbol{x}^{t+1} = \Pi_X\left(\eta \cdot \sum_{\tau=1}^{t} \boldsymbol{u}^{\tau}\right)$$

a.k.a. gradient descent!

Follow the leader will always play deterministic actions

Follow the regularized leader will mix

Follow the *Regularized* Leader

Idea: Add a strictly convex *regularizer* $R : X \to \mathbb{R}$

$$\boldsymbol{x}^{t+1} = \arg\max_{\boldsymbol{x} \in X} \sum_{\tau \le t} \langle \boldsymbol{u}^{\tau}, \boldsymbol{x} \rangle - \frac{1}{\eta} R(\boldsymbol{x})$$

Example 2: *negative entropy*

$$R(\boldsymbol{x}) = \sum_{a} x[a] \log x[a]$$

Closed-form optimization:

$$\boldsymbol{x}^{t+1} \propto \exp\left(\eta \cdot \sum_{\tau=1}^{t} \boldsymbol{u}^{\tau}\right)$$

a.k.a. multiplicative weights update (MWU), hedge, (discrete-time) replicator dynamics, randomized weighted majority, …

Follow the leader will always play deterministic actions

Follow the regularized leader will mix

A Common Template for Regret Minimizers

- Given utility vectors $\boldsymbol{u}^1, \ldots, \boldsymbol{u}^t$, we compute the empirical regrets up to time t of each action:

$$r^t[a] \coloneqq \sum_{\tau=1}^{t} \left(u^\tau[a] - \langle \boldsymbol{u}^\tau, \boldsymbol{x}^\tau \rangle\right)$$

- Then, intuitively the next strategy $\boldsymbol{x}^{t+1}$ gives mass to actions in a manner related to how much regret they have accumulated

A Common Template for Regret Minimizers

Empirical regret:

Hyperparameter ("learning rate")

$$r^t[a] := \sum_{\tau=1}^{t} (u^\tau[a] - \langle \boldsymbol{u}^\tau, \boldsymbol{x}^\tau \rangle)$$

Algorithm	Rule
Gradient descent	$\boldsymbol{x}^{t+1} = \Pi_X(\eta \cdot \boldsymbol{r}^t)$
Multiplicative weights update (MWU) (aka Hedge, Randomized Weighted Majority, …)	$\boldsymbol{x}^{t+1} \propto \exp(\eta \cdot \boldsymbol{r}^t)$
Regret matching (RM) [Hart & Mas-Collel 2000]	$\boldsymbol{x}^{t+1} \propto \max\{0, \boldsymbol{r}^t\}$

No learning rate.
Scale-invariant!

RM Regret Bound Proof

$\mathbf{x}^{t+1} \propto [\boldsymbol{r}^t]^+$ where

$$\boldsymbol{r}^t := \boldsymbol{r}^{t-1} + \boldsymbol{g}^t$$
$$\boldsymbol{g}^t := \boldsymbol{u}^t - \langle \boldsymbol{u}^t, \boldsymbol{x}^t \rangle \cdot \mathbf{1}$$

Note: $\langle \boldsymbol{g}^t, \boldsymbol{x}^t \rangle = 0$

$$\|[\boldsymbol{r}^{t+1}]^+\|_2^2 \leq \|[\boldsymbol{r}^t]^+ + \boldsymbol{g}^{t+1}\|_2^2$$

using inequality $[x+y]^+ \leq |[x]^+ + y|$ for $x, y \in \mathbb{R}$

$$= \|[\boldsymbol{r}^t]^+\|_2^2 + \|\boldsymbol{g}^{t+1}\|_2^2 + 2(\boldsymbol{g}^{t+1})^\top [\boldsymbol{r}^t]^+$$

→ **0**

induction

$$\|[\boldsymbol{r}^T]^+\|_2^2 \leq \sum_{t=1}^{T} \|\boldsymbol{g}^t\|_2^2 \leq nT$$

since $\boldsymbol{g}^t \in [-1,1]^n$

$$R^T := \max_a r^T[a] \leq \|[\boldsymbol{r}^T]^+\|_2 \leq \sqrt{nT} \qquad \square$$

A Common Template for Regret Minimizers

Empirical regret: $\boldsymbol{r}^t \coloneqq \boldsymbol{r}^{t-1} + \boldsymbol{g}^t$

Simple modification: $\boldsymbol{r}_+^t \coloneqq [\boldsymbol{r}_+^{t-1} + \boldsymbol{g}^t]^+$

(Floor regrets at 0 after every iteration)

Algorithm	Rule
Gradient descent	$\boldsymbol{x}^{t+1} = \Pi_X(\eta \cdot \boldsymbol{r}^t)$
Multiplicative weights update (MWU) (aka Hedge, aka Randomized Weighted Majority)	$\boldsymbol{x}^{t+1} \propto \exp\{\eta \cdot \boldsymbol{r}^t\}$
Regret matching (RM) [Hart & Mas-Collel 2000]	$\boldsymbol{x}^{t+1} \propto [\boldsymbol{r}^t]^+$
Regret matching plus (RM+) [Tammelin 2014]	$\boldsymbol{x}^{t+1} \propto [\boldsymbol{r}_+^t]^+$

(Regret bound proof is identical)

A Common Template for Regret Minimizers

All of these algorithms guarantee that after seeing any number T of utilities $\boldsymbol{u}^1, \dots, \boldsymbol{u}^T$, the regret cumulated by the algorithm satisfies

$$R^T \leq C\sqrt{T}$$

Constant that depends on number of actions

MWU: $C = \sqrt{\log n}$

RM, RM+, GD: $C = \sqrt{n}$

Remember:
This holds without any assumption about the way the utilities are selected by the environment!

Consequence: when using these algorithms in self-play in 2-player 0-sum games, the **average strategy** converges to a Nash equilibrium at a rate of $C/\sqrt{T}$

Reminder: Self-play

for $t = 1, \dots, T$:

- $\boldsymbol{x}^t \leftarrow$ request strategy from P1's regret minimizer
- $\boldsymbol{y}^t \leftarrow$ request strategy from P2's regret minimizer
- Pass utility $\boldsymbol{A}\boldsymbol{y}^t$ to P1's regret minimizer
- Pass utility $-\boldsymbol{A}^\top \boldsymbol{x}^t$ to P2's regret minimizer

State-of-the-art variant in practice: Discounted RM (DRM)

- Linear RM (LRM)
  - Weight iteration t by t (in regrets and averaging)
  - RM+ floors regrets at 0. Can we combine this with linear RM? Theory: Yes. Practice: No! Does very poorly.
- But less-aggressive combinations do well: **Discounted RM**
  - On each iteration, multiply positive regrets by $t^\alpha / (t^\alpha+1)$
  - On each iteration, multiply negative regrets by $t^\beta / (t^\beta+1)$
  - Weight contributions toward average strategy on iteration t by t^γ
  - Worst-case convergence bound only a small constant worse than that of RM
  - RM: $\alpha = \beta = +\infty$
  - RM+: $\alpha = +\infty, \ \beta = -\infty$
  - For $\alpha = 1.5$, $\beta = 0$, $\gamma = 2$, consistently outperforms RM+ in practice

Brown & Sandholm (*AAAI* 2019), "Solving Imperfect-Information Games via Discounted Regret Minimization"

What Regret Minimizers are Used in Practice?

Follow the Regularized Leader (FTRL)
(*e.g.*, gradient descent, multiplicative weights)

- ✔ Works for general convex sets
- ✔ Widely used & understood
- ✘ Slow in practice
- ✘ Has hyperparameters (stepsize)

- ✔ Can incorporate optimism about future losses to converge faster in 2-player 0-sum games

Regret Matching (RM)
& Regret Matching+ (RM+)

- ✘ Only for **simplex** domains
- ✘ Not as well studied theoretically
- ✔ Fast in practice
- ✔ No hyperparameters

Modern variants of this, such as DCFR, are the standard in extensive-form game solving!

- ? Unknown

...until recently ✔

Optimistic (Predictive) Regret Minimizers

Algorithm	Standard (non-optimistic) rule	Optimistitic (aka Predictive) rule
GD	$\boldsymbol{x}^{t+1} = \Pi_X(\eta \cdot \boldsymbol{r}^t)$	Replace $\boldsymbol{r}^t$ with $\boldsymbol{r}^t + \boldsymbol{g}^t$
MWU	$\boldsymbol{x}^{t+1} \propto \exp\{\eta \cdot \boldsymbol{r}^t\}$	
RM	$\boldsymbol{x}^{t+1} \propto [\boldsymbol{r}^t]^+$	
RM+	$\boldsymbol{x}^{t+1} \propto [\boldsymbol{r}_+^t]^+$	

Typically, one-line change in implementation

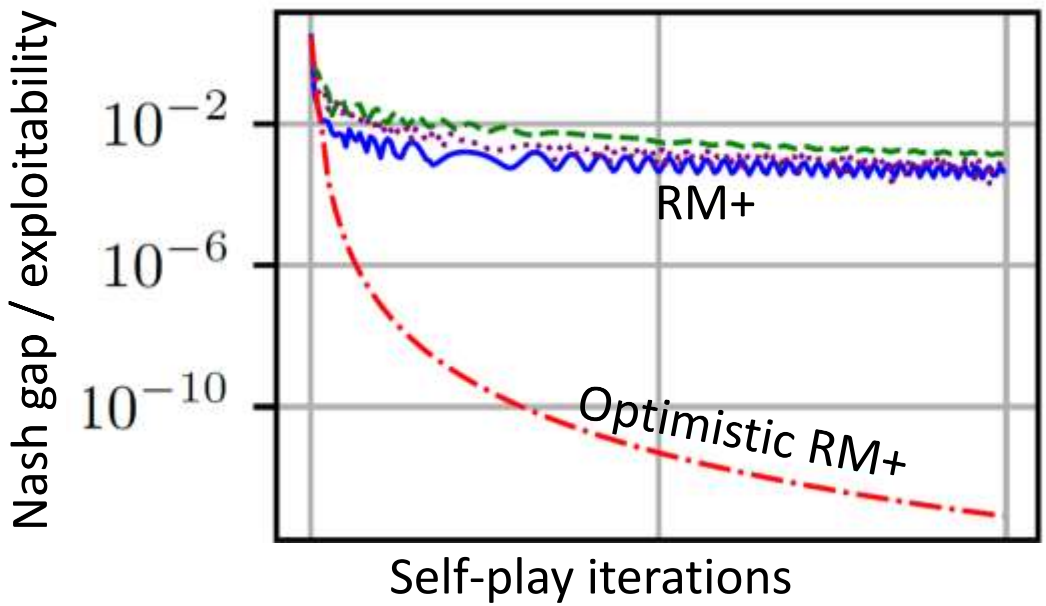
All of these algorithms guarantee that after seeing any number T of utilities $\boldsymbol{u}^1, \dots, \boldsymbol{u}^T$, the regret cumulated by the algorithm satisfies

$$R^T \leq C \sqrt{\sum_{t=1}^{T} \|\boldsymbol{g}^t - \boldsymbol{g}^{t-1}\|_2^2} \qquad \text{(where } \boldsymbol{g}^0 \coloneqq \boldsymbol{0}\text{)}$$

Remember:
This holds without any assumption about the way the utilities are selected by the environment!

Takeaway message: still $\approx \sqrt{T}$ regret, but much smaller when there is little change to the utilities over time

Empirical Performance



Green dashed: Linear RM
Violet dotted: Discounted RM

(RM was omitted as it is typically much slower than RM+)

[Farina, Kroer, and Sandholm; Faster Game Solving via Predictive Blackwell Approachability: Connecting Regret Matching and Mirror Descent, AAAI 2021]

Practical State-of-the-Art

- In general, Discounted RM and Optimistic RM+ are the fastest in practice
  - For some games, like poker, Discounted RM is empirically consistently faster than Optimistic RM+
  - For many other games, Optimistic RM+ is significantly faster

[Farina, Kroer, and Sandholm; Faster Game Solving via Predictive Blackwell Approachability: Connecting Regret Matching and Mirror Descent, AAAI 2021]

Beyond Zero-Sum Games

Correlated strategy profile:

Note: not $\Delta(A_1) \times \cdots \times \Delta(A_n)$

$$\mu^T := \frac{1}{T} \sum_{t=1}^{T} \underbrace{(x_1^t \otimes x_2^t \otimes \cdots x_n^t)} \in \Delta(A_1 \times \cdots \times A_n)$$

the product distribution in $\Delta(A_1) \times \cdots \times \Delta(A_n)$
whose marginal on A_i *is* $x_i^t \in \Delta(A_i)$

Regret guarantee: for all players i:

$$\max_{x_i^*} \frac{1}{T} \sum_{t=1}^{T} \left[u_i(x_i^*, x_{-i}^t) - u_i(x_i^t, x_{-i}^t)\right] \le O_n\left(\frac{1}{\sqrt{T}}\right)$$

$$= \max_{x_i^*} \mathop{\mathbb{E}}_{x \sim \mu^T} \left[u_i(x_i^*, x_{-i}) - u_i(x_i, x_{-i})\right]$$

μ^T is an ϵ-*"coarse-correlated equilibrium" (CCE)* where $\epsilon = O_n(1/\sqrt{T})$

Note: A CCE that happens to be a product distribution ($\mu^T \in \Delta(A_1) \times \cdots \times \Delta(A_n)$) is a Nash equilibrium

References

Fictitious play:

- J Robinson (*Ann. Math.* 1951), "An iterative method of solving a game"
- C Daskalakis, Q Pan (*FOCS* 2014), "A Counter-Example to Karlin's Strong Conjecture for Fictitious Play"
- S Karlin (1959), *Mathematical Methods and Theory in Games, Programming, and Economics*

Blackwell Approachability (used in the original correctness proof of RM/RM+):

- D Blackwell (*Pacific J. of Math*. 1956), "An analog of the minmax theorem for vector payoffs"

Regret Matching and Regret Matching Plus:

- S Hart, A Mas-Colell (*Econometrica* 2000), "A Simple Adaptive Procedure Leading to Correlated Equilibrium"
- O Tammelin (*arXiv* 2014), "Solving large imperfect information games using CFR+"
- N Brown, T Sandholm (*AAAI* 2019), "Solving Imperfect-Information Games via Discounted Regret Minimization"
- **Simple proof of correctness presented in this lecture due to** G Farina (2023), https://www.mit.edu/~gfarina/2023/6S890f23_L05_learning_algorithms/L05.pdf

Predictivity:

- CK Chiang et al. (*COLT* 2012), "Online optimization with gradual variations"
- A Rakhlin, K Sridharan (*COLT* 2013), "Online Learning with Predictable Sequences"
- G Farina, C Kroer, T Sandholm (*AAAI* 2021), "Faster Game Solving via Predictive Blackwell Approachability: Connecting Regret Matching and Mirror Descent"