Learning and Correlated Equilibria

Algorithmic Game Theory

Winter 2023/24

Imitating Experts

No-Regret Algorithms and Coarse-Correlated Equilibria

Zero-Sum Games

Correlated Equilibria

Expert Problem: Example

In many applications we make decisions not once but repeatedly, say, every day, without knowing the behavior of other actors or "nature" on that day. We consider learning algorithms that enable us to cope with such problems.

Example:

- Suppose you are commuting day by day from home to Campus Bockenheim and back.
- The traveling time per day is between 30 and 60 minutes depending on the chosen route and the traffic situation.
- Suppose you know, say, three **experts** that are also commuting from your area to campus and use different strategies for choosing the route.

We will show that you can become almost as fast as the best expert just by imitating the expert choices.

Expert Problem: Definition

Assume an adversarial online model with discrete time steps $1, \dots, T$. Let $[T]$ denote $\{1, \dots, T\}$.

Experts and Losses

- There are N *experts* numbered from 1 to N.
- In step $t \in [T]$, expert $i \in [N]$ experiences a *loss* of $\ell_i^t \in [0, 1]$ (as chosen by an adversary or "nature").
- Let $L_i^t = \sum_{k=1}^{t} \ell_i^k$.

Combining Experts

- In step t, an *online algorithm* H chooses expert $i \in [N]$ with probability p_i^t.
- The vector p^t might depend on the loss vectors $\ell^1, \dots, \ell^{t-1}$.
- The (expected) loss of H in step t is $\ell_H^t = \sum_{i \in [N]} p_i^t \ell_i^t$.
- Let $L_H^t = \sum_{k=1}^{t} \ell_H^k$.

Imitating Experts

No-Regret Algorithms and Coarse-Correlated Equilibria

Zero-Sum Games

Correlated Equilibria

Greedy Algorithm

In the following, let $L_{min}^{t-1} = \min_{i \in [N]} L_i^{t-1}$, for $1 \leq t \leq T$.

Greedy Algorithm

At every time t,

- let $S^{t-1} = \{i : L_i^{t-1} = L_{min}^{t-1}\}$;
- let $j = \min\{S^{t-1}\}$;
- set $p_j^t = 1$, and $p_i^t = 0$, for $i \neq j$.

In the analysis of the Greedy algorithm, we assume for simplicity that all losses are either 0 or 1 instead of real numbers from $[0, 1]$.

Greedy Algorithm

Example:

ℓ_1	1	0	0	1	0	0	1	0	0	1	0	0	1	0
L_1	1	1	1	2	2	2	3	3	3	4	4	4	5	5
ℓ_2	0	1	0	0	1	0	0	1	0	0	1	0	0	1
L_2	0	1	1	1	2	2	2	3	3	3	4	4	4	5
ℓ_3	0	0	1	0	0	1	0	0	1	0	0	1	0	0
L_3	0	0	1	1	1	2	2	2	3	3	3	4	4	4
j	1	2	3	1	2	3	1	2	3	1	2	3	1	2
ℓ_{Greedy}	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L_G	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Greedy Algorithm

Theorem

The Greedy algorithm, for any sequence of losses from $\{0,1\}$*, has*

$$L_G^T \leq N \cdot L_{min}^T + (N-1).$$

Proof:

- Partition the sequence into phases $0, \ldots, L_{min}^T$ such that

  Every step t with $L_{min}^{t-1} = i$ belongs to phase i.

- In each phase $i < L_{min}^T$, the Greedy algorithm incurs a loss of at most N.
- In phase L_{min}^T, the loss of Greedy is at most $N-1$.

□

Lower Bound for Deterministic Algorithms

Theorem

For any deterministic online algorithm D and every $T \geq 1$, there exists a sequence of T losses such that

$$L_D^T = T \quad \textit{and} \quad L_{min}^T \leq \lfloor T/N \rfloor.$$

This lower bound can be shown quite easily by generalizing the example that we have given for the Greedy algorithm. (How?)

The lower bound shows that one cannot get better than the Greedy algorithm without using randomization.

Randomized Weighted Majority (RWM) Algorithm

Let $\eta \in (0, \frac{1}{2}]$ be a suitably chosen parameter.

Randomized Weighted Majority (RWM) Algorithm

Initially, set $w_i^1 = 1$, for every $i \in [N]$.
At every time t,

- let $W^t = \sum_{i=1}^{N} w_i^t$;
- choose expert i with probability $p_i^t = w_i^t / W^t$;
- set $w_i^{t+1} = w_i^t \cdot (1-\eta)^{\ell_i^t}$.

Randomized Weighted Majority (RWM) Algorithm

Theorem (Littlestone, Warmuth, 1994)

The RWM algorithm, for any sequence of losses from $[0, 1]$*, has*

$$L_{RWM}^T \leq (1+\eta) L_{min}^T + \frac{\ln N}{\eta} \ .$$

Setting $\eta = \sqrt{\frac{\ln N}{T}}$ yields

$$L_{RWM}^T \leq L_{min}^T + 2\sqrt{T \ln N} \ .$$

Randomized Weighted Majority (RWM) Algorithm

The *regret* of a learning algorithm H is defined as $L_H^T - L_{min}^T$.

Corollary

The RWM algorithm with $\eta = \sqrt{\frac{\ln N}{T}}$ *has regret at most* $2\sqrt{T \ln N}$.

The *average regret per step* is thus only $2\sqrt{\frac{\ln N}{T}}$.

Observe that this quantity is going to zero when increasing T.

Algorithms with this property are called

no-regret learning algorithms.

Thus, in contrast to the simple greedy algorithms is RWM a no-regret algorithm.

Randomized Weighted Majority (RWM) Algorithm

Proof of the theorem:

- Let us analyze how the sum of weights W^t decreases over time. It holds

$$W^{t+1} = \sum_{i=1}^{N} w_i^{t+1} = \sum_{i=1}^{N} w_i^t (1-\eta)^{\ell_i^t} .$$

- Observe that $(1-\eta)^\ell = (1-\ell\eta)$, for both $\ell = 0$ and $\ell = 1$.
- Furthermore, $(1-\eta)^\ell$ is a convex function in ℓ.
- For $\ell \in [0,1]$ this implies $(1-\eta)^\ell \leq (1-\ell\eta)$.
- This gives

$$W^{t+1} \leq \sum_{i=1}^{N} w_i^t (1-\ell_i^t \eta) .$$

Randomized Weighted Majority (RWM) Algorithm

- Let F^t denote the expected loss of RWM in step t.
- It holds $F^t = \sum_{i=1}^{N} \ell_i^t w_i^t / W^t$.
- Substituting this into the bound for W^{t+1} gives

$$W^{t+1} \leq W^t - \eta F^t W^t = W^t(1 - \eta F^t) .$$

- As a consequence,

$$W^{T+1} \leq W^1 \prod_{t=1}^{T}(1 - \eta F^t) = N \prod_{t=1}^{T}(1 - \eta F^t) .$$

- The sum of weights after step T can be upper bounded in terms of the expected loss of RWM.

Randomized Weighted Majority (RWM) Algorithm

- On the other hand, the sum of weights after step T can be lower bounded in terms of the loss of the best expert as follows:

$$W^{T+1} \geq \max_{1 \leq i \leq N} (w_i^{T+1}) = \max_{1 \leq i \leq N} \left((1-\eta)^{\sum_{t=1}^{T} \ell_i^t} \right) = (1-\eta)^{L_{min}^T} .$$

- Combining the bounds and taking the logarithm on both sides gives

$$L_{min}^T \ln(1-\eta) \leq (\ln N) + \sum_{t=1}^{T} \ln(1-\eta F^T) .$$

- In order to simplify, we will now use the following estimation

$$-z - z^2 \leq \ln(1-z) \leq -z$$

holding for every $z \in [0, \frac{1}{2}]$.

Randomized Weighted Majority (RWM) Algorithm

- This gives

$$\begin{aligned} L_{min}^T(-\eta - \eta^2) &\leq (\ln N) + \sum_{t=1}^{T}(-\eta F^t) \\ &= (\ln N) - \eta L_{RWM}^T \ . \end{aligned}$$

- Finally, solving for L_{RWM}^T gives

$$L_{RWM}^T \leq (1+\eta)L_{min}^T + \frac{\ln N}{\eta} \ .$$

Learning Equilibria in Games

Regret minimization is a natural model for behavior in cases where we have to make repeated decisions with incomplete information.

We consider regret learning when a game $\Gamma = (\mathcal{N}, (\Sigma_i)_{i\in\mathcal{N}}, (c_i)_{i\in\mathcal{N}})$ is played over and over again for T rounds (called *repeated game*).

Initially, no player $i \in \mathcal{N}$ knows the game. In each round t he picks a pure strategy $s_i^t \in \Sigma_i$ using a no-regret algorithm. The algorithm of player i is based *only on the costs observed by i in previous rounds*.

Does the system converge to (approx.) Nash equilibrium in this case?

This would be a nice and plausible explanation how Nash equilibria can evolve in practice. Unfortunately, in general, the answer is "No".

Learning and Equilibria

For every player the average regret over time is going to 0. Based on this property, we can derive a (more general) equilibrium concept.

Definition

Let $\mathcal{V}$ be a probability distribution over the states of a finite game. $\mathcal{V}$ is called coarse-correlated equilibrium if for every player $i \in \mathcal{N}$ and every strategy $s_i' \in S_i$ it holds

$$\mathbb{E}_{s\sim\mathcal{V}}[c_i(s)] \quad \leq \quad \mathbb{E}_{s\sim\mathcal{V}}[c_i(s_i', s_{-i})] \ .$$

$\mathcal{V}$ is called (additive) ε-approximate coarse-correlated equilibrium if

$$\mathbb{E}_{s\sim\mathcal{V}}[c_i(s)] \quad \leq \quad \mathbb{E}_{s\sim\mathcal{V}}[c_i(s_i', s_{-i})] + \varepsilon \ .$$

No-Regret and Coarse-Correlated Equilibria

Consider the history of play $s^1, s^2, \ldots, s^T$ in a repeated game over T rounds. We interpret the history as a distribution over states by choosing $k \in [T]$ uniformly at random.

If player i has regret $R_i(T)$, then for every strategy $s_i' \in S_i$

$$\begin{aligned} \mathbb{E}_{k\in[T]}[c_i(s^k)] &= \sum_{t=1}^{T} \frac{1}{T} \cdot c_i(s^t) \leq \sum_{t=1}^{T} \frac{1}{T} \cdot c_i(s_i', s_{-i}^t) + \frac{R_i(T)}{T} \\ &= \mathbb{E}_{k\in[T]}[c_i(s_i', s_{-i}^k)] + \frac{R_i(T)}{T} \quad . \end{aligned}$$

Proposition

After T rounds if every player has regret at most R, then the history of play represents a $\frac{R}{T}$-approximate coarse-correlated equilibrium.

Suppose all players are using RWM, then after at most $T = \frac{4}{\varepsilon^2} \cdot \log(\max_i |S_i|)$ rounds the history of play represents a ε-approximate coarse-correlated equilibrium.

Imitating Experts

No-Regret Algorithms and Coarse-Correlated Equilibria

Zero-Sum Games

Correlated Equilibria

Recall 2-Player Zero-Sum Games

- A 2-player zero-sum game is a strategic game with 2 players, where $c_{\mathtt{I}}(s) + c_{\mathtt{II}}(s) = 0$ for every state s.
- Matrix A with $|\Sigma_{\mathtt{I}}|$ rows and $|\Sigma_{\mathtt{II}}|$ columns.
  Player I is row player, player II is column player.
- a_{ij} is utility for player I in state (i, j),
  a_{ij} is cost or loss for player II in state (i, j).
- We here normalize A to have $a_{ij} \in [0, 1]$:

  Make A non-negative by adding $\max |a_{ij}|$ to every entry. Then divide by the resulting largest entry scaling all a_{ij} to $[0, 1]$.

  Observe that this does not alter the optimal strategies (and thereby the Nash equilibria) of the game.

Examples

Matching Pennies (normalized)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Rock-Paper-Scissors (normalized)

$$\begin{pmatrix} 1/2 & 0 & 1 \\ 1 & 1/2 & 0 \\ 0 & 1 & 1/2 \end{pmatrix}$$

A game with $|\Sigma_{\text{I}}| \neq |\Sigma_{\text{II}}|$:

$$\begin{pmatrix} 0 & 1/2 & 1 \\ 1/4 & 1/2 & 3/4 \end{pmatrix}$$

Maximin Strategies

- Gain-Floor for player I: $v_{\mathtt{I}}^* = \max_x \min_y x^T A y$.

  Optimal strategy x^* guarantees the gain-floor for I (maximin strategy).

- Loss-Ceiling for player II: $v_{\mathtt{II}}^* = \min_y \max_x x^T A y$.

  Optimal strategy y^* guarantees the loss-ceiling for II (minimax strategy).

Lemma

It holds that $v_{\mathtt{I}}^* \leq v_{\mathtt{II}}^*$.

Theorem (Minimax Theorem)

In every 2-player zero-sum game it holds that $v = v_{\mathtt{I}}^* = v_{\mathtt{II}}^*$.

Mixed Nash equilibrium

Corollary

State (x, y) in a 2-player zero-sum game is a mixed Nash equilibrium

$$\Leftrightarrow$$

x and y are optimal strategies.

Corollary

Every 2-player zero-sum game has at least one mixed Nash equilibrium. All mixed Nash equilibria in such a game yield the same expected utility for player `I`.

Theorem

In 2-player zero-sum games a mixed Nash equilibrium can be computed in polynomial time.

Learning in Zero-Sum Games

- Players do not know the game they are playing. They use no-regret learning algorithms to make their strategy choice.

  Can they learn to play optimally (i.e., learn a Nash equilibrium) ?

Learning in Zero-Sum Games

- Players do not know the game they are playing. They use no-regret learning algorithms to make their strategy choice.

  Can they learn to play optimally (i.e., learn a Nash equilibrium) ?

- Consider player `II`, experts are pure strategies, adversary is player `I`.
- In each step t learning algorithm H of player `II` picks mixed strategy y^t against an unknown adversary strategy x^t of player `I`.

Learning in Zero-Sum Games

- Players do not know the game they are playing. They use no-regret learning algorithms to make their strategy choice.

  Can they learn to play optimally (i.e., learn a Nash equilibrium) ?

- Consider player II, experts are pure strategies, adversary is player I.
- In each step t learning algorithm H of player II picks mixed strategy y^t against an unknown adversary strategy x^t of player I.
- Loss in round t for strategy (expert) i is

$$\ell_i^t = \sum_{j \in \Sigma_{\mathrm{I}}} x_j^t a_{ji} \ .$$

- Total loss in round t of learning algorithm H is

$$\ell_H^t = c_{\mathrm{II}}(x^t, y^t) = \sum_{i \in \Sigma_{\mathrm{II}}} \sum_{j \in \Sigma_{\mathrm{I}}} x_j^t a_{ji} y_i^t \ .$$

No-Regret and Optimal Strategies

- No-regret learning algorithm H:

$$\frac{L_H^T - L_{min}^T}{T} \longrightarrow 0 \qquad \text{when } T \to \infty \ .$$

No-Regret and Optimal Strategies

- No-regret learning algorithm H:

$$\frac{L_H^T - L_{min}^T}{T} \longrightarrow 0 \qquad \text{when } T \to \infty \ .$$

- By definition, the average loss per round of any no-regret learning algorithm becomes as small as the best average loss of any pure strategy in hindsight.
- Is the average loss L_H^T/T as small as the value of the game?

No-Regret and Optimal Strategies

- No-regret learning algorithm H:

$$\frac{L_H^T - L_{min}^T}{T} \longrightarrow 0 \qquad \text{when } T \to \infty \ .$$

- By definition, the average loss per round of any no-regret learning algorithm becomes as small as the best average loss of any pure strategy in hindsight.
- Is the average loss L_H^T/T as small as the value of the game?

Theorem

For a 2-player zero-sum game with gain floor $v_{\mathtt{I}}^*$, if player `II` *plays for T steps using algorithm H with regret R, then the average loss*

$$\frac{L_H^T}{T} \ \leq \ v_{\mathtt{I}}^* + \frac{R}{T} \ .$$

The result applies similarly for player `I` *and the loss-ceiling.*

Learning the Gain Floor

Proof:

- We will show that the best pure strategy in hindsight has total loss at most $L_{min}^T \leq T \cdot v_{\mathtt{I}}^*$.
- Consider the history of play of the adversary player $\mathtt{I}$, i.e., strategies $x^1, x^2, \ldots, x^T$, and combine them to an "average strategy"

$$\hat{x}_j = \frac{1}{T} \sum_{t=1}^{T} x_j^t \qquad \text{for all } j \in \Sigma_{\mathtt{I}}.$$

- Total loss L_i^T of a single strategy $i \in \Sigma_{\mathtt{II}}$ in hindsight is the same if player $\mathtt{I}$ had always played $\hat{x}$ in all time steps:

$$L_i^t = \sum_{t=1}^{T} \sum_{j \in \Sigma_{\mathtt{I}}} x_j^t \cdot a_{ji} = \sum_{j \in \Sigma_{\mathtt{I}}} \left(\sum_{t=1}^{T} x_j^t \right) \cdot a_{ji} = T \cdot \sum_{j \in \Sigma_{\mathtt{I}}} \hat{x}_j \cdot a_{ji} \ .$$

Learning the Gain Floor

- If we assume I plays always $\hat{x}$ and consider the best pure strategy for II in hindsight, then the scenario reduces to a one-step game.
- In this one-step game, player I first determines the average history of play $\hat{x}$ and then player II picks best pure strategy against $\hat{x}$ – i.e., I moves first, then II answers.
- By definition of gain floor, there is always $i \in \Sigma_{\text{II}}$ such that gain of I/loss of II is reduced to at most v_{I}^*, i.e., $c_{\text{II}}(\hat{x}, i) \leq v_{\text{I}}^*$.
- Hence, there is a pure strategy $i \in \Sigma_{\text{II}}$ such that

$$L_{min}^T \leq L_i^T \leq T \cdot v_{\text{I}}^* \ .$$

- Combining these insights:

$$L_H^T \ \leq \ L_{min}^T + R \ \leq \ T \cdot v_{\text{I}}^* + R \ .$$

A Simple Proof of the Minimax Theorem

Theorem (Minimax Theorem)

In every 2-player zero-sum game it holds that $v = v_{\mathtt{I}}^* = v_{\mathtt{II}}^*$.

Proof:

- For contradiction, assume $v_{\mathtt{I}}^* + \gamma = v_{\mathtt{II}}^*$ for some $\gamma > 0$.
- Let both players play the game iteratively for T steps with a learning algorithm that has regret $R/T < \gamma/3$.
- Using the average history of play as before we note that $L_{min}^T \leq v_{\mathtt{I}}^*$ for player II, and $L_{min}^T \leq -v_{\mathtt{II}}^*$ for player I ("$-$" because of loss).
- But this means the algorithms yield at most $v_{\mathtt{I}} + \gamma/3$ average cost for player II and at least $v_{\mathtt{II}} - \gamma/3$ average gain for player I.
- Average cost of II is average gain of I $\rightarrow$ Contradiction. □

Convergence

Corollary

If both players use a no-regret learning algorithm, the average histories of play $(\hat{x}, \hat{y})$ converge to optimal strategies and, thus, to a mixed Nash equilibrium of the game.

This shows convergence only for the history of play, but not for the actual behavior in the distributions x^t and y^t!

Theorem

There are no-regret algorithms for players `I` *and* `II` *such that the actual behavior x^t and y^t does not converge to optimal strategies.*

Convergence for General No-Regret Algorithms

Matching Pennies (normalized)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Proof: A "weird" no-regret algorithm H:

- Rule: `I` and `II` pick pure strategies, `I` moves to the other pure strategy in rounds 1, 3, 5, ..., `II` moves to the other pure strategy in rounds 2, 4, 6,...
- If one player deviates from the rule, the other invokes the RWM algorithm. (trick to ensure no-regret property for every possible sequence of play).

- $L_i^T/T \to 0.5$ for both strategies $i = 1, 2$ of `II`, average loss of algorithm $L_H^T/T \to 0.5$. No-regret algorithm for `II`! (similar argument for `I`).
- None of the distributions y^t is close to optimal strategy $(0.5, 0.5)$, no single round loss ℓ_H^t is close to $v = 0.5$. □

An Adaptive RWM Algorithm (Freund, Schapire, 1999)

Let $\eta_0 \in (0, \frac{1}{2}]$ and u be an upper bound on the value of the game.

Variable Randomized Weighted Majority (vRWM) Algorithm

Initially, set $w_i^1 = 1$, for every $i \in [N]$.
At every time t,

- let $W^t = \sum_{i=1}^N w_i^t$;
- choose expert i with probability $p_i^t = w_i^t / W^t$;
- **if** $\ell_{vRWM}^t \leq u$ **then** set $w_i^{t+1} = w_i^t$;
- **else** set

$$\eta_t = 1 - \frac{u(1 - \ell_{vRWM}^t)}{(1-u)\ell_{vRWM}^t}$$

and $w_i^{t+1} = w_i^t \cdot (1 - \eta_t)^{\ell_i^t}$.

Comparison of Distributions

Definition

The relative entropy or Kullback-Leibler divergence of two distributions y and y' is defined as

$$RE(y \mid\mid y') = \sum_{i=1}^{n} y_i \cdot \ln\left(\frac{y_i}{y'_i}\right) \quad .$$

To compare $a, b \in [0,1]$ we use distributions $(a, 1-a)$ and $(b, 1-b)$:

$$\begin{aligned} RE(a \mid\mid b) &= RE((a,1-a) \mid\mid (b,1-b)) \\ &= a \ln\left(\frac{a}{b}\right) + (1-a)\ln\left(\frac{1-a}{1-b}\right) \quad . \end{aligned}$$

The relative entropy for distributions is always non-negative and $RE(y \mid\mid y') = 0$ if and only if $y = y'$.

Convergence of vRWM

Theorem

Let y' be any mixed strategy for `II` *which generates a loss of at most u against every best response of* `I`*. Then in any iteration t of vRWM in which $\ell^t_{vRWM} \geq u$ the relative entropy between y' and y^{t+1} satisfies*

$$RE(y' \mid\mid y^{t+1}) \leq RE(y' \mid\mid y^t) - RE(u \mid\mid \ell^t_{vRWM}) \ .$$

In every step, in which the loss of vRWM is too high, the adjustment moves the next distribution closer to a good strategy.

Proof of the Theorem

Proof:
For completeness, we provide a proof of the last theorem. Consider a step, in which $\ell^t_{vRWM} > u$ and bound

$$\begin{aligned} & RE(y' \mid\mid y^{t+1}) - RE(y' \mid\mid y^t) \\ = & \sum_{i \in \Sigma_{II}} y'_i \ln \frac{y'_i}{y_i^{t+1}} - \sum_{i \in \Sigma_{II}} y'_i \ln \frac{y'_i}{y_i^t} \\ = & \sum_{i \in \Sigma_{II}} y'_i \ln \frac{y_i^t}{y_i^{t+1}} \\ \leq & \sum_{i \in \Sigma_{II}} y'_i \ln \frac{1 - \eta_t \ell^t_{vRWM}}{(1-\eta_t)^{\ell_i^t}} \quad , \end{aligned}$$

where we use that $y_i^{t+1} = w_i^t (1-\eta_t)^{\ell_i^t} / W^{t+1}$ and $W^{t+1} \leq W^t(1 - \eta_t F^t) = W^t(1 - \eta_t \ell^t_{vRWM})$ as observed above.

Proof of the Theorem

$$
\begin{aligned}
& RE(y' \mid\mid y^{t+1}) - RE(y' \mid\mid y^t) \\
\leq & \sum_{i \in \Sigma_{\text{II}}} y'_i \ln \frac{1 - \eta_t \ell^t_{vRWM}}{(1-\eta_t)^{\ell^t_i}} \\
= & \sum_{i \in \Sigma_{\text{II}}} y'_i \ln \left(\frac{1}{1-\eta_t} \right)^{\ell^t_i} + \ln(1 - \eta_t \ell^t_{vRWM}) \\
= & \left(\ln \frac{1}{1-\eta_t} \right) \cdot \sum_{i \in \Sigma_{\text{II}}} y'_i \ell^t_i + \ln(1 - \eta_t \ell^t_{vRWM}) \\
\leq & \left(\ln \frac{1}{1-\eta_t} \right) u + \ln(1 - \eta_t \ell^t_{vRWM}) \ ,
\end{aligned}
$$

because strategy y' never generates more loss than u.

Proof of the Theorem

We take the derivative of

$$\left(\ln \frac{1}{1-\eta_t}\right) u + \ln(1-\eta_t \ell^t_{vRWM})$$

for η_t and equate it with 0. This implies a minimum is attained at

$$\eta_t = 1 - \frac{u(1-\ell^t_{vRWM})}{(1-u)\ell^t_{vRWM}}$$

as desired. Plugging in this expression for η_t yields

$$\begin{aligned} & - u \ln \left(\frac{u}{\ell^t_{vRWM}} \cdot \frac{1-\ell^t_{vRWM}}{1-u} \right) + \ln \frac{1-\ell^t_{vRWM}}{1-u} \\ = & - RE(u \mid\mid \ell^t_{vRWM}) \ . \end{aligned}$$

□

Convergence of vRWM

Corollary

For any sequence of strategies $y^1, y^2, \ldots$ the number of rounds in which the loss $\ell^t_{vRWM} \geq u + \varepsilon$ is at most

$$\frac{\ln |\Sigma_{\text{II}}|}{RE(u \mid\mid u + \varepsilon)} .$$

For fixed ε this time is independent of T. Thus, the loss suffered in time steps t must get closer to u when t gets larger and larger. This is a much more desirable behavior than, e.g., the weird no-regret algorithm, which yields a loss of 1 every second round, even for arbitrarily large t.

Imitating Experts

No-Regret Algorithms and Coarse-Correlated Equilibria

Zero-Sum Games

Correlated Equilibria

Correlated Equilibria

When players use no-regret algorithms, they converge to (approximate) coarse-correlated equilibria. These equilibria provide a weaker stability guarantee than mixed Nash equilibria. Are coarse-correlated equilibria the strongest equilibrium notion that can be approximated quickly?

Definition

Let $\mathcal{V}$ be a probability distribution over the states of a finite game. $\mathcal{V}$ is called a correlated equilibrium if for every player $i \in \mathcal{N}$ and every swap function $\sigma : S_i \to S_i$ it holds

$$\mathbb{E}_{s\sim\mathcal{V}}[c_i(s)] \quad \leq \quad \mathbb{E}_{s\sim\mathcal{V}}[c_i(\sigma(s_i), s_{-i})] \ .$$

$\mathcal{V}$ is called (additive) ε-approximate correlated equilibrium if

$$\mathbb{E}_{s\sim\mathcal{V}}[c_i(s)] \quad \leq \quad \mathbb{E}_{s\sim\mathcal{V}}[c_i(\sigma(s_i), s_{-i})] + \varepsilon \ .$$

Correlated Equilibria

In a mixed Nash equilibrium every player chooses her own distribution x_i over her own strategies. A distribution $\mathcal{V}$ over states emerges by independent combination of player strategies x_i. No player wants to unilaterally switch to a different (pure) strategy.

In correlated and coarse-correlated equilibria we directly specify a distribution $\mathcal{V}$ over states of the game. The distribution can combine and correlate the strategy choices of the players (e.g. when resulting from a joint learning process over time).

No player wants to unilaterally switch to a different (pure) strategy:

- Correlated: Suppose i knows that the random draw from $\mathcal{V}$ gives strategy s_i for her. Given this knowledge, she does not want to deviate from s_i.
- Coarse-correlated: No player i wants to deviate to a fixed strategy s_i before knowing the random draw from $\mathcal{V}$. Hence, in this case i cannot condition her deviation on the strategy s_i chosen for her in the draw from $\mathcal{V}$.

Example

Two drivers are heading towards an intersection.

	(A)ccelerate	(B)reak
(A)ccelerate	101, 101	2, 0
(B)reak	0, 2	1, 1

Equilibria:

- Pure: States (A,B) und (B,A)
- Mixed (not pure): Both players $x_A = 0.01$ und $x_B = 0.99$.
- Correlated (not mixed):
  $\Pr_{s\sim\mathcal{V}}[s = (B, A)] = 0.5$ and
  $\Pr_{s\sim\mathcal{V}}[s = (A, B)] = 0.5$

Example

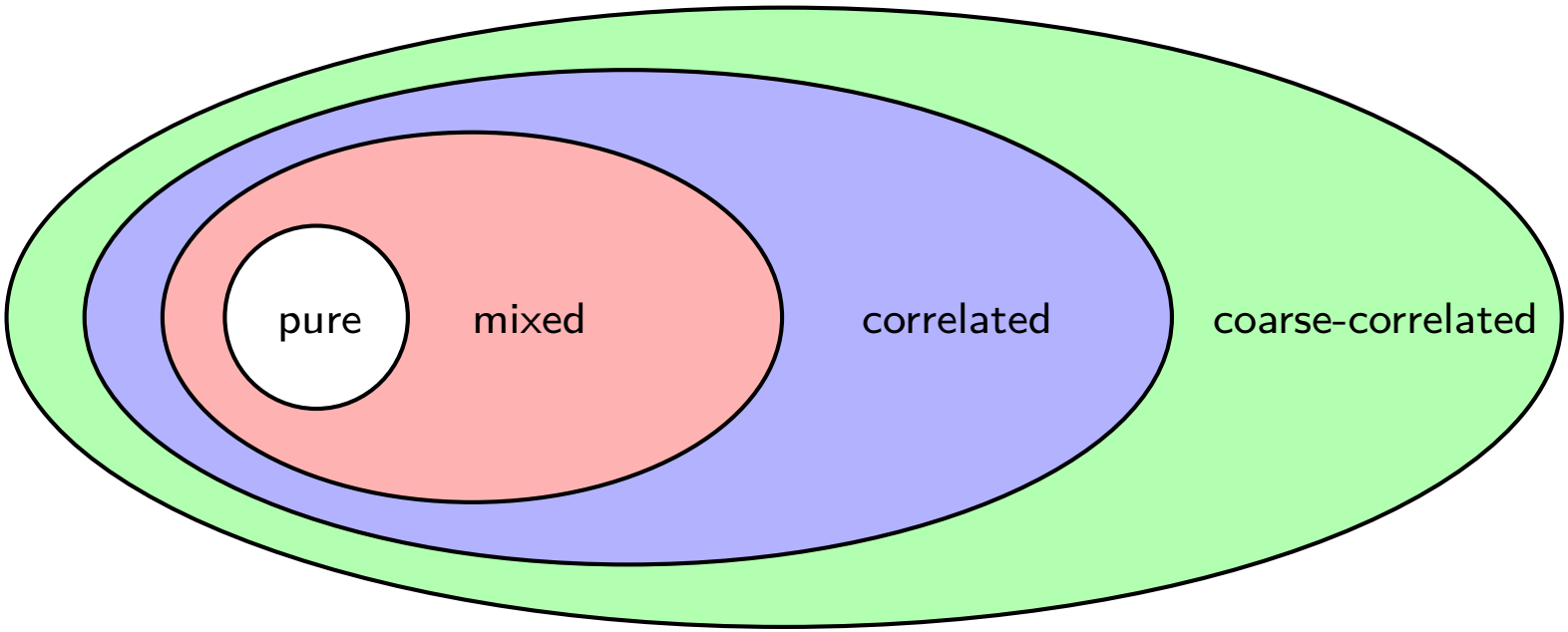
Two drivers are heading towards an intersection.

	(A)ccelerate	(B)reak
(A)ccelerate	101 101	2 0
(B)reak	0 2	1 1

Equilibria:

- A correlated equilibrium is like a traffic light: Every player gets a strategy suggestion drawn from the known distribution $\mathcal{V}$ over states. No player wants to unilaterally change her suggested strategy, given that all others stick to their suggestion.
- Coarse-correlated (not correlated): Not possible in symmetric 2×2-games (every coarse-correlated equilibrium is correlated), see exercises.

A Hierarchy of Equilibrium Concepts



Learning Correlated Equilibria

Can the players learn a correlated equilibrium?

For this task we concentrate on algorithms with a stronger guarantee in the expert problem. For algorithm H let $H(t) \in [N]$ be the expert chosen in step t. To measure regret, our benchmark was the total loss of the best single expert in hindsight. Here we change this benchmark to the best function of experts

$$\sigma^* \in \arg \min_{\sigma:[N]\to[N]} \sum_{t=1}^{T} \ell^t_{\sigma(H(t))} \ ,$$

with $L^T_{s\min} = \sum_{t=1}^{T} \ell^t_{\sigma^*(H(t))}$ the total loss of the best function of experts.

Swap Regret

Definition

The swap regret of H is given by

$$SR(T) \quad = \quad L_H^T - L_{s\min}^T \quad = \quad \sum_{t=1}^{t} \ell_H^t - \sum_{t=1}^{T} \ell_{\sigma^*(H(t))}^t \ .$$

Swap regret captures the loss we would suffer if we could swap expert choices.
If the optimal swap function sets $\sigma^*(i) = j$, then this implies

Whenever H chose expert i, it should have better chosen expert j.

A no-swap-regret algorithm obtains $SR(T)/T \to 0$ for $T \to \infty$.

Observe: $SR(T) \geq R(T)$, and no-swap-regret $\Rightarrow$ no-regret. (Why?)

Swap Regret and Correlated Equilibrium

Consider a history of play $s^1, s^2, \ldots, s^T$ over T rounds and interpret the history as distribution over states.

If player i has swap regret $SR_i(T)$, then for every strategy s_i and every function $\sigma : S_i \to S_i$

$$\begin{aligned}
\mathbb{E}_{k\in[T]}[c_i(s^k)] &= \sum_{t=1}^{T} \frac{1}{T} \cdot c_i(s_i^t, s_{-i}^t) \\
&\leq \sum_{t=1}^{T} \frac{1}{T} \cdot c_i(\sigma(s_i^t), s_{-i}^t) + \frac{SR_i(T)}{T} \\
&= \mathbb{E}_{k\in[T]}[c_i(\sigma(s_i^k), s_{-i}^k)] + \frac{SR_i(T)}{T} \quad .
\end{aligned}$$

Proposition

After T rounds if every player has swap regret at most R, then the history of play represents a $\frac{R}{T}$-approximate correlated equilibrium.

No-Swap-Regret Algorithms

We can turn no-regret algorithms into no-swap-regret algorithms!

Theorem

If H is an algorithm with regret $R(T)$, then there is an algorithm M with swap reget $SR(T) = N \cdot R(T)$.

Proof:

We use N copies $H_1, \ldots, H_N$ of algorithm H. In some sense, H_j makes sure that no function σ that maps expert j to $\sigma(j)$ generates too much swap regret.

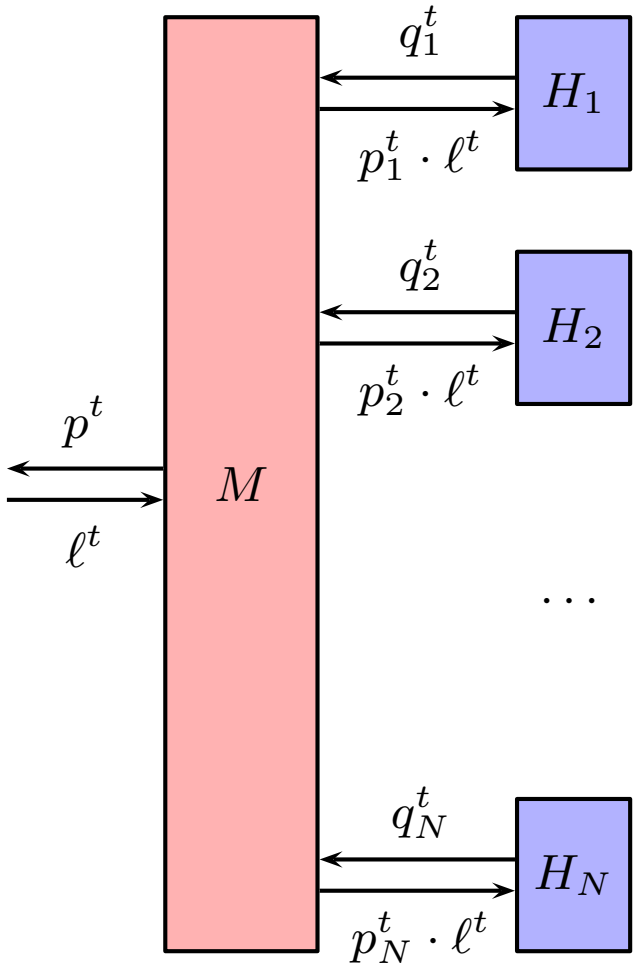
A master algorithm M coordinates the copies of H.

Master Algorithm

Master Algorithm

In every step t,

1. obtain distributions $q_1^t, \ldots, q_n^t$ of $H_1, \ldots, H_N$
2. compute consensus distribution p^t
3. choose expert i with probability p_i^t
4. report to H_j a loss of $p_j^t \cdot \ell_i^t$, for every expert $i \in N$



Master Algorithm: No Swap Regret

We defer the details of computing p^t in line 2. to the end of the proof.

For swap regret we use the following notation:

- Total loss of master algorithm: $L_M^T = \sum_{t=1}^{T} \sum_{i \in [N]} p_i^t \ell_i^t$
- Total loss of best function σ^*: $L_{s\min}^T = \sum_{t=1}^{T} \sum_{i \in [N]} p_i^t \ell_{\sigma^*(i)}^t$

Algorithm H_j "believes", we are choosing expert i with probability $q_{j,i}^t$ and the loss for this choice would be $p_j^t \ell_i^t$.

- Total "internal" loss of H_j: $L_{H_j}^T = \sum_{t=1}^{T} \sum_{i \in [N]} q_{j,i}^t \cdot (p_j^t \ell_i^t)$
- Total loss of best expert k_j for H_j: $L_{\min,j}^T = \sum_{t=1}^{T} p_j^t \ell_{k_j}^t$

Regret for H_j is $L_{H_j}^T - L_{\min,j}^T \leq R(T)$.

Master Algorithm: No Swap Regret

Let σ^* be the best function of experts. Then, by summing up,

$$\begin{aligned}\sum_{j\in[N]} L^T_{H_j} &\leq \sum_{j\in[N]} L^T_{\min,j} + R(T) \\ &\leq \sum_{j\in[N]}\sum_{t=1}^{T} p^t_j \ell^t_{\sigma^*(j)} + \sum_{j\in[N]} R(T) \\ &= L^T_{s\,\min} + N \cdot R(T) \ .\end{aligned}$$

Hence, for the sum of internal losses we obtained a bound on the swap regret.

But how does the real total loss L^T_M relate to the sum of internal losses $\sum_j L^T_{H_j}$?

To obtain this relation we now specify how to compute the consesus distribution p^t.

Master-Algorithmus: Consensus Distribution

Choose the consensus distribution p^t by

$$p_i^t = \sum_{j \in [N]} q_{j,i}^t p_j^t \ .$$

Then it holds:

$$\begin{aligned} L_M^T &= \sum_{t=1}^{T} \sum_{i \in [N]} p_i^t \ell_i^t \\ &= \sum_{t=1}^{T} \sum_{i \in [N]} \sum_{j \in [N]} q_{j,i}^t p_j^t \ell_i^t \\ &= \sum_{j \in [N]} L_{H_j}^T \\ &\leq L_{s\,\min}^T + N \cdot R(T) \ , \end{aligned}$$

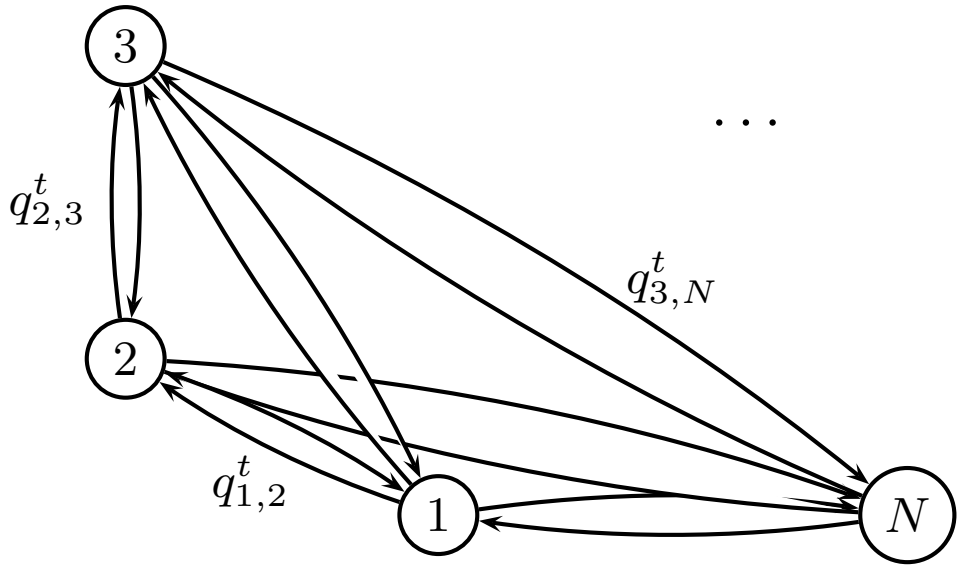
and the theorem is shown. □

Master Algorithm: Markov Chain

p^t ist die stationary distribution of the following Markov chain:

The states are the experts. If we arrive at state/expert i, then the probability to move to state/expert j is given by $q^t_{i,j}$. Matrix Q^t with $Q^t(i,j) = q^t_{i,j}$ contains the transtition probabilities of all states/experts i.

Choose p^t as the stationary distribution of this Markov chain, i.e., the dominant eigenvector of matrix Q^t. p^t can be computed efficiently.



Overview: Equilibria, Computation, Convergence

Pure Nash Equilibrium

- Existence not guaranteed, only in some games (e.g. weakly acyclic games)
- Computation in poly-time in the size of the cost matrix (which is huge)
- PLS-hard in compactly representable congestion games
- Potential Games:

  Convergence time of best-response dynamics can be exponential

  Poly-time convergence for subclasses

Mixed Nash Equilibrium

- Existence in finite games (Nash's Theorem)
- Computation in PPAD (Sperner's Lemma) and PPAD-complete
- 2-player Zero Sum:

  Computation in poly-time via LP

  Convergence in poly-time with No-Regret algorithms

  Convergence of play with vRWM-Algorithm

Overview: Equilibria, Computation, Convergence

Correlated Equilibrium

- Existence in finite games (mixed NE is correlated equilibrium)
- Distribution over states without profitable "swap"-deviations
- Convergence in poly-time with No-Swap-Regret algorithms

Coarse-Correlated Equilibrium

- Existence in finite games (correated is coarse-correlated)
- Distribution over states without profitable "best expert"-deviations
- Convergence in poly-time with No-Regret algorithms

Note: Convergence guarantees for no-regret or no-swap-regret algorithms apply to ε-approximate variants of equilibria, which emerge on average over the history of play. ε decreases as a function of T.

Literature

- Nisan, Roughgarden, Tardos, Vazirani. Algorithmic Game Theory, 2007. (Chapter 4).
- Littlestone, Warmuth. The Weighted Majority Algorithm. Information & Computation 108(2):212–261, 1994.
- Freund, Schapire. Adaptive Game Playing using Multiplicative Weights. Games and Economic Behavior, 29:79–103, 1999.
- Roughgarden. Twenty Lectures on Algorithmic Game Theory, 2016. (Chapter 17+18).
- Blum, Mansour. From External to Internat Regret. Journal of Machine Learning Research 8:1307–1324, 2007.