ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά μαθήματα ΠΠ

Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Learning, Enforcement and Equilibria

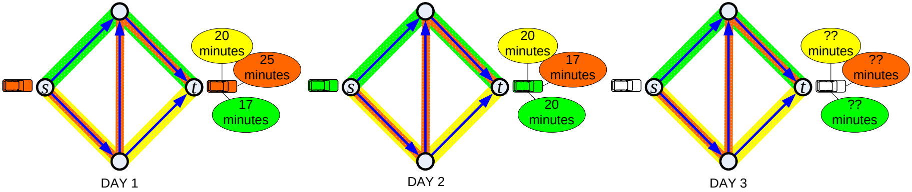
MYA 2015: Case Studies in Decision Making
Invited Lecture

Spyros Kontogiannis

Computer Science & Engineering Department
University of Ioannina -- Greece

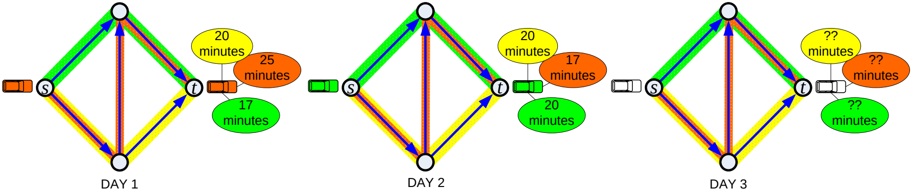
Friday, June 22 2015

A Motivating Example (I)



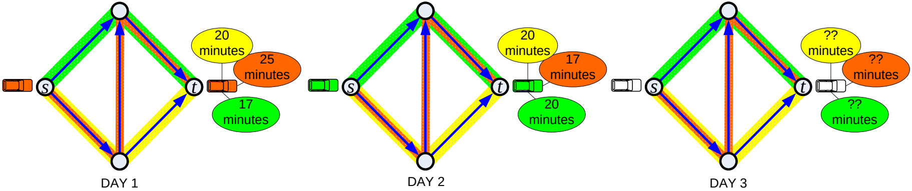
- *Every day* you need to **choose** one of N possible routes to get to work.

A Motivating Example (I)



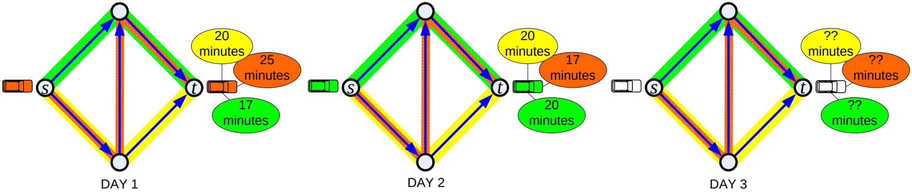
- *Every day* you need to **choose** one of *N* possible routes to get to work.
- The **cost/delay** you suffer depends on *current* traffic conditions.

A Motivating Example (I)



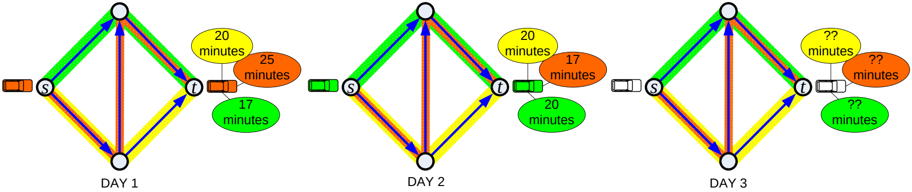
- *Every day* you need to **choose** one of N possible routes to get to work.
- The **cost/delay** you suffer depends on *current* traffic conditions.
- You make a **decision** which may only be based on *history* (not clear a priori which route is the best).

A Motivating Example (I)



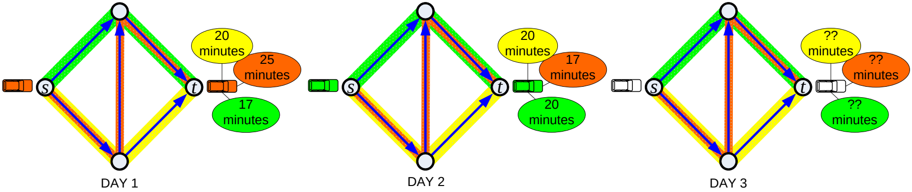
- *Every day* you need to **choose** one of *N* possible routes to get to work.
- The **cost/delay** you suffer depends on *current* traffic conditions.
- You make a **decision** which may only be based on *history* (not clear a priori which route is the best).
- You become **informed** of *your own actual cost* (how long your route took) only when you get to your office, possibly along with the *actual costs of alternatives* (how long some colleagues' routes took).

A Motivating Example (II)



Q1 **What can be learnt?**

Is there an *online algorithm* which learns how to pick routes so that, in the long run, whatever the sequence of traffic patterns occurred, you have done not much worse than the **best fixed choice**, in retrospective?



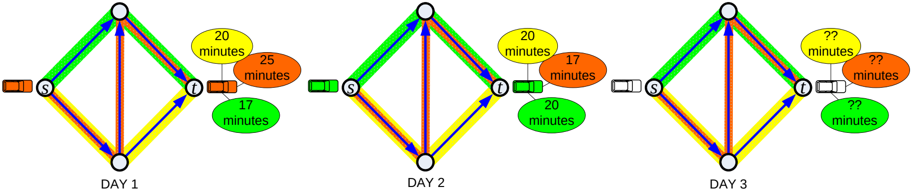
Q1 What can be learnt?

Is there an *online algorithm* which learns how to pick routes so that, in the long run, whatever the sequence of traffic patterns occurred, you have done not much worse than the **best fixed choice**, in retrospective?

Q2 What can be enforced?

Given a collection of routes that is good (above the **minmax values**) but *not necessarily the best* for every individual day, is it possible *for the system* to enforce it as a stable solution, in the long run, for everyone?

A Motivating Example (II)



Q1 What can be learnt?

Is there an *online algorithm* which learns how to pick routes so that, in the long run, whatever the sequence of traffic patterns occurred, you have done not much worse than the **best fixed choice**, in retrospective?

Q2 What can be enforced?

Given a collection of routes that is good (above the **minmax values**) but *not necessarily the best* for every individual day, is it possible *for the system* to enforce it as a stable solution, in the long run, for everyone?

A YES and YES!!!

Skeleton

We may use *different means* to go to our work every morning. The loss we incur per day is **dependent on weather** of that day:

			
	1	2	
	3		1
		3	1

Matrix of Losses (per weather type)

- Best response to sunny weather = WALK
- Best response to cloudy weather = MOTORBIKE
- Best response to rainy weather = BUS

Agent Against Nature: An Example

We may use *different means* to go to our work every morning. The loss we incur per day is **dependent on weather** of that day:

	1	2	
	3		1
		3	1

Matrix of Losses (per weather type)

- Best response to sunny weather = WALK
- Best response to cloudy weather = MOTORBIKE
- Best response to rainy weather = BUS

GOAL: The agent has to...

- use an online algorithm (OLA) that *decides each day* what to do, based only on history of previous losses due to past decisions.
- suffer *irrevocable losses*, after having made his/her decisions.

Agent Against Nature: Definition

- $[N] = \{1, 2, \ldots, N\}$: The **action space** for a single **agent** against Nature.
- The game is repeated forever, in discrete rounds.
- In each round $t \geq 1$:
  - Agent *OLA* makes a (probabilistic) choice of an action, according to a **strategy** $p^t \in \Delta_N := \{p \in [0,1]^N : \mathbf{1}'p = 1\}$.
  - Nature makes its own move, and reveals a vector $(\ell^t)_{t\in[N]} \in [0,1]^N$ of **losses**, for all the actions.
  - *OLA* incurs irrevocably the *loss for the chosen action* $i^t \in [N]$, only after having made its choice.
  - *OLA* keeps either $(\ell^t)_{t\in[N]}$ (full-info model), or $\ell^t_{i^t}$ (partial-info model) in history.

Agent Against Nature: Definition

- $[N] = \{1, 2, \ldots, N\}$: The **action space** for a single **agent** against Nature.
- The game is repeated forever, in discrete rounds.
- In each round $t \geq 1$:
  - Agent *OLA* makes a (probabilistic) choice of an action, according to a **strategy** $p^t \in \Delta_N := \{p \in [0,1]^N : \mathbf{1}'p = 1\}$.
  - Nature makes its own move, and reveals a vector $(\ell^t)_{t\in[N]} \in [0,1]^N$ of **losses**, for all the actions.
  - *OLA* incurs irrevocably the *loss for the chosen action* $i^t \in [N]$, only after having made its choice.
  - *OLA* keeps either $(\ell^t)_{t\in[N]}$ (full-info model), or $\ell^t_{i^t}$ (partial-info model) in history.

GOAL: *OLA* should **adapt to history** so as to perform as good as possible the *extra loss* per round, in the long run, when compared to **simple alternatives** (e.g., always pick a given action).

Modification Rules

Consider the particular instance of 9 days:

	1	2	1		2		2	2	1
	3		3	1		1			3
		3		1	3	1	3	3	

Matrix of Losses (per day and choice)

Modification Rules

Consider the particular instance of 9 days:

	1	2	1		2		2	2	1
	3		3	1		1			3
		3		1	3	1	3	3	

Matrix of Losses (per day and choice)

Modification Rule: A specific way to differentiate the behavior of *OLA*. Ie, any *family of functions* $f^t : [N]^t \mapsto [N]$ mapping the history of *OLA*'s moves so far to an **alternative move** $f^t(\{i^1, i^2, \ldots, i^t\}) \in [N]$ for each round t.

Modification Rules

Consider the particular instance of 9 days:

	1	2	1		2		2	2	1
	3		3	1		1			3
		3		1	3	1	3	3	

Matrix of Losses (per day and choice)

Modification Rule: A specific way to differentiate the behavior of *OLA*. Ie, any *family of functions* $f^t : [N]^t \mapsto [N]$ mapping the history of *OLA*'s moves so far to an **alternative move** $f^t(\{i^1, i^2, \ldots, i^t\}) \in [N]$ for each round t.

if *OLA* is a *probabilistic* online algorithm

then the history contains *probability distributions* rather than actions, and the alternative move proposed by the modification rule is another *probability distribution* $f^t(p^1, p^2, \ldots, p^t) \in \Delta_N$.

Modification Rules

Consider the particular instance of 9 days:

	1	2	1		2		2	2	1
	3		3	1		1			3
		3		1	3	1	3	3	

Matrix of Losses (per day and choice)

Modification Rule: A specific way to differentiate the behavior of *OLA*. Ie, any *family of functions* $f^t : [N]^t \mapsto [N]$ mapping the history of *OLA*'s moves so far to an **alternative move** $f^t(\{i^1, i^2, \ldots, i^t\}) \in [N]$ for each round t.

if *OLA* is a *probabilistic* online algorithm

then the history contains *probability distributions* rather than actions, and the alternative move proposed by the modification rule is another *probability distribution* $f^t(p^1, p^2, \ldots, p^t) \in \Delta_N$.

Remark: The optimal adaptive modification rule for *this instance* is:

Aggregate Losses

- For **deterministic agents** and first T rounds:
  - Losses of *OLA*: $\boxed{L_{OLA}^T = \sum_{t=1}^T \ell_{i^t}^t}$ where $i^t \in [N]$ is *OLA*'s action for round t.
  - Losses of a family F of modification rules: $\boxed{L_{OLA,F}^T = \sum_{t=1}^T \ell_{f^t}^t}$ where $f^t = f^t(i^1, \ldots, i^t) \in [N]$ is F's action for round t.

Aggregate Losses

- For **deterministic agents** and first T rounds:
  - Losses of *OLA*: $\boxed{L^T_{OLA} = \sum_{t=1}^{T} \ell^t_{i^t}}$ where $i^t \in [N]$ is *OLA*'s action for round t.
  - Losses of a family F of modification rules: $\boxed{L^T_{OLA,F} = \sum_{t=1}^{T} \ell^t_{f^t}}$ where $f^t = f^t(i^1, \ldots, i^t) \in [N]$ is F's action for round t.
- For **probabilistic agents** and first T rounds:
  - Losses of *OLA*: $\boxed{L^T_{OLA} = \sum_{t=1}^{T} \sum_{i=1}^{N} \ell^t_i \cdot p^t_i}$ where $p^t_i \in [0,1]$ is *OLA*'s probability mass for action i in round t.
  - Losses of a family F of modification rules: $\boxed{L^T_{OLA,F} = \sum_{t=1}^{T} \sum_{i=1}^{N} \ell^t_i \cdot f^t_i}$ where $f^t_i \in [0,1]$ is F's probability mass for action i in round t.

Aggregate Losses

- For **deterministic agents** and first T rounds:
  - Losses of OLA: $\boxed{L^T_{OLA} = \sum_{t=1}^T \ell^t_{i^t}}$ where $i^t \in [N]$ is OLA's action for round t.
  - Losses of a family F of modification rules: $\boxed{L^T_{OLA,F} = \sum_{t=1}^T \ell^t_{f^t}}$ where $f^t = f^t(i^1, \ldots, i^t) \in [N]$ is F's action for round t.
- For **probabilistic agents** and first T rounds:
  - Losses of OLA: $\boxed{L^T_{OLA} = \sum_{t=1}^T \sum_{i=1}^N \ell^t_i \cdot p^t_i}$ where $p^t_i \in [0,1]$ is OLA's probability mass for action i in round t.
  - Losses of a family F of modification rules: $\boxed{L^T_{OLA,F} = \sum_{t=1}^T \sum_{i=1}^N \ell^t_i \cdot f^t_i}$ where $f^t_i \in [0,1]$ is F's probability mass for action i in round t.

Remark: At time t, usually the modification rule f^t shifts the probability mass that OLA assign to $j \in [N]$ to action f^t_j.

The Notion of Regret (I)

- A *measure of embarrassment*, in hindsight, for our choice of online algorithm. Ie, the *worst possible diversion* of *OLA*'s total loss from the total loss of an allowable modification rule.

The Notion of Regret (I)

- A *measure of embarrassment*, in hindsight, for our choice of online algorithm. Ie, the *worst possible diversion* of *OLA*'s total loss from the total loss of an allowable modification rule.

DEFINITION: Regret

For any online algorithm *OLA*, set F of allowable amodification rules, and any number of time steps T, the **regret** of *OLA* against F is:

$$R_{OLA,F}(T) = \max_{\ell, f \in F} \left\{ L^{T}_{OLA}(\ell) - L^{T}_{OLA,f}(\ell) \right\}$$

where:

- $\ell = (\ell^{t})_{t \in [T]}$ is the vector of losses for all actions, per round $t \in [T]$.
- $f = (f^{t})_{t \in [T]}$ is an allowable modification rule, from F.
- $L^{T}_{OLA}(\ell), L^{T}_{OLA,f}(\ell)$ are the total (possibly expected) losses of *OLA* and f for the first T time steps, respectively.

No-Regret Algorithms

DEFINITION: No-Regret Algorithms

An online algorithm OLA is **no-regret** wrt a given family F of modification rules, if the average (per round) loss of OLA is *asymptotically equal* to the average loss of the best possible modification rule in F. This implies that the (absolute) regret is $o(T)$, for sufficiently large number T of rounds.

No-Regret Algorithms

DEFINITION: No-Regret Algorithms

An online algorithm OLA is **no-regret** wrt a given family F of modification rules, if the average (per round) loss of OLA is *asymptotically equal* to the average loss of the best possible modification rule in F. This implies that the (absolute) regret is $o(T)$, for sufficiently large number T of rounds.

Remark

Fix a given family of modification rules F and any online algorithm A.

if for any modification rule $f \in F$ and any sequence of *normalized* loss vectors ℓ, ie, all losses in $[0, 1]$, it holds that:

$$L_A^T(\ell) \leq \alpha \cdot L_{OLA,f}^T(\ell) + \beta$$

then $\alpha \in 1 + o(1) \ \wedge \ \beta \in o(T)$ implies that A is no-regret algorithm against F.

Bad News for Adaptive Modification Rules

THEOREM 4.1 (AGT-book): No Hope to Learn Against Adaptive Rules

If we allow the set F_{all} of all possible modification rules, then for any online algorithm OLA there is a vector of losses ℓ (for T rounds of play) such that the regret of OLA against functions mapping time steps to actions, is at least $T(1 - 1/N)$.

THEOREM 4.1 (AGT-book): No Hope to Learn Against Adaptive Rules

If we allow the set F_{all} of all possible modification rules, then for any online algorithm OLA there is a vector of losses ℓ (for T rounds of play) such that the regret of OLA against functions mapping time steps to actions, is at least $T(1 - 1/N)$.

WHY?

- $\forall t \in [T] : \begin{cases} \ell^t_{x^t} = 0; & x^t \in \arg\min_{i\in[N]}\{p^t_i\} \\ \ell^t_i = 1, & \forall i \neq x^t \end{cases}$

Bad News for Adaptive Modification Rules

THEOREM 4.1 (AGT-book): No Hope to Learn Against Adaptive Rules

If we allow the set F_{all} of all possible modification rules, then for any online algorithm *OLA* there is a vector of losses ℓ (for T rounds of play) such that the regret of *OLA* against functions mapping time steps to actions, is at least $T(1-1/N)$.

WHY?

- $\forall t \in [T] : \begin{cases} \ell^t_{x^t} = 0; & x^t \in \arg\min_{i\in[N]}\{p^t_i\} \\ \ell^t_i = 1, & \forall i \neq x^t \end{cases}$
- $L^T_{OLA} \geq T \cdot (1-1/N)$.
- $L^T_{OLA,F_{all}} = 0$. ■

How About Oblivious Rules?

- We wish to compare ourselves against *realistic scenarios* (eg, assuming the same knowledge pattern with *OLA*, but no access to *OLA*).

How About Oblivious Rules?

- We wish to compare ourselves against *realistic scenarios* (eg, assuming the same knowledge pattern with *OLA*, but no access to *OLA*).

- We restrict attention to *simple* oblivious modification rules:

External-Regret Rules: The family of rules f_i that always chooses for all rounds the same action $i \in [N]$ (independently *OLA*).

$$\boxed{R_{OLA,F_{ext}}(T) = \max_{i\in[N]} \left\{ L^T_{OLA} - L^T_{OLA,f_i} \right\}}$$

/∗ N cases to check ∗/

How About Oblivious Rules?

- We wish to compare ourselves against *realistic scenarios* (eg, assuming the same knowledge pattern with *OLA*, but no access to *OLA*).

- We restrict attention to *simple* oblivious modification rules:

External-Regret Rules: The family of rules f_i that always chooses for all rounds the same action $i \in [N]$ (independently *OLA*).

$$\boxed{R_{OLA,F_{ext}}(T) = \max_{i\in[N]}\left\{L^T_{OLA} - L^T_{OLA,f_i}\right\}}$$

$/*$ N cases to check $*/$

Internal-Regret Rules: The family of rules $f_{i,j}$ that mimics *OLA*, but for a single action i which is always substituted by some action j.

$$\boxed{R_{OLA,F_{int}}(T) = \max_{(i,j)}\left\{L^T_{OLA} - L^T_{OLA,f_{i,j}}\right\} = \max_{(i,j)} \sum_{t=1}^{T} p_i^t \cdot (\ell_i^t - \ell_j^t)}$$

$/*$ $N(N-1)$ cases to check $*/$

How About Oblivious Rules?

- We wish to compare ourselves against *realistic scenarios* (eg, assuming the same knowledge pattern with *OLA*, but no access to *OLA*).

- We restrict attention to *simple* oblivious modification rules:

External-Regret Rules: The family of rules f_i that always chooses for all rounds the same action $i \in [N]$ (independently *OLA*).

$$\boxed{R_{OLA,F_{ext}}(T) = \max_{i\in[N]} \left\{ L^T_{OLA} - L^T_{OLA,f_i} \right\}}$$

$/* \ N$ cases to check $*/$

Internal-Regret Rules: The family of rules $f_{i,j}$ that mimics *OLA*, but for a single action i which is always substituted by some action j.

$$\boxed{R_{OLA,F_{int}}(T) = \max_{(i,j)} \left\{ L^T_{OLA} - L^T_{OLA,f_{i,j}} \right\} = \max_{(i,j)} \sum_{t=1}^{T} p_i^t \cdot (\ell_i^t - \ell_j^t)}$$

$/* \ N(N-1)$ cases to check $*/$

Swap-Regret Rules: The family of rules that determine arbitrary maps of the choices of *OLA* to possible actions.

$$\boxed{R_{OLA,F_{sw}}(T) = \max_{f_{sw}} \left\{ L^T_{OLA} - L^T_{OLA,f_{sw}} \right\} = \sum_{i=1}^{N} \max_j \sum_{t=1}^{T} p_i^t \cdot (\ell_i^t - \ell_j^t)}$$

$/* \ N^N$ cases to check $*/$

Agent Against Nature: An Example (contd.)

Consider the online algorithm:

OLA = ``Make the best choice, given the weather of the previous day''.

and suppose that the possible losses are:

	1	2	
	3		1
		3	1

Matrix of Losses (per weather type)

Consider the following instance of 9 days:

	1	2	1		2		2	2	1
	3		3	1		1			3
		3		1	3	1	3	3	

Matrix of Losses (per day and choice)

OLA										16
EXT										16-11
INT										16-13 : b → m
SWP										16-8 : b → m, w → b, m → b

Total Loss and Regrets of OLA

Greediness vs. External Regret (I)

GREEDY (G): Choose as current action the cheapest action so far (for all rounds). Break ties in favor of smallest action index.

1. $x^1 = 1$
2. $\forall t \geq 2$:

   $S^{t-1} \arg\min_{i \in [N]} \left\{ \sum_{\tau \leq t-1} \ell_i^\tau \right\}$ /∗ best responses according to history ∗/

   $\forall i \in [N], x^t = \min\{i : i \in S^{t-1}\}$ /∗ choose smallest-ID best response ∗/

Greediness vs. External Regret (I)

GREEDY (G): Choose as current action the cheapest action so far (for all rounds). Break ties in favor of smallest action index.

1. $x^1 = 1$
2. $\forall t \geq 2$:
   $S^{t-1} \arg\min_{i \in [N]} \left\{ \sum_{\tau \leq t-1} \ell_i^\tau \right\}$ /∗ best responses according to history ∗/
   $\forall i \in [N], x^t = \min\{i : i \in S^{t-1}\}$ /∗ choose smallest-ID best response ∗/

RANDOMIZED GREEDY (RG): Uniformly at random choose as current action any of the cheapest actions so far (for all rounds).

1. $x^1 = 1$
2. $\forall t \geq 2$:
   $S^{t-1} \arg\min_{i \in [N]} \left\{ \sum_{\tau \leq t-1} \ell_i^\tau \right\}$ /∗ best responses according to history ∗/
   $\forall i \in [N], p_i^t = \frac{\mathbb{I}_{[i \in S^{t-1}]}}{|S^{t-1}|}$
   Select x^t according to distribution $\mathbf{p(t)}$. /∗ uniform choice of a best response ∗/

Greediness vs. External Regret (II)

Denote by $L^T_{\min} = \min_{i\in[N]} \left\{ \sum_{t=1}^{T} \ell_i^t \right\}$ the *minimum total loss* that any action may achieve for the first T steps of a sequence of loss vectors $(\ell^t)_{t\in[T]}$.

THEOREMS 4.2-3 (AGT-book): Greediness is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

1. For any sequence of T loss vectors, GREEDY's loss is upper bounded by:
$$\boxed{L^T_G \leq N \cdot L^T_{min} + (N-1)}$$
2. For any sequence of T loss vectors, RG's loss is upper bounded by:
$$\boxed{L^T_{RG} \leq (1 + \ln(N)) \cdot L^T_{min} + \ln(N)}$$

Greediness vs. External Regret (II)

Denote by $L^T_{\min} = \min_{i\in[N]} \left\{\sum_{t=1}^{T} \ell_i^t\right\}$ the *minimum total loss* that any action may achieve for the first T steps of a sequence of loss vectors $(\ell^t)_{t\in[T]}$.

THEOREMS 4.2-3 (AGT-book): Greediness is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

1. For any sequence of T loss vectors, GREEDY's loss is upper bounded by:
$$\boxed{L^T_G \leq N \cdot L^T_{min} + (N-1)}$$
2. For any sequence of T loss vectors, RG's loss is upper bounded by:
$$\boxed{L^T_{RG} \leq (1+\ln(N)) \cdot L^T_{min} + \ln(N)}$$

Remark: These are NOT no-external-regret algorithms!

Greediness vs. External Regret (III)

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:

Greediness vs. External Regret (III)

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:
   - $L^{t-1}_{x^t} \leq L^{t-1}_{x^*}$
   - $L^t_{x^t} = 1 + L^{t-1}_{x^t} > L^{t-1}_{x^*} = L^t_{x^*}$

Greediness vs. External Regret (III)

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:
   - $L^{t-1}_{x^t} \leq L^{t-1}_{x^*}$
   - $L^t_{x^t} = 1 + L^{t-1}_{x^t} > L^{t-1}_{x^*} = L^t_{x^*}$

   $\therefore\ |S^t| \leq |S^{t-1}| - 1$

Greediness vs. External Regret (III)

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:
   - $L^{t-1}_{x^t} \leq L^{t-1}_{x^*}$
   - $L^t_{x^t} = 1 + L^{t-1}_{x^t} > L^{t-1}_{x^*} = L^t_{x^*}$

   $\therefore$ $|S^t| \leq |S^{t-1}| - 1$

   $\therefore$ At most N losses of G between two consecutive losses of L^T_{min}.

Greediness vs. External Regret (III)

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:
   - $L^{t-1}_{x^t} \leq L^{t-1}_{x^*}$
   - $L^t_{x^t} = 1 + L^{t-1}_{x^t} > L^{t-1}_{x^*} = L^t_{x^*}$

   $\therefore\ |S^t| \leq |S^{t-1}| - 1$

   $\therefore$ At most N losses of G between two consecutive losses of L^T_{min}.

2. Analysis for RG: Let $t_j = \min\{t : L^T_{min} \geq j\}$.
   - $\forall t \in (t_j, t_{j+1}]$, **if** $|S^t| = |S^{t-1}| - k$ **then** $L^t_{RG} - L^{t-1}_{RG} \leq \frac{k}{|S^{t-1}|}$
   - $L^{t_{j+1}} - L^{t_j} \leq \frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{2} + 1 \leq 1 + \ln(N)$.

WHY?

1. Analysis for G: For each round t at which $\ell^t_{x^t} - \ell^t_{x^*} = 1$:
   - $L^{t-1}_{x^t} \leq L^{t-1}_{x^*}$
   - $L^{t}_{x^t} = 1 + L^{t-1}_{x^t} > L^{t-1}_{x^*} = L^{t}_{x^*}$

   $\therefore\ |S^t| \leq |S^{t-1}| - 1$

   $\therefore$ At most N losses of G between two consecutive losses of L^T_{min}.

2. Analysis for RG: Let $t_j = \min\{t : L^T_{min} \geq j\}$.
   - $\forall t \in (t_j, t_{j+1}]$, **if** $|S^t| = |S^{t-1}| - k$ **then** $L^t_{RG} - L^{t-1}_{RG} \leq \frac{k}{|S^{t-1}|}$
   - $L^{t_{j+1}} - L^{t_j} \leq \frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{2} + 1 \leq 1 + \ln(N)$.

■

Determinism vs. External Regret (I)

THEOREM 4.4 (AGT-book): Determinism is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

3. For any deterministic online algorithm D, there is a sequence of T loss vectors, such that $L_D^T \geq T$ and $L_{min}^T \leq \left\lfloor \frac{T}{N} \right\rfloor$.

Determinism vs. External Regret (I)

THEOREM 4.4 (AGT-book): Determinism is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

3. For any deterministic online algorithm D, there is a sequence of T loss vectors, such that $L_D^T \geq T$ and $L_{min}^T \leq \left\lfloor \frac{T}{N} \right\rfloor$.

WHY?

- x^t = the choice of D in round t.
- Loss sequence: $\forall t \in [T]$, $\ell_{x^t}^t = 1$ and $\ell_i^t = 0, \ \forall i \neq x^t$.

Determinism vs. External Regret (I)

THEOREM 4.4 (AGT-book): Determinism is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

3. For any deterministic online algorithm D, there is a sequence of T loss vectors, such that $L_D^T \geq T$ and $L_{min}^T \leq \left\lfloor \frac{T}{N} \right\rfloor$.

WHY?

- $x^t =$ the choice of D in round t.
- Loss sequence: $\forall t \in [T]$, $\ell_{x^t}^t = 1$ and $\ell_i^t = 0, \ \forall i \neq x^t$.
- $L_D^T = T$.

Determinism vs. External Regret (I)

THEOREM 4.4 (AGT-book): Determinism is not enough for EXT-REG

Wrt the External-Regret family of modification rules, the following hold:

3. For any deterministic online algorithm D, there is a sequence of T loss vectors, such that $L_D^T \geq T$ and $L_{min}^T \leq \left\lfloor \frac{T}{N} \right\rfloor$.

WHY?

- x^t = the choice of D in round t.
- Loss sequence: $\forall t \in [T]$, $\ell_{x^t}^t = 1$ and $\ell_i^t = 0, \ \forall i \neq x^t$.
- $L_D^T = T$.
- **Pigeonhole Principle:** At least one action x^* is chosen by D at most $\frac{T}{N}$ times.

$\therefore \ L_{min}^T \leq \left\lfloor \frac{T}{N} \right\rfloor$.

Determinism or Randomness?

- Determinism is **hopeless**.

Determinism or Randomness?

- Determinism is **hopeless**.
- Randomness helps: *RG* improved over *G*.

Determinism or Randomness?

- Determinism is **hopeless**.
- Randomness helps: *RG* improved over *G*.
- *RG* is still not no-external-regret algorithm.

Determinism or Randomness?

- Determinism is **hopeless**.
- Randomness helps: *RG* improved over *G*.
- *RG* is still not no-external-regret algorithm.

What did really help? Can it be further exploited?

No-External-Regret Algorithms (I)

RANDOMIZED WEIGHTED MAJORITY (RWM): Smoothly decrease probability masses of actions as they become worse. For some small $\eta \in (0,1)$:

1. $\forall i \in [N],\ w_i(1) = 1;\ p_i(1) = 1/N;$
2. $\forall t \geq 2$:
   $\forall i \in [N],\ w_i(t) = w_i(t-1) \cdot (1-\eta)^{\ell_i(t-1)};$
   $W(t) = \sum_{i\in[N]} w_i(t);\ \forall i \in [N], p_i(t) = \frac{w_i(t)}{W(t)}$
   Select $x(t)$ according to distribution $\mathbf{p(t)}$.

POLYNOMIAL WEIGHTS (PW): Substitute exponentially sensitive or RWM to polynomially sensitive weight updates. For some small $\eta \in (0,1)$:

1. $\forall i \in [N],\ w_i(1) = 1;\ p_i(1) = 1/N;$
2. $\forall t \geq 2$:
   $\forall i \in [N],\ w_i(t) = w_i(t-1) \cdot (1-\eta \cdot \ell_i(t-1));$
   $W(t) = \sum_{i\in[N]} w_i(t);\ \forall i \in [N], p_i(t) = \frac{w_i(t)}{W(t)}$
   Select $x(t)$ according to distribution $\mathbf{p(t)}$.

No-External-Regret Algorithms (II)

THEOREMS 4.5-6 (AGT-book): RWM & PW Learn against EXT-REG

1. For any $\eta \in (0, 1/2]$ and any sequence of *binary losses*, RWM has
$$L^T_{RWM} \leq L^T_{\min} + 2\sqrt{T \cdot \ln(N)}$$
2. For $\eta \in (0, 1/2]$ and any sequence of *normalized losses*, PW has
$$L^T_{PW} \leq L^T_{\min} + 2\sqrt{T \cdot \ln(N)}$$

No-External-Regret Algorithms (III)

WHY? (for RWG only, similar analysis for PW)

- $Z^t = \sum_{i:\ell_i^T=1} \frac{w_i^t}{W^t}$. /* expected loss of RWM at round t */
- $W^{t+1} = W^t \cdot (1 - \eta Z^t) \geq \max_i\{w_i^{t+1} = (1-\eta)^{L^t_{min}}$
- $W^1 = N$.
- $(1-\eta)^{L^T_{\min}} \leq W^{t+1} = W^t(1-\eta Z^t) = \cdots = N \prod_{r=1}^t (1-\eta Z^r)$

$$\Rightarrow L^T_{\min} \ln(1-\eta) \leq \ln(N) + \sum_{r=1}^t \ln(1-\eta Z^r) \leq \ln(N) - \underbrace{\sum_{r=1}^t Z^r}_{=L^t_{RWM}}$$

$$\Rightarrow L^T_{RWM} \leq \frac{\ln(N)}{\eta} - \frac{\ln(1-\eta)}{\eta} L^T_{min} \leq \frac{\ln(N)}{\eta} + (1+\eta) L^T_{min}$$

$$\Rightarrow L^T_{RWM} \leq L^T_{min} + 2\sqrt{T \ln(N)}$$

/* set $\eta = \min\{\sqrt{\ln(N)/T}, 1/2\}$ */

■

Lower Bounds Against External Regret (I)

THEOREMS 4.7-8 (AGT-book): Not Much More Can Be Done

1. For any $T \leq \log_2(N)$, there is a stochastic generation of a loss sequence, s.t. any online algorithm R has $\mathbb{E}\left[L_R^T\right] = \frac{T}{2}$ and yet, $L_{\min}^T = 0$.
2. For $N = 2$ possible actions, there exists a stochastic generation of a loss sequence, s.t. any online algorithm R has $\mathbb{E}\left[L_R^T - L_{\min}^T\right] = \Omega\left(\sqrt{T}\right)$.

Lower Bounds Against External Regret (I)

WHY? (only of first bound, similar analysis for second)

- Proposed sequence of losses:

  t=1: $S^1 \in_{\mathbf{uar}} [N] : |S^1| = N/2$.

  t=2: $S^2 \in_{\mathbf{uar}} S^1 : |S^2| = N/4$.

  ...

  t=k: $S^k \in_{\mathbf{uar}} S^{k-1} : |S^k| = N/2^k$.

Lower Bounds Against External Regret (I)

WHY? (only of first bound, similar analysis for second)

- Proposed sequence of losses:

  t=1: $S^1 \in_{\mathbf{uar}} [N] : |S^1| = N/2.$

  t=2: $S^2 \in_{\mathbf{uar}} S^1 : |S^2| = N/4.$

  ...

  t=k: $S^k \in_{\mathbf{uar}} S^{k-1} : |S^k| = N/2^k.$

- $\forall t \geq 1,\ \forall i \in S^t, \ell_i^t = 0 \ \wedge\ \forall i \notin S^t, \ell_i^t = 1.$

Lower Bounds Against External Regret (I)

WHY? (only of first bound, similar analysis for second)

- Proposed sequence of losses:

  t=1: $S^1 \in_{\mathbf{uar}} [N] : |S^1| = N/2$.

  t=2: $S^2 \in_{\mathbf{uar}} S^1 : |S^2| = N/4$.

  ...

  t=k: $S^k \in_{\mathbf{uar}} S^{k-1} : |S^k| = N/2^k$.

- $\forall t \geq 1,\ \forall i \in S^t, \ell_i^t = 0\ \wedge\ \forall i \notin S^t, \ell_i^t = 1$.

- $T < \log_2(N) \Rightarrow S^T \geq 1 \Rightarrow L_{min}^T = 0$.

Lower Bounds Against External Regret (I)

WHY? (only of first bound, similar analysis for second)

- Proposed sequence of losses:

  t=1: $S^1 \in_{\mathbf{uar}} [N] : |S^1| = N/2$.

  t=2: $S^2 \in_{\mathbf{uar}} S^1 : |S^2| = N/4$.

  ...

  t=k: $S^k \in_{\mathbf{uar}} S^{k-1} : |S^k| = N/2^k$.

- $\forall t \geq 1,\ \forall i \in S^t, \ell_i^t = 0 \ \wedge\ \forall i \notin S^t, \ell_i^t = 1$.
- $T < \log_2(N) \Rightarrow S^T \geq 1 \Rightarrow L_{min}^T = 0$.
- $L_R^T \geq \frac{T}{2}$.

Lower Bounds Against External Regret (I)

WHY? (only of first bound, similar analysis for second)

- Proposed sequence of losses:

  t=1: $S^1 \in_{\mathbf{uar}} [N] : |S^1| = N/2$.

  t=2: $S^2 \in_{\mathbf{uar}} S^1 : |S^2| = N/4$.

  ...

  t=k: $S^k \in_{\mathbf{uar}} S^{k-1} : |S^k| = N/2^k$.

- $\forall t \geq 1, \ \forall i \in S^t, \ell_i^t = 0 \ \wedge \ \forall i \notin S^t, \ell_i^t = 1$.
- $T < \log_2(N) \Rightarrow S^T \geq 1 \Rightarrow L_{min}^T = 0$.
- $L_R^T \geq \frac{T}{2}$.

■

From External-Regret to Swap-Regret Algorithms (I)

THEOREMS 4.15 (AGT-book): No-Swap-Regret Algorithms

Given an algorithm A that has *external-regret* R:

$$\boxed{L_A^T \leq L_{min}^T + R}$$

it is possible to create, via a polynomial reduction using N copies of A, some (master) online algorithm H with *swap-regret* NR:

$$\boxed{L_H^T \leq L_{H,F_{sw}}^T + NR}$$

From External-Regret to Swap-Regret Algorithms (I)

THEOREMS 4.15 (AGT-book): No-Swap-Regret Algorithms

Given an algorithm A that has *external-regret* R:

$$\boxed{L_A^T \leq L_{min}^T + R}$$

it is possible to create, via a polynomial reduction using N copies of A, some (master) online algorithm H with *swap-regret* NR:

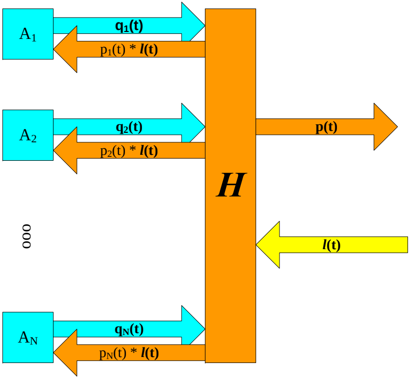
$$\boxed{L_H^T \leq L_{H,F_{sw}}^T + NR}$$

COROLLARY 4.16 (AGT-book): No-Swap-Regret Algorithms

There is an online algorithm H such that, for any (swap) function $f : [N] \mapsto [N]$ it guarantees that:

$$\boxed{L_H^T \leq L_{H,F}^T + \mathrm{O}\left(N\sqrt{T\ln(N)}\right)}$$

- Each copy acts as an **independent expert**.
- The master algorithm H creates a new distribution $\mathbf{p}(t)$ as the outcome of the experts' opinions.
  - $\mathbf{p}(\mathbf{t})' = \mathbf{p}(\mathbf{t})' Q(t)$ is the stationary distribution of the Markov process with transition matrix $Q(t) = [\mathbf{q_1}(\mathbf{t})'; \mathbf{q_2}(\mathbf{t})'; \cdots ; \mathbf{q_N}(\mathbf{t})']$.
- H splits the actual loss vector $\ell(t)$ among the experts, to allow them to *learn*.



Skeleton

How About Multi-agent Environments?

Rather than having an ``agent vs. Nature'' scenario, what if two (or more) agents are *self-interested*, ie, each of them has its own *preferential order* to the states of the whole system, and acts in an attempt to bring about the most preferable states for it?

Game Theoretic Notation: Strategic Games (I)

Strategic Game or **Stage Game**: $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P}\rangle$.

- P is the set of (self--interested) agents (aka **players**).

Game Theoretic Notation: Strategic Games (I)

Strategic Game or **Stage Game**: $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P} \rangle$.

- P is the set of (self--interested) agents (aka **players**).
- $\forall p \in P,\ S_p$ is the set of **actions** for agent p. $S = \times_{p\in P} S_p$ is the system's **state space**.

Game Theoretic Notation: Strategic Games (I)

Strategic Game or **Stage Game**: $G = \langle P, (S_p)_{p \in P}, (c_p)_{p \in P} \rangle$.

- P is the set of (self--interested) agents (aka **players**).
- $\forall p \in P,\ S_p$ is the set of **actions** for agent p. $S = \times_{p \in P} S_p$ is the system's **state space**.
- $\forall p \in P, c_p : S \mapsto [0, 1]$ is the (normalized) **cost function** for agent p, depending on the system state determined by the actions of all agents.

Game Theoretic Notation: Strategic Games (I)

Strategic Game or **Stage Game**: $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P}\rangle$.

- P is the set of (self--interested) agents (aka **players**).
- $\forall p \in P,\ S_p$ is the set of **actions** for agent p. $S = \times_{p\in P} S_p$ is the system's **state space**.
- $\forall p \in P, c_p : S \mapsto [0,1]$ is the (normalized) **cost function** for agent p, depending on the system state determined by the actions of all agents.
- **Strategy** $\mathbf{x}_p \in \Delta(S_p) = \{\mathbf{z} \in [0,1]^{|S_p|} : \sum_{s_p\in S_p} z(s_p) = 1\}$ is a probability distribution used by agent p to determine its action, *independently* of the other agents' choices.

Game Theoretic Notation: Strategic Games (I)

Strategic Game or **Stage Game**: $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P}\rangle$.

- P is the set of (self--interested) agents (aka **players**).
- $\forall p \in P,\ S_p$ is the set of **actions** for agent p. $S = \times_{p\in P} S_p$ is the system's **state space**.
- $\forall p \in P, c_p : S \mapsto [0,1]$ is the (normalized) **cost function** for agent p, depending on the system state determined by the actions of all agents.
- **Strategy** $\mathbf{x}_p \in \Delta(S_p) = \{\mathbf{z} \in [0,1]^{|S_p|} : \sum_{s_p\in S_p} z(s_p) = 1\}$ is a probability distribution used by agent p to determine its action, *independently* of the other agents' choices.
- **Correlated Strategy** $\sigma \in \Delta(S) = \{\mathbf{z} \in [0,1]^{|S|} : \sum_{\mathbf{s}\in S} z(\mathbf{s}) = 1\}$ is a probability distribution for *the system* to determine its own (suggested) state.

Game Theoretic Notation: Strategic Games (II)

- **Loss** of agent $p \in P$: The *expected cost* that p suffers for the actions profile adopted by all the agents. Ie: $\forall (\mathbf{x}_p)_{p \in P} \in \times_{p \in P} \Delta(S_p),\ \forall \sigma \in \Delta(S)$, $\ell_p(\mathbf{x}_1, \ldots, \mathbf{x}_{|P|}) = \mathbb{E}_{(s_q \sim \mathbf{x}_q)_{q \in P}} \left[c_p(s_1, \ldots, s_{|P|}) \right]$ and $\ell_p(\sigma) = \mathbb{E}_{\mathbf{s} \sim \sigma} \left[c_p(\mathbf{s}) \right]$

Game Theoretic Notation: Strategic Games (II)

- **Loss** of agent $p \in P$: The *expected cost* that p suffers for the actions profile adopted by all the agents. Ie: $\forall (\mathbf{x}_p)_{p\in P} \in \times_{p\in P}\Delta(S_p),\ \forall \sigma \in \Delta(S)$, $\ell_p(\mathbf{x}_1, \ldots, \mathbf{x}_{|P|}) = \mathbb{E}_{(s_q \sim \mathbf{x}_q)_{q\in P}}\left[c_p(s_1, \ldots, s_{|P|})\right]$ and $\ell_p(\sigma) = \mathbb{E}_{\mathbf{s}\sim\sigma}\left[c_p(\mathbf{s})\right]$

- $\forall \mathbf{x}_p, \mathbf{y}_p \in \Delta(S_p)$,
  $\mathbf{x}_p$ is **dominated** by $\mathbf{y}_p$ iff $\forall \mathbf{z}_{-p} \in \times_{q\neq p}\Delta(S_q),\ \ell_p(\mathbf{x}_p, \mathbf{z}_{-p}) \leq \ell_p(\mathbf{y}_p, \mathbf{z}_{-p})$.

Game Theoretic Notation: Strategic Games (II)

- **Loss** of agent $p \in P$: The *expected cost* that p suffers for the actions profile adopted by all the agents. Ie: $\forall (\mathbf{x}_p)_{p\in P} \in \times_{p\in P}\Delta(S_p),\ \forall \sigma \in \Delta(S)$, $\ell_p(\mathbf{x}_1, \ldots, \mathbf{x}_{|P|}) = \mathbb{E}_{(s_q\sim \mathbf{x}_q)_{q\in P}}\left[c_p(s_1, \ldots, s_{|P|})\right]$ and $\ell_p(\sigma) = \mathbb{E}_{\mathbf{s}\sim\sigma}\left[c_p(\mathbf{s})\right]$

- $\forall \mathbf{x}_p, \mathbf{y}_p \in \Delta(S_p)$,
  $\mathbf{x}_p$ is **dominated** by $\mathbf{y}_p$ iff $\forall \mathbf{z}_{-p} \in \times_{q\neq p}\Delta(S_q),\ \ell_p(\mathbf{x}_p, \mathbf{z}_{-p}) \leq \ell_p(\mathbf{y}_p, \mathbf{z}_{-p})$.

- **Nash Equilibrium (NE):** A (publicly known) profile of strategies $(\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_{|P|})$ for all the agents, such that no agent can reduce its own loss by *unilaterally deviating* from its strategy, given the strategies of the other agents.

Game Theoretic Notation: Strategic Games (II)

- **Loss** of agent $p \in P$: The *expected cost* that p suffers for the actions profile adopted by all the agents. Ie: $\forall(\mathbf{x}_p)_{p\in P} \in \times_{p\in P}\Delta(S_p),\ \forall\sigma \in \Delta(S)$, $\ell_p(\mathbf{x}_1, \ldots, \mathbf{x}_{|P|}) = \mathbb{E}_{(s_q\sim\mathbf{x}_q)_{q\in P}}\left[c_p(s_1, \ldots, s_{|P|})\right]$ and $\ell_p(\sigma) = \mathbb{E}_{\mathbf{s}\sim\sigma}\left[c_p(\mathbf{s})\right]$

- $\forall\mathbf{x}_p, \mathbf{y}_p \in \Delta(S_p)$,
  $\mathbf{x}_p$ is **dominated** by $\mathbf{y}_p$ iff $\forall\mathbf{z}_{-p} \in \times_{q\neq p}\Delta(S_q),\ \ell_p(\mathbf{x}_p, \mathbf{z}_{-p}) \leq \ell_p(\mathbf{y}_p, \mathbf{z}_{-p})$.

- **Nash Equilibrium (NE):** A (publicly known) profile of strategies $(\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_{|P|})$ for all the agents, such that no agent can reduce its own loss by *unilaterally deviating* from its strategy, given the strategies of the other agents.

- **Correlated Equilibrium (CE):** A (publicly known) correlated strategy $\bar{\sigma} \in \Delta(S)$ such that if the system first chooses an action profile $\mathbf{s} \sim \bar{\sigma}$ and then *suggests secretely* action s_p to each agent $p \in P$, then no agent can reduce its own loss by deviating from s_p, given that the other agents will follow the system's suggestion.

EXAMPLE: Prisoner's Dilemma

- Two individuals are caught for a delinquency (eg, causing a car accident) deserving 1 year of imprisonment.
- There are suspicions for having committed a felony (eg, bank robbery) deserving 10 years of imprisonment. But there are no sufficient evidence.
- Police tries to get their confessions by making the following agreement with both of them, but not allowing them to communicate with each other:

		Prisoner 2: Betray	Prisoner 2: Silent
Prisoner 1	Betray	5,5	0,10
Prisoner 1	Silent	10,0	1,1

Game Theoretic Notation: Strategic Games (III)

EXAMPLE: Prisoner's Dilemma

- Two individuals are caught for a delinquency (eg, causing a car accident) deserving 1 year of imprisonment.
- There are suspicions for having committed a felony (eg, bank robbery) deserving 10 years of imprisonment. But there are no sufficient evidence.
- Police tries to get their confessions by making the following agreement with both of them, but not allowing them to communicate with each other:

``Silent'' is **dominated** by ``Betray'', for both players.

Unique **NE** and CE point: $(\mathit{Betray}, \mathit{Betray})$.

		Prisoner 2	
		Betray	Silent
Prisoner 1	Betray	5,5	0,10
	Silent	10,0	1,1

Game Theoretic Notation: Repeated Games (I)

(Infinitely) Repeated Game G^∞: The infinite realizations of *independent instances* of the stage game G.

- Each player $p \in P$ must determine an **algorithm** M_p that takes as input the *history of the game* for the first $t-1$ rounds, and returns a strategy $\mathbf{x_p^t} \in \Delta(S_p)$ for round t.
- The **loss** $\ell_p^t(M_1, \ldots, M_{|P|})$ of agent p at round t, is the expected cost it suffers for the profile $\mathbf{x}^t$ adopted at round t, according to algorithms $M_1, \ldots, M_{|P|}$.
- The **cumulative loss** $L_p^T(M_1, \ldots, M_{|P|})$ of $p \in P$ up to $T \in \mathbb{N}$ is the sum of losses of p for the first T rounds.

Game Theoretic Notation: Repeated Games (II)

- **Limit-Of-Means** criterion: For any given profile of algorithms M_{-p} for the other agents, two different algorithms M_p, M'_p for agent p are compared according to the **average loss** they produce over T rounds, as $T \to \infty$.
- A collection $(M_p)_{p \in P}$ of algorithms for the agents is **Nash Equilibrium** of G^{∞} iff $\forall p \in P$ no alternative algorithm M'_p can assure *smaller average loss*, given that the other agents keep their algorithms unchanged.

Game Theoretic Notation: Repeated Games (III)

EXAMPLE: Repeated Prisoner's Dilemma

		Prisoner 2	
		Betray	Silent
Prisoner 1	Betray	5,5	0,10
	Silent	10,0	1,1

- The algorithm

  $\boxed{\text{``Always betray''}}$

  adopted by both agents is a NE point of G^{∞}, (considers a NE point of the stage game in each round).

- The algorithm

  $\boxed{\text{``Remain silent until opponent betrays. Then always betray''}}$

  adopted by both agents is a NE point of G^{∞}, although it does NOT create NE points in the rounds, in fear of **future explosion in loss**.

Learning NE in Bimatrix Games (I)

- A stage game $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P}\rangle$ is **constant-sum**, if there is a constant $\gamma \in \mathbb{R}$, such that $\forall \mathbf{s} \in S, \sum_{p\in P} c_p(\mathbf{s}) = \gamma$.
- For each $\gamma-$sum bimatrix game, any NE point assures *exactly the same pair of losses*, $(v_1, v_2) \in [0, 1]^2$ for the two players (their minmax values).

Learning NE in Bimatrix Games (I)

- A stage game $G = \langle P, (S_p)_{p \in P}, (c_p)_{p \in P} \rangle$ is **constant-sum**, if there is a constant $\gamma \in \mathbb{R}$, such that $\forall \mathbf{s} \in S, \sum_{p \in P} c_p(\mathbf{s}) = \gamma$.
- For each γ—sum bimatrix game, any NE point assures *exactly the same pair of losses*, $(v_1, v_2) \in [0, 1]^2$ for the two players (their minmax values).

EXAMPLE: Matching Pennies

		Player 2: Heads	Player 2: Tail
Player 1	Heads	1,0	0,1
Player 1	Tail	0,1	1,0

- Unique NE point: $((0.5, 0.5), (0.5, 0.5))$.
- Values: $v_1 = v_2 = 0.5$.

THEOREM: External-Regret Works For $\gamma-$Sum Bimatrix Games

1. For any $\gamma-$sum bimatrix (stage) game G with values $(v_1, v_2) \in [0,1]^2$, if *one player* p adopts some algorithm ON with external-regret R in the infinite game G^∞, then for any algorithm A adopted by the opponent, its *cumulative loss* after T rounds will be: $L_p^T(ON, A) \leq T \cdot v_p + R$

Learning NE in Bimatrix Games (II)

THEOREM: External-Regret Works For $\gamma-$Sum Bimatrix Games

1. For any $\gamma-$sum bimatrix (stage) game G with values $(v_1, v_2) \in [0, 1]^2$, if *one player* p adopts some algorithm ON with external-regret R in the infinite game G^∞, then for any algorithm A adopted by the opponent, its *cumulative loss* after T rounds will be: $L_p^T(ON, A) \leq T \cdot v_p + R$

2. If *both players* adopt no-external-regret algorithms (ON_1, ON_2) for G^∞, then the profile produced by the *average strategy* per player converges to a NE point of the stage game G.

Learning NE in Bimatrix Games (II)

THEOREM: External-Regret Works For $\gamma-$Sum Bimatrix Games

1. For any $\gamma-$sum bimatrix (stage) game G with values $(v_1, v_2) \in [0, 1]^2$, if *one player* p adopts some algorithm ON with external-regret R in the infinite game G^∞, then for any algorithm A adopted by the opponent, its *cumulative loss* after T rounds will be: $L_p^T(ON, A) \leq T \cdot v_p + R$
2. If *both players* adopt no-external-regret algorithms (ON_1, ON_2) for G^∞, then the profile produced by the *average strategy* per player converges to a NE point of the stage game G.
3. We can use the existence of no-external-regret algorithms to prove the **von Neumann's minimax theorem** for $\gamma-$sum bimatrix games.

THEOREM: External-Regret Works For $\gamma-$Sum Bimatrix Games

1. For any $\gamma-$sum bimatrix (stage) game G with values $(v_1, v_2) \in [0, 1]^2$, if *one player* p adopts some algorithm ON with external-regret R in the infinite game G^∞, then for any algorithm A adopted by the opponent, its *cumulative loss* after T rounds will be: $L_p^T(ON, A) \leq T \cdot v_p + R$
2. If *both players* adopt no-external-regret algorithms (ON_1, ON_2) for G^∞, then the profile produced by the *average strategy* per player converges to a NE point of the stage game G.
3. We can use the existence of no-external-regret algorithms to prove the **von Neumann's minimax theorem** for $\gamma-$sum bimatrix games.
4. For a *non-constant-sum* bimatrix game G, we cannot guarantee convergence of any no-external-regret algorithms to NE point of G.

Convergence to CE points of $N-$person Games

THEOREM: Swap-Regret Works With $N-$Person Games

Let $G = \langle P, (S_p)_{p\in P}, (c_p)_{p\in P}\rangle$ be an $N-$person stage game.

1. If each player $p \in P$ adopts an algorithm ON_p with swap-regret R for the first T time steps of G^∞, then the **empirical distribution** of the joint actions played is an $(R/T)-$correlated equilibrium of the game.

Convergence to CE points of $N-$person Games

THEOREM: Swap-Regret Works With $N-$Person Games

Let $G = \langle P, (S_p)_{p \in P}, (c_p)_{p \in P} \rangle$ be an $N-$person stage game.

1. If each player $p \in P$ adopts an algorithm ON_p with swap-regret R for the first T time steps of G^{∞}, then the **empirical distribution** of the joint actions played is an $(R/T)-$correlated equilibrium of the game.

2. For any player $p \in P$ that uses an algorithm ON_p with swap-regret R for T time steps, the **average probability mass** that p puts on the set of $\varepsilon-$dominated actions is at most $\frac{R}{\varepsilon T}$.

Convergence to CE points of $N-$person Games

THEOREM: Swap-Regret Works With $N-$Person Games

Let $G = \langle P, (S_p)_{p \in P}, (c_p)_{p \in P} \rangle$ be an $N-$person stage game.

1. If each player $p \in P$ adopts an algorithm ON_p with swap-regret R for the first T time steps of G^{∞}, then the **empirical distribution** of the joint actions played is an $(R/T)-$correlated equilibrium of the game.

2. For any player $p \in P$ that uses an algorithm ON_p with swap-regret R for T time steps, the **average probability mass** that p puts on the set of $\varepsilon-$dominated actions is at most $\frac{R}{\varepsilon T}$.

RECAP: Given any algorithm with external-regret R that chooses among N possible states, there is a generic algorithm H for player $p \in P$ with swap-regret at most $N \cdot R$. This implies then that for any *swap-regret modification rule* $f : [N] \mapsto [N]$, p can assure:

$$L_H^T \leq L_{H,f} + \mathrm{O}\left(N\sqrt{T \cdot \log(N)}\right)$$

How About Special Cases Of Games?

A strategic game is **sociall concave** iff it has:

- Closed convex strategy sets.
- A (weighted) social welfare function that is concave.
- Convex utility functions of each player, in the vector of the *other players´ actions*.

Examples of socially concave games:

- Zero-sum games.
- Resource allocation games.
- Selfish routing games.
- Cournot oligopoly.
- TCP congestion control.

How About Special Cases Of Games?

A strategic game is **sociall concave** iff it has:

- Closed convex strategy sets.
- A (weighted) social welfare function that is concave.
- Convex utility functions of each player, in the vector of the *other players´ actions*.

Examples of socially concave games:

- Zero-sum games.
- Resource allocation games.
- Selfish routing games.
- Cournot oligopoly.
- TCP congestion control.

THEOREM: External Regret Works with Socially Concave Games

If each player uses a no-regret procedure in an infinite game G^{∞} whose stage game belongs to some class of interesting games, then their joint play **converges** to Nash equilibrium.

Skeleton

Beyond Learning?

- Learning can help us discover NE points in **constant-sum bimatrix** stage games.

Beyond Learning?

- Learning can help us discover NE points in **constant-sum bimatrix** stage games.
- How about **non-constant-sum** games? How about **more than 2 players**?

Beyond Learning?

- Learning can help us discover NE points in **constant-sum bimatrix** stage games.
- How about **non-constant-sum** games? How about **more than** 2 **players**?

Can we compute NE points for **general stage** games?

Beyond Learning?

- Learning can help us discover NE points in **constant-sum bimatrix** stage games.
- How about **non-constant-sum** games? How about **more than 2 players**?

Can we compute NE points for *general stage* games?

(Chen-Deng (2006), Daskalakis-Goldberg-Papadimitriou (2006)) : Computing NE points is $\mathcal{PPAD}-$hard for stage games, even for two players.

Beyond Learning?

- Learning can help us discover NE points in **constant-sum bimatrix** stage games.
- How about **non-constant-sum** games? How about **more than 2 players**?

Can we compute NE points for general stage games?

☹ (Chen-Deng (2006), Daskalakis-Goldberg-Papadimitriou (2006)) : Computing NE points is $\mathcal{PPAD}-$hard for stage games, even for two players.

How about infinitely repeated games?

DEFINITION: Threat Point

- $G = \langle P, (S_p)_{p\in P}, (U_p : \times_{q\in P} S_q \mapsto \mathbb{Q})_{p\in P}\rangle$: An arbitrary stage game, with rational payoff functions (to be *maximized*).
- G^∞: The *infinitely repeated game* using the stage game G in each round.
- **Threat Point**: The vector of *minimum payoffs* that each player would accept in a realization of G, against a profile of **uncoordinated strategies** for the opponents. Ie:

$$\boxed{\forall p \in P, \theta_p(G) \equiv \min_{\mathbf{x}_{-p}\in\times_{q\neq p}\Delta(S_q)} \max_{\mathbf{x}_p\in\Delta(S_p)} U_p(\mathbf{x}_{-p}, \mathbf{x}_p)}$$

The Folk Theorem and its Consequences

Folk Theorem

*``Any vector of payoffs in a **one-shot** game G which is component-wise larger than the **threat point** of G, can be enforced as a NE point of the corresponding **infinitely repeated** game G^{∞}''.*

- Computation of Equilibrium Points in repeated games **should be (?) easier**.

The Folk Theorem and its Consequences

Folk Theorem

``Any vector of payoffs in a ***one-shot*** *game G which is component-wise larger than the* ***threat point*** *of G, can be enforced as a NE point of the corresponding* ***infinitely repeated*** *game G^{∞}''.*

- Computation of Equilibrium Points in repeated games **should be (?) easier**.

☺ (Littman-Stone (2003)) *Polynomial--time* construction of a *succinctly representable* profile of algorithms for G^{∞}, that induces an **arbitrary rational payoffs point** that is above the threat point, as a NE of G^{∞}, for the case of two players.

The Folk Theorem and its Consequences

Folk Theorem

*``Any vector of payoffs in a **one-shot** game G which is component-wise larger than the **threat point** of G, can be enforced as a NE point of the corresponding **infinitely repeated** game G^{∞}''.*

- Computation of Equilibrium Points in repeated games **should be (?) easier**.

(Littman-Stone (2003)) *Polynomial--time* construction of a *succinctly representable* profile of algorithms for G^{∞}, that induces an **arbitrary rational payoffs point** that is above the threat point, as a NE of G^{∞}, for the case of two players.

(Borgs et al. (2008)) Computing Nash equilibria for infinitely repeated games with at least three players, is $\mathcal{PPAD}-$hard.

The Threat Point (II)

Remark

Two parameters of intractability in (Borgs et al. (2008)) :

1. Computing (even approximately) the *threat point* of a **one-shot game** G among $k \geq 3$ players, is $\mathcal{NP}-$hard.

The Threat Point (II)

Remark

Two parameters of intractability in (Borgs et al. (2008)) :

1. Computing (even approximately) the *threat point* of a **one-shot game** G among $k \geq 3$ players, is $\mathcal{NP}-$hard.

2. Computing an *approximate NE point* of a $(k+1)$-player, infinitely repeated game is as hard as computing an (approximate) NE point in a $k-$player, one-shot game, for any $k \geq 2$.

The Threat Point (II)

Remark

Two parameters of intractability in (Borgs et al. (2008)) :

1. Computing (even approximately) the *threat point* of a **one-shot game** G among $k \geq 3$ players, is $\mathcal{NP}-$hard.

   Crucial knowledge for the approach of (Littman-Stone (2003)) to solve the $2-$players case.

2. Computing an *approximate NE point* of a $(k+1)$-player, infinitely repeated game is as hard as computing an (approximate) NE point in a $k-$player, one-shot game, for any $k \geq 2$.

The Threat Point (II)

Remark

Two parameters of intractability in (Borgs et al. (2008)) :

1. Computing (even approximately) the *threat point* of a **one-shot game** G among $k \geq 3$ players, is $\mathcal{NP}-$hard.

   Crucial knowledge for the approach of (Littman-Stone (2003)) to solve the $2-$players case.

2. Computing an *approximate NE point* of a $(k+1)$-player, infinitely repeated game is as hard as computing an (approximate) NE point in a $k-$player, one-shot game, for any $k \geq 2$.

How much credible can a threat be, when it is not efficiently computable by any of the players?

A Way Out

Our main objectives are to:

1. Find a way to *tackle the intractability* of the threat point.

2. Find a way to *implement the Folk Theorem*, ie, induce **some / any** (rational) payoff point above the (new) threat point as a **succinctly representable** equilibrium of the infinitely repeated game.

A New Notion of Threat

DEFINITION: Correlated Threat Point

The **correlated threat point** of a stage game $G = \langle P, (S_p)_{p \in P}, (U_p)_{p \in P} \rangle$ is a vector of minimum payoffs that each of the players would be *willing to accept*, against any profile of **coordinated strategies** of the opponents against her. Ie:

$$\boxed{\forall p \in [k], \ \varphi_p(G) \equiv \min_{\sigma_{-p} \in \Delta\left(\times_{q \neq p} S_q\right)} \max_{\mathbf{x}_p \in \Delta(S_p)} U_p(\sigma_{-p}, \mathbf{x}_p)}$$

Some Observations on the Correlated Threat Point

1. It constitutes a **credible threat** for any *constant number* of players.

Some Observations on the Correlated Threat Point

1. It constitutes a **credible threat** for any *constant number* of players.

2. It is **more severe** than the traditional notion of threat point, but **not overwhelming** for each player (it is closer to the notion of *worst case* scenario, that is widely used in TCS).

Some Observations on the Correlated Threat Point

1. It constitutes a **credible threat** for any *constant number* of players.

2. It is **more severe** than the traditional notion of threat point, but **not overwhelming** for each player (it is closer to the notion of *worst case* scenario, that is widely used in TCS).

3. It implies, not Nash equilibria, but **almost Nash** equilibria:
   Correlation is only required for the punishments. During normal play the agents act *independently* (but in time--synchrony).

Some Observations on the Correlated Threat Point

1. It constitutes a **credible threat** for any *constant number* of players.

2. It is **more severe** than the traditional notion of threat point, but **not overwhelming** for each player (it is closer to the notion of *worst case* scenario, that is widely used in TCS).

3. It implies, not Nash equilibria, but **almost Nash** equilibria:
   Correlation is only required for the punishments. During normal play the agents act *independently* (but in time--synchrony).

4. It implements the main idea of the Folk Theorem:
   Any *rational* payoff point above it can be induced by the system as equilibrium of the infinite game, by providing *succinctly representable* strategies for the players.

Tractability of Correlated Threat Point (I)

- Player p's **defensive strategy**:

$$\boxed{\mathbf{d}_p \in \arg\max_{\mathbf{x}_p \in \Delta(S_p)} \left\{ \min_{\sigma_{-p} \in \Delta\left(\times_{q \neq p} S_q\right)} U_p\left(\sigma_{-p}, \mathbf{x}_p\right) \right\}}$$

- **Aggressive (correlated) strategy** of the other players against player p:

$$\boxed{\mathbf{a}_p \in \arg\min_{\sigma_{-p} \in \Delta\left(\times_{q \neq p} S_q\right)} \left\{ \max_{\mathbf{x}_p \in \Delta(S_p)} U_p\left(\sigma_{-p}, \mathbf{x}_p\right) \right\}}$$

Tractability of Correlated Threat Point (I)

- Player p's **defensive strategy**:

$$\boxed{\mathbf{d}_p \in \arg\max_{\mathbf{x}_p \in \Delta(S_p)} \left\{ \min_{\sigma_{-p} \in \Delta\left(\times_{q \neq p} S_q\right)} U_p\left(\sigma_{-p}, \mathbf{x}_p\right) \right\}}$$

- **Aggressive (correlated) strategy** of the other players against player p:

$$\boxed{\mathbf{a}_p \in \arg\min_{\sigma_{-p} \in \Delta\left(\times_{q \neq p} S_q\right)} \left\{ \max_{\mathbf{x}_p \in \Delta(S_p)} U_p\left(\sigma_{-p}, \mathbf{x}_p\right) \right\}}$$

THEOREM: Computability of defensive & aggressive strategies

(Kontogiannis-Spirakis (2008))

For any *fixed constant* natural number $k \geq 2$, any finite $k-$person stage game $G = \langle P, (S_p)_{p \in P}, (U_p)_{p \in P} \rangle$ with *rational payoffs*, and any player $p \in P$, the correlated threat value $\varphi_p(G)$, the defensive strategy $\mathbf{d}_p$ and the aggressive strategy $\mathbf{a}_p$ of the other players against p, are **succinctly representable** and **polynomial time computable**, wrt $size(G)$.

Tractability of Correlated Threat Point (II)

WHY?

For each player $p \in P$:

- Consider the following $|S_p| \times |\times_{q \neq p} S_q|$ payoff matrix P_p:

$$\forall (s_p, \mathbf{s}_{-p}) \in S_p \times S_{-p}, P_p[s_p, \mathbf{s}_{-p}] = U_p(s_p, \mathbf{s}_{-p})$$

- Any Nash equilibrium of the *zero sum bimatrix game* $\langle P_p, -P_p \rangle$ determines player p's threat value, her defensive strategy, and the aggressive strategy against her:

$$(V_p, \mathbf{d}_p) \in \arg\max \left\{ \bar{V}_p : \forall \mathbf{s}_{-p} \in S_{-p}, \bar{\mathbf{d}}_p \cdot P_p[\star, \mathbf{s}_{-p}] \geq \bar{V}_p ; \bar{\mathbf{d}}_p \in \Delta(S_p) \right\}$$

$$(V_p, \mathbf{a}_p) \in \arg\min \left\{ \bar{V}_p : \forall s_p \in S_p, P_p[s_p, \star] \cdot \bar{\mathbf{a}} \leq \bar{V}_p ; \bar{\mathbf{a}}_p \in \Delta(S_{-p}) \right\}$$

$$\varphi_p(G) = V_p$$

- Given the rationality of the payoff functions, $V_p, \mathbf{a}_p, \mathbf{d}_p$ are rational vectors and numbers, of size polynomial in $size(G)$. ■

The Strictly Individually Rational Region (I)

- $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$: A $k-$person stage game, with *rational* payoff functions $U_p : S \mapsto \mathbb{Q}$.
- $Z = \left\{ \mathbf{z} \in \mathbb{Q}^k : \exists \mathbf{s} \in S \text{ s.t. } \forall p \in [k], U_p(\mathbf{s}) = \mathbf{z}[p] \right\}$ is the set of all the rational vectors that are payoff points of *some* actions profile $\mathbf{s} \in S$ of G.
- $conv(Z) = \left\{ \sum_{\mathbf{s}\in S} \lambda_{\mathbf{s}} \cdot \mathbf{U}(\mathbf{s}) \in \mathbb{R}^k : \sum_{\mathbf{s}\in S} \lambda_{\mathbf{s}} = 1;\ \forall \mathbf{s} \in S, \lambda_{\mathbf{s}} \geq 0 \right\}$

DEFINITION: Strictly Individual Rational Region

The **strictly individual rational region** of G is the set of all payoff points that are *point-wise greater* than the correlated threat point of G:

$$\boxed{sirr(G) = conv(Z) \cap \left\{ \mathbf{z} \in \mathbb{R}^k : \mathbf{z} > \varphi(G) \right\}}$$

The Strictly Individually Rational Region (II)

Following the terminology of (Littman--Stone (2003)) :

- **Mutual Advantage Case**: $sirr(G) \neq \emptyset$.
- **No Mutual Advantage Case**: $sirr(G) = \emptyset$.

We shall handle these two cases separately.

The Strictly Individually Rational Region (III)

LEMMA 1: Checking emptiness of $sirr(G)$

For any *fixed* integer $k \geq 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$, we can determine in time $poly(size(G))$ whether $sirr(G) \neq \emptyset$.

The Strictly Individually Rational Region (III)

LEMMA 1: Checking emptiness of $sirr(G)$

For any *fixed* integer $k \geq 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$, we can determine in time $poly(size(G))$ whether $sirr(G) \neq \emptyset$.

WHY?

- For any correlated strategy $\sigma \in \Delta(\times_{p\in[k]} S_p)$, the payoff point $\mathbf{U}(\sigma)$ belongs to $conv(Z)$, and vice versa.
- Look for a **minimum-payoff maximizing** point in the boundary of $conv(Z)$: For each of the $\binom{|S|}{k}$ $k-$subsets of vertices $\forall \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\} \subseteq Z$,

$$\boxed{MAC(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)} \quad \begin{array}{lll} \text{maximize} & \zeta & \\ \text{s.t.} & \sum_{i=1}^{k} \mathbf{z}_i[p] \cdot \lambda_i \geq \zeta, & \forall p \in [k] \\ & \sum_{i=1}^{k} \lambda_i = 1 & \\ & \forall i \in [k], \lambda_i \geq 0;\ \zeta \geq 0 & \end{array}$$

■

Enforcing a Payoff for the Mutual Advantage Case (I)

THEOREM: Existence & Construction of NE point of G^{∞}

For any constant $k \geq 2$, and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ such that $sirr(G) \neq \emptyset$, there is a profile of algorithms $M = (M_p)_{p\in[k]}$ for the players that is an equilibrium of G^{∞}, whose description size is $poly(size(G))$.

Enforcing a Payoff for the Mutual Advantage Case (II)

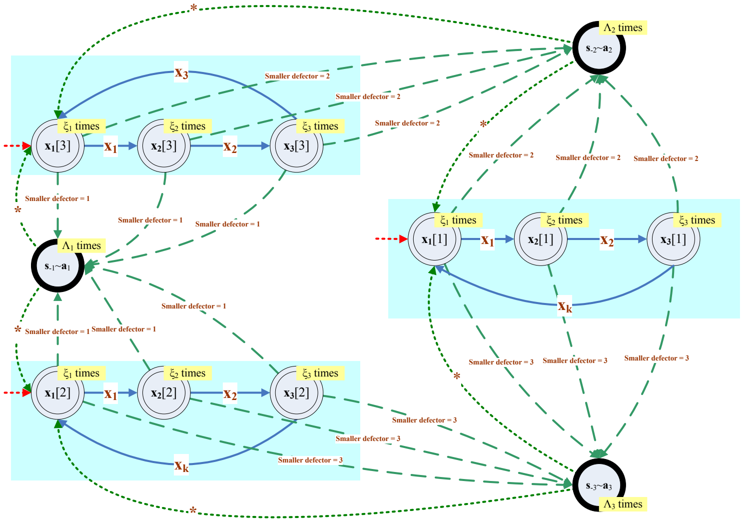
WHY?

- $\mathbf{z}^* = \sum_{i=1}^{k} \hat{\lambda}_i \hat{\mathbf{z}}_i$: The payoff point of *sirr*(G), chosen in LEMMA 1.
- $\forall i \in [k],\ \hat{\lambda}_i = \frac{\gamma_i}{\Gamma_i} = \frac{\gamma_i \prod_{j \neq i} \Gamma_j}{\prod_{j \in [k]} \Gamma_j} = \frac{\xi_i}{\Xi};\ \sum_{i=1}^{k} \hat{\lambda}_i = 1 \Leftrightarrow \sum_{i=1}^{k} \xi_i = \Xi.$
- **Protocol Abiding Phase:** $p \in [k]$ behaves as described by a *finite state automaton* M_p determined by a cycle of actions, of length Ξ. The expected payoff of p during the whole cycle is $\mathbf{z}^*[p] > \varphi_p(G)$.
- **Punishment Phase:** Upon discovery of a defection from the protocol abiding behavior, each agent $p \neq q$ gives up control, for Λ_q consecutive rounds, to a *punishment correlation device* that implements the aggressive strategy $\mathbf{a}_q$ against the defector q of minimum ID. ■

Enforcing a Payoff for the Mutual Advantage Case (III)

- Suppose that we have 3 players, and $\mathbf{z}^* = \hat{\lambda}_1 \mathbf{U}(\mathbf{x}_1) + \hat{\lambda}_2 \mathbf{U}(\mathbf{x}_2) + \hat{\lambda}_3 \mathbf{U}(\mathbf{x}_3)$.
- The profile that induces $\mathbf{z}^*$ as the equilibrium of G^∞ is:

- Suppose that we have 3 players, and $\mathbf{z}^* = \hat{\lambda}_1 \mathbf{U}(\mathbf{x}_1) + \hat{\lambda}_2 \mathbf{U}(\mathbf{x}_2) + \hat{\lambda}_3 \mathbf{U}(\mathbf{x}_3)$.
- The profile that induces $\mathbf{z}^*$ as the equilibrium of G^∞ is:



Enforcing a Payoff for the Mutual Advantage Case (IV)

Remark

Any *rational* payoff point that is an element of $sirr(G)$ can be induced as an equilibrium of G^{∞}, by a similar construction. The profile will have polynomial description in the size of representation of this payoff point, but not necessarily in $size(G)$.

How About The No--Mutual Advantage Case?

- (WLOG) Assume that $\varphi(G) = \mathbf{0}$.
- $\mu(G)$: The maximum number of players having *concurrently positive payoffs*.

How About The No--Mutual Advantage Case?

- (WLOG) Assume that $\varphi(G) = \mathbf{0}$.
- $\mu(G)$: The maximum number of players having *concurrently positive payoffs*.

LEMMA 2: Max #Players with Concurrently Positive Payoffs

For any *constant* $k \geq 2$ and any game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with rational payoffs, $\mu(G)$ is computable in time $poly(size(G))$.

How About The No--Mutual Advantage Case?

- (WLOG) Assume that $\varphi(G) = \mathbf{0}$.
- $\mu(G)$: The maximum number of players having *concurrently positive payoffs*.

LEMMA 2: Max #Players with Concurrently Positive Payoffs

For any *constant* $k \geq 2$ and any game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with rational payoffs, $\mu(G)$ is computable in time $poly(size(G))$.

WHY?

- Exploit the **constant** number of players.
- Starting from $k-$subsets, down to $1-$subsets of points from Z, keep solving LPs similar to the MAC LP of the Mutual-Advantage case, until the first solvable instance with positive value.

Partial Answer for No--Mutual Advantage Case (I)

THEOREM: Construction of NE for G^∞ when $\mu(G) \le 2$

For any *constant* $k \ge 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with $sirr(G) = \emptyset$, there is an efficiently computable equilibrium point for G^∞, when at most two players may have concurrently positive payoffs, ie, $\mu(G) \le 2$.

Partial Answer for No--Mutual Advantage Case (I)

THEOREM: Construction of NE for G^{∞} when $\mu(G) \leq 2$

For any *constant* $k \geq 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with $sirr(G) = \emptyset$, there is an efficiently computable equilibrium point for G^{∞}, when at most two players may have concurrently positive payoffs, ie, $\mu(G) \leq 2$.

WHY?

if $\mu(G) = 0$

then the profile $(\mathbf{d}_p)_{p\in[k]}$ of defensive strategies is NE point of G.

Partial Answer for No--Mutual Advantage Case (I)

THEOREM: Construction of NE for G^∞ when $\mu(G) \leq 2$

For any *constant* $k \geq 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with $sirr(G) = \emptyset$, there is an efficiently computable equilibrium point for G^∞, when at most two players may have concurrently positive payoffs, ie, $\mu(G) \leq 2$.

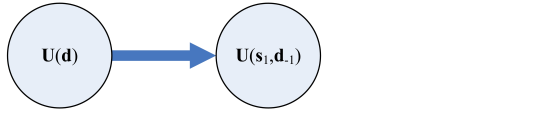
WHY?

if $\mu(G) = 0$

then the profile $(\mathbf{d}_p)_{p\in[k]}$ of defensive strategies is NE point of G.

else if $\mu(G) = 1$

then any *pure best response* defection from the defensive profile leads to a NE of G:



Partial Answer for No--Mutual Advantage Case (I)

THEOREM: Construction of NE for G^∞ when $\mu(G) \leq 2$

For any *constant* $k \geq 2$ and one-shot game $G = \langle [k], (S_p)_{p\in[k]}, (U_p)_{p\in[k]} \rangle$ with $sirr(G) = \emptyset$, there is an efficiently computable equilibrium point for G^∞, when at most two players may have concurrently positive payoffs, ie, $\mu(G) \leq 2$.

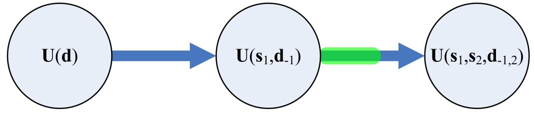
WHY?

if $\mu(G) = 0$

then the profile $(\mathbf{d}_p)_{p\in[k]}$ of defensive strategies is NE point of G.

else if $\mu(G) = 1$

then any *pure best response* defection from the defensive profile leads to a NE of G:



Partial Answer for No--Mutual Advantage Case (II)

WHY? (contd.)

The case $2 = \mu(G) < k$.

- Locate a payoff point in $conv(Z)$ such that exactly two players (eg, players $1, 2$) deviate from their defensive strategies (to pure strategies) and get *positive payoffs*.

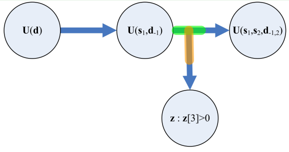
Partial Answer for No--Mutual Advantage Case (II)

WHY? (contd.)

The case $2 = \mu(G) < k$.

- Locate a payoff point in $conv(Z)$ such that exactly two players (eg, players $1, 2$) deviate from their defensive strategies (to pure strategies) and get *positive payoffs*.

- The *defensive strategies profile* $\mathbf{d}_{-1,2}$ for the other $k-2$ players is **weakly dominant**:

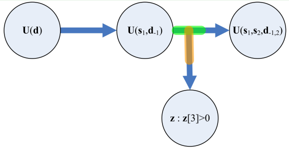


Partial Answer for No–Mutual Advantage Case (II)

WHY? (contd.)

The case $2 = \mu(G) < k$.

- Locate a payoff point in $conv(Z)$ such that exactly two players (eg, players $1, 2$) deviate from their defensive strategies (to pure strategies) and get *positive payoffs*.
- The *defensive strategies profile* $\mathbf{d}_{-1,2}$ for the other $k - 2$ players is **weakly dominant**:



- Lock the $k - 2$ players' strategies to the weakly dominant profile $\mathbf{d}_{-1,2}$ and inductively solve (using correlated threats) the *infinitely repeated subgame* between players 1 and 2.

Skeleton

Recap

- Learning is helpful for stage games. In particular, for:
  - Computing NE points in special classes of stage games (eg, socially concave games, constant-sum bimatrix games).
  - Computing CE points of arbitrary stage games.
  - Eliminating dominated strategies in arbitrary stage games.
- Enforcement is helpful for repeated games. In particular, we proposed a new, credible, notion of **Correlated Threat Point**, that is capable of implementing the essence of the Folk Theorem, for the case of more than 2 players.

Open Issues

- What else can be learnt for stage games?
- How can we exploit learning in repeated games (eg, computing more efficient NE points than the ones of the stage game)?
- How should we deal with the general No-Mutual-Advantage case?
- How can we handle non-constant number of players?
- How can we implement the correlation devices in a decentralized way (eg, as in (Barany (1992)))?
- What can be done for asynchronous plays of agents?

Some Related Bibliography

Related to Learning:

- A. Blum, Y. Mansour: *Learning, Regret Minimization and Equilibria*. Chapter 4 in N. Nisan, E. Tardos, T. Roughgarden (editors) Algorithmic Game Theory, Cambridge Univ. Press, 2007.
- N. Cesa-Bianchi, G. Lugosi: *Prediction, Learning and Games*. Cambridge Univ. Press, 2006.

Related to Enforcement:

- I. Bárány: *Fair distribution protocols or how the players replace fortune*. Mathematics of Operations Research, 17(2):327--340, 1992.
- C. Borgs, J. Chayes, N. Immorlica, A. Tauman Kalai, V. Mirrokni, C. Papadimitriou: *The myth of the folk theorem*. STOC '08, pp. 365--372, 2008.
- D. Fudenberg, E. Maskin: *Folk theorems for repeated games with discounting and incomplete information*. Econometrica, 54:533--554, 1986.
- S. Kontogiannis, P. Spirakis: *Equilibrium points in fear of correlated threats*. WINE '08, pp. 210--221, 2008.
- M. Littman and P. Stone: *A polynomial-time nash equilibrium algorithm for repeated games*. Decision Support Systems, 39(1):55--66, 2005.

Thank you
for your attention!

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρών, Σπύρος Κοντογιάννης «Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων: Learning, Enforcement and Equilibria». Έκδοση: 1.0. Πάτρα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:

https://eclass.upatras.gr/courses/MATH959/

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει) μαζί με τους συνοδευόμενους υπερσυνδέσμους.