Reinforcement learning in a prisoner's dilemma

Arthur Dolgopolov
(slides by J.Luis Martínez)

Introduction

- The author propose a characterization [1] of the set of Stochastically Stable States (SS).
- His results is a generalization of the solution proposed by Newton and Sawa (2015).
- He use his solution to explore collusion under no memory algorithms in a Prisoner's Dilemma context.

[1] a property or a condition to define a certain notion

Outline

Playground Specifications:

- Description of the Space-Game
- Assumptions on the Dynamics

Characterization of equilibrium:

- Notion of Cost
- Required Tools
- Results

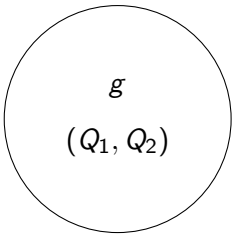
One Application: Prisoner's Dilemma

Description of the Space-Game

- Consider a **2-players symmetric** matching game.
- **Same** Set of actions: $A = \{a_1, a_2, ..., a_n\}$
- **Same** Set of Pay-offs: $\Pi = \{(\pi_{a_i,a_j}) \mid a_i, a_j \in A\}$
- **Reinforcement Learning**:
  - **Decisions based** only in **Present Q-value** *("No-Memory")*
  - **Q-values updates** after a decision
- The whole **game's flow** can be structured in **states**.

Description of the Space-Game

This, is a state:

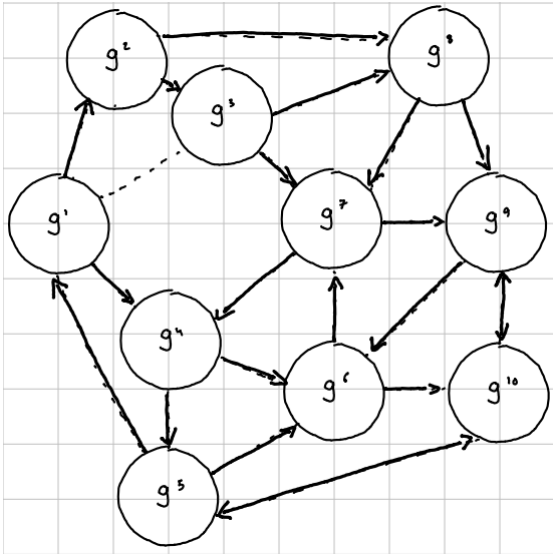


$$\Longrightarrow \quad g = \left[\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} Q_1^{a_1}(g) & Q_1^{a_2}(g) & \cdots & Q_1^{a_n}(g) \\ Q_2^{a_1}(g) & Q_2^{a_2}(g) & \cdots & Q_2^{a_n}(g) \end{pmatrix} \right]$$

where:

$Q_i^a(g) \equiv$ **Q-value** at state **g**, given for player **i** to action **a**.

Description of the Space-Game



Assumptions on Dynamics

- Over **Learning rule** ($\equiv \mathcal{F}_i^{a_i, a_{-i}}$) :

  How the Q-values updates?

  1. **Only** the Q-value related to the **action taken** is updated.
  2. Q-values **Must** be updated towards the 'true' payoff in full or in part. (**'get closer' to** $\pi_{\mathbf{a_i}, \mathbf{a_j}}$)

Assumptions on Dynamics

- Over **Perturbed Dynamics** ($\equiv \{P_\eta\}_{\eta \in (0,\hat{\eta})}$):

  How players experiment?
  - **Moving out from exploit** strategy ($\arg\max_a Q_i^a(g)$), which means;
  - **Introducing Noise** in decisions.
    Experimentation parameter ($\equiv \eta$)

    Caution: This experimentation came with a **Cost**! ($\equiv c(g, g')$)

Assumptions on Dynamics

- Over **Perturbed Dynamics** ($\equiv \{P_\eta\}_{\eta \in (0,\hat{\eta})}$):

1. Regularity Conditions: (So Process and Cost limiting distribution are well-define):
   - (i) $P_\eta \xrightarrow{\eta \to 0} P_0$ (in the limit you behave with no error)
   - (ii) Irreducibility on P_η (everywhere is reachable from everywhere)
   - (iii) Continuity of P_η over η
   - (iv) Existence of a cost $c(g, g') = c$ for any experimentation.

2. Additional Conditions:
   - (v) If the transition to a state is possible under some experimentation η, then for a combination of action a_i, a_j the Learning Rule can drive you to that state. (There is a possible Update that brings you there)
   - (vi) The Cost of both players experimenting at same time is the same as each player experimenting alone.
   - (vii) The relative Cost of experimentation between two different states is related to how painful (how stupid you are) it is to deviate from exploit action in each of this states,(the bigger the loss the bigger the cost).

Outline

Playground Specifications:

- Description of the Space-Game
- Assumptions on the Dynamics

Characterization of equilibrium:

- Notion of Cost
- Required Tools
- Results

One Application: Prisoner's Dilemma

Notion of Cost

Cost tell us **how probable are transition** from one state to another. It has the following form:

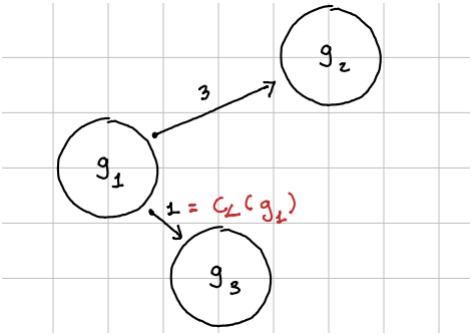
$$c(g, g') := \lim_{\eta \to 0} -\eta \quad \cdot \quad \underbrace{\log P_\eta(g, g')}_{\text{lower prob.} \to \text{higher costs}}$$

Moreover we describe the least cost transitions ('easiest deviation') as:

$$c_L(g) := \min_{g' \neq g} c(g, g')$$

This concept will be **key** to derive the **characterization of the equilibrium**.

Notion of Cost



Required Tools

1. Modified Cost
2. Minimum Cost Spanning Trees
3. Stochastically Stable State (SS)
4. Central State

Required Tools: Modified Cost

The following costs represents the total cost of moving from state g_1 to state g_r, once we take into account (subtract) the 'easiest deviations' of the intermediate states:

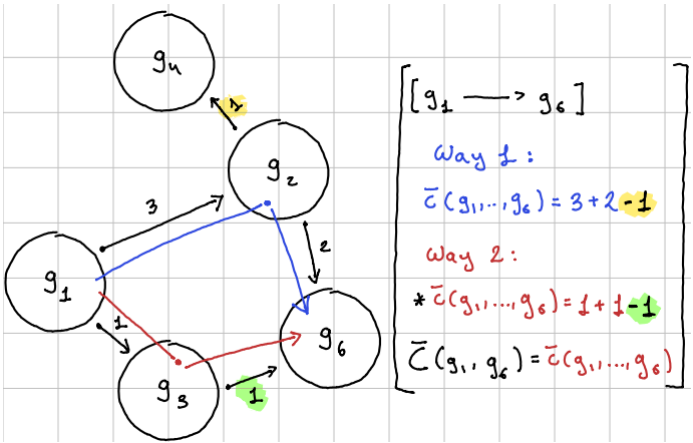
$$\bar{c}(g_1, g_2, ..., g_r) = c(g_1, g_2, ..., g_r) - \sum_{l=2}^{r-1} c_L(g_l).$$

The minimum adjusted cost among all possibles paths from state g_1 to state g_r is our Modified Cost:

$$\bar{C}(g_1, g_r) = \min_{g_1,...,g_r \in S(g_1,g_r)} \left(\bar{c}(g_1, g_2, ..., g_r)\right).$$

(Kind of 'the smartest' way to move from g_1 *to* g_r)

Required Tools: Modified Cost



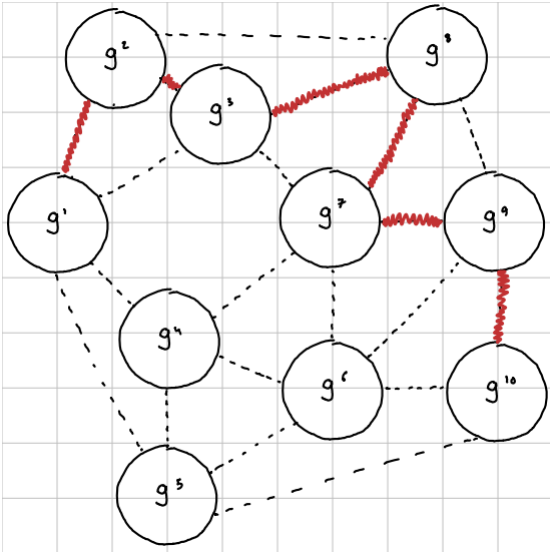
Required Tools: Minimum Cost Spanning Tree

This concept is crucial.

Our **equilibria** $\rightarrow$ will be the **Root Minimum Cost Spanning Tree**
What it is a ___ ?:

(1) Tree
(2) Spanning Tree
(3) Spanning Tree Rooted at state $\hat{g}$
(4) Minimum Cost Spanning Tree Rooted at state $\hat{g}$

Required Tools: Tree



Required Tools: Spanning Tree

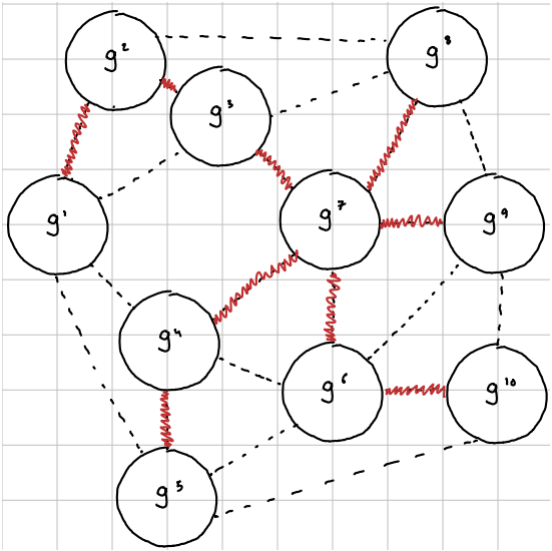
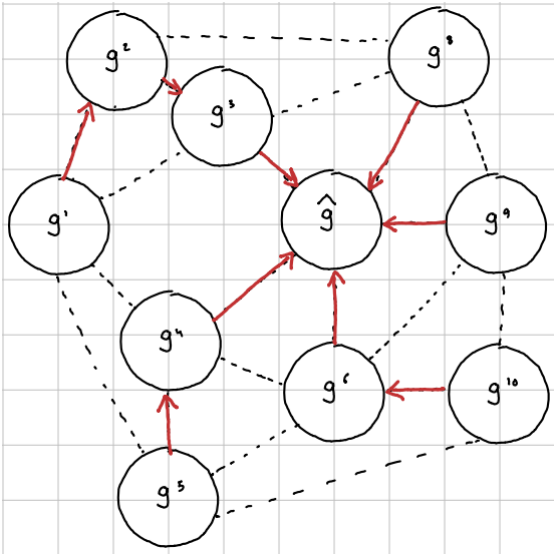
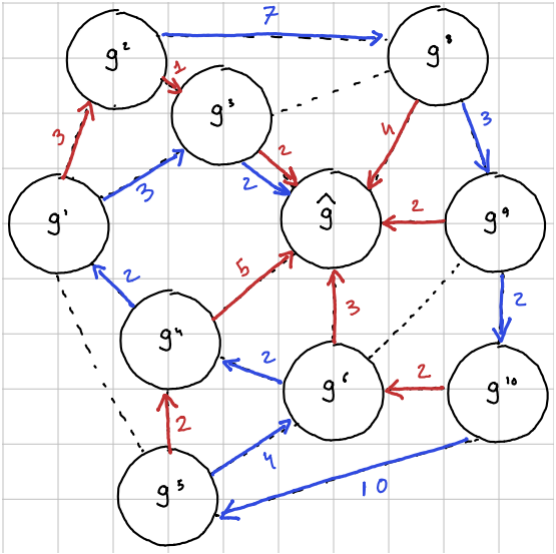


Figure: Graph with four node disjoint spanning trees

Required Tools: Spanning Tree Rooted at $\hat{g}$



Required Tools: Minimum Coat Spanning Tree Rooted at $\hat{g}$



Required Tools: Stochastically Stable State (SS)

Stochastically Stable (SS) states are **states** that in the **presence of** small random perturbations (**noise** (η)), are more likely to be visited and **persist in the long term**.

- **Young (1993):** A state $\hat{g}$ is stochastically stable only if there exists a minimum-cost spanning tree rooted at $\hat{g}$.

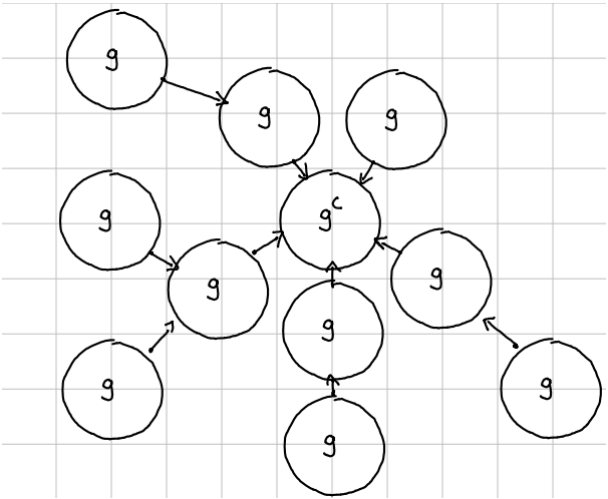
Required Tools: Central State

The characterization will be describe **relative** to the presence if this **central state**.

A given state g^c is a central state [2] if from all the others states exists a path formed by sequent of 'easiest deviations'.

[2] Formal Definition is in **Definition 1.** of the paper.

Required Tools: Central State



Results

If a central state g^c is know to exist, then any minimum cost spanning tree has to be formed by minimum cost edges ('easiest deviations') except (maybe) for the path between g^c and the tree root $\hat{g}$. [3]

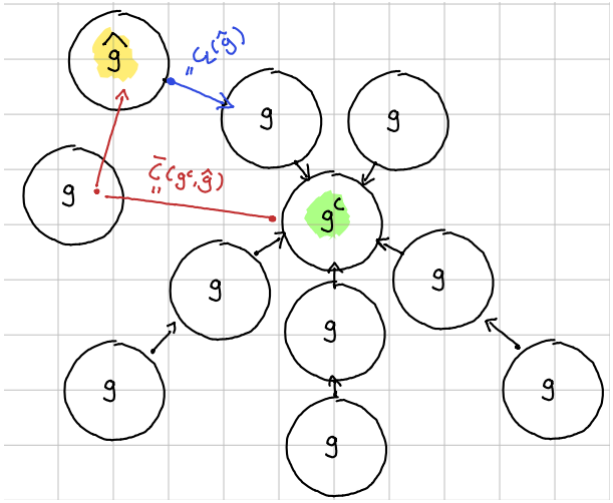
(This is powerfull becouse the existance of g^c allow us say something about the shape of minimum cost spanning trees, in which root resides our SS equilibrium).

Then the whole **analysis reduce** to study whether it is **easier to move from** g^c **to** $\hat{g}$ **or from** $\hat{g}$ **to** g^c. The winner will 'absorb' the looser and will become the stochastic stable equilibrium. The answer to that question is capture here:

$$\bar{C}(g^c, \hat{g}) - c_L(\hat{g})$$

[3] Intuition derived from the Formal Definition in **Lemma 3.** of the paper.

Results



Results

Then we end up with two possible cases [4]:

(i) It is more costly to move from $\hat{g}$ to g^c.
(Also $\hat{g}$ has to minimize this difference: **'best competitor'**)
($\hat{g}$ 'absorbs' g^c) $\rightarrow$ A minimum-cost tree is rooted in state $\hat{g}$

$$\bar{C}(g^c, \hat{g}) \leq c_L(\hat{g})$$

(ii) It is more costly to move from g^c to $\hat{g}$.
(g^c 'absorbs' $\hat{g}$) $\rightarrow$ A minimum-cost tree is rooted in state g^c

$$\bar{C}(g^c, \hat{g}) \geq c_L(\hat{g})$$

[4] Intuition derived from the Final Characterization in **Proposition 2.** of the paper.

Outline

Playground Specifications:

- Description of the Space-Game
- Assumptions on the Dynamics

Characterization of equilibrium:

- Notion of Cost
- Required Tools
- Results

One Application: Prisoner's Dilemma

Solving Prisoner's Dilemma using Proposition 2

Main point of the characterization:

Whole problem focus on locking for central states (g^c), and their 'best competitor' ($\hat{g}$).

The author makes a discussion and determines that:

- Central State $(g^c) : g^* \Rightarrow [Q_i^N(g^*) = \pi_{NN}, Q_i^C(g^*) = \pi_{CN}]$
- Best Competitor $(\hat{g}) : g^{**} \Rightarrow [Q_i^N(g^{**}) = \pi_{NN}, Q_i^C(g^{**}) = \pi_{CC}]$

Solution:[5]

Cooperation	**Defection**	**Both Possible**
• $\bar{C}(g^*, g^{**}) < c_L(g^{**})$ • Then $g^* \notin SS$ • Converge to cooperation in any SS.	• $\bar{C}(g^*, g^{**}) > c_L(g^{**})$ • Then $SS = \{g^*\}$ • Converge to defection.	• $\bar{C}(g^*, g^{**}) = c_L(g^{**})$ • SS may include states with defection, cooperation, or both.

[5] Taken from **Corollary 3.** in the paper

Recapitulation and Conclusions

- We have saw a **characterization for the SS** set, in **matching games**[6] for **value-based** learning algorithms with **no memory**.
- Results are based on the **existence of at least one** of this **central states** (g^c).

[6] symmetric games with underlying match intention, e.g Prisoner's Dilemma

Recapitulation and Conclusions

- Good Things (**Generalization**):

  (The results) "includes other **dynamics beyond reinforcement learning** and other **games beyond the prisoner's dilemma**: central states appear ubiquitous in matching scenarios"

- Drawbacks (**Not so Generalizable (?)**):

  1. **Learning Rules Problems:** "While my approach applies directly after appropriate transformation of the state space, the **central state** may **not always exist** for more **complex learning rules**".

  2. **Central State Problems:** " Unfortunately, not all results extend in a straightforward manner to games without a dominant strategy equilibrium as the minimum-cost path to a **central state may no longer exist**"

The End