Learning Correlation in Multi-Player General-Sum Games with Regret Minimization

Tommaso Bianchi

Advisor: Professor Nicola Gatti

CSE Track

September 30, 2019

POLITECNICO MILANO 1863

HONOURS PROGRAMME HP-SR in Information Technology

Goal

Develop novel algorithms to efficiently compute game theoretical equilibria that enable **correlation** among players.

General approach for all **multi-player**, **general-sum** games.

Online and decentralized computation via **regret minimization**.

Game representations - Normal-form game

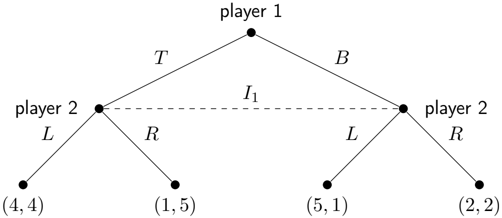
player 2

player 1		L	R
	T	$4, 4$	$1, 5$
	B	$5, 1$	$2, 2$

Model **simultaneous**, one-shot interactions.

Each player's goal is to play as to maximize its own utility.

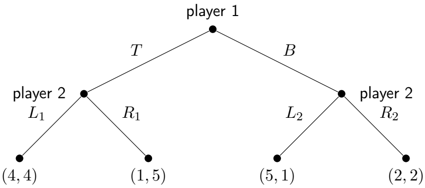
Game representations - Extensive-form game



Model **sequential** interactions among players.

Can explicitly model **imperfect information** through information sets, which are sets of indistinguishable nodes of a player.

Game representations - Equivalence



player 2

player 1	L_1L_2	L_1R_1	R_1L_2	R_1R_2
T	4, 4	4, 4	1, 5	1, 5
B	5, 1	2, 2	5, 1	2, 2

Equivalence by enumerating all the possible **action plans**, which specify an action for each information set.

The set of action plans has a cardinality which is **exponential** in the size of the extensive-form game.

Strategy representations - Normal-form strategies

player 2

L_1L_2	L_1R_1	R_1L_2	R_1R_2
0.1	0.4	0	0.5

1

A **normal-form strategy** x_i for player i is a probability distribution over the actions in A_i.

Strategy representations - Behavioural strategies

Information set I_1

L_1	R_1
0.5	0.5

$\underbrace{\qquad\qquad}_{1}$

Information set I_2

L_2	R_2
0.7	0.3

$\underbrace{\qquad\qquad}_{1}$

A **behavioural strategy** π_i for player i is a function specifying a probability distribution for each information set $I \in \mathcal{I}_i$.

In extensive-form games, behavioural strategies allow for a much more compact representation than the normal-form strategies of the equivalent normal-form game.

Strategy representations - Joint strategies

player 1	T				B			
player 2	L_1L_2	L_1R_1	R_1L_2	R_1R_2	L_1L_2	L_1R_1	R_1L_2	R_1R_2
	0.1	0.1	0	0.2	0.4	0	0	0.2

(sum = 1)

A **normal-form joint strategy** x is a probability distribution over the set $A = \bigtimes_{i \in \mathcal{P}} A_i$ of **action profiles** of the players.

Joint strategies specify how players **correlate** their play.

It is always possible to construct a joint strategy from a set of marginal normal-form strategies (one for each player); the opposite is not always true.

Solution concepts - Nash equilibrium

A **Nash equilibrium** (Nash, 1951) is a strategy profile $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)$ such that no player has any incentive to deviate (*i.e.*, to change its strategy), given that all the other players do not deviate themselves.

Nash equilibria models the way in which perfectly rational, selfish agents will act given they are completely isolated from each other.

Introducing correlation in solution concepts

Correlation is introduced through a **mediator**, a central device with the role of sending recommendations to the players on how to play.

The mediator takes a sample from a publicly known joint strategy, and privately communicates to each player how they should play.

Players are free to play according to the recommendation or to deviate and play differently.

Solution concepts - Coarse-correlated equilibrium

In a **Coarse-correlated equilibrium** (Moulin and Vial, 1978), players have no incentive to deviate given the knowledge *a-priori* of the probability distribution from which recommendations will be sampled, given that also the other players commit to following the correlation plan.

Coarse-correlated equilibria are well-suited for scenarios where the players have limited communication capabilities and can only communicate before the game starts.

Regret minimization

Regret is a measure of how much a player would have preferred to play a different strategy with respect to the one he actually used.

$$R_i^T := \max_{a_i \in A_i} \sum_{t=1}^{T} u_i(a_i, x_{-i}^t) - \sum_{t=1}^{T} u_i(x^t)$$

A **regret minimizer** is a device providing player i's strategy x_i^{t+1} for the next iteration $t + 1$ on the basis of the past history of play.

Regret matching (Hart and Mas-Colell, 2001)

$$x_i^{T+1}(a_i) = \begin{cases} \frac{[R_i^T(a_i)]_+}{\sum_{a_i' \in A_i} [R_i^{T,+}(a_i')]_+} & \text{if } \sum_{a_i' \in A_i} [R_i^T(a_i')]_+ > 0 \\ \frac{1}{|A_i|} & \text{otherwise} \end{cases}$$

Regret matching is a regret minimizer for normal-form games based on the simple idea that the probability to play an action is proportional to how ‘good’ it would have been to play it in the past (*i.e.*, on the regret of not having played it).

CFR - Counterfactual regret minimization (Zinkevich et al., 2008)

Counterfactual regret minimization (CFR) is a regret minimizer for extensive-form games.

Regret is **decomposed** into local terms at each information set, so as to guarantee that minimizing the local regrets implies the minimization the overall regret.

CFR uses simpler regret minimizers at each information set, such as regret matching.

Empirical frequency of play (Hart and Mas-Colell, 2000)

Definition

The *empirical frequency of play* $\bar{x}$ is the joint probability distribution defined as $\bar{x}(\sigma) := \frac{|t \leq T : \sigma^t = \sigma|}{T}$ for each normal-form action plan σ.

Proposition

If $\limsup_{T \to \infty} \frac{1}{T} R_i^T \leq 0$ *almost surely for each player* i*, then the empirical frequency of play* $\bar{x}$ *approaches almost surely as* $T \to \infty$ *the set of CCE.*

Framework - General idea

Use a **regret minimizer** for each player to ensure that their play approaches over time the set of CCE.

Combine it with a **polynomial-time oracle** that maps players' strategies in the space of normal-form strategies so as to explicitly keep track of the empirical frequency of play.

CCE computation with a sampling oracle

Use a **sampling oracle** to generate at each iteration a normal-form **action plan** from the more compact strategies of the players.

Sampled action plan can be stored to explicitly keep track of the empirical frequency of play.

Polynomial-time sampling is often trivial, but can be dispersive if the strategies to sample from have some symmetries.

CCE computation with a marginal reconstruction oracle

Use a **reconstruction oracle** to generate normal-form **strategies** that are equivalent to the compact strategies of the players.

Reconstructed strategies are multiplied together to get a joint strategy.

We proved that the time average of the reconstructed joint strategies behaves like the empirical frequency of play.

CFR with Sampling (CFR-S)

Use **CFR** as a regret minimizer, which employs behavioural strategies as compact strategy representation.

Sampling a normal-form action plan from a behavioural strategy simply requires sampling one action at each information set.

Fast iterations, but a lot of them might be required before reaching a good approximation of the empirical frequency of play.

Marginal reconstruction oracle

Algorithm 1 Reconstruct x_i from π_i

function Nf-reconstruct(π_i)
  $\mathbf{X} \leftarrow \varnothing$ ▷ $\mathbf{X}$ is a dictionary defining x_i
  $\omega_z \leftarrow \rho_z^{\pi_i} \quad \forall z \in Z$
  while $\omega > 0$ **do**
    $\bar{\sigma}_i \leftarrow \arg\max_{\sigma_i \in \Sigma_i} \min_{z \in Z(\sigma_i)} \omega_z$
    $\bar{\omega} \leftarrow \min_{z \in Z(\bar{\sigma}_i)} \omega_i(z)$
    $\mathbf{X} \leftarrow \mathbf{X} \cup (\bar{\sigma}_i, \bar{\omega})$
    $\omega \leftarrow \omega - \bar{\omega}\, \rho^{\bar{\sigma}_i}$
  return x_i built from the pairs in $\mathbf{X}$

Main idea: assign probability to normal-form action plans σ_i so as to match the probability ω_z of reaching terminal node z induced by behavioural strategy π_i.

CFR with Joint reconstruction (CFR-Jr)

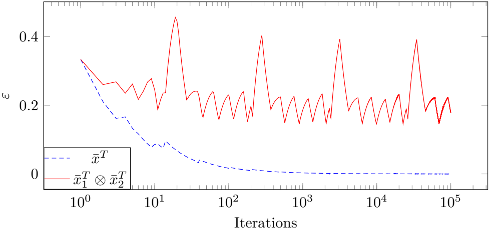
Use **CFR** as a regret minimizer, which employs behavioural strategies as compact strategy representation.

Use the **reconstruction oracle** to build normal-form realization equivalent strategies from the behavioural strategies built by CFR.

Iterations are slower due to the more complex oracle, but usually even a few reconstruction steps are sufficient to build a good approximation of the empirical frequency of play.

Non-convergence of product of marginal strategies

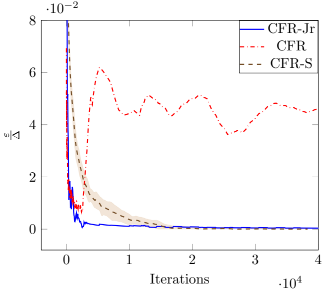
1,0	0,1	0,0
0,0	2,0	0,1
0,1	0,0	1,0

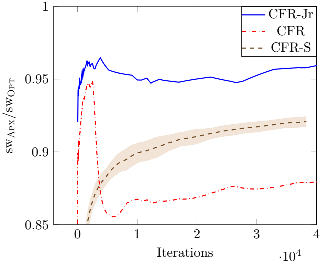


The naïve solution of keeping track of each players' marginal strategy and building the product of the average strategies might lead to **cyclic** behaviours.

For example, by employing regret matching (right figure) in a variant of the **Shapley game** (Shapley, 1964; left figure).

Non-convergence of product of marginal strategies





Cyclic behaviours for the product of marginal strategies in an instance of the **Goofspiel** (Ross, 1971) card game.

CFR-Jr clearly outperforms CFR-S in terms of convergence speed (left figure) and in terms of attained social welfare (right figure).

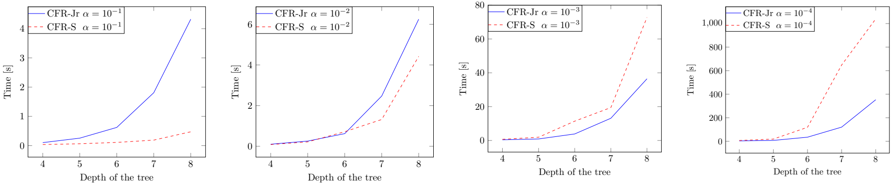
Comparison with state of the art algorithm

Game	Tree size #infosets	CFR-S $\alpha = 0.05$	CFR-S $\alpha = 0.005$	CFR-S $\alpha = 0.0005$	CFR-S $\mathrm{sw}_{\mathrm{APX}}/\mathrm{sw}_{\mathrm{OPT}}$	CFR-Jr $\alpha = 0.05$	CFR-Jr $\alpha = 0.005$	CFR-Jr $\alpha = 0.0005$	CFR-Jr $\mathrm{sw}_{\mathrm{APX}}/\mathrm{sw}_{\mathrm{OPT}}$	CG
K3-6	72	1.41s	9h15m	> 24h	-	1.03s	13.41s	11m21s	-	3h47m
K3-7	84	4.22s	17h11m	> 24h	-	2.35s	14.33s	51m27s	-	14h37m
K3-10	120	22.69s	> 24h	> 24h	-	7.21s	72.78s	4h11m	-	> 24h
L3-4	1200	10m33s	> 24h	> 24h	-	1m15s	6h10s	> 24h	-	> 24h
L3-6	2664	2h5m	> 24h	> 24h	-	2m40s	11h19m	> 24h	-	> 24h
L3-8	4704	13h55m	> 24h	> 24h	-	20m22s	> 24h	> 24h	-	> 24h
G3-4-A*	98508	1h33m	> 24h	> 24h	0.996	1h3m	4h13m	> 24h	0.999	> 24h
G3-4-DA*	98508	1h13m	> 24h	> 24h	0.987	12m18s	1h50m	> 24h	1.000	> 24h
G3-4-DH*	98508	47m33s	19h40m	> 24h	0.886	16m38s	4h8m	15h27m	1.000	> 24h
G3-4-AL*	98508	32m34s	15h32m	17h30m	0.692	1h21m	5h2s	> 24h	0.730	> 24h

Comparison with the prior state of the art technique, a column generation algorithm (Celli et al., 2019).

Both CFR-Jr and CFR-S vastly outperform it, and can be effectively used in much larger game instances.

Comparison between CFR-S and CFR-Jr



Comparison between the running time of CFR-S and CFR-Jr on random game instances.

Faster iterations lead CFR-S to reach a rough approximation of a solution in a shorter time, but as we require a higher accuracy CFR-Jr performs better.

Conclusions

There exist **general regret minimization approaches** that guarantee convergence to the set of CCE in general-sum, multi-player games.

The best algorithm derived through this method is able to vastly **outperform** the prior state of the art in reasonably-sized extensive-form games.

No optimality guarantee, but **high social-welfare** in practice.

Future works

Compute approximate Coarse-correlated equilibria in other classes of structured games by employing our regret minimization framework.

Employ a CCE strategy profile as a starting point to approximate tighter solution concepts that admit some form of correlation.

Give theoretical guarantees on the approximation of the optimal social welfare.

Define regret-minimizing procedures for general, multi-player extensive-form games leading to refinements of CCE, such as Correlated equilibria and Extensive-form Correlated equilibria.

Bibliography

Andrea Celli, Stefano Coniglio, and Nicola Gatti. Computing optimal ex ante correlated equilibria in two-player sequential games. *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 2019.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 2000.

Sergiu Hart and Andreu Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 2001.

H. Moulin and J-P Vial. Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 1978.

John Nash. Non-cooperative games. *Annals of mathematics*, 1951.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Proceedings of the Annual Conference on Neural Information Processing*, 2008.

Lloyd Shapley. Some topics in two-person games. *Advances in game theory*, 1964.

Sheldon M Ross. Goofspiel – the game of pure strategy. *Journal of Applied Probability*, 1971.