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

Tommaso Bianchi
Advisor: Professor Nicola Gatti
CSE Track

September 30, 2019
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

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>4,4</td>
<td>1,5</td>
</tr>
<tr>
<td>B</td>
<td>5,1</td>
<td>2,2</td>
</tr>
</tbody>
</table>

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

Equivalence by enumerating all the possible \textbf{action plans}, which specify an action for each information set.

The set of action plans has a cardinality which is \textit{exponential} in the size of the extensive-form game.
Strategy representations - Normal-form strategies

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

<table>
<thead>
<tr>
<th></th>
<th>$L_1 L_2$</th>
<th>$L_1 R_1$</th>
<th>$R_1 L_2$</th>
<th>$R_1 R_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>player 2</td>
<td>0.1</td>
<td>0.4</td>
<td>0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The sum of each row is 1.
A behavioural strategy $\pi_i$ 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

A normal-form joint strategy $\pi$ is a probability distribution over the set $A = \prod_{i \in P} A_i$ 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 \( \hat{x} = (\hat{x}_1, \ldots, \hat{x}_n) \) 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.

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

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 $\sigma$.

**Proposition**
If $\limsup_{T \to \infty} \frac{1}{T} R^T_i \leq 0$ 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.
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 $\pi_i$

1: function Nf-reconstruct($\pi_i$)
2:     $X \leftarrow \emptyset$ \hspace{1cm} $\triangleright$ $X$ is a dictionary defining $x_i$
3:     $\omega_z \leftarrow \rho^z_{\pi_i}$ \hspace{0.5cm} $\forall z \in Z$
4:     while $\omega > 0$ do
5:         $\bar{\sigma}_i \leftarrow \arg \max_{\sigma_i \in \Sigma_i} \min_{z \in Z(\sigma_i)} \omega_z$
6:         $\bar{\omega} \leftarrow \min_{z \in Z(\sigma_i)} \omega_i(z)$
7:         $X \leftarrow X \cup (\bar{\sigma}_i, \bar{\omega})$
8:     $\omega \leftarrow \omega - \bar{\omega} \rho^\bar{\sigma}_i$

return $x_i$ built from the pairs in $X$

Main idea: assign probability to normal-form action plans $\sigma_i$ so as to match the probability $\omega_z$ of reaching terminal node $z$ induced by behavioural strategy $\pi_i$. 
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

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

<table>
<thead>
<tr>
<th>Game</th>
<th>Tree size</th>
<th>CFR-S</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#infosets</td>
<td>α = 0.05</td>
<td>α = 0.005</td>
</tr>
<tr>
<td>K3-6</td>
<td>72</td>
<td>1.41s</td>
<td>9h15m</td>
</tr>
<tr>
<td>K3-7</td>
<td>84</td>
<td>4.22s</td>
<td>17h11m</td>
</tr>
<tr>
<td>K3-10</td>
<td>120</td>
<td>22.69s</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>L3-4</td>
<td>1200</td>
<td>10m33s</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>L3-6</td>
<td>2664</td>
<td>2h5m</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>L3-8</td>
<td>4704</td>
<td>13h55m</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>G3-4-A*</td>
<td>98508</td>
<td>1h33m</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>G3-4-DA*</td>
<td>98508</td>
<td>1h13m</td>
<td>&gt; 24h</td>
</tr>
<tr>
<td>G3-4-DH*</td>
<td>98508</td>
<td>47m33s</td>
<td>19h40m</td>
</tr>
<tr>
<td>G3-4-AL*</td>
<td>98508</td>
<td>32m34s</td>
<td>15h32m</td>
</tr>
</tbody>
</table>

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 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


