Equilibrium's Action Bound in Extensive Form Games with Many Actions

Martin Schmid
Charles University In Prague
lifrordi@gmail.com

Matej Moravcik
Charles University In Prague
moravcikmatej42@gmail.com

Abstract

Recently, there has been great progress in computing optimal strategies in large extensive games. But there are still many games where computing an equilibrium remains intractable, mostly due to the size of the game tree. In many of them, such as no-limit poker, huge game tree size is caused by too many actions available in information sets. We show that it's possible to limit number of used actions in Nash equilibirum. We bound this number by a private information and show how to compute this Nash equilibirum.

Introduction

We focus on extensive form games with imperfect information and publicly observable actions. These games model sitations where players recieve some private information, and players sequently take actions.

This is the case of many card games, where players are dealt some private cards. If players make bets, there may be many different bet sizes available, making the game tree large. Fore example in the no-limit poker played in the *Annual Computer Poker Competition* 2010 (ACPC), there were up to 20 000 bet sizes available in every information set, and the game consiquently contains 6.3×10^{164} game states (Johanson 2013).

In games with no hidden information, players don't need to mix their strategies to play optimaly - in every game node, they just select the one, best action (best response). This is not the case in games with imperfect information, where player may need to mix his actions. But how many actions does he need to mix?

Consider simple card game, where each players recieves only one private card from the deck (each player has his own deck). If there's one card in the deck (for example, only A♠), opponents know what the player will be dealt, so this is game with no hidden information and player may use only one action in every information set.

But what if there are two, three or four cards in the deck? Does he need to mix all his actions, only two/three/four actions or some other number? We show that this number of used actions is indeed equal to the size of private information.

We show that the number of actions that player needs to mix is indeed equal to the size of his private information.

Background

Extensive form games - for a complete definition see (Osborne and Rubinstein 1994, p. 200). Briefly, extensive form game G consist of

- *A finite set N (the set of **players**)*
- *A set H of sequences (possible **histories**)*
- *A function p that assigns to each non-terminal history an acting player (member of $N \cup c$, where c stands for chance)*
- *A function f_c that associates with every history for which $p(h) = c$ a probability measure on $A(h)$*
- *For each player $i \in N$ a partition $\mathcal{I}_i$ of $h \in H : p(h) = i$. $\mathcal{I}_i$ is the **information partition** of player i; a set $I_i \in \mathcal{I}_i$ is an **information set** of player i.*
- *For each player $i \in N$ an utility function $u_i : H \to \mathcal{R}$*

We suppose that the game is **finite** and satisfies **perfect recall** (Osborne and Rubinstein 1994, p. 203).

A **mixed strategy** of player i in extensive form game G is a probability measure over the set of player's pure strategies.

A **behavior strategy** of player i is a collection $(\beta_i(I_i))$ of independent probability measures, where $\beta_i(I_i)$ is the probability measure over over $A_i(I)$. For any history $h \in I_i \in H$ a any action $a \in A(h)$ we denote by $\beta_i(h, a)$ the probability assigned by $\beta_i(I)$to the action a.

If the game satisfies perfect recall, mixed and behavior strategies are equivalent (Osborne and Rubinstein 1994, p. 203).

We refer to the strategy of player i as σ_i. A strategy profile σ consists of a strategy for every player $\sigma = (\sigma_i)_{i \in N}$. For any history h, let $\pi(h)$ be the probability that history occurs if all players play according to the profile σ. We denote $\pi_{-i}(h)$ product of all players' contribution to this probability (including chance), except of player i. For all I, define $\pi^\sigma(I) = \sum_{h \in I} \pi^\sigma(h)$

Finally, for any profile σ, we define **outcome** $O_i(\sigma) = \sum_{h \in H} u_i(h) \pi^\sigma(h)$

Nash equilibrium

A Nash equilibrium in behavior strategies (Osborne and Rubinstein 1994, p. 203) is a profile σ^* with the property that for every player $i \in N$ we have

$$O_i(\sigma^*_{-i}, \sigma^*_i) \geq O_i(\sigma^*_{-i}, \sigma_i) \tag{1}$$

for every strategy σ_i of player i.

Sequential equilibrium

Inituitively, sequential equilibrium states that player plays optimally in every infromation set. Because player doesn't know in what history the game actually is, his selects his action according to what he *beliefs* that the history is.

An **assessment** in an extensive game is a pair (σ, μ), where σ is a profile of behavioral strategies and μ is a function that assigns to every information set a probability measure on the set of histories in the information set (**belief system**).

Let $O_i(\sigma, \mu|I)$ denote the **outcome** of (σ, μ) **conditional on** I **being reached**.

Sequential rationality The assessment (σ, μ) is sequentially rational, if for every player $i \in N$ and every information set $I \in \mathcal{I}_i$ we have

$$O_i(\sigma, \mu|I) \geq O_i((\sigma_{-i}, \sigma'_i), \mu|I) \tag{2}$$

for every strategy σ'_i of player i.

Consistency Basically, consistency means that belief system is derived from strategy profile using Bayes' rule (so it's consistent with the strategy). For information sets where $\pi(I) = 0$, one can't straigtforwadly use Bayes' - for details and formal definition see (Osborne and Rubinstein 1994, p. 224)

An assessment is a **sequential equilibrium** of a finite extensive game with perfect recall if it is *sequentially rational* and *consistent*. We know that every finite extensive game with perfect recall has a sequential equilibrium (Osborne and Rubinstein 1994, p. 225)

Public State Tree

*We call a partition of the histories, $\mathcal{P}$, a **public partition** and $P \in \mathcal{P}$ a **public state** if (Johanson et al. 2011)*

- *No two histories in the same information set are in different public states*
- *Two histories in different public states have no descendants in the same public state*
- *No public state contains both terminal and non-terminal historie*

For this public state tree, we also define:

- *A set of actions available in every $P \in \mathcal{P}$*

$$A(P) := \cup_{I \in P} A(I)$$

- *Acting player in P*

$$p(P) := p(I) \text{ for any } I \in P$$

- *$\lambda(P)$ - information sets of player $p(P)$ grouped in this public node*
- *$\nu(P, I)$ - Probability of being in information set $I \in P$, conditional on $P \in \mathcal{P}$ being reached (consistent with σ)*
- *Probability measure over $A(PI)$*

$$\gamma(P, a) = \sum_{I \in P} \nu(P, I)\beta(I, a)$$

- *$P_a \in \mathcal{P}$, public state that follows $P \in \mathcal{P}$ if action $a \in A(P)$ is taken.*

$$(h, a) \in P_a \in \mathcal{P} \iff (h) \in P \in \mathcal{P}$$

- ***Counterfactual information set** CI^a corresponding to $I \in \lambda(P)$. We refer to these information sets as counterfactual, because they would be information sets if the player $p(P)$ was to play again (and not an opponent).*

$$(h) \in I \iff (h, a) \in CI^a$$

For counterfactual information sets, we also analogicaly define $\lambda_c(P_a)$, $\nu_c(P_a, CI^a)$, and μ_c.

Main theorem

Theorem 0.1. *There's an sequential equilibrium with property that for any $P \in \mathcal{P}$:*

$$|\{a \mid \gamma(P, a) > 0\}| \leq |\lambda(P)| \tag{3}$$

In other words, there's an sequential equilibrium using no more than $|\lambda(P)|$ actions in any P. Thus we bound number of actions needed in every information set grouped in this P.

No-limit poker corrolary *There's a sequential equilibrium using no more than $\binom{52}{2}$ **bet sizes** in every information set.*

Outline of approach

Let's suppose there's an sequential equilibrium (σ, μ) using more than $|\lambda(P)|$ actions in any $I \in P$ (if there's no such equilibrium, we are done). We create new assessment (σ', μ), that **differs only** in information sets in P, so that this assesment satisfies:

1. sequential rationality
2. consistency
3. $|$actions used by new strategy in $P| \leq |\lambda(P)|$

So we get a new sequential equilibrium, using no more than $|\lambda(P)|$ actions in every $I \in P$. Iteratively, we take this new equilibrium and if there's another $I' \in P'$ that uses too many actions, we just repeat steps above.

Finally, since the game is finite, we get Nash equilibrium using no more than $|\lambda(P)|$ actions in every information set in P.

Proof

We denote $\beta, \mu, \nu, \pi, \gamma$ of the new strategy σ' as $\beta', \mu', \nu', \pi', \gamma'$.

Given an assetment (σ, μ), we find P that violates action bound. Now we want to compute new strategy profile σ', but since we don't change beliefs, (σ', μ) must be *consistent*. First step is to show that.

Lemma. *If for all P_a, for all $CI^a \in \lambda_c(P_a)$: $\nu_c = \nu'_c$, then (σ', μ) is consistent* $^{(*)}$

With this result in mind, we write down some simple equations, where each variable x_i correspond to $\gamma'(P, a_i)$ so that $\nu_c = \nu'_c$.

$$\begin{pmatrix} \sum_i \nu_c(P_{a_i}, CI_1^{a_i} \in \lambda_c(P_{a_i}))\, x_i \\ \sum_i \nu_c(P_{a_i}, CI_2^{a_i} \in \lambda_c(P_{a_i}))\, x_i \\ \vdots \\ \sum_i \nu_c(P_{a_i}, CI_{|\lambda(P)|}^{a_i} \in \lambda_c(P_{a_i}))\, x_i \end{pmatrix} = \begin{pmatrix} \nu(P, I_1) \\ \nu(P, I_2) \\ \vdots \\ \nu(P, I_{|\lambda(P)|}) \end{pmatrix}$$

$$x_i \geq 0 \; \forall i \tag{4}$$

See that for any solution, $\sum_i x_i = 1$. $^{(*)}$

New strategy

Because x_i correspond to new $\gamma'(PI, a_i)$, we set $\beta'(I, a)$ to:

$$\beta'(I, a_i) = \frac{\nu_c(P_{a_i}, CI^{a_i}) x_i}{\nu(P, I)} \tag{5}$$

Strategy properties

1. β' is valid distribution: $^{(*)}$

$$\sum_{a_i \in A(I)} \beta'(I, a_i) = 1$$

2. Beliefs remain consistent: $^{(*)}$

$$\forall a \in A(P), \forall I \in \lambda(P): \; \nu'_c(P_a, CI^{a_i}) = \nu_c(P_a, CI^{a_i})$$

So any solution to (4) gives us new strategy σ', so that assessment $((\sigma'_i, \sigma_{-i}), \mu)$ remains consistent. Since beliefs remain unchanged, we know that all players except of i are *sequentially rational*.

Sequential rationality

To satisfy sequential rationality for player $p(P)$, we simply maximize his expected value:

$$f(x) = \sum_{a_i \in A(P)} x_{a_i} O_{p(P)}(\sigma', \mu \big| (P_{a_i})) \tag{6}$$

Action elimination

Maximizing function (6) over conditions (4) gives us new sequential equilibrium. But both, the conditions and function are linear! Thus there must be some optimal *basic* solution, using no more than $rank(A)$ non-zero variables (Dantzig and Thapa 1997). Finally, because there are only $|\lambda(P)|$ rows in (4), this concludes our proof.

Conclusion

We showed that there's an sequential equilibrium using no more than limited number of actions. We also showed a simple way how to compute this equilibrium from any equilibrium that uses too many actions. Unfortunately, we are not aware of any usage for better equilibrium-finding algorithms (except for straight-forward upper bound for some support finding techniques). But once we find some equilibrium, it's easy to compute more compact one.

We also believe that this idea could be used to enhance game abstraction techniques. However, these abstractions tend to be very game-specific, so this needs to be further evaluated.

*for proof see appendix

References

Dantzig, G. B., and Thapa, M. N. 1997. *Linear Programming: 1: Introduction*, volume 1. Springer.

Johanson, M.; Waugh, K.; Bowling, M.; and Zinkevich, M. 2011. Accelerating best response calculation in large extensive games. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume One*, 258–265. AAAI Press.

Johanson, M. 2013. Measuring the size of large no-limit poker games. *arXiv preprint arXiv:1302.7008*.

Osborne, M. J., and Rubinstein, A. 1994. *A course in game theory*. MIT press.

Appendix

Lemma.

Proof. The key idea is to notice that no matter how we change strategy in P, setting μ'_c to

$$\mu'_c(CI^a, (h,a)) = \mu_c(CI^a, (h,a)) \ \forall CI \in P_a$$

Is consistent with the new strategy σ'. Now we compute the probability of game being in the state (h,a) given that P_a is reached:

$$\begin{aligned}\pi'((h,a)|P_a) &= \mu'_c(CI^a,(h,a))\nu'_c(P_a,CI^a)\\ &= \mu_c(CI^a,(h,a))\nu_c(P_a,CI^a) \quad = \pi((h,a)|P_a)\end{aligned}$$

So all these probability remains the same. And finally, beliefs:

$$\begin{aligned}\mu(I,(h,a)) &= \frac{\pi'((h,a)|P_a)}{\sum_{(h',a)\in I}\pi'((h',a)|P_a)}\\ &= \frac{\pi((h,a)|P_a)}{\sum_{(h',a)\in I}\pi((h',a)|P_a)} = \mu(I,(h,a))\end{aligned}$$

And we see that if $\nu = \nu'$, beliefs remain consistent. □

0) $\sum_i x_i = 1$**.**

Proof. Sum all rows in (4):

$$\begin{aligned}\sum_{I\in\lambda(P)}\sum_i \nu_c(P_{a_i}, CI_1^{a_i}\in\lambda_c(P_{a_i}))\,x_i &= \sum_{I\in\lambda(P)}\nu(P,I)\\ \sum_i\sum_{I\in\lambda(P)}\nu_c(P_{a_i}, CI_1^{a_i}\in\lambda_c(P_{a_i}))\,x_i &= 1\\ \sum_i x_i &= 1\end{aligned}$$

□

1) β' **sums to** 1

Proof.

$$\begin{aligned}\sum_{a_i\in A(I)}\beta'(I,a_i) &= \sum_{a_i\in A(I)}\frac{\nu_c(P_{a_i},CI^{a_i})x_i}{\nu(P,I)}\\ &= \frac{1}{\nu(P,I)}\sum_{a_i\in A(I)}\nu_c(P_{a_i},CI^{a_i})x_i\\ &= \frac{1}{\nu(P,I)}\nu(P,I) = 1\end{aligned}$$

□

2) $\nu'(P_a, CI^a) = \nu(P_a, CI^a)\ \forall a, CI$

Proof.

$$\begin{aligned}\nu'(P_{a_i},CI^a) &= \frac{\nu'(P,I)\beta'(I,a)}{\sum_{I'\in\lambda(P)}\nu'(P,I')\beta'(I',a)}\\ &= \frac{\nu(P,I)\frac{\nu_c(P_a,CI)x_a}{\nu(P,I)}}{\sum_{I'\in\lambda(P)}\nu(P,I')\frac{\nu_c(PI_a,CI')x_a}{\nu(P,I')}}\\ &= \frac{\nu_c(P_a,CI)x_a}{\sum_{I'\in\lambda(P)}\nu_c(P_a,CI')x_a}\\ &= \nu_c(P_a,CI^a)\end{aligned}$$

□