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

Martin Schmid
Charles University In Prague
lifordi@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 equilibrium. We bound this number by a private information and show how to compute this Nash equilibrium.

Introduction
We focus on extensive form games with imperfect information and publicly observable actions. These games model situations where players receive some private information, and players sequentially 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. For 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 consequently contains $6.3 \times 10^{164}$ game states (Johnson 2013).

In games with no hidden information, players don’t need to mix their strategies to play optimally - 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 receives 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 $u_i : H \rightarrow \mathbb{R}$

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 $A_i(I)$.

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 $\sigma_i$. A strategy profile $\sigma$ 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 $\sigma$. We denote $\pi_{-i}(h)$ product of all players’ contribution to this probability (including chance), except of player $i$. For all $I$, define $\pi^I(h) = \sum_{h \in I} \pi^I(h)$

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

Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
Nash equilibrium
A Nash equilibrium in behavior strategies (Osborne and Rubinstein 1994, p. 203) is a profile $\sigma^*$ with the property that for every player $i \in N$ we have
\[ O_i(\sigma_{i-1}^*, \sigma_i^*) \geq O_i(\sigma_{i-1}^*, \sigma_i) \]
for every strategy $\sigma_i$ of player $i$.

Sequential equilibrium
Intuitively, sequential equilibrium states that player plays optimally in every information 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 $(\sigma, \mu)$, where $\sigma$ is a profile of behavioral strategies and $\mu$ 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 $(\sigma, \mu)$ conditional on $I$ being reached.

Sequential rationality The assessment $(\sigma, \mu)$ is sequentially rational, if for every player $i \in N$ and every information set $I \in \mathcal{I}_i$ we have
\[ O_i(\sigma_i, \mu|I) \geq O_i((\sigma_{i-1}, \sigma_i'), \mu|I) \]
for every strategy $\sigma_i'$ 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 straightforwardly 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 histories

For this public state tree, we also define:
- A set of actions available in every $P \in \mathcal{P}$
\[ A(P) := \bigcup_{I \in P} A(I) \]
- 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 $I \in \mathcal{P}$, conditional on $P \in \mathcal{P}$ being reached (consistent with $\sigma$)
- Probability measure over $A(P)$
\[ \gamma(P, a) = \sum_{I \in P} \nu(P, I) \beta(I, a) \]
- $P_a \in \mathcal{P}$, public state that follows $P \in \mathcal{P}$ if action $a \in A(P)$ is taken.
\[ (h, a) \in P_a \iff (h) \in P \]

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).

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

Main theorem
Theorem 0.1. There’s an sequential equilibrium with property that for any $P \in \mathcal{P}$:
\[ |\{ a \mid \gamma(P, a) > 0 \}| \leq |\lambda(P)| \]

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 $(\sigma, \mu)$ using more than $|\lambda(P)|$ actions in any $I \in P$ (if there’s no such equilibrium, we are done). We create new assessment $(\sigma', \mu)$, that differs only in information sets in $P$, so that this assessment 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 $\sigma'$ as $\beta', \mu', \nu', \pi', \gamma'$.

Given an assessment $(\sigma, \mu)$, we find $P$ that violates action bound. Now we want to compute new strategy profile $\sigma'$, but since we don’t change beliefs, $(\sigma', \mu)$ 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 $(\sigma', \mu)$ 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 $\nu_c = \nu'_c$.

$$
\begin{align*}
\sum_i \nu_c(P_a, CI^a_1 \in \lambda_c(P_a)) x_i & = \nu(P, I_1) \\
\sum_i \nu_c(P_a, CI^a_2 \in \lambda_c(P_a)) x_i & = \nu(P, I_2) \\
\vdots \\
\sum_i \nu_c(P_a, CI^a_{|\lambda(P)|} \in \lambda_c(P_a)) x_i & = \nu(P, I_{|\lambda(P)|})
\end{align*}
$$

$$
\begin{align*}
\sum_i x_i & \geq 0 \forall i
\end{align*}
$$

(4)

See that for any solution, $\sum_i x_i = 1$, (*)

New strategy

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

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

(5)

Strategy properties

1. $\beta'$ 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 $\sigma'$, so that assessment $((\sigma', \sigma_{-i}), \mu)$ 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:

$$f(x) = \sum_{a_i \in A(P)} x_i O_{p(P)}(\sigma', \mu\{P_a\})$$

(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 $\text{rank}(A)$ 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 straightforward 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


Lemma.

Proof. The key idea is to notice that no matter how we change strategy in \( P \), setting \( \mu_c' \) to

\[
\mu_c'(CI^a, (h, a)) = \mu_c(CI^a, (h, a)) \forall CI \in P_a
\]

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

\[
\pi((h, a)|P_a) = \mu_c'(CI^a, (h, a))\nu_c(P_a, CI^a) = \mu_c(CI^a, (h, a))\nu_c(P_a, CI^a) = \pi((h, a)|P_a)
\]

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

\[
\mu(I, (h, a)) = \frac{\pi'((h, a)|P_a)}{\sum_{(h', a') \in I} \pi'((h', a')|P_a)} = \frac{\pi((h, a)|P_a)}{\sum_{(h', a') \in I} \pi((h', a')|P_a)} = \mu(I, (h, a))
\]

And we see that if \( \nu = \nu' \), beliefs remain consistent.

\[\boxed{0) \sum_i x_i = 1}\]

Proof. Sum all rows in (4):

\[
\sum_{i \in \lambda(P)} \nu_c(P_{a_i}, CI^a_{1}) \sum_{i \in \lambda(P)} x_i = \nu(P, I) \]

\[
\sum_{i \in \lambda(P)} \nu_c(P_{a_i}, CI^a_{1}) x_i = 1
\]

\[\sum_i x_i = 1\]