Simplifying Optimal Strategies in Positive Stochastic Games

Bill Sudderth
based on joint work with
János Flesch and Arkadi Predtetchinski
Main question for this talk

Does the existence of an optimal strategy imply the existence of a stationary optimal strategy?
Two-Person Zero-Sum Positive Stochastic Game

Five ingredients:  $S, A, B, r, p$.

$S$ - state space

$A, B$ - action sets for players 1 and 2

$p(\cdot|s, a, b)$ - law of motion

$r(s, a, b)$ - nonnegative daily reward player 2 pays to player 1
Play of the game

Play begins at some state $s_0 \in S$, player 1 chooses $a_0 \in A$, player 2 chooses $b_0 \in B$, and player 2 pays player 1 the amount $r_0 = r(s_0, a_0, b_0)$.

The next state $s_1$ has distribution $p(\cdot|s_0, a_0, b_0)$, player 1 chooses $a_1 \in A$, player 2 chooses $b_1 \in B$, and player 2 pays player 1 the amount $r_1 = r(s_1, a_1, b_1)$.

The next state $s_2$ has distribution $p(\cdot|s_2, a_2, b_2)$, and play continues.

Player 2 pays player 1 the amount $\sum_{n=0}^{\infty} r_n$.
Strategies

A strategy $\pi$ for player 1 selects each action $a_n$ with a probability distribution $\pi(h) \in \Delta(A)$ for every history $h = (s_0, a_0, b_0, \ldots, s_{n-1}, a_{n-1}, b_{n-1}, s_n)$.

A strategy $\sigma$ for player 2 selects each $b_n$ with a probability distribution $\sigma(h) \in \Delta(B)$ for all $h$.

The strategy $\pi$ is stationary if there is a function $x : S \mapsto \Delta(A)$ such that $\pi(h) = x(s_n)$ for all $h$. The stationary strategy $\pi$ is identified with the function $x$.

Similarly, the strategy $\sigma$ is stationary if there is a function $y : S \mapsto \Delta(B)$ such that $\sigma(h) = y(s_n)$ for all $h$, and $\sigma$ is identified with $y$. 
Expected payoff and value

A pair $\pi, \sigma$ together with an initial state $s_0 = s$ and the law of motion $p$ determine the distribution $P_{s,\pi,\sigma}$ of the stochastic process $s_0, a_0, b_0, s_1, a_1, b_1, \ldots$.

The **expected payoff** from player 2 to player 1 is $E_{s,\pi,\sigma}[\sum_{n=0}^{\infty} r_n]$.

The game with initial state $s$ has **value** $v(s)$ if

$$v(s) = \sup_{\pi} \inf_{\sigma} E_{s,\pi,\sigma}[\sum_{n=0}^{\infty} r_n] = \inf_{\sigma} \sup_{\pi} E_{s,\pi,\sigma}[\sum_{n=0}^{\infty} r_n].$$

If $S, A, B$ are countable and either $A$ or $B$ is finite, (or if $S, A, B$ are Borel, either $A$ or $B$ is compact, and there are measurability and continuity conditions on $p$ and $r$,) then the value $v(s)$ exists for all $s$.

**Assumption.** The value $v(s)$ exists and is finite for all $s$. 
Optimal Strategies

The strategy $\pi$ is \textbf{optimal for player 1 at $s$} if

$$E_{s,\pi,\sigma} \left[ \sum_{n=0}^{\infty} r_n \right] \geq v(s)$$

for all $\sigma$; $\pi$ is \textbf{optimal for player 1} if it is optimal at every $s$.

Similarly $\sigma$ is \textbf{optimal for player 2 at $s$} if

$$E_{s,\pi,\sigma} \left[ \sum_{n=0}^{\infty} r_n \right] \leq v(s)$$

for all $\pi$; $\sigma$ is \textbf{optimal for player 2} if it is optimal at every $s$. 
The Shapley Equation.

\[ v(s) = \sup_{\alpha} \inf_{\beta} \sum_{a,b} [r(s, a, b) + \sum_{t} v(t) p(t|s, a, b)] \alpha(a) \beta(b) \]

\[ = \inf_{\beta} \sup_{\alpha} \sum_{a,b} [r(s, a, b) + \sum_{t} v(t) p(t|s, a, b)] \alpha(a) \beta(b). \]

The infimum (respectively, the supremum) is over \( \beta \in \Delta(B) \)
(respectively, \( \alpha \in \Delta(A) \)).

The Shapley equation says that \( v(s) \) is the value of the **one-shot game** \( M(s) \) with the same action sets \( A, B \) and payoff

\[ f(s, a, b) = r(s, a, b) + \sum_{t} v(t) p(t|s, a, b). \]
Lemma 1. If a strategy is optimal for player 1 or player 2 in the stochastic game starting at \( s \), then its first action is optimal in the one-shot game \( M(s) \).

Definition. A strategy \( \pi \) is **locally optimal** for player 1 if \( \pi(h) \) is optimal for player 1 in \( M(s_n) \) for every history \( h = (s_0, a_0, b_0, \ldots, s_{n-1}, a_{n-1}, b_{n-1}, s_n) \). A **locally optimal** strategy \( \sigma \) for player 2 is defined similarly.

Locally optimal strategies correspond to the strategies called **thrifty** by Dubins and Savage (1965) in their theory of one-person gambling problems.
Theorem 1. If $\sigma$ is locally optimal for player 2, then $\sigma$ is optimal for player 2.

Proof. For all $s$ and $\pi$,

$$
E_{s, \pi, \sigma}[r_0 + v(s_1)] = \sum_{a_0, b_0} \left[ r(s, a_0, b_0) + \sum_{s_1} v(s_1)p(s_1|s, a_0, b_0) \right] \pi(s)(a_0)\sigma(s)(b_0) \leq v(s).
$$

Induction shows that, for all $n$,

$$
E_{s, \pi, \sigma} \left[ \sum_{k=0}^{n} r_k + v(s_{n+1}) \right] \leq v(s).
$$
Because $v \geq 0$,

$$E_{s,\pi,\sigma} \left[ \sum_{k=0}^{\infty} r_k \right] = \lim_{n} E_{s,\pi,\sigma} \left[ \sum_{k=0}^{n} r_k \right]$$

$$\leq \lim_{n} E_{s,\pi,\sigma} \left[ \sum_{k=0}^{n} r_k + v(s_{n+1}) \right] \leq v(s).$$

Thus $\sigma$ is optimal.

**Corollary 1.** If $y(s)$ is optimal for player 2 in $M(s)$ for every $s$, then the stationary strategy $y$ is optimal for player 2.

By Lemma 1, every optimal strategy must begin with a locally optimal action. Hence:

**Corollary 2.** If player 2 has an optimal strategy, then player 2 has an optimal stationary strategy.

Theorem 1 and Corollary 1 both fail for player 1.
A game where player 1 has a locally optimal strategy but no optimal strategy

Let $S = \{s, t\}$. State $t$ is absorbing with no payoff. Each player has 2 actions at state $s$ with rewards and motion as below:

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$1 \rightarrow t$</td>
<td>$0 \rightarrow s$</td>
</tr>
<tr>
<td>$B$</td>
<td>$0 \rightarrow t$</td>
<td>$1 \rightarrow t$</td>
</tr>
</tbody>
</table>

Player 1 wins 1 with probability at least $1 - \epsilon$ by playing $(1 - \epsilon, \epsilon)$ repeatedly. So $v(s) = 1$. The unique optimal play for 1 in $M(s)$ is $(1, 0)$, but there is no optimal strategy for 1 in the stochastic game.
Optimal strategies for player 1

**Assumption.** The state space $S$ and action set $B$ are finite and $E_{s,\pi,\sigma}[\sum_{n=0}^{\infty} r_n] < \infty$ for all $s, \pi, \sigma$.

**Theorem 2.** *If there exists an optimal strategy for player 1, then there also exists an optimal stationary strategy for player 1.*

The proof is more difficult than for negative games.
Lemma 2. Assume that
(1) $\pi$ is a locally optimal strategy for player 1, and
(2) $\sigma$ is a strategy for player 2 such that $P_{s,\pi,\sigma}[\lim_n v(s_n) = 0] = 1$
for a given $s \in S$.
Then $E_{s,\pi,\sigma}[\sum_{n=0}^{\infty} r_n] \geq v(s)$.

Proof. By an argument like that for Theorem 1, it follows from
(1) that
\[
v(s) \leq \lim_n E_{s,\pi,\sigma}[\sum_{k=0}^{n} r_k + v(s_{n+1})].
\]
And by (2)
\[
\lim_n E_{s,\pi,\sigma}[\sum_{k=0}^{n} r_k + v(s_{n+1})] = \lim_n E_{s,\pi,\sigma}[\sum_{k=0}^{n} r_k]
= E_{s,\pi,\sigma}[\sum_{k=0}^{\infty} r_k].
\]
Corollary 3. For a positive game, $x$ is an optimal stationary strategy if (1) $x(s)$ is optimal for player 1 in $M(s)$ for every $s$, and (2) $P_{s,x,\sigma}[\lim_n v(s_n) = 0] = 1$ for all $s, \sigma$.

The converse is also true.

Therefore, to find an optimal stationary strategy $x$ for player 1, we can restrict attention to locally optimal $x$ and look for one that satisfies condition (2).

The two conditions are similar to those called "thrifty" and "equalizing" by Dubins and Savage (1965).
The key idea in the proof of Theorem 2 goes back to Flesch, Thuijsman, and Vrieze (1999), who studied long run average games.

For each $s \in S$, define

$$X^*(s) = \{x \in \Delta(A) \mid x \text{ is optimal for player 1 in } M(s)\}$$

$$A^*(s) = \{a \in A \mid x(a) > 0 \text{ for some } x \in X^*(s)\}$$

$$X^{**}(s) = \{x \in X^*(s) \mid x(a) > 0 \text{ for all } a \in A^*(s)\}.$$ 

**Definition.** A stationary strategy $x$ is **maximally mixed** if, for all $s$, $x(s) \in X^{**}(s)$. 
Lemma 3. Every maximally mixed stationary strategy for player 1 is optimal.

Theorem 2 follows from this lemma. The proof is in several steps and is too long for this talk.

Some intuition: A maximally mixed stationary strategy $x$ is locally optimal by its definition, and, if there is any locally optimal strategy $\pi$ that reaches the set where $v = 0$ with positive probability against every $\sigma$, then so does $x$. This is because $x$ assigns positive probability to every action used by $\pi$.

Note: $\lim v(s_n) = 0$ if and only if $s_n$ is eventually in the set $[v = 0]$. 
Positive games with $S$ infinite. Blackwell (1970) showed that Theorem 2 remains true for $S$ countable if player 2 is a dummy. However, there is the following counterexample in the two-person case.

Player 1 chooses $c$ or $s$ at states $p_0, p_2, \ldots$ and player 2 chooses $c$ or $s$ at states $p_1, p_3, \ldots$. When, if ever, $s$ is chosen at some state $p_n$, player 1 receives the payoff $n/(n + 1)$ and the game terminates. At each state $p_n$ where Player 1 has the move, his unique optimal first move is $c$. However, if he plays $c$ at every stage, player 2 can also play all $c$’s and player 1 receives 0.
References


