Note: On A Proof Of Positionality Of Mean-Payoff Stochastic Games

Hugo Gimbert, Edon Kelmendi

Note: On A Proof Of Positionality Of Mean-Payoff Stochastic Games

Hugo Gimbert
CNRS, LaBRI
Université de Bordeaux
hugo.gimbert@labri.fr

Edon Kelmendi
LaBRI
Université de Bordeaux
edon.kelmendi@labri.fr

December 4, 2014

1 Introduction

Two-player stochastic games played on finite arenas are considered, where to each state a reward is attached. The players objectives are to maximize, and respectively minimize the average reward on the long run. The purpose of this note is to point out an error in the proof given by Ligget and Lippman in [6] that both players have optimal strategies that are positional, and to provide one possible resolution.

The arena consists of a finite set of states S_i controled by player $i \in \{1,2\}$, for all $s \in S_1 \cup S_2 = S$ a set of actions denoted $A(s)$, and a transition probability $p : S \times A(S) \to \Delta(S)$ where $A(S) = \cup_{s \in S} A(s)$ and $\Delta(S)$ the set of probability distributions on S, i.e functions $d : S \to \mathbb{R}^+$ with the property that $\sum_{s \in S} d(s) = 1$. The game starts at some state $s_0 \in S_i$ whence player i chooses an action a in $A(s_0)$ after which the chance of being in state s_1 is $p(s_0, a)(s_1)$, and so on for an infinite duration. Denote by S^* (S^ω) the set of finite (infinite) sequences of elements of S. Strategies for player i are functions $\sigma_i : S^* S_i \to A(S)$. The positional strategies are of the type $\sigma_i : S_i \to A(S)$. Fixing an initial state $s_0 \in S$ and two strategies σ_1, σ_2 for each player, gives rise to a unique probability measure on the sigma-algebra generated by the cylinders $\{pS^\omega \mid p \in S^*\}$, denoted $\mathbb{P}_{s_0}^{\sigma_1,\sigma_2}$, with the property

$$\mathbb{P}_{s_0}^{\sigma_1,\sigma_2}(s_1 s_2 \cdots s_n S^\omega) = \prod_{i=0}^{n-1} p(s_i, \sigma_{k(s_i)}(s_0 \cdots s_i))(s_{i+1}),$$

where $k(s_i) = j$ if $s_i \in S_j$. A payoff functions is a function $f : S^\omega \to \mathbb{R}$. We will deal mainly with the following payoff functions:

$$D_\beta(s_0 s_1 \cdots) = \sum_{i=0}^{\infty} \beta^i r(s_i)$$

$$\underline{M}(s_0 s_1 \cdots) = \liminf_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} r(s_i)$$

$$\overline{M}(s_0 s_1 \cdots) = \limsup_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} r(s_i),$$

where $\beta \in [0,1)$ and $r : S \to \mathbb{R}^+$ describes the rewards attached to states. Let T_i for $i \in \mathbb{N}$ be the random variable defined as: $T_i(s_0 \cdots s_i \cdots) = s_i$. We abbreviate the random variable $D_\beta(T_0 T_1 \cdots)$ as D_β, and the same for the two other payoff functions.

Having fixed the notation, we proceed by presenting the proof of the following theorem, to which [6] is devoted.

Theorem 1 ([6])**.** *There exist a pair of positional strategies σ_1^*, σ_2^* such that for all strategies σ_1, σ_2 and $s \in S$,*

$$\mathbb{E}_s^{\sigma_1,\sigma_2^*}[\overline{M}] \leq \mathbb{E}_s^{\sigma_1^*,\sigma_2^*}[M] \leq \mathbb{E}_s^{\sigma_1^*,\sigma_2}[\underline{M}],$$

where $M(s_0 s_1 \cdots) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} r(s_i)$.

We introduce two results on which the proof is based.

Theorem 2 ([1])**.** *For one player games (when $S_i = \emptyset$ for one $i \in \{1,2\}$), there exists a positional strategy σ^* and $\beta^* \in [0,1)$ such that for all $\beta \geq \beta^*$, strategies σ, and $s_0 \in S$,*

$$\mathbb{E}_{s_0}^{\sigma^*}[D_\beta] \geq \mathbb{E}_{s_0}^{\sigma}[D_\beta],$$

if $S_2 = \emptyset$, and the opposite inequality if $S_1 = \emptyset$.

Theorem 3 ([7])**.** *For all $\beta \in [0,1)$ there exists positional strategies σ_1^*, σ_2^* such that for all strategies σ_1, σ_2 and $s_0 \in S$,*

$$\mathbb{E}_{s_0}^{\sigma_1,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2}[D_\beta].$$

Fixing a positional strategy for one player results in a Markov decision process. Let σ_2 be a positional strategy for player 2, then according to Theorem 2 there exists a positional strategy σ_1^* and $\beta_{\sigma_2} \in [0,1)$ such that for all strategies σ_1 for player 1 and $\beta \geq \beta_{\sigma_2}$, $\mathbb{E}_s^{\sigma_1^*,\sigma_2}[D_\beta] \geq \mathbb{E}_s^{\sigma_1,\sigma_2}[D_\beta]$ for all $s \in S$. By fixing a positional strategy of player 1, we get a symmetric statement. Let $\beta^* = \max\{\beta_\sigma \mid \sigma \text{ a positional strategy}\}$, i.e the largest β resulting from Theorem 2 by fixing positional strategies for either player. Let σ_1^*, σ_2^* be the pair positional strategies for $\beta = \beta^*$ in Theorem 3. The erroneous claim in [6] is that theorems 2 and 3 imply the following: for all $\beta \geq \beta^*$ and $s_0 \in S$

$$\mathbb{E}_{s_0}^{\sigma_1,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2}[D_\beta], \tag{1}$$

as shown by the example in the next session.

2 An example

Consider the following one-player game with two states.

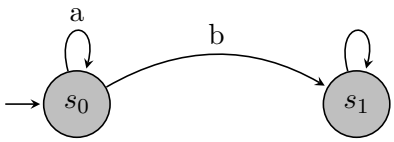

Figure 1: A one player game

Assume that $r(s_0) = 0$ and $r(s_1) = 1$ and that $S_1 = \{s_0, s_1\}$. For all $\beta \geq 0$ the optimal strategy is to play b in s_0 with a gain of $\sum_k \beta^k$. But regard that for $\beta = 0$ the strategy that plays a in s_0 is also optimal. Therefore β^* can be equal to 0 and σ_1^* in Theorem 3 can be the strategy that plays a in s_0. But for this pair (1) is not true, since for $\beta > 0$, the strategy that plays b in s_0 is strictly better.

We provide a different proof of the existence of positional strategies which are optimal for $\beta \geq \beta^*$. This proof follows closely the one given in [3] for the discrete case, which in turn follows the one given in [4] for Markov decision processes.

3 A different proof

Lemma 1 ([3],[4]). *There exist a pair of positional strategies σ_1^*, σ_2^* and $\beta^* \in [0,1)$ such that for all $\beta \geq \beta^*$ and $s_0 \in S$,*

$$\mathbb{E}_{s_0}^{\sigma_1,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_\beta] \leq \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2}[D_\beta]. \tag{2}$$

Proof. Let σ_1, σ_2 be two positional strategies and $s_0 \in S$. We argue that $f(\beta) = \mathbb{E}_{s_0}^{\sigma_1,\sigma_2}[D_\beta]$ is a rational function (a ratio between two polynomials in β). Indeed since the fixed strategies are positional, the dynamics of the game are described by the Markov chain given in the form of the $S \times S$ stochastic matrix

$$P_{\sigma_1,\sigma_2}(s,s') = \begin{cases} p(s,\sigma_1(s),s') \text{ if } s \in S_1 \\ p(s,\sigma_2(s),s') \text{ if } s \in S_2 \end{cases}.$$

For $s \in S$ let $\mathbf{s}$ be the Dirac distribution on s given in the form of a row matrix (i.e $1 \times S$ matrix with all components 0 except in position s), and let $\mathbf{r}$ be the column vector of rewards. Then

$$\mathbb{E}_{s_0}^{\sigma_1,\sigma_2}[D_\beta] = \mathbf{s_0} \sum_{k=0}^{\infty} (\beta P_{\sigma_1,\sigma_2})^k \mathbf{r},$$

where $P_{\sigma_1,\sigma_2}^0 = I$, the identity matrix. It is a theorem (cf. [5]) that if $\lim_{k\to\infty}(\beta P_{\sigma_1,\sigma_2})^k = 0$, then the matrix $I - \beta P_{\sigma_1,\sigma_2}$ is invertible and $\sum_{k=0}^{\infty}(\beta P_{\sigma_1,\sigma_2})^k = (I - \beta P_{\sigma_1,\sigma_2})^{-1}$, from where we conclude that $f(\beta) = \mathbb{E}_{s_0}^{\sigma_1,\sigma_2}[D_\beta]$ is a rational function, with both polynomials having a finite degree.

Let (β_n') be a sequence with elements in $[0,1)$ such that $\lim_{n\to\infty} \beta_n' = 1$. By Theorem 2 for all β_n' we have a pair of positional strategies that are optimal, and since the set of positional strategies is finite we may assume that there exists σ_1^*, σ_2^* positional, such that (2) holds for a subsequence $(\beta_n) \subseteq (\beta_n')$ and the pair of strategies σ_1^*, σ_2^*. We argue that there exists $\beta^* \in [0,1)$ such that (2) holds for all $\beta \geq \beta^*$. Assume on the contrary that for all $\beta^* \in [0,1)$ there exist $\beta \geq \beta^*$, a pair of positional strategies τ_1, τ_2, and a state $s_0 \in S$ such that, either $\mathbb{E}_{s_0}^{\tau_1,\sigma_2^*}[D_\beta] > \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_\beta]$ or $\mathbb{E}_{s_0}^{\sigma_1^*,\tau_2}[D_\beta] < \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_\beta]$. Consequently there exists a sequence (μ_n) with $\mu_n \in [0,1)$ and $\lim_{n\to\infty} \mu_n = 1$ such that either $\mathbb{E}_{s_0}^{\tau_1,\sigma_2^*}[D_{\mu_n}] > \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_{\mu_n}]$ or $\mathbb{E}_{s_0}^{\sigma_1^*,\tau_2}[D_{\mu_n}] < \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_{\mu_n}]$ for all $n \in \mathbb{N}$. Assume the latter. But then $g(\beta) = \mathbb{E}_{s_0}^{\sigma_1^*,\tau_2}[D_{\mu_n}] - \mathbb{E}_{s_0}^{\sigma_1^*,\sigma_2^*}[D_{\mu_n}]$ would have infinitely many zeros, which is not possible since we showed that it is a rational function. Similarly if we assume the former inequality. □

We give the rest of the proof of Theorem 1 found in [6] for the sake of completeness. For all $n \in \mathbb{N}$ define $H^n(s_0 s_1 \dots) = \sum_{k=0}^{n} r(s_k)$, and abbreviate by H^n the random variable $H^n(T_0 T_1 \cdots)$. Let σ_1^*, σ_2^* be a pair of positional strategies resulting from Lemma 1. Now from Theorem 4.2 in [2] we have for all $s \in S$:

$$\mathbb{E}_s^{\sigma_1^*,\sigma_2^*}[H^n] - \sup_{\sigma} \mathbb{E}_s^{\sigma,\sigma_2^*}[H^n] \text{ is bounded uniformely in } n.$$

Therefore for all strategies σ,

$$\mathbb{E}_s^{\sigma_1^*,\sigma_2^*}[M] = \lim_{n\to\infty} \frac{1}{n} \sup_{\sigma'} \mathbb{E}_s^{\sigma',\sigma_2^*}[H^n] \geq \limsup_{n\to\infty} \frac{1}{n} \mathbb{E}_s^{\sigma,\sigma_2^*}[H^n] = \mathbb{E}_s^{\sigma,\sigma_2^*}[\overline{M}],$$

and symmetrically for the inequality $\mathbb{E}_s^{\sigma_1^*,\sigma_2^*}[M] \leq \mathbb{E}_s^{\sigma_1^*,\sigma}[\underline{M}]$.

References

[1] David Blackwell. Discrete dynamic programming. *The Annals of Mathematical Statistics*, 33(2):719–726, 06 1962.

[2] Barry W Brown. On the iterative method of dynamic programming on a finite space discrete time markov process. *The annals of mathematical statistics*, pages 1279–1285, 1965.

[3] Hugo Gimbert and Wiesaw Zielonka. Applying blackwell optimality: Priority mean-payoff games as limits of multi-discounted games.

[4] Arie Hordijk and Alexander A Yushkevich. Blackwell optimality. In *Handbook of Markov decision processes*, pages 231–267. Springer, 2002.

[5] John G Kemeny and James Laurie Snell. *Finite markov chains*, volume 356. van Nostrand Princeton, NJ, 1960.

[6] Thomas M Liggett and Steven A Lippman. Stochastic games with perfect information and time average payoff. *Siam Review*, 11(4):604–607, 1969.

[7] Lloyd S Shapley. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10):1095, 1953.