Weak Determinacy and Decidability of Reachability Games with Partial Observation

Nathalie Bertrand, Blaise Genest, Hugo Gimbert

December 10, 2008

Abstract

We consider two-players stochastic reachability games with partial observation on both sides and finitely many states, signals and actions. We prove that in such games, either player 1 has a strategy for winning with probability 1, or player 2 has a strategy for winning with probability 1, or both players have strategies that guarantee winning with strictly positive probability (positively winning strategies). We give a fix-point algorithm for deciding in doubly-exponential time which of the three cases holds.

Introduction

We prove two determinacy and decidability results about two-players stochastic reachability games with partial observation on both sides and finitely many states, signals and actions. Player 1 wants the play to reach the set of target states, while player 2 wants to keep away the play from target states. Players take their decisions based upon *signals* that they receive all along the play, but they cannot observe the actual state of the game, nor the actions played by their opponent, nor the signals received by their opponent. Each player only observes the signals he receives and the actions he plays. Players have common knowledge of the initial state of the game.

Our determinacy result is of a special kind, it concerns two notions of solutions for stochastic games. The first one is the well known notion of *almost-surely* winning strategy, which guarantees winning with probability 1 against any strategy of the opponent. The second one is the notion of *positively winning* strategy: a strategy is positively winning if it guarantees a non-zero winning probability against any strategy of the opponent. This notion is less known, to our knowledge it appeared recently in [Hor08]. The notion of positively winning strategy is different from the notion of positive value, because the non-zero winning probability can be made arbitrarily small by the opponent, hence existence of a positively winning strategy does not give any clue for deciding whether the value is zero or not. Existence of a positively winning strategy guarantees that the opponent does not have an almost-surely winning strategy, however there is no straightforward reason that one of these cases should always holds. Actually,

if we consider more complex classes of games than reachability games, there are various examples where neither player 1 has a positively winning strategy nor player 2 has an almost-surely winning strategy.

Our first result (Theorem 2) states that, in reachability games with partial observation on both sides, either player 1 has a positively winning strategy or player 2 has an almost-surely winning strategy. Moreover which case holds is decidable in *exponential* time. Notice that an almost-surely winning strategy for player 2 in a reachability game is *surely winning* as well.

Our second result (Theorem 3) states that either player 1 has an almost-surely winning strategy or player 2 has a positively winning strategy, and this is decidable in *doubly-exponential* time.

Both these results strengthen and generalize in several ways results given in [CDHR07]. Actually, in this paper is addressed only the particular case where player 2 has perfect information and target states are observable by player 1. Moreover in [CDHR07] no determinacy result is established, the paper "only" describes an algorithm for deciding whether player 1 has an almost-sure winning strategy.

1 Reachability games with partial observation on both sides

We consider zero-sum stochastic games with partial observation on both sides, where the goal of Player 1 is to reach a certain set of target states. Players only partially observe the state of the game, via signals. Signals and state transitions are governed by probability transitions: when the state is k and two actions i and j are chosen, player 1 and 2 receive respectively signals c and d and the new state is l with probability $p(c, d, l \mid k, i, j)$.

1.1 Notations

We use the following standard notations [Ren00].
The game is played in steps. At each step the game is in some state $k \in K$. The goal of player 1 is to reach target states $T \subseteq K$. Before the game starts, the initial state is chosen according to the initial distribution $\delta \in \mathcal{D}(K)$, which is common knowledge of both players. Players 1 and 2 choose actions $i \in I$ and $j \in J$, then player 1 receives a signal $c \in C$, player 2 receives a signal $d \in D$, and the game moves to a new state l. This happens with probability $p(c, d, l \mid k, i, j)$ given by fixed transition probabilities $p : K \times I \times J \to \mathcal{D}(C \times D \times K)$, known by both players. We denote $p(l \mid k, i, j) = \sum_{c,d} p(c, d, l \mid k, i, j)$. Players observe and remember their own actions and the signals they receive, it is convenient to suppose that in the signal they receive is encoded the action they just played, formally their exists $\mathrm{act} : C \cup D \to I \cup J$ such that $(p(c, d, k' \mid k, i, j) > 0) \implies ((i = \mathrm{act}(c) \text{ and } j = \mathrm{act}(d))$. We denote $p(c, d, l \mid k) = p(c, d, l \mid k, \mathrm{act}(i), \mathrm{act}(j))$. This way, plays can be described by sequences

of states and signals for both players, without mentioning which actions were played. A sequence $p = (k_0, c_1, d_1, \ldots, c_n, d_n, k_n) \in (KCD)^*K$ is a finite play if for every $0 \leq m < n$, $p(c_{m+1}, d_{m+1}, k_{m+1} \mid k_m, \mathrm{act}(c_{m+1}), \mathrm{act}(d_{m+1})) > 0$. An infinite play is a sequence $p \in (KCD)^\omega$ whose prefixes are finite plays.

A strategy of player 1 is a mapping $\sigma : \mathcal{D}(K) \times C^* \to \mathcal{D}(I)$ and a strategy of player 2 is $\tau : \mathcal{D}(K) \times D^* \to \mathcal{D}(J)$.

In the usual way, an initial distribution δ and two strategies σ and τ define a probability measure $\mathbb{P}_\delta^{\sigma,\tau}(\cdot)$ on the set of infinite plays, equipped with the σ-algebra generated by cylinders.

We use random variables K_n, I_n, J_n, C_n, D_n for designing respectively the n-th state, action of player 1, action of player 2, signal of player 1, signal of player 2. The probability to reach a target state someday is:

$$\gamma_1(\delta, \sigma, \tau) = \mathbb{P}_\delta^{\sigma,\tau}(\exists m \in \mathbb{N}, K_m \in T) \quad ,$$

and the probability to never reach the target is $\gamma_2(\delta, \sigma, \tau) = 1 - \gamma_2(\delta, \sigma, \tau)$. Player 1 seeks maximizing γ_1 while player 2 seeks maximizing γ_2.

1.2 Winning almost-surely or positively

Definition 1 (Almost-surely and positively winning)**.** *A distribution δ is* almost-surely winning *for player 1 if there exists a strategy σ such that*

$$\forall \tau, \gamma_1(\delta, \sigma, \tau) = 1 \ . \tag{1}$$

A distribution δ is positively winning *for player 1 if there exists a strategy σ such that*

$$\forall \tau, \gamma_1(\delta, \sigma, \tau) > 0 \ . \tag{2}$$

If the uniform distribution on a set of states $L \subseteq K$ is almost-surely or positively winning then L itself is said to be almost-surely or positively winning. If there exists σ such that (1) holds for every almost-surely winning distribution then σ is said to be almost-surely winning .

Positively winning strategies for player 1 and almost-sure winning and positively winning strategies for player 2 are defined similarly.

2 Winning almost-surely and positively with finite memory

Of special algorithmic interest are strategies with finite memory.

Definition 2 (Strategies with finite memory)**.** *A strategy σ with finite memory is described by a finite set M called the memory, a strategic function $\sigma_M : M \to \mathcal{D}(I)$, an update function $\mathrm{update}_M : M \times C \to M$, an initialization function $\mathrm{init}_M : \mathcal{P}(K) \to M$.*

For playing with σ, player 1 proceeds as follows. Let L be the support of the initial distribution, then initially player 1 puts the memory in state $\mathrm{init}_M(L)$.

When the memory is in state m, player 1 chooses his action according to the distribution $\sigma_M(m)$. When player 1 receives a signal c and its memory state is m, he changes the memory state to $\text{update}_M(m, c)$.

A crucial tool for establishing decidability and determinacy result is the class of finite memory strategy whose finite memory if based on the notions of beliefs or pessimistic beliefs.

2.1 Beliefs and pessimistic beliefs

The belief of a player at some moment of the play is the set of states he thinks the game could possibly be, according to the signals he received up to now. The pessimistic belief is similar, except the player assumes that no final state has been reached yet. One of the motivations for introducing beliefs and pessimistic beliefs is Proposition 1.

Beliefs of player 1 are defined by mean of the operator $\mathcal{B}_1$ that associates with $L \subseteq K$ and $c \in C$,

$$\mathcal{B}_1(L, c) = \{k \in K \mid \exists l \in L, \exists d \in D, p(k, c, d \mid l) > 0\} \ . \tag{3}$$

We defined inductively the belief after signals $c_1, \ldots, c_n$ by $\mathcal{B}_1(L, c_1, \ldots, c_n, c) = \mathcal{B}_1(\mathcal{B}_1(L, c_1, \ldots, c_n), c)$.

Pessimistic beliefs of player 1 are defined by

$$\mathcal{B}_1^p(L, c) = \mathcal{B}_1(L \backslash T, c) \ .$$

Beliefs $\mathcal{B}_2$ and pessimistic beliefs $\mathcal{B}_2^p$ for player 2 are defined similarly. We will use the following properties of beliefs and pessimistic beliefs.

Proposition 1. *Let σ, τ be strategies for player 1 and 2 and δ an initial distribution with support L. Then for every $n \in \mathbb{N}$,*

$$\begin{aligned}
&\mathbb{P}_\delta^{\sigma,\tau}\left(K_{n+1} \in \mathcal{B}_1(L, C_1, \ldots, C_n)\right) = 1 \ , \\
&\mathbb{P}_\delta^{\sigma,\tau}\left(K_{n+1} \in \mathcal{B}_2(L, D_1, \ldots, D_n)\right) = 1 \ , \\
&\mathbb{P}_\delta^{\sigma,\tau}\left(K_{n+1} \in \mathcal{B}_1^p(L, C_1, \ldots, C_n) \text{ or } K_m \in T \text{ for some } 1 \leq m \leq n\right) = 1 \ , \\
&\mathbb{P}_\delta^{\sigma,\tau}\left(K_{n+1} \in \mathcal{B}_2^p(L, D_1, \ldots, D_n) \text{ or } K_m \in T \text{ for some } 1 \leq m \leq n\right) = 1 \ .
\end{aligned}$$

Suppose τ and δ almost-surely winning for player 2, then for every $n \in \mathbb{N}$,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\mathcal{B}_2(L, D_1, \ldots, D_n) \text{ is a.s.w. for player } 2\right) = 1 \ .$$

Suppose σ and δ almost surely winning for player 1, then for every $n \in \mathbb{N}$,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\mathcal{B}_1^p(L, C_1, \ldots, C_n) \text{ is a.s.w. for player } 1 \text{ or } \exists 1 \leq m \leq n, K_m \in T\right) = 1 \ .$$

Proof. Almost straightforward from the definitions. □

2.2 Belief and pessimistic belief strategies

A strategy σ is said to be a *belief strategy* for player 1 if it has finite memory $M = \mathcal{P}(K)$ and

1. the initial state of the memory is the support of the initial distribution,
2. the update function is $(L, c) \to \mathcal{B}_1(L, c)$,
3. the strategic function $\mathcal{P}(K) \to \mathcal{D}(I)$ associates with each memory state $L \subseteq K$ the uniform distribution on a non-empty set of actions $I_L \subseteq I$.

The definition of a pessimistic belief strategy for player 1 is the same, except the update function is $\mathcal{B}_1^p$.

3 Determinacy and decidability results

In this section, we establish our main result, a determinacy result of a new kind. Usual determinacy results in game theory concern the existence of a value. Here the determinacy refers to positive and almost-sure winning:

Theorem 1 (Determinacy)**.** *Every initial distribution is either almost-surely winning for player* 1*, surely winning for player* 2 *or positively winning for both players.*

Theorem 1 is a corollary of Theorems 2 and 3, in which details are given about the complexity of deciding whether an initial distribution is positively winning for player 1 and whether it is positively winning for player 1.

Deciding whether a distribution is positively winning for player 1 is quite easy, because player 1 has a very simple strategy for winning positively: playing randomly any action.

Theorem 2 (Deciding positive winning for player 1)**.** *Every initial distribution is either positively winning for player* 1 *or surely winning for player* 2*.*

The strategy for player 1 *which plays randomly any action is positively winning. Player* 2 *has a belief strategy which is surely winning.*

The partition of supports between those positively winning for player 1 *and those surely winning for player* 2 *is computable in time exponential in* $|K|$*, together with an almost-surely winning belief strategy for player* 2*.*

Proof of Theorem 2. Let $\mathcal{L}_\infty \subseteq \mathcal{P}(K \backslash T)$ be the greatest fix-point of the monotonic operator $\Phi : \mathcal{P}(\mathcal{P}(K \backslash T)) \to \mathcal{P}(\mathcal{P}(K \backslash T))$ defined by:

$$\Phi(\mathcal{L}) = \{L \in \mathcal{L} \mid \exists j \in J, \forall d \in D, \text{ if } j = \mathrm{act}(d) \text{ then } \mathcal{B}_2(L, d) \in \mathcal{L}\},$$

and let σ_R be the strategy for player 1 that plays randomly any action. To establish Theorem 2 we are going to prove that:

(A) every support in $\mathcal{L}_\infty$ is surely winning for player 2, and

(B) σ_R is positively winning from any support $L \subseteq K$ which is not in $\mathcal{L}_\infty$.

We start with proving (A). For winning surely from any support $L \in \mathcal{L}_\infty$, player 2 uses the following belief strategy: if the current belief of player 2 is $L \in \mathcal{L}_\infty$ then player 2 chooses an action j_L such that whatever signal d player 2 receives (with $\text{act}(d) = j_L$), his next belief $\mathcal{B}_2(L, d)$ will be in $\mathcal{L}_\infty$ as well. By definition of Φ there always exists such an action j, and this defines a belief-strategy $\sigma : L \to j_L$ for player 2. When playing with this strategy, beliefs of player 2 never intersect T hence according to Proposition 1, against any strategy σ of player 1, the play stays almost-surely in $K \backslash T$, hence it stays surely in $K \backslash T$.

Conversely, we prove (B). We fix the strategy for player 1 which consists in playing randomly any action with equal probability, and the game is a one-player game where only player 2 has choices to make: it is enough to prove (B) in the special case where the set of actions of player 1 is a singleton $I = \{i\}$. Let $\mathcal{L}_0 = \mathcal{P}(K \backslash T) \supseteq \mathcal{L}_1 = \Phi(\mathcal{L}_0) \supseteq \mathcal{L}_2 = \Phi(\mathcal{L}_1) \dots$ and $\mathcal{L}_\infty$ be the limit of this sequence, the greatest fixpoint of Φ. We prove that for any support $L \in \mathcal{P}(K)$, if $L \notin \mathcal{L}_\infty$ then:

$$L \text{ is positively winning for player 1 } . \tag{4}$$

If $L \cap T \neq \emptyset$, (4) is obvious. For delaing with the case where $L \in \mathcal{P}(K \backslash T)$, we define for every $n \in \mathbb{N}$, $\mathcal{K}_n = \mathcal{P}(K \backslash T) \backslash \mathcal{L}_n$, and we prove by induction on $n \in \mathbb{N}$ that for every $L \in \mathcal{K}_n$, then for every initial distribution δ_L with support L, for every strategy τ,

$$\mathbb{P}^\tau_{\delta_L} \left(\exists m \in \mathbb{N}, K_m \in T, 2 \leq m \leq n+1 \right) > 0 \ . \tag{5}$$

For $n = 0$, (5) is obvious because $\mathcal{K}_0 = \emptyset$. Suppose that for some $n \in \mathbb{N}$, (5) holds for every $L \in \mathcal{K}_n$, and let $L \in \mathcal{K}_{n+1}$. If $L \in \mathcal{K}_n$ then by inductive hypothesis, (5) holds. Otherwise by definition of $\mathcal{K}_{n+1}$, $L \in \mathcal{L}_n \backslash \Phi(\mathcal{L}_n)$ hence by definition of Φ, whatever action j is played by player 2 at the first round, there exists a signal d_j such that $\text{act}(d_j) = j$ and $\mathcal{B}_2(L, d_j) \notin \mathcal{L}_n$. Let τ be a strategy for player 2 and j an action such that $\tau(\delta_L)(j) > 0$. If $\mathcal{B}_2(L, d_j) \cap T \neq \emptyset$ then according to Proposition 1, $\mathbb{P}^\tau_{\delta_L} (K_2 \in T) > 0$. Otherwise $\mathcal{B}_2(L, d_j) \in \mathcal{P}(K \backslash T) \backslash \mathcal{L}_n = \mathcal{K}_n$ hence according to the inductive hypothesis $\mathbb{P}^{\tau[d_j]}_{\mathcal{B}_2(L,d_j)} (\exists m \in \mathbb{N}, 2 \leq m \leq n+1, K_m \in T) > 0$. Since player 1 has only one action, by definition of beliefs, for every state $l \in \mathcal{B}_2(L_d, j)$, $\mathbb{P}^\tau_{\delta_L} (K_2 = l) > 0$. Together with the previous equation, we obtain
$\mathbb{P}^\tau_{\delta_L} (\exists m \in \mathbb{N}, 3 \leq m \leq n+2, K_m \in T) > 0$. This achieves the inductive step.

The computation of the partition of supports between those positively winning for player 1, and those surely winning for player 2 and a surely winning strategy for player 2 amounts to the computation of the largest fixpoint of Φ. since Φ is monotonic, and each application of the operator can be computed in exponential time, the overall computation can be achieved in exponential time and space. □

Deciding whether an initial distribution is positively winning for player 1 is easy because player 1 has a very simple strategy for that: playing randomly.

Figure 1: A game where player 2 needs a lot of memory.

Player 2 does not have such a simple strategy for winning positively: he has to make hypotheses about the beliefs of player 1, as is shown in the example depicted by fig. 1.

Theorem 3 (Deciding positive winning for player 2)**.** *Every initial distribution is either almost-surely winning for player* 1 *or positively winning for player* 2*.*

Player 1 *has an almost-surely winning strategy which is pessimistic belief. Player* 2 *has a positively winning strategy with finit ememory* $\mathcal{P}(\mathcal{P}(K) \times K)$*.*

The partition of supports between those almost-surely winning for player 1 *and those positively winning for player* 2 *is computable in time doubly-exponential in* $|K|$*, together with the winning strategies for both players.*

The finite memory $\mathcal{P}(\mathcal{P}(K) \times K)$ of the positively winning strategy of player 2 is used by player 2 to remember what are the possible pairs of current state and pessimistic belief of player 1.

The proof of Theorem 3 is based on the following intuition. First, if player 2 wins surely from a support L then, *a fortiori*, he wins positively from that support L. Now suppose L is a support positively winning for player 2. If from another support L' player 2 can force the pessimistic belief of player 1 to be L with positive probability, it can be shown that the support L' is positively winning for player 2 as well. Hence if player 1 wishes to win almost-surely, he should surely avoid his pessimistic belief from being L'. However, doing so, player 1 may prevent the play from reaching target states, which may create another positively winning support L for player 2, and so on...

The reader familiar with fix-point characterizations of winning sets should easily translate this intuition into a fix-point characterization of beliefs positively winning for player 2. Moreover, since there are finitely many supports, this fix-point is computable.

There are a few technical details we take care of in the following lemmatas.

We start with formalizing what it means for player 1 to force his pessimistic beliefs to stay in a certain set.

Definition 3. *Let* $\mathcal{L} \subseteq \mathcal{P}(K)$ *be a set of supports. We say that player* 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L}$ *if player* 1 *has a strategy* σ *such that for every strategy* τ *of player* 2 *and every initial distribution* δ *whose support is not in* $\mathcal{L}$*,*

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\forall n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \notin \mathcal{L}\right) = 1 \ .$$

Equivalently, for every $L \notin \mathcal{L}$*, the set:*

$$I(L) = \{i \in I \text{ such that } \forall c \in C, \text{ if } i = \mathrm{act}(c) \text{ then } \mathcal{B}_1^p(L, c) \notin \mathcal{L}\} \ ,$$

is not empty.

Proof. The equivalence is straightforward from definitions and Proposition 1. On one hand, if I_L is not empty for every $L \notin \mathcal{L}$ then σ consists in playing any action in I_L when the pessimistic belief is L. Conversely, any action played with positive probability by σ with the property above is necessarily in I_L. □

The following proposition provides a fix-point characterization of almost-surely winning supports for player 1.

Proposition 2 (Fix-point characterization of almost-surely winning supports)**.** *Let $\mathcal{L} \subseteq \mathcal{P}(K)$ be a set of supports. Suppose player 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L}$. Then,*

(i) either every support $L \notin \mathcal{L}$ is almost-surely winning for player 1,

(ii) or there exists a set of supports $\mathcal{L}' \subseteq \mathcal{P}(K)$ and a strategy τ for player 2 such that:

(a) $\mathcal{L}'$ is not empty and does not intersect $\mathcal{L}$,

(b) player 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L} \cup \mathcal{L}'$,

(c) for every strategy σ and initial distribution δ with support in $\mathcal{L}'$,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\forall n \in \mathbb{N}, K_n \notin T \mid \forall n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \notin \mathcal{L}\right) > 0 \ . \quad (6)$$

There exists an algorithm running in time doubly-exponential in the size of K for deciding which of cases (i) or (ii) holds. In case (i) holds, the algorithm computes at the same time a pessimistic-belief almost-surely winning strategy for player 1. In case (ii) holds, the algorithm computes at the same time $\mathcal{L}'$ and a finite memory strategy τ with memory $\mathcal{P}(\mathcal{L}' \times K) \backslash \{\emptyset\}$ such that (6) holds for every σ.

The proof of Proposition 2 is based on the notion of $\mathcal{L}$-games.

Definition 4 ($\mathcal{L}$-games)**.** *Let $\mathcal{L}$ be a set of supports such that player 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L}$. For every support $L \notin \mathcal{L}$, let $I(L)$ be the set of actions given by definition 3. The $\mathcal{L}$-game has same actions, transitions and signals than the original partial observation game, only the winning condition changes: player 1 wins if the play reaches a target state and moreover player 1 does not use actions other than I_L whenever his pessimistic belief is L, formally given an initial distribution δ with support L and two strategies σ and τ the winning probability of player 1 is:*

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\exists n \geq 1, K_n \in T \text{ and } \forall n \in \mathbb{N}, I_n \in I(\mathcal{B}_1^p(L, C_1, \ldots, C_n))\right) \ .$$

Actually $\mathcal{L}$-games are special cases of reachability games, as illustrated by the following lemma and its proof, which are based on Theorem 2.

Proposition 3 ($\mathcal{L}$-games)**.** *Let $\mathcal{L}$ be a set of supports such that player 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L}$.*

(i) *In the $\mathcal{L}$-game, every support is either positively winning for player* 1 *or surely winning for player* 2*. The set of surely winning supports for player* 2 *in the $\mathcal{L}$-game contains $\mathcal{L}$. We denote $\mathcal{L}''$ the set of supports that are not in $\mathcal{L}$ but are surely winning for player* 2 *in the $\mathcal{L}$-game.*

(ii) *Suppose $\mathcal{L}''$ is empty. Then every support not in $\mathcal{L}$ is almost-surely winning for player* 1 *both in the $\mathcal{L}$-game and in the original game. Moreover, the strategy σ for player* 1 *which consists in chosing randomly any action in $I(L)$ when its pessimistic belief is L is almost-surely winning in the $\mathcal{L}$-game.*

(iii) *Suppose $\mathcal{L}''$ is* not *empty. Then player* 2 *has a strategy τ for winning surely the $\mathcal{L}$-game from any support in $\mathcal{L}''$, and τ has finite memory $\mathcal{P}(\mathcal{L}' \times K)$.*

(iv) *There is an algorithm running in doubly-exponential time in K for computing $\mathcal{L}''$ and σ or τ.*

Proof. We make the synchronized product $G_{\mathcal{L}}$ of the original game with pessimistic beliefs of player 1. Pessimistic beliefs of player 1 are hidden to both players, signals, actions and transitions remain the same, only an extra sink is added for punishing player 1 whenever the current state is $(l, L) \in K \times \mathcal{P}(K)$ and player 1 plays an action i which is not in $I(L)$.

Applying Theorem 2 to the reachability game $G_{\mathcal{L}}$, we get (i) and (iii).

Now we suppose $\mathcal{L}''$ is empty and prove (ii). According to Theorem 2, any support not in $\mathcal{L}$ is positively winning for player 1 in $G_{\mathcal{L}}$ and moreover the strategy consisting in playing randomly any action is positively winning for player 1. Since playing an action i which is not in $I(L)$ leads immediatly to a non-accepting sink state, the pessimistic belief strategy σ for player 1 which consists in playing randomly any action in $I(L)$ when the pessimistic belief of player 1 is L is positively winning as well.

To prove (ii) it is enough to show that:

$$\sigma \text{ is almost-surely winning for player 1 } . \tag{7}$$

For proving (7), we start with proving that for each $L \notin \mathcal{L}$ there exists $N_L \in \mathbb{N}$ such that for every strategy τ, for every distribution δ with support L,

$$\mathbb{P}_{\delta}^{\sigma,\tau}\left(\exists n \leq N_L, K_n \in T\right) \geq \frac{1}{N_L} \ . \tag{8}$$

We suppose such an N_L does not exist and seek for a contradiction. Suppose for every N there exists τ_N and δ_N such that (8) does not hold. Without loss of generality, we can choose strategies τ_N that are deterministic i.e. $\tau_N : D^* \to J$, and such that δ_N converges to some distribution δ, whose support is included in L. Using Koenig's lemma, it is easy to build a strategy $\tau : D^* \to J$ such that for infinitely many N,

$$\mathbb{P}_{\delta_N}^{\sigma,\tau}\left(\exists n \leq N, K_n \in T\right) < \frac{1}{N} \ .$$

Taking the limit when $N \to \infty$, we get:

$$\mathbb{P}_{\delta}^{\sigma,\tau}\left(\exists n \geq 1, K_n \in T\right) = 0 \ ,$$

which contradicts the fact that σ is positively winning from L, since the support of δ is included in L. This proves the existence of N_L such that (8) holds.

Now we can achieve the proof of (ii). Let $N = \max\{N_L \mid L \notin \mathcal{L}\}$. Then according to (8), when playing σ, every N steps there is probability at least $\frac{1}{N}$ to reach a target state, knowing that a target state was not reached before. Hence there is probability 0 of never reaching a target state. Consequently, σ is almost-surely winning from any support $L \notin \mathcal{L}$. This proves (7) and the last statement of (iii). □

Proof of Proposition 2. Let $\mathcal{L}''$ be the set of supports surely winning for player 2 in the $\mathcal{L}$-game. Let τ_U be the strategy for player 2 playing randomly any action. Let $\mathcal{L}'$ be the set of supports L such that,

$$\forall \sigma, \mathbb{P}_{\delta_L}^{\sigma,\tau_U}\left(\exists n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \in \mathcal{L}'' \cup \mathcal{L}\right) > 0 \ , \tag{9}$$

where δ_L is the uniform distribution on L.

Suppose first that $\mathcal{L}''$ is empty. Since player 1 can enforce his pessimistic beliefs to stay outside $\mathcal{L}$, then $\mathcal{L}'$ is empty as well. Moreover, according to (ii) of Proposition 3, every support not in $\mathcal{L}$ is almost-surely winning for player 1 in the original game, hence we are in case (i) of Proposition 2.

Suppose now that $\mathcal{L}''$ is *not* empty, and let us prove (ii)(a), (ii)(b) and (ii)(c) of Proposition 2. Since $\mathcal{L}'' \subseteq \mathcal{L}'$, then $\mathcal{L}'$ is not empty either, hence (ii)(a). Property (ii)(b) holds by definition (9) of $\mathcal{L}'$.

It remains to prove (ii)(c). According to (iii) of Proposition 3, there exists a strategy τ' for player 2 which is surely winning in the $\mathcal{L}$-game from any support in $\mathcal{L}''$. Let τ be the strategy for player 2 which consists in the following. At each step, player 2 throws a coin. As long as the result is "tail", then player 2 plays randomly any action: he uses σ_U. If the result is "head" then player 2 pick randomly a support $L \in \mathcal{L}''$, forgets all its signals up to now, switches definitively to strategy τ' with initial support L, and stops throwing the coin.

Let us prove that τ guarantees property (6) to hold. Let σ be a strategy of player 1 and δ an initial distribution whose support is in $L \in \mathcal{L}'$. By definition of $\mathcal{L}'$ and τ_U, there exists $c_1, \ldots, c_N$ and a support $L'' \in \mathcal{L}''$ such that $L'' = \mathcal{B}_1^p(L, c_1, \ldots, c_n)$ and

$$\mathbb{P}_{\delta}^{\sigma,\tau_U}\left(C_1 = c_1, \ldots, C_n = c_N\right) > 0 \ .$$

Moreover, since τ_U plays any sequence of actions with positive probability, then,

$$\forall l \in L'', \mathbb{P}_{\delta}^{\sigma,\tau_U}\left(K_n = l, C_1 = c_1, \ldots, C_N = c_N\right) > 0 \ .$$

Now, by definition of τ, there is positive probability that τ plays like τ_U up to stage n hence:

$$\forall l \in L'', \mathbb{P}_{\delta}^{\sigma,\tau}\left(K_n = l, C_1 = c_1, \ldots, C_N = c_N\right) > 0 \ . \tag{10}$$

Moreover, there is positive probability that at stage n, τ switches to strategy τ' in initial state L''. Let δ'' be the uniform distribution on L''. Since τ' is surely winning in the $\mathcal{L}$-game from L'', it guarantees that:

$$\forall \sigma, \mathbb{P}_{\delta''}^{\sigma,\tau'}\left(\forall n \in \mathbb{N}, K_n \notin T \mid \forall n \in \mathbb{N}, I_n \in \mathcal{B}_1^p(L'', C_1, \ldots, C_n)\right) = 1 \ .$$

By definition of pessimistic beliefs,
$\mathcal{B}_1^p(L'', c'_1, \ldots, c'_m) = \mathcal{B}_1^p(L, c_1, \ldots, c_N, c'_1, \ldots, c'_m)$, hence according to (10),

$$\forall \sigma, \mathbb{P}_{\delta}^{\sigma,\tau}\left(\forall n \in \mathbb{N}, K_n \notin T \mid \forall n \geq N, I_n \in (\mathcal{B}_1^p(L, C_1, \ldots, C_n))\right) > 0 \ . \qquad (11)$$

According to the definition of $I(L)$, for every σ and $n \in \mathbb{N}$,

$$\mathbb{P}_{\delta}^{\sigma,\tau_U}\left(\mathcal{B}_1^p(L, C_1, \ldots, C_n, C_{n+1}) \in \mathcal{L} \mid I_n \notin (\mathcal{B}_1^p(L, C_1, \ldots, C_n))\right) > 0 \ ,$$

hence by definition of τ,

$$\mathbb{P}_{\delta}^{\sigma,\tau}\left(\forall n \in \mathbb{N}, I_n \in (\mathcal{B}_1^p(L, C_1, \ldots, C_n)) \mid \forall n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \notin \mathcal{L}\right) > 0 \ ,$$

and together with (11) we get (6), which proves (ii)(c) of Proposition 2.

To achieve the proof of Proposition 2, we have to describe the doubly-exponential algorithm. This algorithm uses the algorithm provided by (iv) in Proposition 3 as its main subprocedure, to obtain $\mathcal{L}''$ and σ or τ'. In case $\mathcal{L}''$ is empty, it simply outputs σ. In case $\mathcal{L}''$ is not empty, it computes $\mathcal{L}'$, which is easy, and outputs strategy τ obtained from strategy τ' as described above, compared to τ', strategy τ requires only one extra memory state. □

The proof of Theorem 3 illustrates how to compose the various finite memory strategies of Proposition 2 to obtain a strategy for player 2 which is positively winning and has finite memory $\mathcal{P}(\mathcal{P}(K) \times K)$.

Proof of Theorem 3. According to Proposition 2, starting with $\mathcal{L}_0 = \emptyset$, there exists a sequence $\mathcal{L}'_0, \mathcal{L}'_1, \ldots, \mathcal{L}'_n$ of disjoint non-empty sets of supports such that for every $m \leq n$,

- if $0 \leq m < M$ then $\mathcal{L}_m = \mathcal{L}'_0 \cup \cdots \cup \mathcal{L}'_{m-1}$, matches case (ii) of Proposition 2. We denote τ_m the corresponding finite memory strategy.
- $\mathcal{L}_M$ matches case (i) of Proposition 2.

Then according to Proposition 2, the set of supports positively winning for player 2 is exactly $\mathcal{L}_M$, and supports that are not in $\mathcal{L}_M$ are almost-surely winning for player 1. This proves the determinacy.

The sequence $\mathcal{L}'_0, \mathcal{L}'_1, \ldots, \mathcal{L}'_n$ is computable in doubly-exponential time, because each application of Proposition 2 involves running the doubly exponential-time algorithm, and the length of the sequence is at most doubly-exponential in K.

The only thing that remains to prove is the existence and computability of a positively winning strategy τ for player 2, with finite memory $\mathcal{P}(\mathcal{P}(K) \times K)$.

Strategy τ consists in playing randomly any action as long as a coin gives result "head". When the coin gives result "tail", then strategy τ chooses randomly an integer $0 \leq m < M$ and a support $L \in \mathcal{L}'_m$ and switches to strategy τ_m. Since each strategy τ_m has memory $\mathcal{P}(\mathcal{L}'_m \times K) \backslash \{\emptyset\}$ and the $\mathcal{L}'_m$ are distincts, strategy τ has memory $\mathcal{P}(\mathcal{P}(K) \times K)$ with $\emptyset$ used as the initial memory state.

We prove that τ is positively winning for player 2 from $\mathcal{L}_M$. Let σ be a strategy for player 1, $L \in \mathcal{L}_M$ and δ an initial distribution with support L. Let m_0 be the smallest index m such that

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\exists n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \in \mathcal{L}'_m\right) > 0 \ .$$

Since $L \in \mathcal{L}_M$ and $\mathcal{L}_M = \bigcup_{m<M} \mathcal{L}'_m$, the set in the definition of m_0 is non-empty and m_0 is well defined. Let $n_0 \in \mathbb{N}$ and $c_1, c_2, \ldots, c_{n_0} \in C^{n_0}$ such that $\mathcal{B}_1^p(L, c_1, \ldots, c_{n_0}) \in \mathcal{L}'_{m_0}$ and

$$\mathbb{P}_\delta^{\sigma,\tau}\left(C_1 = c_1, \ldots, C_{n_0} = c_{n_0} \text{ and } \forall n \leq n_0, K_n \notin T\right) > 0 \ .$$

According to the definition of τ, there is positive probability that τ plays randomly until step n_0, any sequence of actions is played by τ with positive probability, hence according to the definition of pessimistic beliefs, for every state $l \in \mathcal{B}_1^p(L, c_1, \ldots, c_{n_0})$,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(C_1 = c_1, \ldots, C_{n_0} = c_{n_0} \text{ and } \forall n \leq n_0, K_n \notin T \text{ and } K_n = l\right) > 0 \ . \quad (12)$$

According to the definition of τ again, there is positive probability that τ switches to strategy τ_{m_0} at instant n_0, hence according to (12) and to (6) of Proposition 2,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\forall n \in \mathbb{N}, K_n \notin T \mid \forall n \geq n_0, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \notin \mathcal{L}_{m_0}\right) > 0 \ . \quad (13)$$

By definition of m_0 and since $\mathcal{L}_{m_0} = \mathcal{L}'_0 \cup \cdots \cup \mathcal{L}'_{m_0-1}$,

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\forall n \in \mathbb{N}, \mathcal{B}_1^p(L, C_1, \ldots, C_n) \notin \mathcal{L}_{m_0}\right) = 1 \ ,$$

then together with (13),

$$\mathbb{P}_\delta^{\sigma,\tau}\left(\forall n \in \mathbb{N}, K_n \notin T\right) > 0 \ ,$$

which proves that τ is positively winning. □

Conclusion

We considered stochastic reachability games with partial observation on both sides. We established a determinacy result: such a game is either almost-surely winning for player 1, surely winning for player 2 or positively winning for both players. Despite its simplicity, this result is not so easy to prove. Also we gave algorithms for deciding in doubly-exponential time which of the three cases hold.

A natural question is whether these results extend are true for Büchi games as well? The answer is "partially".

One one hand, it is possible to prove that a game is either almost-surely winning for player 1 or positively winning for player 2 and to decide in doubly-exponential time which of the two cases hold. This can be done by techniques almost identical to the ones in this paper.

On the other hand, it was shown recently that the question "does player 1 has a *deterministic* strategy for winning positively a Büchi game?" is undecidable [BBG08], even when player 1 receives no signals and player 2 has only one action. It is quite easy to see that "deterministic" can be removed from this question, without changing its answer. Hence the only hope for solving positive winning for Büchi games is to consider subclasses of partial observation games where the undecidability result fails, an interesting question.

References

[BBG08] Christel Baier, Nathalie Bertrand, and Marcus Größer. On decision problems for probabilistic büchi automata. In *FoSSaCS*, pages 287–301, 2008.

[CDHR07] K. Chatterjee, L. Doyen, T. A. Henzinger, and J.-F. Raskin. Algorithms for omega-regular games of incomplete information. *Logical Methods in Computer Science*, 3(3:4), 2007.

[Hor08] Florian Horn. *Random Games*. PhD thesis, Université Denis-Diderot, 2008.

[Ren00] Jérôme Renault. 2-player repeated games with lack of information on one side and state independent signalling. *Mathematics of Operations Research*, 25:552–572, 2000.