



# Determinacy and Decidability of Reachability Games with Partial Observation on Both Sides

Nathalie Bertrand, Blaise Genest, Hugo Gimbert

## ► To cite this version:

Nathalie Bertrand, Blaise Genest, Hugo Gimbert. Determinacy and Decidability of Reachability Games with Partial Observation on Both Sides. 2008. hal-00341288

HAL Id: hal-00341288

<https://hal.archives-ouvertes.fr/hal-00341288>

Preprint submitted on 24 Nov 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Determinacy and Decidability of Reachability Games with Partial Observation on Both Sides

Nathalie Bertrand, Blaise Genest, Hugo Gimbert

November 24, 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 such a strategy, or both players have strategies that guarantee winning with non-zero probability (positively winning strategies). We give a fix-point algorithm for deciding which of the three cases holds, which can be decided in doubly-exponential time.

## 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 hold. 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  $p : K \times I \times J \rightarrow \mathcal{D}(C \times D \times K)$ , 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 there exists  $\text{act} : C \cup D \rightarrow I \cup J$  such that  $p(c, d, k' \mid k, i, j) > 0 \iff (i = \text{act}(c) \text{ and } j = \text{act}(d))$ . We denote  $p(c, d, l \mid k) = p(c, d, l \mid k, \text{act}(i), \text{act}(j))$ . This way, plays can be described by sequences

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

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

In the usual way, an initial distribution  $\delta$  and two strategies  $\sigma$  and  $\tau$  define a probability measure  $\mathbb{P}_\delta^{\sigma, \tau}(\cdot)$  on the set of infinite plays, equipped with the  $\sigma$ -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  $\gamma_2(\delta, \sigma, \tau) = 1 - \gamma_1(\delta, \sigma, \tau)$ . Player 1 seeks maximizing  $\gamma_1$  while player 2 seeks maximizing  $\gamma_2$ .

## 1.2 Winning almost-surely or positively

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

$$\forall \tau, \gamma_1(\delta, \sigma, \tau) = 1 \quad . \quad (1)$$

*A distribution  $\delta$  is positively winning for player 1 if there exists a strategy  $\sigma$  such that*

$$\forall \tau, \gamma_1(\delta, \sigma, \tau) > 0 \quad . \quad (2)$$

*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  $\sigma$  such that (1) holds for every almost-surely winning distribution then  $\sigma$  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  $\sigma$  with finite memory is described by a finite set  $M$  called the memory, a strategic function  $\sigma_M : M \rightarrow \mathcal{D}(I)$ , an update function  $\text{update}_M : M \times C \rightarrow M$ , an initialization function  $\text{init}_M : \mathcal{P}(K) \rightarrow M$ .*

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

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 is 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  $c \in C$ ,

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

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

Pessimistic beliefs of player 1 are defined by

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

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  $\sigma, \tau$  be strategies for player 1 and 2 and  $\delta$  an initial distribution with support  $L$ . Then for every  $n \in \mathbb{N}$ ,*

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

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

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

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

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

*Proof.* Almost straightforward from the definitions.  $\square$

## 2.2 Belief and pessimistic belief strategies

A strategy  $\sigma$  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  $(L, c) \rightarrow \mathcal{B}_1(L, c)$ ,
3. the strategic function  $\mathcal{P}(K) \rightarrow \mathcal{D}(I)$  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 2.

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  $\mathcal{L}_\infty \subseteq \mathcal{P}(K \setminus T)$  be the greatest fix-point of the monotonic operator  $\Phi : \mathcal{P}(\mathcal{P}(K \setminus T)) \rightarrow \mathcal{P}(\mathcal{P}(K \setminus T))$  defined by:

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

and let  $\sigma_R$  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)  $\sigma_R$  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  $\Phi$  there always exists such an action  $j$ , and this defines a belief-strategy  $\sigma : L \rightarrow j_L$  for player 2. When playing with this strategy, beliefs of player 2 never intersect  $T$  hence according to Proposition 1, against any strategy  $\sigma$  of player 1, the play stays almost-surely in  $K \setminus T$ , hence it stays surely in  $K \setminus T$ .

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  $\mathcal{L}_0 = \mathcal{P}(K \setminus T) \supseteq \mathcal{L}_1 = \Phi(\mathcal{L}_0) \supseteq \mathcal{L}_2 = \Phi(\mathcal{L}_1) \dots$  and  $\mathcal{L}_\infty$  be the limit of this sequence, the greatest fixpoint of  $\Phi$ . We prove that for any support  $L \in \mathcal{P}(K)$ , if  $L \notin \mathcal{L}_\infty$  then:

$$L \text{ is positively winning for player 1 .} \quad (4)$$

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

$$\mathbb{P}_{\delta_L}^\tau (\exists m \in \mathbb{N}, K_m \in T, 2 \leq m \leq n + 1) > 0 . \quad (5)$$

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  $\mathcal{K}_{n+1}$ ,  $L \in \mathcal{L}_n \setminus \Phi(\mathcal{L}_n)$  hence by definition of  $\Phi$ , 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  $\tau$  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,  $\mathbb{P}_{\delta_L}^\tau (K_2 \in T) > 0$ . Otherwise  $\mathcal{B}_2(L, d_j) \in \mathcal{P}(K \setminus T) \setminus \mathcal{L}_n = \mathcal{K}_n$  hence according to the inductive hypothesis  $\mathbb{P}_{\mathcal{B}_2(L, d_j)}^{\tau[d_j]} (\exists m \in \mathbb{N}, 2 \leq m \leq n + 1, K_m \in T) > 0$ . Since player 1 has only one action, by definition of beliefs, for every state  $l \in \mathcal{B}_2(L, d_j)$ ,  $\mathbb{P}_{\delta_L}^\tau (K_2 = l) > 0$ . Together with the previous equation, we obtain

$$\mathbb{P}_{\delta_L}^\tau (\exists m \in \mathbb{N}, 3 \leq m \leq n + 2, K_m \in T) > 0. \text{ This achieves the inductive step.}$$

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  $\Phi$ . since  $\Phi$  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.  $\square$

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 finite memory strategy such that each memory state is a pair of a state and a pessimistic belief of player 1.*

*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 proof of Theorem 3 is based on the following intuition. The easiest way of winning for player 2 is to reach with positive probability a state from where he wins surely. Hence player 1 will try to prevent the play from reaching such surely winning states, in other words player 1 should prevent his pessimistic belief to contain such surely winning states. However, doing so, player 1 may prevent the play to reach a target state: it may hold that player 2 has a strategy for winning positively under the hypothesis that pessimistic beliefs of player 1 never contains surely winning states. This adds new beliefs of player 1 to the collection of pessimistic beliefs that player 1 should avoid. And so on...

For formalizing these intuitions, we make use of  $\mathcal{L}$ -games.

**Definition 3 ( $\mathcal{L}$ -games).** *Let  $\mathcal{L} \subseteq K$  be a collection of supports. The  $\mathcal{L}$ -game associated with  $\mathcal{L}$  is the game with same actions, transitions and signals than the original partial observation game, only the winning condition changes: player 1 loses if either the play never reaches a target state or if at some moment the pessimistic belief of player 1 is in  $\mathcal{L}$  and the play has never visited a target state previously. Formally given an initial distribution  $\delta$  with support  $L$  and two strategies  $\sigma$  and  $\tau$  the winning probability of player 1 is:*

$$\mathbb{P}_{\delta}^{\sigma, \tau} (\exists n \geq 1, K_n \in T \text{ and } \forall m < n, \mathcal{B}_1(L, C_1, \dots, C_m) \notin \mathcal{L}) \ .$$

Actually  $\mathcal{L}$ -games are special cases of reachability games, as shown in the next proposition and its proof.

**Proposition 2.** *In a  $\mathcal{L}$ -game, every support is either positively winning for player 2 or almost-surely winning for player 1. This partition can be computed in time doubly-exponential in  $|K|$ . Player 2 has a positively winning strategy whose states are pairs of states and pessimistic beliefs of player 1. Player 1 has an almost-surely winning pessimistic-belief strategy.*



*Proof.* We define a reachability game  $G_{\mathcal{L}}$  associated with  $\mathcal{L}$  in the following way. Make the synchronized product of the original game with pessimistic beliefs of player 1: each state is a pair  $(k, L)$  with  $k \in K$  and  $L \subseteq K \setminus T$ . Transitions are inherited from the original game except that every state whose second component is in  $\mathcal{L}$  is absorbing. The set of target states is the set of pairs whose first component is in  $T$ . According to Theorem 2 in the reachability game  $G_{\mathcal{L}}$  every state is either positively winning for player 2 or almost-surely winning for player 1. Moreover according to Theorem 2, player 2 has a positively winning belief strategy  $\tau_{\mathcal{L}}$  in  $G_{\mathcal{L}}$  from which it is easy to construct a positively winning strategy in the  $\mathcal{L}$ -game, with finite memory, whose memory states are sets of states of  $G_{\mathcal{L}}$ . Also according to Theorem 2, player 1 has an almost-surely pessimistic belief strategy  $\sigma_{\mathcal{L}}$  in  $G_{\mathcal{L}}$ . Notice that pessimistic beliefs of player 1 in  $G_{\mathcal{L}}$  cannot take all the possible values in  $\mathcal{P}((K \setminus T) \times \mathcal{P}(K \setminus T))$  because intuitively player 1 has perfect knowledge about his own pessimistic beliefs and formally such a pessimistic belief is always of the type  $\cup_{l \in L} \{(l, L)\}$  for some  $L \subseteq K \setminus T$ . As a consequence, it is easy to extract from  $\sigma_{\mathcal{L}}$  a pessimistic belief almost-surely winning strategy in the  $\mathcal{L}$ -game.  $\square$

The heart of the proof of Theorem 3 is based on the two next propositions.

**Proposition 3.** *Let  $\mathcal{L}$  be an upward-closed collection of supports. Suppose that every support in  $\mathcal{L}$  is positively winning for player 2 in the original game.*

*Then any support positively winning for player 2 in the  $\mathcal{L}$ -game is positively winning in the original game as well.*

*If, apart from supports in  $\mathcal{L}$ , there are no supports positively winning for player 2 in the  $\mathcal{L}$ -game, then every support  $L \notin \mathcal{L}$  is almost-surely winning for player 1 in the original game.*

*Proof.* Let  $\mathcal{L}_p$  be the set of supports positively winning in the  $\mathcal{L}$ -game, that are not in  $\mathcal{L}$ .

We start with the case where  $\mathcal{L}_p$  is not empty. Let  $\tau$  be a strategy for player 2 positively winning in the original game. Let  $\tau'$  be a strategy for player 2 positively winning in the  $\mathcal{L}$ -game. Let  $\tau''$  be the following strategy for player 2. Player 2 starts playing totally randomly any action with equal probability. At each step of the play, player 2 throws a dice with three sides to decide whether he should:

- keep playing randomly,
- pick randomly a support  $L \in \mathcal{L}$ , forget the past observations and switch definitively to strategy  $\tau$  with initial support  $L$ ,
- pick randomly a support  $L \in \mathcal{L}_p$ , forget the past observations and switch definitively to strategy  $\tau'$  with initial support  $L$ .

Let us prove that  $\tau''$  is positively winning in the original game, i.e. for every strategy  $\sigma$  and initial distribution  $\delta$  with support  $L \in \mathcal{L}_p$ ,

$$\mathbb{P}_{\delta}^{\sigma, \tau''} (\exists n \geq 1, K_n \in T) < 1 . \quad (6)$$

By definition of  $\tau''$ , there is non-zero probability that the play is consistent with  $\tau'$  i.e.

$$\mathbb{P}_\delta^{\sigma, \tau''} (\forall n \geq 1, J_n = \tau'(L, D_1, \dots, D_{n-1})) > 0 . \quad (7)$$

Since  $\tau'$  is positively winning in the  $\mathcal{L}$ -game,

$$\mathbb{P}_\delta^{\sigma, \tau'} (\exists n \geq 1, K_n \in T \text{ and } \forall m < n, \mathcal{B}_1(L, C_1, \dots, C_m) \notin \mathcal{L}) < 1 . \quad (8)$$

If  $\mathbb{P}_\delta^{\sigma, \tau'} (\exists n \geq 1, K_n \in T) < 1$  then together with (7) this gives (6).

If  $\mathbb{P}_\delta^{\sigma, \tau'} (\exists n \geq 1, K_n \in T) = 1$  then according to (8), there exists  $N \geq 1$  and a pessimistic belief  $B \in \mathcal{L}$  such that:

$$\mathbb{P}_\delta^{\sigma, \tau'} (\mathcal{B}_1(L, C_1, \dots, C_N) = B \text{ and } \forall 1 \leq m \leq N, K_m \notin T) > 0 .$$

Since every sequence of actions is played with positive probability by  $\tau''$ , then:

$$\forall l \in B, \mathbb{P}_\delta^{\sigma, \tau''} (K_N = l \text{ and } \forall 1 \leq m \leq N, K_m \notin T) > 0 . \quad (9)$$

By definition of  $\tau''$ , there is positive probability that  $\tau''$  picks randomly the support  $B \in \mathcal{L}$  and switches to  $\tau$  with initial support  $B$ . By definition,  $\tau$  is positively winning from  $B$  hence there exists  $l \in B$  such that:

$$\forall \sigma', \mathbb{P}_l^{\sigma', \tau} (\forall n \geq 1, K_n \notin T) > 0 ,$$

together with (9) it proves  $\mathbb{P}_\delta^{\sigma, \tau''} (K_N = l \text{ and } \forall m \geq 1, K_m \notin T) > 0$  hence (6).

Now we consider the case where  $\mathcal{L}_p$  is empty. According to Proposition 2, player 1 has a pessimistic belief strategy  $\sigma$  which is almost-surely winning in the  $\mathcal{L}$ -game from every support  $L \notin \mathcal{L}$ . This ensures, for every  $\delta$  whose support is  $L \in \mathcal{L}$ , for every strategy  $\tau$ ,

$$\mathbb{P}_\delta^{\sigma, \tau} (\forall n \geq 1, \mathcal{B}_1^p(L, C_1, \dots, C_n) \notin \mathcal{L} \text{ or } \exists m \leq n, K_m \in T) = 1 . \quad (10)$$

We start with proving for each  $L \notin \mathcal{L}$  there exists  $N_L \in \mathbb{N}$  such that for every strategy  $\tau$ , for every distribution  $\delta$  with support  $L$ ,

$$\mathbb{P}_\delta^{\sigma, \tau} (\exists n \leq N_L, K_n \in T) \geq \frac{1}{2} . \quad (11)$$

We suppose such an  $N_L$  does not exist and seek for a contradiction. Suppose for every  $N$  there exists  $\tau_N$  and  $\delta_N$  such that (11) does not hold. We can suppose  $\tau_N$  is deterministic i.e.  $\tau_N : D^* \rightarrow J$ , and  $\delta_N$  converges to some distribution  $\delta$ , whose support is included in  $L$ . Using Koenig's lemma, it is easy to build a strategy  $\tau : D^* \rightarrow J$  such that for infinitely many  $N$ ,

$$\mathbb{P}_{\delta_N}^{\sigma, \tau} (\exists n \leq N, K_n \in T) \leq \frac{1}{2} .$$

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

$$\mathbb{P}_\delta^{\sigma, \tau} (\exists n \geq 1, K_n \in T) \leq \frac{1}{2} ,$$

which contradicts the fact that  $\sigma$  is almost-surely winning from  $L$ , since the support of  $\delta$  is included in  $L$ . This proves the existence of  $N_L$  such that (11) holds.

Now let  $N = \max\{N_L \mid L \notin \mathcal{L}\}$  and let  $\sigma'$  be the pessimistic belief strategy for player 1 similar to  $\sigma$ , except every  $N$  steps the memory is reset, formally:  $\sigma'(L)(c_1, \dots, c_n) = \sigma(\mathcal{B}_1^p(L, c_1, \dots, c_{(n/N)*N}))(c_{(n/N)*N} \cdots c_n)$ . Then whatever be the strategy played by player 2, according to (10) as long as a target state is not reached, the memory of  $\sigma'$  will stay outside  $\mathcal{L}$ . Then according to (11), when playing  $\sigma'$ , every  $N$  steps there is probability at least  $\frac{1}{2}$  to reach a target state, knowing that it was not reached before, hence there is probability 0 of never reaching a target state. Consequently,  $\sigma$  is almost-surely winning from any support  $L \notin \mathcal{L}$ .  $\square$

Now we can prove Theorem 3.

*Proof of Theorem 3.* Let  $\mathcal{L}_0, \mathcal{L}_1, \dots$  be the sequence defined by  $\mathcal{L}_0 = \emptyset$  and for every  $n \in \mathbb{N}$ ,  $\mathcal{L}_{n+1} \subseteq \mathcal{P}(K)$  is the set of supports positively winning for player 2 in the  $\mathcal{L}_n$ -game. Then  $\mathcal{L}_0 \subseteq \mathcal{L}_1 \subseteq \dots$  and  $\mathcal{P}(\mathcal{P}(K))$  is finite hence there is a limit  $\mathcal{L}_\infty$  to this sequence.

Every  $\mathcal{L}_n$  is upward-closed hence according to Proposition 3, every support in  $\mathcal{L}_\infty$  is positively winning for player 2. Moreover, according to Proposition 2, player 2 has a positively winning strategy with finite memory whose memory states are sets of pairs a state and a pessimistic belief of player 1.

By definition of  $\mathcal{L}_\infty$ , the only support positively winning in the  $\mathcal{L}_\infty$ -game are in  $\mathcal{L}_\infty$ . Hence according to Proposition 3 again, every support not in  $\mathcal{L}_\infty$  is almost-surely winning for player 1. Moreover, according to Proposition 2, player 1 has a pessimistic belief almost-surely winning strategy.

The computation of  $\mathcal{L}_\infty$  can be achieved in doubly-exponential time, because according to Proposition 2 each step can be carried on in time doubly exponential in  $K$  and since the sequence  $(\mathcal{L}_n)_{n \in \mathbb{N}}$  is monotonic its length is at most exponential in  $|K|$ .  $\square$

## 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.