The Frontier of Decidability in Partially Observable Recursive Games

David Auger, Olivier Teytaud

International Journal of Foundations of Computer Science

The Frontier of Decidability in Partially Observable Recursive Games

DAVID AUGER*,***, OLIVIER TEYTAUD*,**
**TAO (INRIA),Lri,CNRS UMR 8623*
bat 490 U. Paris-Sud 91405 Orsay, France
*** OASE Lab, National University of Tainan, Taiwan*
**** Laboratoire PRiSM, Université de Versailles St-Quentin-en-Yvelines, 45 avenue des États-Unis, 78035 Versailles, France*
david.auger@prism.uvsq.fr, olivier.teytaud@inria.fr

The classical decision problem associated with a game is whether a given player has a winning strategy, i.e. some strategy that leads almost surely to a victory, regardless of the other players' strategies. While this problem is relevant for deterministic fully observable games, for a partially observable game the requirement of winning with probability 1 is too strong. In fact, as shown in this paper, a game might be decidable for the simple criterion of almost sure victory, whereas optimal play (even in an approximate sense) is not computable.

We therefore propose another criterion, the decidability of which is equivalent to the computability of approximately optimal play. Then, we show that (i) this criterion is undecidable in the general case, even with deterministic games (no random part in the game), (ii) that it is in the jump 0', and that, even in the stochastic case, (iii) it becomes decidable if we add the requirement that the game halts almost surely whatever maybe the strategies of the players.

Keywords: Game theory; Partial Information.

1. Introduction: partially observable recursive games

We consider two-player stochastic games where payoffs only occur when the game reaches absorbing states, which we call *final states*. The class of games that we consider is quite general and can be also viewed as an extension of the framework introduced by Everett [4], where we add stochasticity in transitions; on the other hand we will only consider games where players always choose their actions from a finite set, as opposed to Everett. We refer to these games as recursive games as in the terminology of Everett to insist on the fact that, contrary to general stochastic games, payoffs only occur at the end of the game.

More precisely, we consider two-player games G described by the following elements:

- a finite set $\mathcal{S}$ of *states* whose cardinal is denoted $n \geq 1$;
- a finite set $\mathcal{A}$ of *actions* whose cardinal is denoted $k \geq 1$;
- a nonempty subset $\mathcal{F}$ of $\mathcal{S}$, the set of *final states* ; with every $f \in \mathcal{F}$ is associated a rational number $r_f \in [-1; 1]$;

- a *transition function* q which associates to every element of (s, a^1, a^2) of $\mathcal{S} \times \mathcal{A}^2$ a rational probability distribution on states;
- an *initial state* $s_0 \in \mathcal{S}$;
- a finite set $\mathcal{O}$, the set of observations;
- two *observation filters*, one for each player, which associate to every element (a^1, a^2, s) of $\mathcal{A}^2 \times \mathcal{S}$ an element of $\mathcal{O}$.

Informally, the game starts in the initial state s_0 and then moves from state to state until a final state is reached. If the current state s is not a final state, both players must choose an action from $\mathcal{A}$: then the transition function q gives a probability distribution $q(a^1, a^2)$ on states, according to which the new state is picked. Finally, when a final state f is reached, the game ends and the first player earns a *score* r_f (and the second $-r_f$ since our games are zero-sum). In case the game never ends, we consider the score to be 0.

To this simple rules is added a layer modelling the partiality of observation. We will consider that every time two actions a^1, a^2 have been chosen and a new state s_{t+1} is consequently reached, both players forget about their decisions and the new state, but receive an observation (or signal) which is in some sense the triple (a^1, a^2, s_{t+1}) distorted by their respective observation filter. When the time comes when a player must choose an action, all the information available on the current play is the time t and the sequence $o_0, o_1, \cdots, o_{t-1}$ of observations received in the past.

This framework for rules and observability is enough to model different levels of randomization and observability, for instance:

- *deterministic* games, in which all transitions probabilities $q(s, a^1, a^2)$ are Dirac measures on a single state;
- games with *full observability*, where both observation filters are the identity map from $\mathcal{A}^2 \times \mathcal{S}$ to itself;
- games with *no observability*, where the observation filter of player i associates to a triple (a^1, a^2, s) the action a^i, i.e. a player only observes his own action and nothing else.
- *one-player* games, where player II has no role to play, i.e. for all states s, actions a^1 for I and a^2, a'^2 for II the probabilities $q(s, a^1, a^2)$ and $q(s, a^1, a'^2)$ are the same. In this case observations and strategies of II play no role hence it can be considered that II has only one strategy.

We now state a few definitions in order to describe the game more formally.

- A *history* of length $t \geq 0$ for a player is a finite sequence of t observations $(o_0, o_1, \cdots, o_{t-1}) \in \mathcal{O}^t$ - for $t = 0$ it is the empty sequence. Hence $\mathcal{O}^t$ is the set of histories of length t, and we denote by $\mathcal{H}$ the union of all histories of any length.

- A *strategy* for a player is then defined as a map from the set of histories to the set of probability distributions on actions. We denote by Σ_i the set of strategies for player i. If all probability distributions defining a strategy are Dirac measures on a single action (i.e. all decisions are based only on past observations and never randomized), the strategy is said to be *pure*, otherwise it is *mixed*.

Given two strategies σ^1, σ^2 for players the issue of the game only depends on random choices due to the probability distributions in strategies and to stochastic transitions in the game. Precisely, with each $t \geq 0$ is associated a *current state* until the game ends. The game behaves as follows :

- initialization : for $t = 0$ the current state s_t is defined as the initial state s_0 and current histories h_0^i of players are empty;
- while the current state is not a final state repeat :
  - choose randomly an action a_t^i for each player $i = 1, 2$ according to the probability distribution $\sigma^i(h_t^i)$;
  - choose randomly the new current state s_{t+1} according to the probability distribution $q(s_t, a_t^1, a_t^2)$ on states;
  - define observations o_t^i as the images by player's respective observability filters of (a_t^1, a_t^2, s_{t+1});
  - let the histories h_{t+1}^i at time $t+1$ be $h_t^i \times o_t^i$;
- let the score $sc(\sigma^1, \sigma^2)$ be r_f, where f is the final current state.

The score is 0 if the game never halts. As we just said, once strategies σ^1, σ^2 have been chosen by the players the issue of the game, and in particular the score $sc(\sigma^1, \sigma^2)$, is random. We are thus interested in maximizing the average score of the first player, regardless of the opponent's strategy, which we call the *value* v^1 of the first player and is defined by

$$\sup_{\sigma^1 \in \Sigma^1} \inf_{\sigma^2 \in \Sigma^2} \mathbb{E} sc(\sigma^1, \sigma^2).$$

If $r_f \in \{0, 1\}$ for all finite states (i.e. the issue is only win / loose), the problem of deciding whether an almost surely winning strategy exists for the first player exactly amounts to the question: does $v^1 = 1$?

Compacity of pure strategies

Since the sets of actions and observations are finite, for all $t \geq 0$ the set $\mathcal{O}^t$ of possible histories of length t is finite. From this easily follows the following lemma which will turn out to be useful.

Lemma 1. *Consider a finite recursive game G with a partial observability setting (general case). For any infinite set S of pure strategies of a player, we can find a*

sequence $(\sigma^t)_{t\geq 0}$ *with values in* S *and a pure strategy* σ^∞ *(possibly not in* S*) such that for all* $t \geq 0$

$$\sigma^\infty_{|t} = \sigma^t_{|t}, \tag{2}$$

where $\sigma_{|t}$ *denotes the restriction of* σ *to* $\mathcal{O}^t$*.*

Proof. This is similar to Koenig's Lemma. Since strategies in S are pure, the possible restrictions $\sigma_{|0}$ of the $\sigma \in S$ to $\mathcal{O}^0$ (histories of length 0, i.e. the empty set) are in finite number. But since S is infinite, there must be an infinite subset of S composed with strategies having the same restriction, i.e. strategies that agree on time step 0. Now in this set S_0, for similar reasons we can can find an infinite subset S_1 of strategies which agree on time step 1 – so that they agree in fact for $t = 0, 1$. We can inductively repeat the process and pick some $\sigma_t \in S_t$. Since σ^t and σ^{t+1} agree on time steps $0 \dots t$, we can *define* the strategy σ^∞ by (2). □

Outline of the paper

Section 2 will show the main undecidability result, namely the undecidability of optimal play in partially observable games with finite state space, even in the deterministic setting. The section also contains the proof that this problem is in 0', the Turing degree of the halting problem. Section 3 then shows that for games which halt almost surely the problem becomes decidable. The conclusion summarizes the paper and illustrates it on existing games; an open problem around the phantom version of the Game of Go is proposed.

2. Undecidability of approximability for deterministic stochastic games with no observability

The classical definition of decidability in games is the existence of a sure win (∃SW, as in Table 1): the question, for which the decidability is investigated, is the existence of a strategy winning with probability 100 %, whatever may be the strategy of the opponent. This makes sense for fully observable games, because for such games, answering this question is sufficient for optimal play. But, for partially observable games, winning with probability 1 is usually impossible, whenever the situation is very good: partially observable games, as well as games with simultaneous actions, involve mixed strategies, introducing stochasticity in the analysis even when the game is deterministic[8]. In particular, there are situations in which no strategy ensures a win with probability 100%; and, yet, there is a win with probability 99% or more. A good alternate decision problem is therefore an approximation of value function, as presented in Table 2.

Theorem 2. *There is an algorithm which, given a finite recursive game* G *with either full observability, partial observability or no observability, builds a finite deterministic recursive game* G' *with the same observability level and where both players have same values in* G *and* G'*.*

($\exists$SW): Existence of an almost sure win
Instance: a game. Question: is there a strategy for winning with probability 1 whatever may be the strategy of the opponent ?

Table 1. The Usual Definition of the decision problem for games. With this definition, decidability holds for 2-player games and there is (Hearn et al, 2009) undecidability for 3-player games (one team of two players, without communication, against a third player).

($VFA_{X,\epsilon}$): Value function approximation for $X \in]0,1[$ and $\epsilon < \min(X, 1-X)$.
Instance: a game with value $\geq X+\epsilon$ or $\leq X-\epsilon$ for the first player. Question: is there a strategy for ensuring a win with probability $\geq X$ whatever may be the strategy of the opponent ?

Table 2. The Value Function Approximation problem for X and ϵ. The decision algorithm is allowed to be wrong if the value of the game is in $]X-\epsilon, X+\epsilon[$.

Proof. We describe the transformation in high level language. First, duplicate states so that each transition probability $q(s, a^1, a^2)$ is a uniform distribution on a finite set of states – this clearly can be done since probabilities are supposed to be rational. From now on, suppose that G satisfies this property.

The new game G' will have the same set $\mathcal{S}'$ of states as G with initial and final states remaining the same. Let n be the number of states in G ; we define the set of actions of G' as $\mathcal{A}' = \mathcal{A} \times \{1, 2, \cdots, n!\}$. Let us describe the transition function q' of G' - since the game is deterministic, we define its values as states (instead of Dirac measures on states). For every state s and couple $((a^1, i)(a^2, j)) \in {\mathcal{A}'}^2$ define

$$q'(s, (a^1, i), (a^2, j)) = s_{i+j[p]}$$

where $s_0, s_1, \cdots, s_{p-1}$ is the set of p states which is the support of the uniform distribution $q(s, a^1, a^2)$. To ensure coherence one has to keep the same labelling of these states for a given couple (a^1, a^2).

Finally the observability framework is the same as in G, with any rule about observations concerning i and j - hence one can ensure that G' has no observability, or full observability, if G is in one of these cases.

To see that G and G' have the same values (or value if G has one) consider for instance an ϵ-optimal strategy σ^1 for player I and let v^1 be the value of this player. We claim that the strategy consisting of playing the product strategy $\sigma^1 \times u$, where u is the strategy consisting of a random uniform choice of i in $\{1, 2, \cdots, n!\}$ ensures $v - \epsilon$.

It simply follows from the fact that for all probability distributions v on $\{1, 2, \cdots, n!\}$, the distribution $u + v$ will be uniform on $\{1, 2, \cdots, n!\}$, so $u + v[k]$ will be uniform on $\{1, 2, \cdots, k\}$ since k divides $n!$. Hence the transition in G' will mimic what happens in G when forgetting about i, j. Conversely, if a strategy in G' ensures a value $v' - \epsilon$, it ensures $v' - \epsilon$ against all strategies of II that play uniformly on j, hence the corresponding strategy in G ensures $v - \epsilon$. □

It is proved in [7] that the problem of approximating the maximum probability of acceptance for a probabilistic finite-state automaton (PFA) is undecidable; this is proved by reduction to known undecidability results on probabilistic finite automata[9]. Classically, one uses for PFA a different formalism than the one we use here for games. It would be pointless to introduce here this notation thus we restate the theorem in our setting. The reader can easily check by the definition given in [7] that with our notation for games, a PFA exactly amounts to a one-player recursive game with stochastic transitions and no observability, where all rewards equal 1. The emptiness problem for PFA is then to decide whether for a given threshold τ the (only) player has a value greater than τ. Furthermore, they prove:

Theorem 3. [7] *Consider $C \in]0, 1[$ and $0 < \delta < \min(C, 1 - C)$. Given a one-player recursive game with stochastic transitions, all rewards equal to 1 and no observability, such that one of the following cases holds:*

(1) either the player has a value at least $C + \delta$;
(2) or the player has a value at most $C - \delta$

it is undecidable to compute which case hold.

From this, using Theorem 2 we directly deduce:

Corollary 4. *For all $X \in]0, 1[$ and all $\epsilon \in]0, \min(X, 1 - X)[$, the approximation value problem $VFA_{X,\epsilon}$ for recursive games with no observation and deterministic transitions is undecidable.*

2.1. *VFA is in the Turing degree $0'$ of the halting problem*

Here we prove that the problem of approximating the value for the first player in deterministic recursive games with partial observability where all rewards are

positive, while being undecidable, belongs to the undecidability class O', jump of the class of decidable problems [10]. More precisely, we show that :

Theorem 5. *There is a Turing machine which, when given the description of a deterministic recursive game with partial observability and positive rewards, and a rational number $c \geq 0$, halts if and only if the first player has a value $v^1 > c$.*

The proof relies on the following lemma. We define sc_t as the expected score, under the additional requirement that all games longer than t time steps have a reward 0. sc_t is the non-asymptotic approximation of sc, and we will see below in which sense it is a reasonably good approximation.

Lemma 6. *Let G be a P.O. recursive game and σ^1 be a strategy for the first player. Denote*

$$v^1(\sigma^1) = \inf_{\sigma^2 \in \Sigma^2} \mathbb{E} sc(\sigma^1, \sigma^2)$$

and

$$v_t^1(\sigma^1) = \inf_{\sigma^2 \in \Sigma^2} \mathbb{E} sc_t(\sigma^1, \sigma^2).$$

We have

$$\lim_{t \to \infty} v_t^1(\sigma^1) = v^1(\sigma^1). \tag{6}$$

Proof. Consider a fixed $\sigma^1 \in \Sigma^1$. We show Eq. 6. Suppose (in order to get a contradiction) that there exists an $\epsilon > 0$ and for infinitely many $t \geq 0$ some strategy σ_t^2 for II such that

$$\mathbb{E}[sc_t(\sigma^1, \sigma_t^2)] \leq v^1 - \epsilon.$$

The strategies σ_t^2 can be supposed pure. Apply then Lemma 1 to obtain a strategy σ^2 which coincides for all $t \geq 0$ with some $\sigma_{t'}^2$, where $t' \geq t$. Since we have

$$\mathbb{E}[sc(\sigma^1, \sigma^2)] = \sum_{t \geq 0} \sum_{F \in \mathcal{F}} \mathbb{P}_{\sigma^1, \sigma^2} (\text{the current state is } F \text{ at time } t) \cdot r_F$$

and the sum of all probabilities in the above sum is at most 1, there exist t_0 such that

$$\mathbb{E}[sc(\sigma^1, \sigma^2)] \leq \sum_{0 \leq t \leq t_0} \sum_{F \in \mathcal{F}} \mathbb{P}_{\sigma^1, \sigma^2} (\text{the current state is } F \text{ at time } t) \cdot r_F + \frac{\epsilon}{2} \tag{9}$$

Now by construction of σ^2 there is some strategy $\sigma_{t'}^2$ with $t' \geq t_0$ which coincides with σ^2 on $t = 0, 1, \cdots, t_0$, so that we can replace σ^2 by $\sigma_{t'}^2$ for the definition of the probability in (9). From this we deduce that

$$\mathbb{E}[sc(\sigma^1, \sigma^2)] \leq \mathbb{E}[sc_{t_0}(\sigma^1, \sigma_{t'}^2)] + \epsilon \leq v^1(\sigma^1) - \frac{\epsilon}{2},$$

which is impossible by definition of $v^1(\sigma^1)$. □

Lemma 7. *Let v^1 be the value of the first player and v_t^1 be the value of the game restricted to the t first time steps (value for the first player). Then*

$$\lim_{t\to\infty} v_t^1 = v^1.$$

Proof. Consider $\epsilon > 0$.

There clearly exist strategies ensuring v_t^1 at time t since the game G_t (restriction of G to t time steps) has finite horizon t and therefore boils down to a finite matrix game. By playing such a strategy in G, since rewards are positive we see that the sequence $(v_t^1)_t$ is non-decreasing and that $v^1 \geq v_t^1$ for all t. Let now σ^1 be a strategy ensuring $v^1 - \epsilon$ for the first player. By Lemma 6 we see that

$$v_t^1 \geq v_t^1(\sigma^1) \geq v^1(\sigma^1) - \epsilon$$

for t big enough, so that

$$v_t^1 \geq v^1 - 2\epsilon.$$

□

With this last lemma the theorem follows quite easily: just build a machine which enumerates the values $(v_t^1)_{t\geq 0}$, and stops if for some t $(v_t^1) > c$.

3. Games that end almost surely

Lots of games, at least those that we really play with our friends and family, have particular rules ensuring that they will end. A notable exception is the game of Go with Japanese rules, in which loops are possible.

If a game has finite horizon, then it is surely decidable to solve $VFA_{X,\delta}$. Another case is when the game can end at every time step (by a transition to a final state with reward zero) with a given probability - then the game ends with probability 1 but with small probability can last for a long time.

In this section, we consider a game which is supposed to end with probability 1, regardless of what players do. We include in the analysis games with stochastic transitions. More precisely, we make the following hypothesis : for all strategies σ^1, σ^2, there almost surely exists t such that the game is over at time t.

Theorem 8. *The value approximation problem VFA is decidable for a stochastic game with reward in $[-1, 1]$ and with partial observability that ends almost surely.*

Remark: The proof is much simpler if we consider only deterministic games; we here consider the general (stochastic) case.

Proof.

We show that, once $\epsilon > 0$ is fixed, a brute-force search algorithm suffices to approximate the value of player 1 up to ϵ. More precisely, the algorithm can be described as follows:

1: $t \leftarrow 0$
2: $p \leftarrow 1$
3: **while** $p > \frac{\epsilon}{2}$ **do**
4: $t \leftarrow t + 1$
5: compute the set $\Sigma^1_{|t}$ of all pure strategies of player 1 from timesteps 0 to t;
6: compute the set $\Sigma^2_{|t}$ of all pure strategies of player 2 from timesteps 0 to t ;
7: for each couple $(\sigma^1_{|t}, \sigma^2_{|t}) \in \Sigma^1_{|t} \times \Sigma^2_{|t}$, compute the probability that the game is not over at the end of timestep t under strategies $(\sigma^1_{|t}, \sigma^2_{|t})$;
8: update p as the *maximum* of all values computed just above;
9: **end while**
10: compute the value of the game restricted to timesteps 0 to t and deliver this value as an ϵ-approximation.

Let us denote by T the random time (depending on the choice of strategies of both players) when the game ends (i.e. the first timestep where a final step is reached), with $T = +\infty$ if the game never ends. By assumption, for all strategies we have $T < +\infty$ almost surely. Note that this is equivalent to require this property for all choices of pure strategies, since mixed strategies are randomizations of pure strategies.

First, let us observe that if the algorithm given above ends, i.e. if condition on line 3 becomes false at some point, then t ends up with a value t_0 such that that for all choices of pure strategies the game will be over before t_0 with probability $1 - \frac{\epsilon}{2}$:

$$\forall \sigma^1, \sigma^2 \text{ strategies, } P(T > t_0) \leq 1 - \frac{\epsilon}{2}. \tag{13}$$

Clearly, if (13) is true for all couples (σ^1, σ^2) of pure strategies, then it is also true for all couples of *mixed* strategies.

Step 1: the brute-force algorithm, if it terminates, finds the correct answer with precision ϵ. So let us suppose for now that the algorithm terminates, hence a t_0 satisfying (13) is found. We will show that the value $v^1_{t_0}$ of the first player for the game stopped at time t_0, which we denote by $G_{|t_0}$ (all plays unfinished at t_0 get a reward 0, as in Lemma 6) is an ϵ-approximation of the value v^1. Below, we will denote the restriction to $G_{|t_0}$ of a strategy σ in G by $\sigma_{|t_0}$, and conversely if σ_{t_0} is a strategy in $G_{|t_0}$ we will denote by σ any strategy consisting of playing according to $\sigma_{|t_0}$ at timesteps prior to t_0 and then playing arbitrarily.

Suppose that there is a mixed strategy $\sigma^1_{|t_0}$ ensuring a value v to the first player in $G_{|t_0}$. Then in G the first player can use any strategy σ^1 consisting of playing $\sigma^1_{|t_0}$ until time t_0 and then playing arbitrarily. If $v \geq 0$, then against any strategy σ^2, player 1 ensures:

$$\begin{aligned}\mathbb{E}[sc(\sigma^1,\sigma^2)] &= P(T\leq t_0)\mathbb{E}[sc(\sigma^1,\sigma^2)\mid T\leq t_0] + P(T>t_0)\mathbb{E}[sc(\sigma^1,\sigma^2)\mid T>t_0]\\ &\geq (1-\frac{\epsilon}{2})v - \frac{\epsilon}{2}\\ &\geq v-\epsilon.\end{aligned}$$

If $v<0$, then we also have

$$\begin{aligned}\mathbb{E}[sc(\sigma^1,\sigma^2)] &= P(T\leq t_0)\mathbb{E}[sc(\sigma^1,\sigma^2)\mid T\leq t_0] + P(T>t_0)\mathbb{E}[sc(\sigma^1,\sigma^2)\mid T>t_0]\\ &\geq v-\frac{\epsilon}{2}\\ &\geq v-\epsilon.\end{aligned}$$

From these two cases, we deduce that $v^1 \geq v^1_{|t_0} - \epsilon$.

On the other hand, for any choice of a strategy σ^1 for the first player, there is a strategy $\sigma^2_{|t_0}$ for the second player such that in $G_{|t_0}$

$$\mathbb{E}[sc_{t_0}(\sigma^1_{|t_0},\sigma^2_{|t_0})] \leq v^1_{|t_0} + \frac{\epsilon}{2}.$$

By playing this strategy σ^2 against σ^1 in G (playing according to $\sigma^2_{|t_0}$ before t_0 and arbitrarily after) it is easy to see by a similar analysis that $v^1 \leq v^1_{|t_0} + \epsilon$.

Step 2: The brute force algorithm always terminates. So our brute-force search algorithm, if it terminates, delivers an ϵ-approximation of v^1. To prove that it always terminates, suppose the opposite: there is an $\epsilon>0$, and pure strategies (σ^1_t,σ^2_t) for all t such that the probability that the game is over at time t_0 if players play accordingly to strategies $(\sigma^1_{t_0},\sigma^2_{t_0})$ is less than $1-\frac{\epsilon}{2}$. By a minor variation on Lemma 1 it is easy to show the existence of pure strategies $(\sigma^1_\infty,\sigma^2_\infty)$ and of a subsequence $(\sigma^1_{k^1_t},\sigma^2_{k^2_t})_{t\geq 0}$ (with $k^1_t,k^2_t\geq t$) of $(\sigma^1_t,\sigma^2_t)_{t\geq 0}$ such that for $i=1,2$ and all $t\geq 0$, strategies $\sigma^i_{k^i_t}$ and σ^i_∞ coincide on times prior to t.

Now if players play according to $(\sigma^1_\infty,\sigma^2_\infty)$, then for all t, the probability that the game is not over at time k_t is at least $\frac{\epsilon}{2}$, which contradicts the assumption that the game ends almost surely for any choice of strategies.

□

4. Conclusion

We have proposed an alternate decision problem for partially observable games (Table 2). This decision problem is based on the approximation of the probability of winning when playing optimally, instead of the existence of a strategy for winning surely.

This criterion:

- directly extends the classical decision problem in matrix games[8],

- is known as more relevant for partially observable games as it is closely related to optimal play (see examples in [12], showing that for some natural situations in real world games there is no move with winning probability 100% but there are strategies with winning probability 99% and other strategies with winning probability 1% - and therefore a refined decision criterion as our Value Function Approximation is required);
- is significantly different from the classical decision problem in papers about decidability in games.

The third point is clear when comparing with published results with the classical decision problem in games ($\exists$SW, recalled in Table 1): whereas [5] shows decidability with 2 players and undecidability with 3 players (two players playing as a team against a third player, without communication inside the team), we show undecidability (Corollary 4) with 2 players.

We have also shown that $VFA_{X,\epsilon}$ is in O', and that $VFA_{X,\epsilon}$ becomes approximable if we consider only games which halt almost surely.

Open problem.

An interesting open question is the game of Go. The game of Go is already famous for various mathematical results:

- PSPACE-completeness of Go ladders[6, 3],
- EXP-completeness of the complete game[11],
- NP-completeness of some Tsume-Go[2].

These results suggest that all restricted forms of Go are as complicated as possible for their category (i.e. NP-completeness when one player has only one reasonable move and the horizon is polynomial as in the Tsume-Go in [2], EXP-completeness in the general case, PSPACE-completeness with polynomial horizon as in the case of ladders).

The game of Go with Japanese rules (in the less usual version of Go termed "Go with Chinese rules", the superko rule forbids loops) has the particularity that it can have loops, and not only in theory; this sometimes happen in real professional games. As phantom-Go is the partially observable variant of Go[1], it's a good candidate for undecidability results for a natural game. To the best of our knowledge, an undecidability result would be the first known case of real-world undecidable game really played by humans.

Acknowledgements

Author #2 is very grateful to the BIRS seminar on Combinatorial Game Theory, to the Dagstuhl seminar on the Theory of Evolutionary Algorithms, and to the Bielefeld seminar on Search Methods. Author #2 is also grateful to National Science

Council (Taiwan) for grant NSC97-2221-E-024-011-MY2 and NSC 99-2923-E-024-003-MY3.

References

[1] T. Cazenave. A phantom-go program. In H. J. van den Herik, S.-C. Hsu, T.-S. Hsu, and H. H. L. M. Donkers, editors, *Proceedings of Advances in Computer Games*, volume 4250 of *Lecture Notes in Computer Science*, pages 120–125. Springer, 2006.

[2] M. Crasmaru. On the complexity of Tsume-Go. In H. Jaap and H. Iida, editors, *Proceedings of the First International Conference on Computer Games*, volume 1558 of *Lecture Notes in Computer Science*, pages 222–231, Tsukuba, Japan, Nov. 1998. Springer.

[3] M. Crasmaru and J. Tromp. Ladders are PSPACE-complete. In *Computers and Games*, pages 241–249, 2000.

[4] H. Everett. *Recursive games*, volume 3 of *Annals of Mathematical Studies AM-39*, pages 47–78. Princeton University Press, 1957.

[5] R. A. Hearn and E. Demaine. *Games, Puzzles, and Computation*. AK Peters, 2009.

[6] D. Lichtenstein and M. Sipser. Go is polynomial-space hard. *J. ACM*, 27(2):393–401, 1980.

[7] O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2):5–34, 2003.

[8] J. Nash. Some games and machines for playing them. Technical Report D-1164, Rand Corporation, 1952.

[9] A. Paz. *Introduction to probabilistic automata*. Academic Press, Inc., Orlando, FL, USA, 1971.

[10] E. L. Post. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society*, 50:284–316, 1944.

[11] J. M. Robson. The complexity of go. In *IFIP Congress*, pages 413–417, 1983.

[12] O. Teytaud. *Artificial Intelligence with parallelism, with Applications to Games, Planning and Optimization*. PhD thesis, TAO, Université Paris-Sud, May 2011. Habilitation thesis.