ELSEVIER

Theoretical Computer Science 157 (1996) 161–183

Theoretical
Computer Science

On the complexity of partially observed Markov decision processes

Dima Burago[a,1], Michel de Rougemont[b], Anatol Slissenko[c,d,e,*]

[a] *Laboratory for Theory of Algorithms, SPIIRAN [2], St. Petersburg, Russia and LRI, Université Paris-Sud, France*

[b] *Laboratoire de Recherche en Informatique, Université Paris-Sud, Bât. 490, F-91405 Orsay, France*

[c] *Université Paris-12, Bât.P3, Informatique, 61, Ave. du Général de Gaulle, 94010 Créteil, France*

[d] *LITP, Institut Blaise Pascal, Paris, France*

[e] *Laboratory for Theory of Algorithms, SPIIRAN [2], St. Petersburg, Russia*

Abstract

In the paper we consider the complexity of constructing optimal policies (strategies) for some type of partially observed Markov decision processes. This particular case of the classical problem deals with finite stationary processes, and can be represented as constructing optimal strategies to reach target vertices from a starting vertex in a graph with colored vertices and probabilistic deviations from an edge chosen to follow. The colors of the visited vertices is the only information available to a strategy. The complexity of Markov decision in the case of perfect information (bijective coloring of vertices) is known and briefly surveyed at the beginning of the paper. For the unobservable case (all the colors are equal) we give an improvement of the result of Papadimitriou and Tsitsiklis, namely we show that the problem of constructing even a very weak approximation to an optimal strategy is NP-hard. Our main results concern the case of a fixed bound on the multiplicity of coloring, that is a case of partially observed processes where some upper bound on the unobservability is supposed. We show that the problem of finding an optimal strategy is still NP-hard, but polytime approximations are possible. Some relations of our results to the Max-Word Problem are also indicated.

1. Introduction

1.1. We consider a particular case of Markov decision processes (e.g. see [17]) from the point of view of computational complexity. This case concerns stationary processes

[*] Correspondence address: Université Paris-12, Bât.P3, Informatique, 61, Ave. du Général de Gaulle, 94010 Créteil, France. E-mail: sliss@litp.ibp.fr. The research of this author was partially supported by DRET contract 91/1061.

[1] Present address: Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: dburago@math.upenn.edu. The research of this author was supported by DRET and Armines contract 92-0171.00.1013.

[2] St. Petersburg Institute for Informatics and Automation of the Academy of Sciences of Russia.

SSDI 0304-3975(95)00158-1

with imperfect information (partially observed) with a finite number of states and actions, and under a concrete cost criterion. Our motivation is, on the whole, standard, i.e. the analysis of situations where the processes entailed by our actions are predictable only with some probability. What is common in these problems is that we consequently make decisions to undertake certain actions that change the state of the system, with a goal to reach some desirable state or to realize some behavior. As neither the exact result of the action nor the current state are known precisely, we are in a situation of twofold uncertainty: we are subjected to probabilistic deviations from planned results, and we get only partial information about the state where we arrive at.

The traditional formalization considers a finite set of states, a finite set of actions (or decisions) permissible at a state, with every action implying a transition of the system to another state with a known probability. The traditional terminology (e.g. see [17]) seems to be too cumbersome for our particular case, so we slightly deviate from that system of notions, giving, however, references. The states can be interpreted as vertices of a graph whose directed edges go from a vertex to all other ones reachable by some action with nonzero probability. In other words, we act on a colored digraph supplied with a function describing the probability to deviate from an edge chosen to go along. A strategy, or a policy, is a function from strings of colors (histories of realizations) to actions. While processing, the strategy traverses vertices, and the color of a reached vertex is the only new information available at this vertex. The problem is to construct a strategy fulfilling some task. One of the simplest tasks is to reach a target vertice from a source vertex with maximum probability.

Our specific motivations go back to robotics (e.g. [6, 8]) and to analysis of some probabilistic models. The first goal was to analyze the *complexity* of constructing strategies optimal in different classes, and as one of the further goals, to look at the complexity of optimal strategies for situations with more diverse uncertainty. Different models of uncertainty (e.g. [4, 18, 14, 16, 7, 19]) remain separated.

1.2. In Section 2 we give the basic notions from the field of Markov decision processes related to the problems under consideration, and then specify the criteria of optimality of strategies interesting from the point of view of our motivations, and make precise some computational aspects. Here we also introduce a type of graphs convenient for describing concrete processes. In this paper, as criterion we use the probability to reach target states from a starting state.

Then in Section 3 the complexity of the case of perfect information (bijective coloring) is briefly surveyed.

In short Section 4 for the case of total uncertainty (unobservability) we strengthen Corollary 2 from [15], and show that even very weak approximations to optimal strategies are NP-hard.

The main results are contained in Section 5 where we treat the case of unobservability bounded by a fixed parameter. In terms of colors this means that the number of vertices of the same color is bounded by the parameter. In other words, the set of states is partitioned into classes, the number of elements in every class is bounded by the parameter, and at any moment of execution of a strategy we know only the class which the actual state belongs to. The parameter, say m, is called the *multiplicity* of coloring. We show that even for $m = 3$ constructing an optimal strategy is NP-hard. But for any m, polytime approximations are possible. Finally, relations with the Max-Word problem are discussed.

2. Main notions

2.1. Uncertainty model. We consider only finite stationary Markov decision processes, e.g. see [17].

Let V be a finite set. Its elements are interpreted as states of a system to control. The set V is supplied with the following additional structure.

(a) $clr : V \rightarrow C$ is the coloring function where C is a finite set. It defines a partition of the states into classes $clr^{-1}(c)$, $c \in C$, which characterizes the uncertainty of determining current state. In traditional terms the coloring defines partial observability of the process.

(b) $\mu : D \times V \times V \rightarrow [0,1]$, where D is a finite set, is such that for all $\lambda \in D$ and $u \in V$

$$\sum_{v \in V} \mu(\lambda, u, v) = 1. \qquad (1)$$

For brevity $\mu(\lambda, uv)$ is used for $\mu(\lambda, u, v)$.

The set D may be interpreted as a set of *actions* (or *moves* or even *decisions*), and the function μ describes the probabilities of results: $\mu(\lambda, uv)$ is the probability to arrive at v from u if the action λ has been made. For example, one can think of D as local names of outgoing edges, and μ gives the probability to follow an edge other than the chosen one.

To avoid some trivialities, we assume $|D|$ is polynomially bounded with respect to $|V|$.

When treated as a part of input of algorithms, μ is supposed to have rational values and to be represented as a usual table of its values.

Thus, using the notations introduced above, the input for the algorithmic problems to analyze is of the form (V, D, C, clr, μ), or (V, D, C, clr, μ, s) when a starting vertex s is fixed. Such an object will be called a *CU-graph* (*CU* stays for *Control* under *Uncertainty*). It is convenient to interpret this structure as a graph with the set of vertices V and edges uv defined by the condition $\exists \lambda \in D\, \mu(\lambda, uv) > 0$, especially for describing examples, and we will use it below.

When treating a decision process as a graph we use the following notations. Let $G = (V, E)$ be a directed graph with vertices V and edges E. Loops, i.e. edges of the form vv, $v \in V$, are permitted. An edge with the *tail* u and the *head* v will be denoted by uv or (u, v). By $OUT(v)$, resp. $IN(v)$, we denote the set of all edges of G outgoing from, resp. incoming to, $v \in V$.

2.2. Strategies. Given a CU-graph G, a *strategy* is a function $\sigma : C^+ \to D$, where we use the notation A^+ for the set of all nonempty strings over alphabet A. So, we consider policies remembering history in the terminology of the theory of Markov decision processes.

Below we define the notion of *universal strategy* that may seem to be a bit cumbersome. To get some intuition, imagine we got lost in a forest or a city, and are seeking to reach some goal. On what information would we base our decision where to go? We would use a map (CU-graph in our context) that, however, does not allow us to recognize directions for sure. Evidently, our decisions depend on our purpose, that is on the criterion to value possible results of our actions (criterion $\mathscr{R}_r$ below), and that may be rather complex and contain, say, a description of regions that we would not cross, or the time at our disposal. We would also take into account the history of our wandering (a string W of colors).

We assume that possible criteria $\mathscr{R}_r$ are encoded as strings r of some language $\mathscr{X}$, concrete criteria will be described in Section 2.4.

Given a class of CU-graphs $\mathscr{G}$ and a class of criteria $\mathscr{X}$, a *universal strategy* (for $\mathscr{G}$ and $\mathscr{X}$) is an algorithm σ whose input is of the form $(G = (V, E, D, C, clr, \mu), r, W)$, where $G \in \mathscr{G}$, $r \in \mathscr{X}$ and $W \in C^+$, and whose output is an action from D. For fixed G and r, a universal strategy σ determines a strategy $\sigma_{G,r} : C^+ \to D$.

Denote by $\mathscr{P}^k_x$ the set of all k-vertex paths in the graph G starting from x, and by $\mathscr{P}^k(T)$ the set of all paths having k vertices and containing a vertex from $T \subseteq V$.

Assume that a starting vertex $s \in V$ is fixed.

The "semantics" of a strategy σ is given by the probability distributions $\boldsymbol{p}^\sigma$ on $\mathscr{P}^k_s$ defined as follows:

$$\boldsymbol{p}^\sigma(v_1 \dots v_{k-1}v_k) = \prod_{i=1}^{k-1} \mu(\sigma(clr(v_1)\dots clr(v_i)), v_iv_{i+1}).$$

Informally speaking, $\boldsymbol{p}^\sigma(P)$ is the probability to follow a given path P of the length k when executing σ.

Actually, we denote by $\boldsymbol{p}^\sigma$ many different probability distributions on different discrete spaces. It will be clear from the context what set $\boldsymbol{p}^\sigma$ is being considered.

One can treat the semantics of a strategy from another point of view, namely, considering a strategy as a family of transformations of the set $\mathscr{D}(V)$ of probability distributions on V. If we have a probabilistic distribution Δ of the initial location then the probability of being at a vertex v after exactly k steps of executing σ is

$$\sigma^k(\Delta)(v) = \sum_{u\in V} \Delta(u) \cdot \sum_{P\in\mathscr{P}^k_u \,\&\, last(P)=v} \boldsymbol{p}^\sigma(P),$$

where $last(P)$ denotes the last character of a string P.

For a fixed string of colors $c_1 \dots c_k$ we define also the conditional probability

$$\sigma_{|c_1\dots c_k}(\Delta)(v) = \sum_{u\in V} \Delta(u) \cdot \sum_{P\in\mathscr{P}^k_u \,\&\, last(P)=v \,\&\, clr(P)=c_1\dots c_k} \boldsymbol{p}^\sigma(P).$$

The semantics of a universal strategy σ is the family of semantics of strategies $\sigma_{G,r}$.

Remark. If we were to follow our motivations one can notice that the history of actions, i.e. the sequence of chosen actions, is an available information, and thus may be included into the argument of σ. One can define the semantics of this type of strategies in a similar way as above. However, it is easy to show that for every strategy of this "generalized" type there exists a strategy that depends only on the colors of visited vertices and determines the same probability distribution on the set of paths.

Remark. As a starting position we could consider not a fixed vertex s but an initial probability distribution on V. However, we can get this distribution by adding a purely probabilistic first move.

2.3. Reliable moves and vertices. Simple graphs. An action (move) $\lambda \in D$ is called *reliable* at v along $e \in OUT(v)$ if $\mu(\lambda, e) = 1$. This edge e will be denoted by $lbl(\lambda, v)$. Such edges will be also called *reliable*. A vertex is said to be *reliable* if every action is reliable at this vertex. A vertex is *random* if the function μ does not depend on action on all the edges outgoing from this vertex.

A CU-graph where every vertex is either random or reliable will be called *simple*. Such graphs are convenient to describe them, and, in particular, they will be used in our examples. We use the following notations, shown in Fig. 1, for our drawings (sometimes we omit redundant information):

(1) reliable vertex colored by color c;

(2) random vertex colored by color c;

(3) reliable edge that corresponds to actions λ and θ;

(4) edge outgoing from random vertex; p is the value of μ (that does not depend on actions);

(5) *trap*, i.e. a vertex where all actions lead back to itself.

One can show that the simple model (even under stronger constraints) is as powerful as the original one.

2.4. Criteria of quality of strategies. General definitions of criteria can be found in texts on Markov decision processes, e.g. [17, 3]. Here, by a *criterion* we mean a function from the set of strategies to real numbers that depends only on the semantics of strategies (i.e. on the probability distribution defined by a strategy). We define below the particular criterion considered in the paper with a generalization studied in a related paper, and just mention a criterion that probably was not considered and that may be of theoretical interest.

(1) *Probability to reach the target in not more than k steps*: Let $T \subseteq V$ be a target set to reach. This criterion, denoted by $\boldsymbol{R}_k^{s,T}(\sigma)$, is defined as the probability to reach

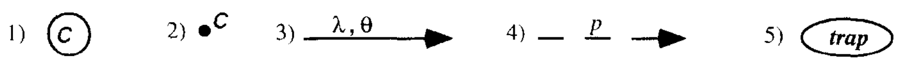


Fig. 1.

any vertex from T starting at s in not more than k steps of execution of σ. When s and T are clear from the context we drop them and use the notation $\boldsymbol{R}_k(\sigma)$.

(2) *Probability of realizing a given behavior*: Let L be a set of paths interpreted as a set of allowed realizations. The criterion $\boldsymbol{R}_k^L(\sigma)$ is the probability to follow only realizations from L (cf. [2] where finite automaton L's are studied).

For criterion $\boldsymbol{R}_k^{s,T}(\sigma)$ one can consider also its limit version

$$\boldsymbol{R}_\infty^{s,T}(\sigma) = \lim_{k\to\infty} \boldsymbol{R}_k^{s,T}(\sigma).$$

Clearly, the criterion $\boldsymbol{R}_k^{s,T}(\sigma)$ is nondecreasing on k, and hence the limit does exist.

We mention also the H^k-criterion, which can be interesting from a theoretical point of view:

$$H^k(\sigma) = \sum_{v\in V} \sigma^k(\delta_s)(v) \cdot \ln \sigma^k(\delta_s)(v),$$

where δ_s is the distribution concentrated in s. To maximize this criterion means to minimize the entropy (i.e. the uncertainty) of the location after k steps of executing σ.

When speaking about a universal strategy, we write $\mathcal{R}_r$ for the criterion encoded by a string r. Hereafter we consider only the criterion $\boldsymbol{R}_k^{s,T}$, $k \in \boldsymbol{N} \cup \{\infty\}$, i.e. the probability to reach T from s in not more than k steps (for natural k) or its limit version ($k = \infty$). We assume that in the input of a universal strategy these criteria are encoded in the form $r = (k,s,T)$ where k is a natural number in the unary notation or the symbol "∞". If instead of a starting vertex we use a starting distribution Δ, we write $\boldsymbol{R}_k^{\Delta,T}$. If values of some of the parameters k, s, T or Δ are clear from the context they will be omitted.

We will denote $\sup\{\boldsymbol{R}_k^{v,T}(\sigma) : \sigma \text{ is a strategy}\}$ by $p_k^{\text{opt}}(v,T)$. Thus, $p_k^{\text{opt}}(v,T)$ is the "optimal" probability to reach T from v in not more than k steps.

2.5. Optimal strategies. A strategy σ is *optimal* with respect to a criterion $\mathcal{R}_r$, or *$\mathcal{R}_r$-optimal* if $\mathcal{R}_r(\sigma') \leqslant \mathcal{R}_r(\sigma)$ for every strategy σ'.

We say that a universal strategy σ (for a class of CU-graphs $\mathcal{G}$ and a class of criteria $\mathcal{X}$) is *optimal* if for every $G \in \mathcal{G}$ and $r \in \mathcal{X}$ the strategy $\sigma_{G,r}$ is $\mathcal{R}_r$-optimal.

Obviously, an $\boldsymbol{R}_k$-optimal strategy does exist for every finite k since the number of strategies different on the first k steps is finite. However, there is no $\boldsymbol{R}_\infty$-optimal strategy in the example described by Fig. 2.

Indeed, the actions after an odd number of steps are made at random vertices, and they do not influence the further behavior. Before we make the action *right* after an even number of steps for the first time, we observe only the color a, and after this action we arrive either at *trap* or at *target*. Thus, any strategy is characterized by one integer $2n$: the number of steps after which we decide to go *right*. One can see that the $\boldsymbol{R}_\infty$-quality of this strategy is $1 - 2^{-n}$.

In a known example given by Fig. 3 the first action of an $\boldsymbol{R}_k$-optimal strategy differs from the first moves of $\boldsymbol{R}_m$-optimal strategies for all $m < k$ with k being exponentially greater than the size of the graph.

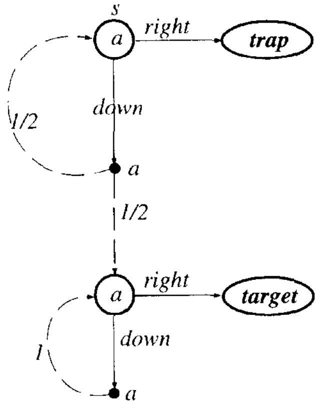


Fig. 2.

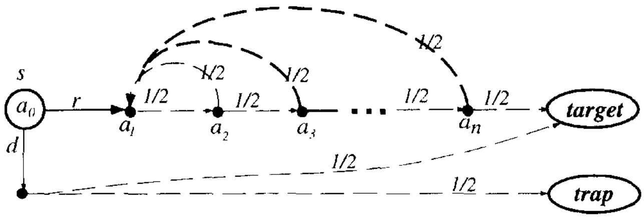


Fig. 3.

Execution of k steps of a strategy σ determines a probabilistic distribution of the current position $\sigma_{|c_1 \dots c_k}(\delta_s)$ (under the condition that colors of visited vertices constitute the sequence $c_1 \dots c_k$), see Section 2.2. In many cases this distribution is the only information needed for strategy. We say that a strategy σ is a ***PT**-strategy* (***PT*** stands for *probability* and *time* dependent) if there exists a function $f : \mathscr{D}(V) \times \boldsymbol{N} \to D$ such that

$$\sigma(c_1 \dots c_k) = f(\sigma_{|c_1 \dots c_k}(\delta_s), k).$$

For example, one can prove the following proposition:

Proposition 1. *For every strategy σ and for every k there exists a **PT**-strategy σ' such that $\boldsymbol{R}_k(\sigma') \geqslant \boldsymbol{R}_k(\sigma)$.*

3. Perfect information (bijective coloring)

3.1. *M*- and *T*-strategies. Here we recall the notion of Markov policies, stationary and nonstationary, and we call them, for brevity, ***M***- and ***T***-strategies.

A strategy σ is called an ***M**-strategy* if it depends on the last color of the argument only, i.e. if there exists a function $\sigma' : C \to D$ such that

$$\forall W: \ \sigma(W) = \sigma'(last(W)).$$

A strategy σ is called a *$\boldsymbol{T}$-strategy* if it depends on the last color and the length of its input only, i.e. there exists a function $\sigma' : C \times \boldsymbol{N} \to D$ such that

$$\forall W:\ \sigma(W) = \sigma'(last(W), |W|),$$

where $|W|$ denotes the length of W.

When speaking about an $\boldsymbol{M}$- or $\boldsymbol{T}$-strategy we assume that its argument is of the form (v) or (v, m), respectively, where $v \in C$ and $m \in \boldsymbol{N}$. The underlying interpretation is $v = last(W)$ and $m = |W|$.

We say that a universal strategy σ is a universal $\boldsymbol{M}$- or $\boldsymbol{T}$-strategy if every strategy $\sigma_{G,r}$ is an $\boldsymbol{M}$- or $\boldsymbol{T}$-strategy, respectively. As above, the input of a universal $\boldsymbol{M}$- or $\boldsymbol{T}$-strategy is assumed to be of the form (G, r, v) or $(G, r, (v, m))$, respectively.

It is clear that $\boldsymbol{M}$-strategies correspond to stationary Markov chains, and $\boldsymbol{T}$-strategies to nonstationary ones. Sufficient information on Markov chains can be found in [13, 9].

In this section we review the case of *bijective coloring* (that is called the case of *perfect information* in the theory of Markov decision processes), and assume $C = V$ and $clr = \mathrm{id}$.

3.2. Optimal M-strategies. The following theorem is known (see [17, Theorem 7.7] or [12]) even for the general case of positive/negative gains. In our case it can be proven by a direct combinatorial argument.

Theorem 1. *For every CU-graph with bijective coloring an $\boldsymbol{R}_\infty^{s,T}$-optimal strategy does exist among $\boldsymbol{M}$-strategies.*

Using the standard technique of the theory of Markov chains, one can show that, given a CU-graph G and an $\boldsymbol{M}$-strategy σ, the value $\boldsymbol{R}_\infty(\sigma)$ can be computed in polytime (in the size of G), see [13, Propositions 3.3.5 and 3.3.8]. Since for every CU-graph the number of $\boldsymbol{M}$-strategies is finite (although exponentially large), this allows us to reformulate the above theorem as follows:

Theorem 2. *For the class of CU-graphs with bijective coloring and for the class of $\boldsymbol{R}_\infty^{s,T}$-criteria, there exists an optimal universal $\boldsymbol{M}$-strategy.*

The following theorem is actually known (see [12, 3.5]), in our case it can be proven rather simply (see [2]).

Theorem 3. *For the class of CU-graphs with bijective coloring and for the class of $\boldsymbol{R}_\infty^{s,T}$-criteria, there exists an optimal universal $\boldsymbol{M}$-strategy with polynomial running time.*

This means that *an optimal $\boldsymbol{M}$-strategy can be computed in polytime.*

3.3. Optimal T-strategies. The following result is known (e.g. see [3]) and easily provable by usual dynamic programming which proceeds backward in time starting from the target set T.

Proposition 2. *For the class of CU-graphs with bijective coloring and the class of* $\boldsymbol{R}_k^{s,T}$*-criteria,* $k \in \boldsymbol{N}$*, there exists a polytime optimal universal* $\boldsymbol{T}$*-strategy.*

We fix now a CU-graph G and a goal set T, and will be interested in the behavior of the $\boldsymbol{R}_k$-optimal $\boldsymbol{T}$-strategy when k grows. It is straightforward that for a $\boldsymbol{T}$-strategy σ the value of the criterion $\boldsymbol{R}_k(\sigma)$ can be computed in time polynomial in k and the size of G. We mean here that σ is represented as a table of its actions up to the kth step, since the later steps do not matter for $\boldsymbol{R}_k$. Indeed, the probability distribution of being at vertices after i steps of executing σ can be computed by multiplying i transition matrices determined by σ in a standard way.

Notice that $p_k^{\text{opt}}(s,T)$ is the value of the $\boldsymbol{R}_k^{s,T}$-criterion for an $\boldsymbol{R}_k^{s,T}$-optimal $\boldsymbol{T}$-strategy, and $p_\infty^{\text{opt}}(s,T)$ is the value of the $\boldsymbol{R}_\infty^{s,T}$-criterion for an optimal $\boldsymbol{M}$-strategy. It is clear that $p_k^{\text{opt}}(s,T)$ converges to $p_\infty^{\text{opt}}(s,T)$ when k tends to infinity. The actions of $\boldsymbol{R}_k^{s,T}$-optimal $\boldsymbol{T}$-strategies also converge to the actions of an optimal $\boldsymbol{M}$-strategy in the following weak sense.

Denote by $D_k(v)$, $v \in V$, the set of actions of all $\boldsymbol{R}_k^{v,T}$-optimal $\boldsymbol{T}$-strategies on the input $(v,1)$. Thus $D_k(v)$ is the set of possible first moves of strategies that lead to T from v in not more than k steps with optimal probability. Then there exists a natural N such that for every $k \geqslant N$ there exists an optimal $\boldsymbol{M}$-strategy σ such that $\sigma(v) \in D_k(v)$ for every vertex v. Such minimum N is not more than exponentially large on the size of G (that can be proved by using estimations on root separation for the characteristic polynomials). However, this convergence actually may be exponentially slow, see Fig. 3.

Moreover, the sequence of sets $D_k(v)$ not necessarily stabilizes when k grows, see Fig. 4.

It is not hard to see that the first action of an $\boldsymbol{R}_k$-optimal $\boldsymbol{T}$-strategy depends on k mod 4: $D_{4l+1}(s) = D_{4l+3}(s) = \{r, l\}$, $D_{4l}(s) = \{r\}$, $D_{4l+2}(s) = \{l\}$.

In this example $D_k(v)$ depends on k (ultimately) periodically. However, this is not the general case:

Theorem 4 (Beauquier et al. [1]). *There exists a CU-graph G such that the sequence $D_k(s)$ is not (ultimately) periodic on k.*

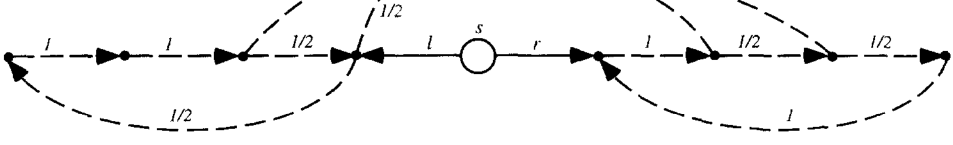


Fig. 4.

4. Unobservable processes

4.1. Statement of results. In this section we consider the class of CU-graphs with one color, i.e. with the set of colors consisting of one element. We will call such graphs *noncolored.* The argument of a strategy is a string of one and the same character, and hence it contains only the information on the number of executed steps in unary notation. Thus, the action of a strategy depends only on time, and we consider a strategy σ as a function $\sigma : \boldsymbol{N} \to D$ that may be represented also by the string of its values $d_1 d_2 \ldots$ So, this is a particular case of the $\boldsymbol{T}$-strategy. The following result was proven in [15] as a corollary of a more general theorem on the complexity of partially observed Markov decision processes.

Theorem 5 (Papadimitriou and Tsitsiklis [15, Corollary 2]). *The following problem is* NP-*complete*: Given *a noncolored CU-graph with k vertices, a starting vertex s, and a set of target vertices T,* recognize *whether there exists a strategy σ with* $\boldsymbol{R}_k^{s,T}(\sigma) = 1$.

The following theorem shows that the problem of computing an optimal strategy in the case of total unobservability does not admit even very weak approximations.

Theorem 6. *The following problem is* NP-*hard:* Given *a noncolored CU-graph with k vertices, a starting vertex s and a set of target vertices T (such that $p_k^{\mathrm{opt}}(s,T)$ equals* 1 *or is less than* $\exp(-\sqrt{k})$), recognize *whether there exists a strategy σ which leads from s to T in k steps with probability not less than* $\exp(-\sqrt{k})$.

We prove Theorem 6 below.

4.2. Notations on the $3SAT$**-problem.** The proof is based on a polytime reduction of the $3SAT$-problem, which is a classical NP-complete problem, see [11]. Let

$$F = \bigwedge_{1 \leqslant i \leqslant m} \bigvee_{1 \leqslant j \leqslant 3} z_{i,j} \tag{2}$$

be a $3CNF$-formula over n variables $x_1, \ldots, x_n$, where $z_{i,j}$ are literals, i.e. elements of the set

$$Z = \{x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n\}, \quad n \leqslant 3m.$$

To visualize the formula we represent it as a table of height 3 and length m; the ith column of the table corresponds to the ith clause (disjunction) of the formula, see Fig.5.

A pair z_1, z_2 of literals is said to be *contrary* iff $z_1 \Leftrightarrow \bar{z}_2$.

A *path* in F is a horizontal path P in the table composed by picking up one literal of every clause, in other words, P is a list of literals of the form

$$z_{1,j_1}, z_{2,j_2}, \ldots, z_{m,j_m}, \quad 1 \leqslant j_i \leqslant 3, \quad 1 \leqslant i \leqslant m,$$

$z_{1,1}$	$z_{2,1}$	....	$z_{m,1}$
$z_{1,2}$	$z_{2,2}$	....	$z_{m,2}$
$z_{1,3}$	$z_{2,3}$	....	$z_{m,3}$

Fig. 5. $3CNF$-formula F as a table.

determined by the sequence (string) $j_1 j_2 \dots j_m$. We interpret such a path as an assignment of its literals by the value ***true*** . If such a path does not contain a contrary (and thus contradictory) pair of literals, it gives a boolean model of the $3CNF$-formula. We call a path in F without contrary pairs an *open* or *satisfying* path, and a path with a contrary pair of literals a *closed* or *contradictory* path.

4.3. Proof of Theorem 6. Given a formula F, we construct the following simple CU-graph H_F.

- The set of actions $D = \{1,2,3\}$.
- Reliable vertices: $\{t, trap, 1,2,\dots,m-1\} \cup (Z \times \{1,2,\dots,m-1\} \times \{1,2\})$, $s = 1$.
- Random vertices: $D \times \{1,2,\dots,m-1\}$.
- Reliable edges for action λ:
  $lbl(\lambda,t) = (t,t)$, $lbl(\lambda, trap) = (trap, trap)$,
  $lbl(\lambda,i) = (i,(\lambda,i))$,
  $lbl(\lambda,(z,i,a)) =$
  case 1.1: $a = 1 \;\&\; z = \overline{z_{i+1,\lambda}} \implies$ edge to *trap*;
  case 1.2: $a = 1 \;\&\; z \neq \overline{z_{i+1,\lambda}} \implies$ edge to $(z,i,2)$;
  case 2.1: $a = 2 \;\&\; i < m-1 \implies$ edge to $(z,i+1,1)$;
  case 2.2: $a = 2 \;\&\; i = m-1 \implies$ edge to t.
- Two random edges from a vertex (λ,i), $i < m-1$: an edge "right" to $i+1$ with probability $(m-i-1)/(m-i)$ *and* an edge "down" to $(z_{i,\lambda}, i, 1)$ with probability $1/(m-i)$; and one edge "down" to $(z_{m-1,\lambda}, m-1, 1)$ from vertex $(\lambda, m-1)$.

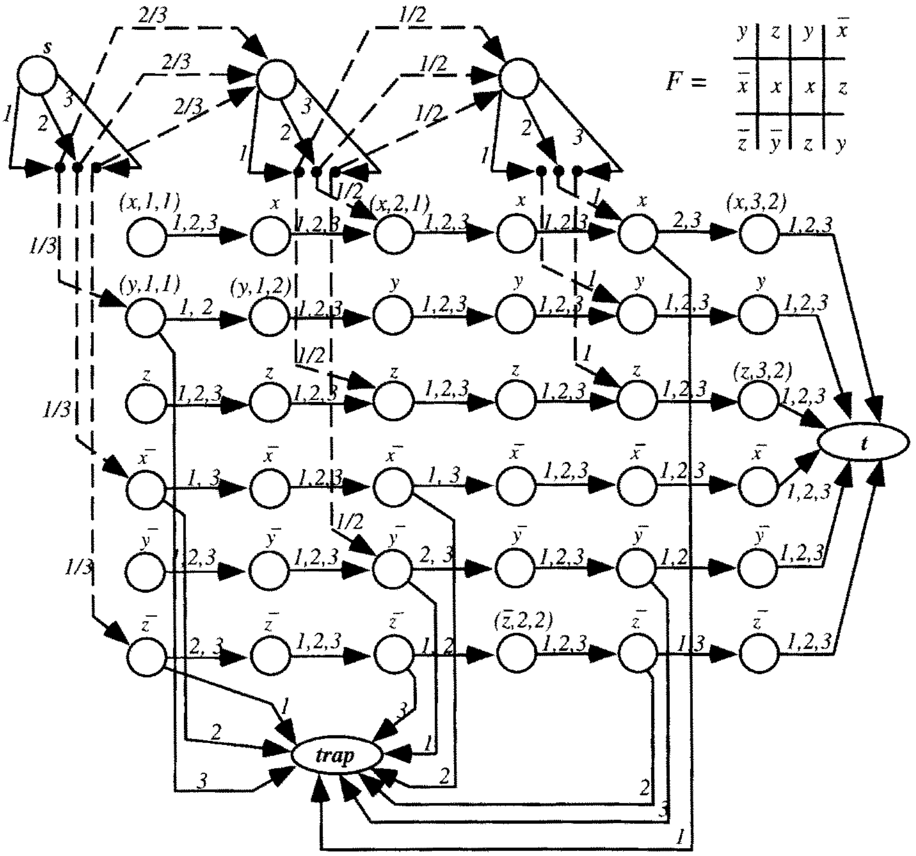
For an example of graph H_F see Fig. 6.

Claim. *If the path* $P = z_{1,d_1} z_{2,d_3} \dots z_{m,d_{2m-1}}$ *determined by a strategy* $\sigma = d_1 d_2 \dots d_{2m}$ *is contradictory then* σ *traverses* H_F *from* s *to* t *with probability not more than* $1 - 1/(m-1)$.

If P *is open then* $\boldsymbol{R}^{s,\{t\}}_{2m+1}(\sigma) = 1$.

Indeed, if P is a contradictory path we have $z_{i,d_{2i-1}} = \bar{z}_{j,d_{2j-1}}$ for some $1 \leqslant i < j \leqslant m$. When executing σ, we follow from the vertex 1 to $D \times \{1\}$ then "right" to 2 and to $D \times \{2\}$ etc. until the first "down" move. The probability to go "down" from the $D \times \{i\}$ is exactly $1/(m-1)$. Then at the $(2j-1)$th step we arrive at the condition of case 1.1 and go to *trap*. Thus, such a strategy traverses H_F successfully from s to t with probability not more than $1 - 1/(m-1)$.

The proof of the second assertion of the claim is similar. □



$$F = \begin{array}{c|c|c|c} y & z & y & \bar{x} \\ \hline \bar{x} & x & x & z \\ \hline \bar{z} & \bar{y} & z & y \end{array}$$

Fig. 6. Some of the reliable vertices of the set $Z \times \{1,2,\ldots,m-1\} \times \{1,2\}$ are labeled by all their coordinates and others are marked by their first coordinate.

Graph H_F contains less than $20m^2$ vertices (for m large enough). We construct now the desired graph $\hat{H}_F$ as follows. Take $20m^4$ copies of H_F denoted by $H_F^1,\ldots,H_F^{20m^4}$. We consider the vertex $1 = s$ of H_F^1 as a *starting* vertex for $\hat{H}_F$ and the vertex t of $H_F^{20m^4}$ as a *target* vertex for $\hat{H}_F$ and redefine the reliable edges from vertices t of H_F^i's. We put a unique reliable edge from t of H_F^i to s of H_F^{i+1} for all $i < 20m^4$. Thus we get sequential composition of the initial graphs. Obviously, $\hat{H}_F$ has not more than $k = 20m^2 \cdot 20m^4 = (20m^3)^2$ vertices.

Consider a strategy $\sigma = d_1 \ldots d_{2m \cdot 20m^4}$ for traversing $\hat{H}_F$ from *starting* to *target*. If the paths $z_{1,d_{(2m+1)i+1}} z_{2,d_{(2m+1)i+3}} \cdots z_{m,d_{(2m+1)i+(2m-1)}}$ are contradictory for all $0 \leqslant i < 20m^4$ then the claim implies that σ traverses each of H_F^{i+1} with probability not greater than $1 - 1/(m-1)$, and hence the total probability for σ to traverse $\hat{H}_F$ from *starting* to *target* is not more than $(1 - 1/(m-1))^{20m^4} < \exp(-20m^3) = \exp(-\sqrt{k})$. Thus for any strategy σ whose probability of success is not less than $\exp(-\sqrt{k})$, one of the paths $z_{1,d_{(2m+1)i+1}} z_{2,d_{(2m+1)i+3}} \cdots z_{m,d_{(2m+1)i+(2m-1)}}$ is open, since k is not more than polynomially greater than the size of F. □

5. Bounded unobservability

It was shown in [15] that the problem of computing an optimal strategy for partially observed processes is PSPACE-complete. We consider here the partially observed processes when this uncertainty concerning observability of states is bounded by a fixed parameter.

5.1. Graphs with a fixed multiplicity of colors. We say that a CU-graph has a *coloring of multiplicity* m if the pre-image of each color contains not more than m vertices. That is, when the color is known, the location is determined up to not more than m vertices. Obviously, bijective coloring corresponds to multiplicity 1. As an intermediate case between bijective coloring and total unobservability we consider CU-graphs with a fixed multiplicity of coloring $m > 1$. The notion of $\boldsymbol{PT}$-strategy gives a reasonable generalization of $\boldsymbol{T}$-strategies for this case, and Proposition 1 shows that in some sense it suffices to consider $\boldsymbol{PT}$-strategies only.

Consider the first nontrivial case $m = 2$, and assume for simplicity that the set of moves D is $\{right, left\}$. For a color $v \in C$ we denote by v^+ and v^- the two vertices of this color. When traversing the graph we actually have just one "hidden parameter" ($+$ or $-$) that influences, however, the probabilities of further transitions. Having arrived at a color v after k steps a $\boldsymbol{PT}$-strategy σ makes its next action basing it on k and on the probabilities p^+ and p^-, $p^+ + p^- = 1$, of being at v^+ and v^-, respectively. Thus, σ induces a partition of $[0, 1]$ into two sets L and R such that σ goes *right* if $p^+ \in R$ and goes *left* otherwise. One might expect that if it is more profitable to go *right* from v^+ and to go *left* from v^- then there should exist some boundary probability p_0 such that if $p^+ \geqslant p_0$ then it is better to go *right*, and if $p^+ \leqslant p_0$ then it is better to go *left*. But this is not the case. In fact, the sets R and L may contain exponentially many (on k) intervals that alternate.

5.2. Complexity of optimization. The following theorem shows that computing an optimal strategy for graphs with a small multiplicity of colors is NP-hard.

Theorem 7. *Every optimal universal strategy for the class of CU-graphs with a coloring of multiplicity* 3 *and the class of* $\boldsymbol{R}_k^{s,T}$*-criteria,* $k \in \boldsymbol{N}$, *is universal for* NP (*with respect to polytime Turing reducibility*). *In simpler words, constructing an optimal strategy for CU-graphs with a multiplicity of coloring* 3 *is* NP-*hard.*

It is an interesting open question related to the Max-Word Problem (see Section 5.3) as to whether the theorem holds for multiplicity 2 and/or for a class of CU-graphs containing only one graph.

We can reformulate Theorem 7 as NP-hardness of recognizing whether there exists a strategy with probability of success not less than a given parameter.

However, contrary to the case of total uncertainty, the problem of computing an optimal strategy for graphs with a small multiplicity of colors does admit a reasonable polytime approximation.

A universal strategy σ is said to be *ε-optimal* if it is optimal up to an additive error ε, i.e.

$$\boldsymbol{R}_k^{s,T}(\sigma_{G,(k,s,T)}) \geqslant \boldsymbol{R}_k^{s,T}(\zeta) - \varepsilon$$

for all G, k, s, T and for all strategies ζ.

We can consider the property to be ε-optimal as a criterion with the value 1 on ε-optimal strategies and 0 otherwise.

Theorem 8. *There exists an optimal universal strategy σ with respect to the criterion of ε-optimality such that for the class of CU-graphs with a fixed multiplicity of coloring m it is computable in time polynomial on the size of input graphs and $1/\varepsilon$. In particular, this means that for a fixed multiplicity of colors optimal strategies admit polytime approximations with an additive error.*

Theorem 8 is interesting in the context of Theorem 7, taken as itself it may seem very natural.

The proofs of Theorems 8 and 7 are given in Sections 5.4 and 5.7.

5.3. Relations with the Max-Word Problem for stochastic matrices. Recall that the Max-Word Problem for stochastic matrices is the following one. Given a set $S = \{M_i\}_{1 \leqslant i \leqslant n}$ of stochastic $(m \times m)$-matrices with rational entries, $M_i = (M^i_{\alpha\beta})$, $\sum_\alpha M^i_{\alpha\beta} = 1$, two (row) vectors V,W with positive coordinates and an integer k in unary notation, the problem is to find a sequence $M_{i_1}, \ldots, M_{i_k}$ which maximizes the product $\langle V, (\prod_{j=1}^k WM_{i_j})\rangle$.

It was shown in [5] that the Max-Word Problem for stochastic matrices is NP-hard as well as its approximation version up to any multiplicative factor.

The Max-Word Problem for stochastic $(m \times m)$-matrices can be reduced to the problem of constructing an optimal strategy for CU-graphs with a coloring of multiplicity m (see (i) below in this subsection). Together with Theorem 8 this implies that for every fixed m the Max-Word Problem for stochastic $(m \times m)$-matrices admits polytime approximations with every additive precision.

The problem of constructing an optimal strategy for CU-graphs with one color can be straightforwardly reduced to the Max-Word Problem for stochastic matrices (see (ii) below in this subsection). With Theorem 6 this implies that the Max-Word Problem for stochastic matrices does not admit polytime approximations within additive precision $\exp(-\sqrt{k})$.

The reductions mentioned above are described as follows:

(i) For an input $M_i = (M^i_{\alpha\beta})$, $1 \leqslant i \leqslant n$, $V = (v_\alpha)$, $W = (w_\beta)$, $1 \leqslant \alpha, \beta \leqslant m$ and k of the Max-Word Problem for stochastic $(m \times m)$-matrices we build a CU-graph with vertices $\boldsymbol{s}$, $\{v_{i,\alpha}\}_{1 \leqslant i \leqslant k+1, 1 \leqslant \alpha \leqslant m}$, $\boldsymbol{t}$ and ***trap***, and with the set of actions $\{1, \ldots, n\}$. Every action leads from $\boldsymbol{s}$ to $v_{1,\alpha}$ with the probability $w_\alpha / \sum_{1 \leqslant \beta \leqslant m} w_\beta$ and from $v_{k+1,\beta}$ to $\boldsymbol{t}$ with the probability $v_\beta / \sum_{1 \leqslant \beta \leqslant m} v_\beta$. An action i leads from $v_{j,\alpha}$ to $v_{j+1,\beta}$ with the probability $M^i_{\alpha\beta}$.

A simple consideration shows that the probability of success of a strategy σ which makes the actions $(i_0 i_1 \dots i_k i_{k+1})$ is

$$\frac{1}{\sum_{1 \leqslant \beta \leqslant m} w_\beta} \cdot \left\langle V, W\left(\prod_{j=1}^{k} M_{i_j}\right)\right\rangle .$$

(ii) For a CU-graph with one color and the set of vertices $\{v_1 = \boldsymbol{s}, v_2, \dots, v_m = \boldsymbol{t}\}$ and the set of actions $\{d_1, \dots, d_n\}$ the problem of computing an optimal strategy to reach $\boldsymbol{t}$ from $\boldsymbol{s}$ in k steps is equivalent to the Max-Word Problem for stochastic $(m \times m)$-matrices with the input $M_i = (\mu(d_i, v_\alpha v_\beta))_{\alpha,\beta}$, $W = (1, 0, \dots, 0)$, $V = (0, \dots, 0, 1)$, k.

Remark. Approximabilities with additive and multiplicative errors are equivalent unless the value of an optimization problem under consideration is more than polynomially large or small. So, this difference occurs when either the value under approximation or its inverse are too small.

5.4. Proof of Theorem 8. The proof shows that the partially observed problem is smooth enough, and it may look tedious as compared with the underlying ideas which are usual in the theory of Markov decision processes.

Enumerate the vertices of the graph G by 2 indices i and α such that the first one is a color, so $V = \{v_{i,\alpha}:\ i = 1, \dots, n,\quad \alpha = 1, \dots, m\}$.

We supply $\boldsymbol{R}^m$ with l^1 metric $\|(x_i)_i - (y_i)_i\| = \sum_i |x_i - y_i|$, and will consider Lipschitz property with respect to this metric.

A point of the simplex S (in $\boldsymbol{R}^m$) defined by the inequalities

$$\sum_{i=1}^{m} x_i = 1, \quad x_i \geqslant 0$$

can be treated as a distribution of probabilities over the set $\{v_{i,1}, \dots, v_{i,m}\}$ of vertices of color i.

Let $P^i = (p_1^i, \dots, p_m^i)$ be this probability distribution, i.e. $P^i(v_{jk}) = p_k \delta_{ij}$, where δ_{ij} is Kronecker's delta.

Let $F_{N,i}(P^i)$ be the probability to reach T starting with the distribution P^i in not more than N steps by an optimal strategy.

Lemma 1. *All $F_{N,i}$ are Lipschitz-1 functions, i.e. $|F_{N,i}(P) - F_{N,i}(Q)| \leqslant \|P - Q\|$ for $P, Q \in S$.*

Proof. Extend the functions $F_{N,k}$ onto the points

$$P \in \tilde{S} = \left\{(p_1, \dots, p_m):\ \sum_{i=1}^{m} p_i \leqslant 1\ \&\ p_i \geqslant 0\right\}$$

in the following way. We append a new trap to our graph, and treat $P^k \in \tilde{S}$ as the probability distribution of being at v_{jl} with the probability $p_l \delta_{kj}$ and at the new trap with the probability $1 - \sum_{i=1}^{m} p_i$. Now the function $F_{N,k}$ is defined in all the points of the simplex $\tilde{S}$, again as the optimal probability to reach the target starting with the distribution P^k.

To verify the Lipschitz property of $F_{N,k}$ consider 2 points $P, Q \in S$. Let $d = \|P-Q\|$. Then for some vectors $A_i = (0,\dots,0,a_i,0,\dots,0)$ with the only nonzero ith coordinate we have $\sum |a_i| = d$, $Q = P + \sum_{1 \leqslant i \leqslant m} A_i$. It suffices to check $|F_{N,k}(P+A) - F_{N,k}(P)| \leqslant a$ where $A = (0,\dots,0,a,0,\dots,0)$, $a > 0$ occupies the ith coordinate of A and P, $P + A \in \tilde{S}$.

It is clear that $F_{N,k}(P+A) - F_{N,k}(P) \geqslant 0$. Reasoning by contradiction assume that

$$|F_{N,k}(P+A) - F_{N,k}(P)| > a.$$

Then for some strategy σ we have $\boldsymbol{R}_N^{(P+A)^k}(\sigma) > F_{N,k}(P) + a$. On the other hand (recall that $\mathscr{P}^k(T)$ is the set of all k-vertex paths containing a vertex from T),

$$\begin{aligned}\boldsymbol{R}_N^{(P+A)^k}(\sigma) &= \sum_{w_1 \dots w_N \in \mathscr{P}^N(T)} (P+A)^k(w_1) \cdot p^\sigma(w_1 w_2 \dots w_N)\\ &= \sum_{w_1 \dots w_N \in \mathscr{P}^N(T)} P^k(w_1) \cdot p^\sigma(w_1 w_2 \dots w_N)\\ &\qquad + \sum_{v_{ki} w_2 \dots w_N \in \mathscr{P}^N(T)} a \cdot p^\sigma(v_{ki} w_2 \dots w_N)\\ &= \boldsymbol{R}_N^{P^k}(\sigma) + \sum_{v_{ki} w_2 \dots w_N \in \mathscr{P}^N(T)} a \cdot p^\sigma(v_{ki} w_2 \dots w_N) \leqslant F_{N,k}(P) + a,\end{aligned}$$

which is a contradiction. □

The family of functions $F_{N,i}$ satisfies the following recurrent system of equations:

$$F_{N,i}(p_1,\dots,p_m) = \max_{d \in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot F_{N-1,j}(T_j^{i,d}(P)), \tag{3}$$

where $q_j^{i,d}(P)$ is the probability to arrive at the color j starting with distribution P^i by the action d and $T_j^{i,d}(P)$ is the conditional distribution on the vertices of the color j if this color has been observed after the move d from the distribution P^i. More formally,

$$q_j^{i,d}(P) = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{m} p_\alpha \cdot \mu(d, v_{i,\alpha}, v_{j,\beta}) \tag{4}$$

and

$$(T_j^{i,d}(P))_h = \frac{1}{q_j^{i,d}(P)} \cdot \sum_{\alpha=1}^{m} p_\alpha \cdot \mu(d, v_{i,\alpha}, v_{j,h}). \tag{5}$$

Notice that $q_j^{i,d}(P) \geqslant 0$ and

$$\sum_{j=1}^{m} q_j^{i,d}(P) = 1. \tag{6}$$

5.5. Let $\delta = \varepsilon/2K$, where K is the number of steps and ε is a chosen precision. Let M be the smallest integer greater than $1/\delta$.

We subdivide S into M^{m-1} equal simplices by hyperplanes parallel to the faces of S.

Consider the class $\mathscr{F}$ of continuous functions on S whose restriction onto every tiny simplex of our partition is linear.

For a function f we denote by f^* the unique function from F that coincides with f on all vertices of the simplices of the partition. It is clear that $\sup_S |f - f^*| \leqslant \delta$ for every Lipschitz-1 function f.

5.6. Algorithm. For constructing our strategy we, firstly, define recursively the functions $\tilde F_{N,i} : S \longrightarrow \boldsymbol{R}^m$, $N \geqslant 0$, and $d_{N,i} : S \longrightarrow D$, $n \geqslant 1$.

$N = 0$: $\tilde F_{N,i}(P) = P^i(T \cap \{v_{i,1}, \dots, v_{i,m}\})$ where $P(\emptyset) = 0$.

$N > 0$: Put

$$\hat F_{N,i}(p_1, \dots, p_m) = \max_{d \in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot \tilde F_{N-1,j}(T_j^{i,d}(P)), \quad \tilde F_{N,i} = \hat F^*_{N,i}, \tag{7}$$

and put $d_{N,i}(P)$ to be an element of D maximizing the right-hand side of (7).

Before the description of the desired strategy σ we prove Claims 1–3.

Claim 1. *$F_{0,k} = \tilde F_{0,k}$ for all k.*

Proof. By the definition. □

Claim 2. *$|F_{1,k} - \tilde F_{1,k}| \leqslant \delta$ for all k.*

Proof. We have $\hat F_{1,k} = F_{1,k}$ because both the functions are defined by the same equations as given by Claim 1. Hence $\tilde F_{1,k} = F^*_{1,k}$, and thus $|F_{1,k} - \tilde F_{1,k}| \leqslant \delta$ since $F_{1,k}$ is Lipschitz-1 (Lemma 1). □

Claim 3. *$|F_{N,k} - \tilde F_{N,k}| \leqslant N\delta$ for all k, N.*

Proof (*Induction on N*). As the base of the induction we use Claim 1. Suppose the inequalities are valid for $N - 1$:

$$|F_{N-1,k} - \tilde F_{N-1,k}| \leqslant (N-1)\delta. \tag{8}$$

Consider a point $P = (p_1, \dots, p_m)$. The inequality (8) implies that for some ζ

$$\tilde F_{N-1,k}(P) = F_{N-1,k}(P) + \zeta, \quad |\zeta| \leqslant (N-1)\delta. \tag{9}$$

By definition, we have

$$F_{N,i}(P) = \max_{d \in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot F_{N-1,j}(T_j^{i,d}(P)), \tag{10}$$

$$\hat F_{N,i}(P) = \max_{d \in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot \tilde F_{N-1,j}(T_j^{i,d}(P)). \tag{11}$$

From these equations and (9) we get

$$
\begin{aligned}
&|\hat{F}_{N,i}(P) - F_{N,i}(P)| \\
&= \left| \max_{d\in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot (F_{N-1,j}(T_j^{i,d}(P)) + \zeta) - F_{N,i}(P) \right| \\
&= \left| \max_{d\in D} \left\{ \sum_{j=1}^{n} q_j^{i,d}(P) \cdot F_{N-1,j}(T_j^{i,d}(P)) + \sum_{j=1}^{n} q_j^{i,d}(P) \cdot \zeta \right\} \right. \\
&\quad \left. - \max_{d\in D} \sum_{j=1}^{n} q_j^{i,d}(P) \cdot F_{N-1,j}(T_j^{i,d}(P)) \right| \qquad (12) \\
&\leqslant \left| \zeta \cdot \sum_{j=1}^{n} q_j^{i,d}(P) \right| \leqslant (N-1)\delta \qquad (13)
\end{aligned}
$$

since the coefficients $q_j^{i,d}(P)$ are nonnegative with the sum equal to 1, see (6). Hence,

$$|F_{N,k} - \hat{F}_{N,k}| \leqslant (N-1)\delta. \qquad (14)$$

The point $P = (p_1, \dots, p_m)$ lies in a tiny simplex of our partition, let it be a simplex with vertices $X_1, \dots, X_m$. Then $P = \sum_{1 \leqslant i \leqslant m} \beta_i \cdot X_i$ for some nonnegative β_i, $\sum \beta_i = 1$.

Since $F_{N,k}$ is Lipschitz-1 and the diameter of our tiny simplex is not greater than δ we have $|F_{N,k}(P) - F_{N,k}(X_i)| \leqslant \delta$ for all i. Adding these inequalities with the coefficients β_i we get

$$|F_{N,k}(P) - \sum \beta_i \cdot F_{N,k}(X_i)| \leqslant \delta. \qquad (15)$$

On the other hand,

$$\tilde{F}_{N,k}(P) = \hat{F}^*_{N,k}(P) = \sum \beta_i \cdot \hat{F}_{N,k}(X_i). \qquad (16)$$

Together with (15) and (14) this gives the required inequality $|F_{N,k}(P) - \tilde{F}_{N,k}(P)| \leqslant N\delta$, since the coefficients β_i are nonnegative with the sum equal to one. □

Now we describe our *strategy* σ. Firstly, it computes and stores all the functions $\tilde{F}_{N,i}$, $0 \leqslant N \leqslant K$, $1 \leqslant i \leqslant m$, as tables of their values at the vertices of our partition. This can be done in polytime. After that for every $P \in S$ the value of the function $d_{N,i}(P)$ is computed in polytime due to (7) by trying all the $d \in D$. For a string of colors $W = c_1 \dots c_N$ the strategy computes the probability distribution of being at vertices of the color c_N. This distribution is represented as a point P of S. Then the action to make is defined by $\sigma(W) = d_{N,c_N}(P)$.

Claim 4. *$|\boldsymbol{R}_N^{P^i}(\sigma) - \tilde{F}_{N,i}(P)| \leqslant N\delta$ for all i.*

Proof. Similar to the proof of Claim 3 using the fact that $\boldsymbol{R}_N^{P^i}$ is Lipschitz-1 on the argument P that can be shown as in Lemma 1. □

Claim 4 together with Claim 3 immediately imply

$$\boldsymbol{R}_K^{P^i}(\sigma) \geqslant F_{K,i}(P^i) - \varepsilon,$$

which completes the proof of Theorem 8.

5.7. Proof of Theorem 7. Our proof is based on a reduction of the *Partition Problem* [10], A3.2: *Given* a set $\{z_a\}_{a\in A}$ of natural numbers indexed by natural numbers from A, *to find* whether there exists a subset $A' \subset A$ such that $\sum_{a\in A'} z_a = \sum_{a\in A\setminus A'} z_a$. If such a subset A' exists we say that the instance of the problem *admits* a partition.

As in the proof of Theorem 8 we treat the distributions of probabilities as points of the appropriate simplex and vice versa.

For a given instance of the Partition Problem represented by a set $\{z_a\}_{a\in A}$ we construct a CU-graph G in the following way.

Let $k = |A|$, $p = \sum_{i\in A} z_i$ and $\alpha_i = \pi z_i/p$. Without loss of generality we can assume that $\alpha_i < \pi/2$. Denote by $\hat{R}^i$ the matrix of rotation in $\boldsymbol{R}^3$ with the axis $x = y = z$ and the angle α_i, and by $\hat{H}^i$ the (3×3)-matrix with the eigenvectors $(1,1,1)$, $(1,0,-1)$ and $(1,-2,1)$ and the eigenvalues 1, $e^{-c\alpha_i}$ and $e^{-c\alpha_i}$, where c is a constant that guarantees the elements of the matrices $\hat{M}^i$ defined below being positive. (Recall that the positiveness of elements of a matrix M is equivalent to saying that M maps the positive quadrant into itself.)

Let $\hat{M}^i = \hat{R}^i \cdot \hat{H}^i = \hat{H}^i \cdot \hat{R}^i$.

The graph G is constituted by (the notations are of the same type as in the proof of Theorem 8):

- the vertices: $V = \{v_{i,\alpha}:\ i = 1,\dots,k+1,\quad \alpha = 1,2,3\} \cup \{\boldsymbol{t}\} \cup \{\boldsymbol{trap}\}$, $s = v_{1,1}$;
- the edges go from every vertex $v_{i,\alpha}$ to all vertices $v_{i-1,\alpha}$ with the exception of the last layer with $i = k+1$ from where there are edges to both $\boldsymbol{t}$ and $\boldsymbol{trap}$;
- the set of actions: $D = \{skip, take\}$;
- the function of deviations:

$$\mu(skip, v_{i,\alpha}v_{i+1,\alpha}) = 1, \mu(take, v_{i,\alpha}v_{i+1,\beta}) = \hat{M}^i_{\alpha\beta}, i \neq k+1,$$

$$\mu(skip, v_{k+1,\alpha}\boldsymbol{t}) = \mu(take, v_{k+1,\alpha}\boldsymbol{t}) = l_\gamma,$$

$$\mu(skip, v_{k+1,\alpha}\boldsymbol{trap}) = \mu(take, v_{k+1,\alpha}\boldsymbol{trap}) = 1 - l_\alpha,$$

where l_α will be chosen later.

5.8. For every realization of a strategy σ up to the $(k+1)$th step the observed sequence of colors is $1,2,\dots,k+1$, so a strategy is determined by a sequence of its actions $d_1 \dots d_k$ since the last action does not matter.

After k steps of executing σ the probability distribution of being in vertices $v_{k+1,\alpha}$ is P^{k+1} where $P = \prod_{d_i=take}(1,0,0)\hat{M}^i$. (Recall that we continue to use the notations for P^{k+1} of the previous proof.)

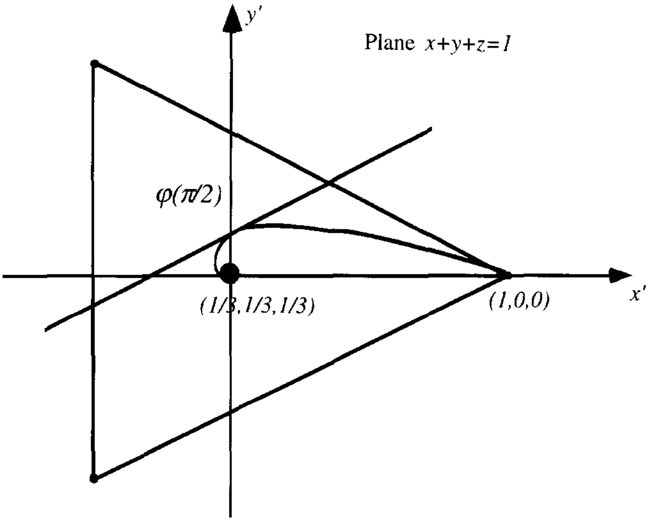


Fig. 7. Geometric interpretation.

To deal with the distribution P^{k+1} we use the following geometric interpretation (see Fig. 7).

Clearly, all our matrices $\hat{R}^i$, $\hat{H}^i$ and $\hat{M}^i$ preserve the plane $x+y+z=1$. Consider the restrictions R^i, H^i and M^i of these matrices onto this plane. The matrices R^i are rotations with angles α_i, and H^i are homotheties with coefficients $e^{-c\alpha_i}$.

Supply the plane $x+y+z=1$ with Cartesian coordinates (x',y') centered at $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$, with the x'-axis containing $(1,0,0)$. Consider the logarithmic spiral $\varphi(t) = e^{-ct}(\cos t, \sin t)$, the parameter t can be taken as the coordinate of a point on the spiral. One can see that our matrices $\hat{M}^i$ preserve the spiral, and being restricted on the spiral they act by adding α_i to the coordinate t. Thus, the point $P = \prod_{d_i=take}(1,0,0)\hat{M}^i$ lies on the spiral and has the coordinate $t = \sum_{\{i:d_i=take\}} \alpha_i$.

Now choose a linear function $L : \boldsymbol{R}^3 \to \boldsymbol{R}$, $L(x) = \langle l,x\rangle$, $l = (l_i) \in \boldsymbol{R}^3$, $0 < l_i \leqslant 1$, such that the point $\varphi(\pi/2)$ maximizes L on the spiral. This can be done along the following lines.

Let T be a tangent vector to our spiral $\varphi(t)$ at the point $\varphi(\pi/2)$. Take a vector l such that $\langle l,T\rangle = 0$, and $L(\varphi(\pi/2)) > L(\frac{1}{3},\frac{1}{3},\frac{1}{3})$.

This vector l can be chosen by a small rotation of vector $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ around T.

We use the coordinates l_α as transition probabilities to arrive at $\boldsymbol{t}$ from the vertices $v_{k+1,\alpha}$. Thus the probability of success of σ is $L(P)$. Notice that $\sum \alpha_i = \pi$. So, if a subset of A with the desired property does exist then every optimal strategy has the sum $\sum_{\{i:d_i=take\}} \alpha_i = \pi/2$, and thus, $\sum_{\{i:d_i=take\}} z_i = p/2$.

5.9. The construction above does not take into consideration the rationality of the probabilities of deviations. For this reason we take appropriate rational approximations to the values defined above.

Suppose that the set A admits a partition. Denote it by $\bar{A}$.

Now we replace the values of the function μ by some rational approximations with a polynomial number of digits. We show that every optimal strategy for this CU-graph also provides us with a partition of the set A.

Assume that $\sum z_i \leqslant \mathrm{e}^n$ and $k \leqslant n$.

To make necessary estimates we need the following inequality:

Lemma 2. *For a constant $\theta > 0$*

$$L\left(\varphi\left(\frac{\pi}{2}\right)\right) - L\left(\varphi\left(\sum_{i\in A'} \alpha_i\right)\right) \geqslant \theta \mathrm{e}^{-n}$$

whenever $\sum_{i\in A'} z_i \neq \sum_{i\in A\setminus A'} z_i$.

Proof. Consider the function $\psi : [0, 2\pi] \to \boldsymbol{R}$, defined by $\psi(t) = L(\varphi(\pi/2) - \varphi(t))$. The inequality to prove can be rewritten as

$$\psi\left(\frac{\pi}{2} - r\right) \geqslant \theta \mathrm{e}^{-n},$$

where $r = \pi/2 - \sum_{i\in A'} \alpha_i$. The bound on $\sum z_i$ implies that $r \geqslant \mathrm{e}^{-n}$.

Clearly,

(i) $\psi(\pi/2) = 0$ and $\psi(t) \neq 0$ for $t \neq \pi/2$;

(ii) $\psi'(\pi/2) = L(-\varphi'(\pi/2)) = q > 0$ (direct computation). Thus, using Taylor expansion, we can state that for some absolute constants $\varepsilon > 0$ and $\eta > 0$ the inequality

$$\left|\psi\left(\frac{\pi}{2} - r\right)\right| \geqslant \left|\psi\left(\frac{\pi}{2}\right) - \frac{1}{2}\psi'\left(\frac{\pi}{2}\right) r\right| - \eta |r|^2$$

holds for all $0 \leqslant |r| \leqslant \varepsilon$ and thus we have

$$\left|\psi\left(\frac{\pi}{2} - r\right)\right| \geqslant \frac{q}{4}|r|$$

for all $0 \leqslant |r| \leqslant \varepsilon$.

If $|r| \geqslant \varepsilon$ then for some absolute constant $\delta > 0$ we have

$$\psi\left(\frac{\pi}{2} - r\right) \leqslant \delta.$$

Put $\theta = \min\{\delta, \frac{1}{4}q\}$. Now the statement of the lemma follows from the above inequalities.

Indeed, if $|r| < \varepsilon$ then

$$\psi\left(\frac{\pi}{2} - r\right) \geqslant \frac{q}{4}|r| \geqslant \theta \mathrm{e}^{-n}.$$

Otherwise, $\psi(\pi/2 - r) \geqslant \delta \geqslant \delta \mathrm{e}^{-n} \geqslant \theta \mathrm{e}^{-n}$. □

To complete the proof we compute in polytime matrices $\tilde{M}^i$ and a linear function $\tilde{L}$ such that

$$\|\tilde{M}^i - M\| \leqslant \tfrac{1}{10}\theta \mathrm{e}^{-n^3},$$

$$\|\tilde{L} - L\| \leqslant \tfrac{1}{10}\theta \mathrm{e}^{-n^2}.$$

Then for every $B \subset A$ we have

$$\left\|\prod_{i\in B}\tilde{M}^i - \prod_{i\in B}M^i\right\| \leqslant \tfrac{1}{10}\theta \mathrm{e}^{-n^2} \tag{17}$$

since $\|M^i\| \leqslant 1$.

Consider a strategy $\bar{\sigma}$ with actions $d_i = \mathit{take}$ whenever $i \in \bar{A}$.

The probability of success of $\bar{\sigma}$ is

$$\begin{aligned}
\boldsymbol{R}(\bar{\sigma}) &= \tilde{L}\left(\prod_{i\in\bar{A}}\tilde{M}^i\right)\cdot(1,0,0)\\
&\geqslant \tilde{L}\left(\prod_{i\in\bar{A}}M^i\cdot(1,0,0)\right) - \tfrac{3}{10}\theta \mathrm{e}^{-n^2}\\
&\geqslant L\left(\prod_{i\in\bar{A}}M^i\cdot(1,0,0)\right) - \tfrac{3}{10}\theta \mathrm{e}^{-n^2} - \tfrac{1}{10}\theta \mathrm{e}^{-n^2}\\
&= L\left(\varphi\left(\frac{\pi}{2}\right)\right) - \tfrac{4}{10}\theta \mathrm{e}^{-n^2}.
\end{aligned}$$

(We used (17), $\|\prod_{i\in\bar{A}}M^i\| \leqslant 1$ and $l_i \leqslant 1$.)

Let an optimal strategy σ have actions $d_i = \mathit{take}$ for $i \in A'$. Then

$$\begin{aligned}
\boldsymbol{R}(\sigma) &= \tilde{L}\left(\prod_{i\in A'}\tilde{M}^i\right)\cdot(1,0,0)\\
&\leqslant \tilde{L}\left(\prod_{i\in A'}M^i\cdot(1,0,0)\right) + \tfrac{3}{10}\theta \mathrm{e}^{-n^2}\\
&\leqslant L\left(\prod_{i\in A'}M^i\cdot(1,0,0)\right) + \tfrac{3}{10}\theta \mathrm{e}^{-n^2} + \tfrac{1}{10}\theta \mathrm{e}^{-n^2}\\
&= L\left(\varphi\left(\sum_{i\in A'}\alpha_i\right)\right) + \tfrac{4}{10}\theta \mathrm{e}^{-n^2}.
\end{aligned}$$

Applying Lemma 2 and comparing $\boldsymbol{R}(\sigma)$ and $\boldsymbol{R}(\bar{\sigma})$ we see that the optimality of σ implies that A' is also a partition.

Acknowledgements

The authors are grateful to Danièle Beauquier for discussions and many remarks on the manuscript. An anonymous referee pointed out known results and methods not referred to in the first version, and the discussion with him/her considerably helped us to revise that version. We are thankful to him then indeed.

References

[1] D. Beauquier, D. Burago and A. Slissenko, First decisions of an optimal T-strategy can be non periodic, Manuscript, 1994.

[2] D. Beauquier, D. Burago and A. Slissenko, On the complexity of finite memory strategies for control under probabilistic deviations, Manuscript, 1994.

[3] D.P. Bertsekas, *Dynamic Programming and Stochastic Control* (Academic Press, New York, 1976).

[4] C.J. Colbourn, *The Combinatorics of Network Reliability* (Oxford Univ. Press, New York, 1987).

[5] A. Condon, The complexity of the Max Word problem, in: *Proc. 8th Symp. on Theoretical Aspects of Computer Science* (1991) 456–465.

[6] M. de Rougemont and J.F. Diaz-Frias, A theory of robust planning, in: *Proc. IEEE Internat. Conf. on Robotics and Automation* (1992) 2453–2459.

[7] X. Deng, T. Kameda and C.H. Papadimitriou, How to learn an unknown environment, in: *Proc. 32nd Annu. Symp. on Foundations of Computer Science* (1991) 298–303.

[8] J.F. Diaz-Frias, *Vérification probabiliste de problèmes de graphes: applications à la robotique mobile*, Ph.D. Thesis, Université Paris-Sud, 1993.

[9] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1 (Wiley, New York, 1968).

[10] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of* NP-*Completeness* (Freeman, San Francisco, 1979).

[11] D.S. Johnson, A catalog of complexity classes, in: J. van Leeuwen, ed., *Handbook of Theoretical Computer Science*, Vol. A (Elsevier, Amsterdam, 1990) 67–161.

[12] L.C.M. Kallenberg, Linear programming and finite Markovian control problems, Technical Report 148, Mathematics Centrum Tract, Amsterdam, 1983.

[13] J.G. Kemeny and J.L. Snell, *Finite Markov Chains* (Van Nostrand Reinhold, Princeton, NJ, 1960).

[14] C.H. Papadimitriou, Games against nature, *J. Comput. System Sci.* **31** (1985) 288–301.

[15] C.H. Papadimitriou and J.N. Tsitsiklis, The complexity of Markov decision procedures, *Math. Oper. Res.* **12** (1987) 441–450.

[16] C.H. Papadimitriou and M. Yannakakis, Shortest paths without a map, *Theoret. Comput. Sci.* **84** (1991) 127–150.

[17] M.L. Puterman, Markov decision processes, in: D.P. Heyman and M.J. Sobel, eds., *Handbooks in Operations Research and Management Science. Stochastic Models*, Vol. 2 (North Holland, Amsterdam, 1990) 331–434.

[18] L.G. Valiant, The complexity of enumeration and reliability problems, *SIAM J. Comput.* **8** (1979) 410–421.

[19] V.A. Zalgaller, A discussion of one question of Bellman, Tech. Report, St. Petersburg State Univ., 1992; registered in VINITI, No. 849-B92, 34 pp (in Russian).