ELSEVIER

Theoretical Computer Science 158 (1996) 343–359

Theoretical Computer Science

Mathematical Games

The complexity of mean payoff games on graphs [1]

Uri Zwick [a,*], Mike Paterson [b]

[a] *Department of Computer Science, School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel*
[b] *Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK*

Received August 1994; revised September 1995
Communicated by A.S. Fraenkel

Abstract

We study the complexity of finding the values and optimal strategies of *mean payoff games* on graphs, a family of perfect information games introduced by Ehrenfeucht and Mycielski and considered by Gurvich, Karzanov and Khachiyan. We describe a pseudo-polynomial-time algorithm for the solution of such games, the decision problem for which is in $NP \cap coNP$. Finally, we describe a polynomial reduction from mean payoff games to the *simple stochastic games* studied by Condon. These games are also known to be in $NP \cap coNP$, but no polynomial or pseudo-polynomial-time algorithm is known for them.

1. Introduction

Let $G = (V,E)$ be a finite directed graph in which each vertex has at least one edge going out of it. Let $w : E \to \{-W,\ldots,0,\ldots,W\}$ be a function that assigns an integral weight to each edge of G. Ehrenfeucht and Mycielski [8] studied the following infinite two-person game played on such a graph. The game starts at a vertex $a_0 \in V$. The first player chooses an edge $e_1 = (a_0,a_1) \in E$. The second player then chooses an edge $e_2 = (a_1,a_2) \in E$, and so on indefinitely. The first player wants to maximise $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} w(e_i)$. The second player wants to minimise $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} w(e_i)$. Ehrenfeucht and Mycielski show that each such game has a value v such that the first player has a strategy that ensures that $\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} w(e_i) \geqslant v$, while the second player has a strategy that ensures that $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} w(e_i)$

* Corresponding author. E-mail address: zwick@math.tau.ac.il.

[1] Supported in part by the ESPRIT Basic Research Action Programme of the EC under contract No. 7141 (project ALCOM II).

SSDI 0304-3975(95)00188-3

$\leqslant v$. Furthermore, they show that both players can achieve this value using a *positional strategy*, i.e., a strategy in which the next move depends only on the vertex from which the player is to move.

As the players in these *mean payoff games* move alternatively, we may assume, without loss of generality, that the graph $G = (V,E)$ on which such a game is played is bipartite, with V_1 and V_2 being the partition of the vertices into the two 'sides' and with $E = E_1 \cup E_2$ such that $E_1 \subseteq V_1 \times V_2$ and $E_2 \subseteq V_2 \times V_1$. If the original graph is not bipartite, we simply duplicate the set of vertices.

To obtain their results for the infinite game, Ehrenfeucht and Mycielski [8] also consider the following finite version of the game. Again the game starts at a specific vertex of the graph $G = (V,E)$, which is assumed to be bipartite. The players alternate in choosing successive edges that form a path, but the game ends as soon as a cycle is formed. The outcome of the game is then the mean weight of the edges on this cycle. The first player wants to maximise and the second player to minimise this outcome. This game is a finite perfect information two-person game and so, by definition, has a value. Ehrenfeucht and Mycielski [8] show that the value v of this finite game is also the value of the infinite game described above. Furthermore, they show, surprisingly perhaps, that both players have positional optimal strategies for the finite game. The positional optimal strategies of the finite game are also positional optimal strategies for the infinite game.

Gurvich et al. [11], unaware of the work of Ehrenfeucht and Mycielski [8], considered a slightly wider class of mean payoff games which they refer to as *cyclic games*. A cyclic game is played on a directed graph $G=(V,E)$ with a weight function $w : E \to \{-W,\ldots,0,\ldots,W\}$. The vertices of G are again divided into two classes V_1 and V_2. This need not be a bipartite partition however. The players again form a path starting at a fixed vertex $a_0 \in V$. Whenever the endpoint of the path formed is in V_1, the first player chooses the next edge; whenever it is in V_2, the second player makes the choice. Note that a player may be able, or may be forced, to play a few times in succession. The goal of the first player is again to maximise, and of the second to minimise, the long-term average weight of the path formed. Gurvich et al. [11] note that a general theorem of Moulin [18] on stationary optimal strategies in stochastic games implies that both players of a cyclic game have positional optimal strategies. The theorem of Moulin is proved non-constructively using a fixed point theorem. Gurvich et al. [11] give an exponential time algorithm for finding such positional optimal strategies thereby giving a constructive proof of their existence. Cyclic games *with prohibitions*, a further generalisation of cyclic games, were considered by Karzanov and Lebedev [13]. Generalizations of mean payoff games to n players, where $n > 2$, have been considered by Alpern [1].

Ehrenfeucht and Mycielski [8] give no efficient algorithm for finding optimal strategies for the finite and infinite games. Gurvich et al. [11] give, as mentioned, an exponential-time algorithm for these tasks. We complement their works by exhibiting an $O(|V|^3 \cdot |E| \cdot W)$ time algorithm for finding the values of the mean payoff games played on a graph $G = (V,E)$ with vertex classes V_1 and V_2. The graph G need not

be bipartite so our algorithm applies to the slightly wider class of games considered by Gurvich et al. [11]. The algorithm finds the values of all the vertices of the graph; games starting at different vertices may have different values, of course. We also give an $O(|V|^4 \cdot |E| \cdot \log(|E|/|V|) \cdot W)$ time algorithm for finding positional optimal strategies for both players. Our algorithm is polynomial in the size of the graph but only pseudo-polynomial in the weights. Our algorithm is polynomial if the weights are presented in unary notation. In particular, our algorithms work in polynomial time if the weights are taken from, say, $\{-1, 0, +1\}$. This is already a non-trivial case.

At the end of [11], there is a claim that there exists a polynomial-time algorithm for finding values and optimal strategies of cyclic games. According to Karzanov (personal communication), this claim was made by mistake. Lozovanu [14, 15] also considers cyclic games and claims a strongly polynomial-time algorithm for them. He describes a simple reduction from cyclic games to simple acyclic games. Unfortunately, his reduction is not valid: the first player in his acyclic games gains some control over the length of the cycles formed in the cyclic games. His reduction fails, for example, on the complete bipartite graph with vertex sets $V_1 = \{v_0, v_2\}$, $V_2 = \{v_1, v_3\}$ and edge weights given by $w = 0$ for edges $(0,1), (1,2), (2,3), (3,0)$, $w = 1$ for $(1,0), (3,2)$, and $w = -1$ for $(0,3), (2,1)$.

We also consider situations in which one player knows in advance the positional strategy the other player is going to use. Using a result of Karp [12] we show that an optimal counter-strategy can be found in strongly polynomial time. This immediately implies that the decision problem associated with the game is in NP $\cap$ co-NP. Similar observations were made by Karzanov and Lebedev [13].

The decision problem corresponding to mean payoff games (MPGs) is thus in NP $\cap$ co-NP as well as in $\tilde{P}$ (pseudo-polynomial time), but is not yet known to be in P. This gives the MPG problem a rare status shared only by a few number-theoretic problems, such as *primality* [22].

Condon [6] has recently studied the complexity of *simple stochastic games* (SSGs) introduced originally by Shapley [23]. Condon shows that the decision problem corresponding to SSGs is also in NP $\cap$ co-NP. While MPGs are deterministic, SSGs are games of chance. We describe a simple reduction from MPGs to SSGs in two steps. We first describe a reduction from MPGs to *discounted payoff games* (DPGs), and then a reduction from DPGs to SSGs.

The reduction from MPGs to SSGs shows that SSGs are at least as hard as MPGs. It also supplies an alternative proof that the MPG problem is in NP $\cap$ co-NP, though we believe that the MPG problem is strictly easier then the SSG problem. As attempts to obtain polynomial-time algorithms for SSG's have not yet borne fruit, it may be interesting to focus attention on the possibly easier problem of obtaining a polynomial-time algorithm for MPGs.

Various path-forming games, such as the many different versions of *geography* were studied by Bodlaender [2], Fraenkel and Simonson [10] and Fraenkel et al. [9]. Many of these games are PSPACE-complete. It is therefore somewhat surprising that the mean payoff games that we are considering do have relatively efficient algorithms.

Mean payoff games arise naturally when trying to design algorithms for various *on-line* problems. Some possible applications of mean payoff games are described in Section 7.

The rest of the paper is organised as follows. In the next section we describe an algorithm for finding the values of a game. In Section 3 we describe an algorithm for finding optimal strategies. In Section 4 we consider the case of playing against a known positional strategy. In Section 5 we introduce discounted payoff games (DPGs) and describe a reduction from MPGs to DPGs. In Section 6 we describe the simple stochastic games (SSGs) studied by Condon [6] and present a reduction from DPGs to SSGs. In Section 7 we describe some applications of mean payoff games. We end in Section 8 with some concluding remarks and open problems.

2. Finding the values of a game

Let $G = (V_1, V_2, E)$ be the graph on which the game is to be played, where V_1 are the vertices of the first player and V_2 are the vertices of the second player, let $w : E \to \{-W, \ldots, 0, \ldots, W\}$ be a weight function on its edges, and $|V| = n$, where $V = V_1 \cup V_2$. Recall that the graph G need not be bipartite, there may be edges between different vertices of V_1 and between different vertices of V_2.

Our first goal is to find, for each vertex $a \in V$, the value $v(a)$ of the finite and infinite games that start at a. The proof, given by Ehrenfeucht and Mycielski, that the values of the finite and infinite games are equal, extends easily to the case in which the graph $G = (V_1, V_2, E)$ is not bipartite. To reach this goal we consider a third version of the game. This time the two players play the game for exactly k steps constructing a path of length k, and the weight of this path is the outcome of the game. The length of the game is known in advance to both players. We let $v_k(a)$ be the value of this game started at vertex $a \in V$, where player I or II plays first according to whether $a \in V_1$ or $a \in V_2$.

Theorem 2.1. *The values $v_k(a)$, for every $a \in V$, can be computed in $O(k \cdot |E|)$ time.*

Proof. It is easy to see that for every $a \in V$ and every $k \geqslant 1$ we have

$$v_k(a) = \begin{cases} \max_{(a,b)\in E}\{w(a,b) + v_{k-1}(b)\} & \text{if } a \in V_1\,, \\ \min_{(a,b)\in E}\{w(a,b) + v_{k-1}(b)\} & \text{if } a \in V_2\,. \end{cases}$$

Clearly, $v_0(a) = 0$ for every $a \in V$. The values $v_k(a)$, for every $a \in V$, can be easily computed using these recursive formulae in $O(k \cdot |E|)$ time. □

It seems intuitively clear that $\lim_{k\to\infty} v_k(a)/k = v(a)$, where $v(a)$ is the value of the infinite game that starts at a. The next theorem states that this is indeed the case. In the proof of this theorem we rely on the result, proved by Ehrenfeucht and Mycielski and by Gurvich, Karzanov and Khachiyan, that both players have positional optimal

strategies. A *positional strategy for player I* is just a mapping $\pi_1 : V_1 \to V$ such that $(a_1, \pi_1(a_1)) \in E$ for every $a_1 \in V_1$. Similarly, a *positional strategy for player II* is a mapping $\pi_2 : V_2 \to V$ such that $(a_2, \pi_2(a_2)) \in E$ for every $a_2 \in V_2$.

Theorem 2.2. *For every $a \in V$ we have*

$$k \cdot v(a) - 2nW \leqslant v_k(a) \leqslant k \cdot v(a) + 2nW .$$

Proof. Let $\pi_1 : V_1 \to V_2$ be a positional optimal strategy for player I in the finite game starting at a. We show that if player I plays using the strategy π_1 then the outcome of a k-step game is at least $(k-n) \cdot v(a) - nW$. Consider a game in which player I plays according to π_1. Push (copies of) the edges played by the players onto a stack. Whenever a cycle is formed, it follows from the fact that π_1 is an optimal strategy for player I in the finite game, that the mean weight of the cycle formed is at least $v(a)$. The edges that participate in that cycle lie consecutively at the top of the stack. They are all removed and the process continues. Note that at each stage the stack contains at most n edges and the weight of each of them is at least $-W$. Player I can therefore ensure that the total weight of the edges encountered in a k-step game starting from a is at least $(k-n) \cdot v(a) - nW$. This is at least $k \cdot v(a) - 2nW$ as $v(a) \leqslant W$.

Similarly, if player II plays according to a positional optimal strategy $\pi_2 : V_2 \to V_1$ of the finite game that starts at a, she can make sure that the mean of each cycle closed is at most $v(a)$. At most n edges are left on the stack and the weight of each of them is at most W. She can therefore ensure that the total weight of the edges encountered in a k-step game starting at a is at most $(k-n) \cdot v(a) + nW \leqslant k \cdot v(a) + 2nW$. □

We can now describe the algorithm for computing the exact values of the finite and infinite games.

Theorem 2.3. *Let $G = (V_1, V_2, E)$ be a directed graph and let $w : E \to \{-W, \ldots, 0, \ldots, W\}$ be a weight function on its edges. The value $v(a)$, for every $a \in V$, corresponding to the infinite and finite games that start at all the vertices of V can be computed in $O(|V|^3 \cdot |E| \cdot W)$ time.*

Proof. Compute the values $v_k(a)$, for every $a \in V$, for $k = 4n^3 W$. This can be done, according to Theorem 2.1, in $O(|V|^3 \cdot |E| \cdot W)$ time. For each vertex $a \in V$, compute an estimate $v'(a) = v_k(a)/k$. By Theorem 2.2, we get that

$$v'(a) - \frac{1}{2n(n-1)} < v'(a) - \frac{2nW}{k} \leqslant v(a) \leqslant v'(a) + \frac{2nW}{k} < v'(a) + \frac{1}{2n(n-1)} .$$

The value $v(a)$ is a rational number with a denominator whose size is at most n. The minimum distance between two possible values of $v(a)$ is at least $1/n(n-1)$. The exact value of $v(a)$ is therefore the unique rational number with a denominator of size at most n that lies in the interval $(v'(a) - 1/[2n(n-1)], v'(a) + 1/[2n(n-1)])$. This number is easily found. □

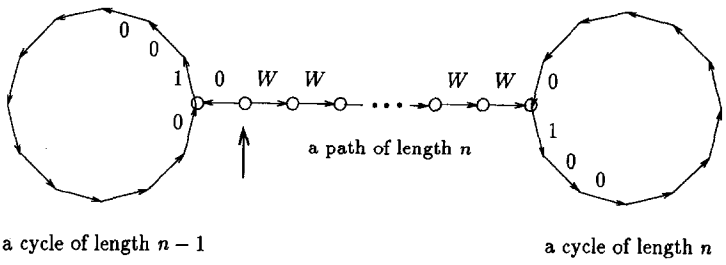

Fig. 1. An example in which $k = \Omega(n^3 \cdot W)$ is needed.

The example given in Fig. 1 shows that to obtain the correct values using the algorithm described above it may be necessary to take $k = \Omega(n^3 W)$. Slightly less accuracy is needed if we just want to know whether the value of each position is negative, zero or positive. This decision problem can therefore be decided more efficiently.

Theorem 2.4. *Let $G = (V_1, V_2, E)$ be a directed graph and let $w : E \to \{-W, \ldots, 0, \ldots, W\}$ be a weight function on its edges. Let T be an integer threshold. A decision whether $v(a) < T$, $v(a) = T$, or $v(a) > T$, for every $a \in V$, can be made in $O(|V|^2 \cdot |E| \cdot W)$ time.*

Proof. The distance between T and the closest rational number with a denominator of size at most n is $1/n$. It is therefore enough to compute the values $v_k(a)$ for $k = 4n^2 W$, and this takes only $O(|V|^2 \cdot |E| \cdot W)$ time. □

3. Finding the optimal strategies

Given an algorithm for finding the value of any vertex of a graph, positional optimal strategies can be found using a simple method, which successively eliminates sets of edges using a 'group testing' technique.

Theorem 3.1. *Let $G = (V_1, V_2, E)$ be a directed graph and let $w : E \to \{-W, \ldots, 0, \ldots, W\}$ be a weight function on its edges. Positional optimal strategies for both players, for games played on this graph, can be found in $O(|V|^4 \cdot |E| \cdot \log(|E|/|V|) \cdot W)$ time.*

Proof. Start by computing the values $v(a)$ for every $a \in V$. If all the vertices $a \in V_1$ have outdegree one, then player I has a unique strategy and this strategy is positional and optimal. Otherwise, consider any vertex $a \in V_1$ with outdegree $d > 1$. Remove any $\lceil d/2 \rceil$ of the edges leaving a, and recompute the value of a, $v'(a)$ say, for the resulting graph. If $v'(a) = v(a)$ then there is a positional optimal strategy for the player I which does not use any of the removed edges; if $v'(a) \neq v(a)$ then there is a positional optimal strategy for this player using one of the removed edges. Whichever is the case, we can

now restrict attention to a subgraph G' with at least $\lfloor d/2 \rfloor$ fewer edges. Let $d(a)$ be the initial outdegree of vertex $a \in V$. After $O(\sum_{a \in V_1} \log d(a))$ such experiments we are left with a positional optimal strategy for player I. A positional optimal strategy for player II is found in a similar way. As $\sum_{a \in V} \log d(a) \leqslant |V| \cdot \log(|E|/|V|)$, we get that the complexity of this algorithm is $O(|V|^4 \cdot |E| \cdot \log(|E|/|V|) \cdot W)$, as required. □

An interesting open problem is whether finding positional optimal strategies is harder than just computing the values of a game. The algorithm we describe calls the full value-finding algorithm repeatedly, but uses only the value at a single vertex and ignores any information about the optimal moves of the players in the truncated games. Unfortunately, optimal moves in the truncated games may not conform to positional strategies. We think however that it should be possible to use the additional information gathered and improve our algorithm.

The running times of all the algorithms described so far depend on the size of the weights. This dependency can be avoided at the (high) price of an exponential running time in the size of the graph.

Theorem 3.2. *Let $G = (V_1, V_2, E)$ be a directed graph with a real weight function $w : E \to R$. Let $V = V_1 \cup V_2$. Positional optimal strategies for both players, for games played on this graph, can be found in $2^{O(|E|)}$ or $2^{O(|V| \log |V|)}$ time.*

Proof. It is easy to see that each player has at most $2^{O(|E|)}$ positional strategies. The values of all the vertices in V when players I and II play according to specific positional strategies, $\pi_1 : V_1 \to V$ and $\pi_2 : V_2 \to V$, are easily found in $O(|E| + |V|)$ time. We can therefore construct, in $2^{O(|E|)}$ time, a $2^{O(|E|)} \times 2^{O(|E|)}$ matrix with all the possible outcomes of the game when both players use positional strategies. Let $v_{\pi_1,\pi_2}(a)$ be the outcome of a game that starts at $a \in V$ in which the two players use strategies π_1 and π_2, respectively. The results of Ehrenfeucht and Mycielski and of Gurvich, Karzanov and Khachiyan imply that $v(a) = \max_{\pi_1} \min_{\pi_2} \{v_{\pi_1,\pi_2}(a)\}$, for every $a \in V$. The values of all the positions can therefore be found in $2^{O(|E|)}$ time. The results of the above-mentioned authors also imply that there exists a positional strategy π_1 for the first player for which $\min_{\pi_2} \{v_{\pi_1,\pi_2}(a)\} = v(a)$, for every $a \in V$. Each such strategy is a positional optimal strategy for player I. A positional optimal strategy for player II is found similarly. The $2^{O(|V| \log |V|)}$ time algorithm is obtained in the same way since the number of positional strategies that each player may have is also bounded by $2^{O(|V| \log |V|)}$. □

4. Playing against a known positional strategy

In this section we consider degenerate games in which there is only one edge out of each vertex for player II, say. This corresponds, for example, to cases in which player I knows in advance the positional strategy according to which player II is going

to play. The simple observations made in this section are very similar to observations made by Gurvich et al. [11] and by Karzanov and Lebedev [13].

An $O(|V|\cdot|E|)$ algorithm of Karp [12] (see also [7, p. 548]) for finding the maximum (or minimum) mean weight cycle of a weighted graph $G = (V,E)$ supplies, almost immediately, an efficient purely combinatorial algorithm for such special cases.

Theorem 4.1. *Let $G = (V_1, V_2, E)$ be a directed graph with a real weight function $w : E \to R$ on its edges, and assume that the outdegree of each vertex $v_2 \in V_2$ is exactly one. Then, the values of all the vertices and a positional optimal strategy $\pi_1 : V_1 \to V$ for player I can be found in $O(|V|\cdot|E|)$ time.*

Proof. The value of vertex $a \in V$ is the maximum mean weight of a cycle reachable from a. We begin therefore by finding the strongly connected components and the component graph of G. This can be done in $O(|E| + |V|)$ time (see, e.g., [7]). Next, we use Karp's algorithm to find the maximum mean weight cycle in each such strongly connected component. This takes $O(|V|\cdot|E|)$ time. We then find, again in $O(|V|\cdot|E|)$ time, the transitive closure of the component graph of G. The maximum mean weight cycle reachable from each vertex of G is then easily found in $O(|V|^2)$ time. □

The maximum (or minimum) mean weight cycle in a graph $G = (V,E)$ with relatively small weights can be found more efficiently using the scaling algorithms of Orlin and Ahuja [20] and Young and Tarjan and Orlin [24]. If there are only two different edge weights then it can be found even faster, using an algorithm of Butkovic and Cuninghame-Green [4].

Could methods used by Karp's algorithm, or by the other maximum mean weight cycle algorithms, be used to obtain a more efficient algorithm for the general case? Could scaling methods be used to speed our algorithm?

The natural decision problem corresponding to MPG's is the following. Given a MPG G and a number v, is the value of G at least v? As a Corollary to Theorem 4.1 we get the following result.

Theorem 4.2. *The decision problem corresponding to mean payoff games is in NP ∩ co-NP.*

Proof. To show that the value of a game is at least v, all we have to do is *guess* a positional optimal strategy for player I. We can then check, using Karp's algorithm, that the value of the game is at least v. To show that the value of the game is less than v, all we have to do is guess a positional optimal strategy for player II, and use Karp's algorithm to check that the value is less than v. □

An alternative, more 'efficient', proof of Theorem 4.2 makes use of *potentials*.

Theorem 4.3. *Let $G = (V_1, V_2, E)$ be a directed graph, let $w : E \to R$ be a weight function on its edges and let $a \in V = V_1 \cup V_2$. Then, the value of the mean payoff*

game that starts at a is at least v if and only if there exist subsets $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$ with $a \in U = U_1 \cup U_2$, and a potential function $h : U \to R$ that satisfy the following two conditions:

(i) $\forall u \in U_1 \ \exists (u,v) \in E \quad v \in U \ \wedge \ h(u) + w(u,v) \geqslant h(v) + v$,

(ii) $\forall u \in U_2 \ \forall (u,v) \in E \quad v \in U \ \wedge \ h(u) + w(u,v) \geqslant h(v) + v$.

Proof. For the 'if' direction, suppose that such sets and potential function exist. Player I then always chooses an edge that satisfies condition (i). This makes sure that the game never leaves the set U and that the mean of every cycle formed is at least v.

For the 'only if' direction, suppose that $v(a) \geqslant v$. If we subtract v from each of the edge weights, we obtain a corresponding game where $v(a) \geqslant 0$, so we may assume $v(a) \geqslant v = 0$ without loss of generality. Let $\pi_1 : V_1 \to V$ be a positional optimal strategy for player I, and consider the subgraph $G' = (U_1, U_2, E')$ of G whose vertices and edges are just those which are reachable from vertex a when player I plays according to the positional strategy π_1 and player II plays arbitrarily. We define $h(u) \in R$ for each $u \in U_1 \cup U_2$ as the minimum weight of a path from a to u, i.e., the *distance* from a to u in the weighted graph formed. Since π_1 is an optimal strategy for player I assuring that $v(a) \geqslant 0$, the weight of any cycle in G' is nonnegative and these distances are well defined. Conditions (i) and (ii) follow immediately from the definitions of G' and h. □

To show that the value of a game is at least v, all we have to do is guess the subsets U_1 and U_2 and the potential function h. The two conditions can then be verified in linear time. A dual condition can be used to verify that the value of the game is at most v.

5. Discounted payoff games

In this section we describe a *discounted* version of mean payoff games. This (fourth) variant, which is also interesting in its own right, will serve in the next section as a link between mean payoff games and simple stochastic games.

Let $0 < \lambda < 1$ be a real number. The weight of the ith edge, e_i, chosen by the players is now multiplied by $(1-\lambda)\lambda^i$ and the outcome of the game is defined to be $(1-\lambda)\sum_{i=0}^{\infty} \lambda^i w(e_i)$. The goal of the first player is again to maximise the outcome of the game and the goal of the second player is to minimise this outcome. The number λ is called the *discounting factor* of the game.

Let $G = (V_1, V_2, E)$ be a directed graph and let $w : E \to R$ be a weight function on its edges. As always, we assume that the outdegree of all the vertices is at least one. Let $V = V_1 \cup V_2 = \{1, 2, \ldots, n\}$. Let $x_i = x_i(\lambda)$ be the value of a discounted game started at i. If $(i,j) \in E$, we use w_{ij} as an abbreviation for $w((i,j))$.

Theorem 5.1. *The value vector* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *of the discounted games played on the graph* $G = (V_1, V_2, E)$ *is the unique solution of the following set of equations*:

$$x_i = \begin{cases} \max_{(i,j)\in E} \{(1-\lambda)w_{ij} + \lambda x_j\} & \text{if } i \in V_1 , \\ \min_{(i,j)\in E} \{(1-\lambda)w_{ij} + \lambda x_j\} & \text{if } i \in V_2 . \end{cases}$$

Proof. Let $\mathscr{F}$ be a mapping that receives a vector $\boldsymbol{x}$ and returns the vector $\boldsymbol{y}$ such that

$$y_i = \begin{cases} \max_{(i,j)\in E} \{(1-\lambda)w_{ij} + \lambda x_j\} & \text{if } i \in V_1 , \\ \min_{(i,j)\in E} \{(1-\lambda)w_{ij} + \lambda x_j\} & \text{if } i \in V_2 . \end{cases}$$

The given set of equations can be expressed in the form $\boldsymbol{x} = \mathscr{F}(\boldsymbol{x})$. If we let $\|\boldsymbol{v}\| = \max_i\{v_i\}$ be the *max* norm, then

$$\forall \boldsymbol{u}, C, \ \|\mathscr{F}(\boldsymbol{u}) - \mathscr{F}(C)\| \leqslant \lambda \|\boldsymbol{u} - C\| .$$

Thus, since $0 < \lambda < 1$, $\mathscr{F}$ is a contraction mapping with respect to the norm. It follows easily that the limit $\boldsymbol{x} = \lim_{n\to\infty} \mathscr{F}^n(\boldsymbol{0})$ exists and is the unique solution to the equation $\boldsymbol{x} = \mathscr{F}(\boldsymbol{x})$.

Let $\boldsymbol{x}$ be the solution of the equation $\boldsymbol{x} = \mathscr{F}(\boldsymbol{x})$. It is easy to verify that if player I plays according to a strategy which at each vertex $i \in V_1$ chooses an edge $(i,j) \in E$ which maximises $(1-\lambda)w_{ij} + \lambda x_j$, then the outcome of the game that starts from each vertex i is at least x_i. Similarly, if player II plays according to a strategy which at each vertex $i \in V_2$ chooses an edge $(i,j) \in E$ which minimises $(1-\lambda)w_{ij} + \lambda x_j$, then the outcome of the game that starts from each vertex i is at most x_i. It follows that the value of the game starting from i is exactly x_i. □

It follows immediately from this theorem that both players of the discounted game again have positional optimal strategies. The proof in this case is much simpler than the proofs given by Ehrenfeucht and Mycielski [8] and by Gurvich, Karzanov and Khachiyan [11] for non-discounted games.

Theorem 5.1 suggests a way of finding the values of the discounted payoff games played on a graph $G=(V_1, V_2, E)$. We are not aware, however, of any strongly polynomial-time algorithm for finding a solution to the set of equations that appear in the theorem. A pseudo-polynomial-time algorithm for finding the values and optimal positional strategies for discounted payoff games, similar to the algorithm presented in the proof of Theorem 2.3, can be easily devised.

Let $v(\lambda)$ be the value of the discounted game with discounting factor λ. As λ tends to 1, we expect $v(\lambda)$ to tend to v, the value of the non-discounted game. This follows from the next theorem.

Theorem 5.2. *Let* $G = (V_1, V_2, E)$ *be a graph on* n *vertices, let* $V = V_1 \cup V_2$, *let* $w : E \to \{-W, \ldots, 0, \ldots, W\}$ *be a weight function on its edges and let* λ *be a real number satisfying* $0 < \lambda < 1$. *If* $v(\lambda)$ *and* v *are the values of the discounted and*

mean payoff games played on the graph $G = (V_1, V_2, E)$ *starting at* $a \in V$, *then*

$$v - 2n(1-\lambda)W \leqslant v(\lambda) \leqslant v + 2n(1-\lambda)W .$$

Proof. Consider the outcome of a discounted game in which player I uses a positional optimal strategy for the *non-discounted* game and player II uses a positional optimal strategy to counter the strategy of player I. The outcome of such a game clearly supplies a lower bound on the value $v(\lambda)$ of the discounted game. The play in such a case consists of a path of length k, followed by a cycle of length ℓ which is repeated indefinitely, where $0 \leqslant k \leqslant n-1$, $1 \leqslant \ell \leqslant n$ and $k + \ell \leqslant n$.

Assume for the moment that all the edge weights are non-negative. Let $w_0, \ldots, w_{\ell-1}$ be the weights of the edges in the cycle formed. As player I uses an optimal strategy for the non-discounted game we get that $\sum_{i=0}^{\ell-1} w_i \geqslant \ell v$. The outcome of the discounted game is then at least

$$(1-\lambda)\lambda^k \left(\sum_{i=0}^{\ell-1} w_i \lambda^i\right)\left(\sum_{j=0}^{\infty} \lambda^{j\ell}\right) = \frac{(1-\lambda)\lambda^k}{1-\lambda^\ell} \cdot \sum_{i=0}^{\ell-1} w_i \lambda^i$$

$$\geqslant \frac{(1-\lambda)\lambda^{k+\ell-1}}{1-\lambda^\ell} \cdot \sum_{i=0}^{\ell-1} w_i \geqslant \frac{\ell(1-\lambda)}{1-\lambda^\ell} \cdot \lambda^{k+\ell-1} \cdot v .$$

As $\ell(1-\lambda)/(1-\lambda^\ell) > 1$ and $\lambda^{k+\ell-1} > \lambda^n > 1 - n(1-\lambda)$, this is at least $(1-n(1-\lambda)) \cdot v$.

We now return to the general case in which the edge weights are not assumed to be non-negative. By adding W to each weight, we can make all the weights non-negative. The value and outcome of the game are changed by exactly W. Applying the previous inequality to the resulting non-negative game we get that

$$(v(\lambda) + W) \geqslant (1 - n(1-\lambda))(v + W) ,$$

or equivalently that

$$v(\lambda) \geqslant v - n(1-\lambda)(v+W) \geqslant v - 2n(1-\lambda)W .$$

The opposite inequality is proved in a similar way. □

In particular, if we choose $\lambda = 1 - 1/(4n^3 W)$, then it is easy to verify that $|v(\lambda) - v| \leqslant 1/(2n(n-1))$, and v can be obtained from $v(\lambda)$ by rounding to the nearest rational with a denominator less than n, as was done in Section 2. We thus obtain a reduction from MPGs to discounted payoff games (DPGs).

6. Reduction to simple stochastic games

In this section we describe a simple polynomial reduction from discounted payoff games (DPGs) to simple stochastic games (SSGs). This reduction, combined with the reduction from MPGs to DPGs, shows that SSGs are at least as hard as MPGs. We believe that MPGs are in fact easier than SSGs.

A *simple stochastic game* is a two-person game played on a directed graph $G = (V, E)$ whose vertex set V is the union of three disjoint sets $V_{\max}$, $V_{\min}$ and V_{average}. The graph also contains a special start vertex and two special vertices called the 0-sink and the 1-sink. Each edge emanating from an 'average' vertex has a rational probability attached to it. The probabilities attached to all the edges from each average vertex add up to 1.

A token is initially placed on the start vertex of the graph. At each step of the game the token is moved from a vertex to one of its neighbours, according to the following rules:

1. At a max vertex, player I chooses the edge along which the token is moved.
2. At a min vertex, player II chooses this edge.
3. At an average vertex, the edge along which the token is moved is chosen randomly according to the probabilities attached to the outgoing edges.

The game ends when the token reaches one of the sink vertices. Player I wins if the token reaches the 1-sink and player II wins otherwise, i.e., if the token reaches the 0-sink or if the game does not end. The *value* of such a game is the probability that player I wins the game when both players play optimally. As was the case for mean payoff games, the two players of a simple stochastic game have positional optimal strategies.

Simple stochastic games were first studied by Shapley [23]. Many variants of them have been studied since then (see Peters and Vrieze [21] for a survey). Condon [6] was the first to study simple stochastic games from a complexity theory point of view. She showed that the natural decision problem corresponding to SSGs (i.e., given a game G and a rational number $0 < \alpha \leqslant 1$, is the value of G at least α?) is in NP $\cap$ co-NP. No polynomial time algorithm for SSGs is yet known. Some exponential algorithms for the problem are described in [17]. A subexponential randomized algorithm for SSGs was recently obtained by Ludwig [16].

Condon [6] actually shows containment in NP $\cap$ co-NP of the decision problem that corresponds to SSGs of the following restricted form. The outdegree of each non-sink vertex is exactly two and the probability attached to each edge that emanates from an average vertex is $\frac{1}{2}$. She then describes a reduction from general SSGs to SSGs of this restricted form. Her reduction, however, is not polynomial. A general SSG on n vertices in which the denominators of all the (rational) probabilities are at most m is transformed into a restricted SSG of size polynomial in n and m, rather than in n and $\log m$. Her transformation can be easily modified however, as we show next, to yield a polynomial reduction.

It is easy to transform a SSG into an equivalent SSG in which the outdegree of each non-sink vertex is exactly two. Each vertex of fan-out k is simply replaced by a binary tree with k leaves. This increases the size of the graph (i.e., the number of vertices and edges) by only a constant factor. The remaining problem is therefore the simulation of binary average vertices with non-equal probabilities. Suppose we want to implement an average vertex u with two emanating edges (u, v_1) and (u, v_2), labelled respectively by the probabilities p/q and $(q-p)/q$, where p and q are integers

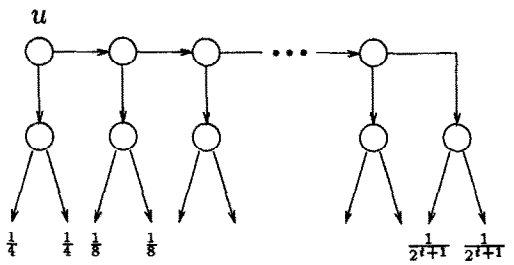

Fig. 2. Implementing an average vertex with arbitrary probabilities.

and $2^{t-1} \leqslant q < 2^t$. Let $a_1 a_2 \ldots a_{t-1} a_t$ and $b_1 b_2 \ldots b_{t-1} b_t$ be the binary representations of p and $q-p$, respectively, where a_1 and b_1 are the most significant digits. We use the construct shown in Fig. 2. All the vertices used are average vertices with equal probabilities. For every i, $2 \leqslant i \leqslant t+1$, there are two emanating edges that are reached from u with probability 2^{-i}. If $a_i = 1$ then connect one of the edges with probability $2^{-(i+1)}$ to v_1, and if $b_i = 1$ then connect one of these edges to v_2. All the unused edges are connected back to u. Is it easy to check that v_1 and v_2 are eventually reached with the appropriate probabilities. The number of vertices used in this construction is proportional to the number of bits needed to represent the transition probabilities. The reduction is therefore polynomial.

A simple stochastic game is said to *halt with probability 1* if, no matter how the players play, the game ends with probability 1. The proof of the following theorem can be found in Condon [6]. Note the similarity of this theorem to Theorem 5.1.

Theorem 6.1. *Let $G = (V, E)$ be a SSG that halts with probability 1, and let $p(u, v)$ denote the probability attached to an edge (u, v) that emanates from an average vertex u. The values $v(v)$ of the vertices of G form the unique solution to the following set of equations:*

$$v(u) = \begin{cases} \max_{(u,v)\in E}\{v(v)\} & \text{if } u \text{ is a max vertex,} \\ \min_{(u,v)\in E}\{v(v)\} & \text{if } u \text{ is a min vertex,} \\ \sum_{(u,v)\in E}\{p(u,v)\cdot v(v)\} & \text{if } u \text{ is an average vertex,} \end{cases}$$

along with the conditions that $v(0\text{-sink}) = 0$ and $v(1\text{-sink}) = 1$.

We are finally in a position to describe a reduction from discounted payoff games (DPGs) to simple stochastic games (SSGs). Recall that we have already described a reduction from MPGs to DPGs.

Let $G = (V_1, V_2, E)$ be a DPG with discounting factor λ. If we add a constant c to all the weights of the game, the value of the game is increased by c. If we multiply all the weights of the game by a constant $c > 0$, the value of the game is multiplied by c. We can therefore scale the weights so that they will all be rational numbers in the interval $[0, 1]$. If the original weights were in the range $\{-W, \ldots, 0, \ldots, W\}$, then the

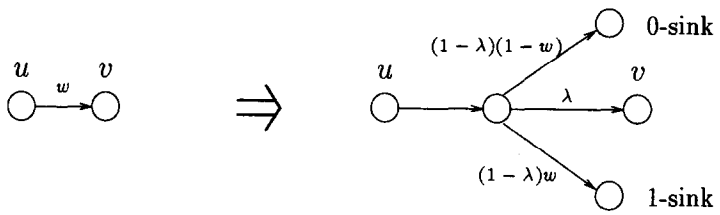

Fig. 3. Simulating a transition of a discounted payoff game.

new weights will be rational numbers with denominators and numerators in the range $\{0,1,\ldots,2W\}$.

We construct in the following way a SSG $G' = (V', E')$, with the same value as the scaled DPG $G = (V_1, V_2, E)$ with discounting factor λ. Each edge (u, v) with weight w in G is replaced by the construct shown in Fig. 3. We let $V' = V_{\max} \cup V_{\min} \cup V_{\text{average}}$, where $V_{\max} = V_1$, $V_{\min} = V_2$ and V_{average} is the set of intermediate vertices added. The simple stochastic game G' halts with probability 1, as in each transition there is a probability of $1 - \lambda$ of reaching a sink vertex. The values of the vertices of the discounted payoff game G satisfy the set of equations given in Theorem 5.1. The values of the vertices of the simple stochastic game G' satisfy the set of equations given in Theorem 6.1. These two sets of equations become *identical* once the intermediate variables, that correspond to the intermediate vertices introduced by the transformation described in Fig. 3, are eliminated. As this set of equations has a unique solution, the values of the two games are equal. The transformation of G to G' can clearly be carried out in polynomial time. This completes the description of the reduction.

7. Some applications

In this section we briefly mention some applications of mean payoff games.

Consider a system with n possible states. At each time unit, the system receives one of k possible requests. The system is allowed to change its state and then it has to serve the request. The transition from state i to state j costs a_{ij}, and serving a request of type t from state i costs b_{it}. What, in the worst-case, is the average cost of serving a request?

Borodin et al. [3] performed a *competitive analysis* of such systems, which they call *on-line metrical task systems*. If we look at the worst-case instead, we get a bipartite mean payoff game $G = (V_1, V_2, E)$ played between the system and an adversary that chooses the requests. The adversary plays from the vertices of $V_1 = \{1,2,\ldots,n\}$. The algorithm plays from the vertices of $V_2 = \{(i,j) \mid 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant k\}$. The edge $i \to (i,j)$ corresponds to the request of task j while the system is in the state i. The edge $(i,j) \to k$ corresponds to the transition of the system to state k before serving this request. The weight of all the edges of the form $i \to (i,j)$ is 0. The weight of an edge $(i,j) \to k$ is $a_{ik} + b_{kj}$.

Consider finite-window on-line string matching algorithms (see [5] for a definition). What, in the worst-case, is the average number of comparisons that an optimal algorithm has to perform per text character? The problem can be formulated as a bipartite mean payoff game played between the designer of a string matching algorithm and an adversary that answers the queries made by an algorithm. The reward (the complement of cost) obtained by the algorithm at each stage is the amount by which it can shift its window. For each pattern string and window size we obtain a mean payoff game, the solution of which yields an optimal string matching algorithm for that pattern and window size.

As a last example, consider the problem of selection with limited storage. Suppose that we are to receive a long stream of numbers. We are supposed to select the kth largest of these numbers. We have however only s storage locations, for some $s > k$, each one of them capable of holding a single number. Each input number must be read into one of these s storage locations before it can be compared to any of the numbers held in the other $s - 1$ locations. The previous value of the cell into which the input number is read is lost. What is the average number of comparisons needed per input element in the worst case? Selection and sorting problems with limited storage were considered by Munro and Paterson [19]. They allowed several passes over the input stream however.

The problem can again be formulated as a bipartite mean payoff game $G = (V_1, V_2, E)$ played by the designer of a selection algorithm and an adversary that answers queries made by an algorithm. The vertices of V_1 correspond to those partial orders of s elements in which no element is known to be smaller than k elements. The vertices of V_2 coresond to such partial orders together with requests for comparing two specific elements in each such partial order. Edges from V_1 to V_2 correspond to comparison requests made by the algorithm. Edges from V_2 to V_1 correspond to the answers of the adversary. Each vertex of V_2 has two edges emanating from it, corresponding to the two possible outcomes of the comparison requested. The weight of all the edges from V_1 to V_2 is 0. The weight of each edge $u \to v$ from V_2 to V_1 is the number of elements that are known, as a result of the last comparison, to be smaller than at least k elements. Such elements are discarded and are replaced by new input elements. Twice the value of the game is the average number of elements that can be discarded as a result of a single comparison. The graph that corresponds to the selection of the second largest element using four storage locations, i.e., $k = 2$ and $s = 4$, is given in Fig. 4. It is not difficult to verify that the value of this game is $v = \frac{1}{3}$, the starting point is this case does not matter. This means that the average number of comparisons needed per input element is $\frac{3}{2}$.

8. Concluding remarks

Mean payoff games form a very natural class of full information games and we think that resolving their complexity is an interesting issue. We conjecture that they lie in P

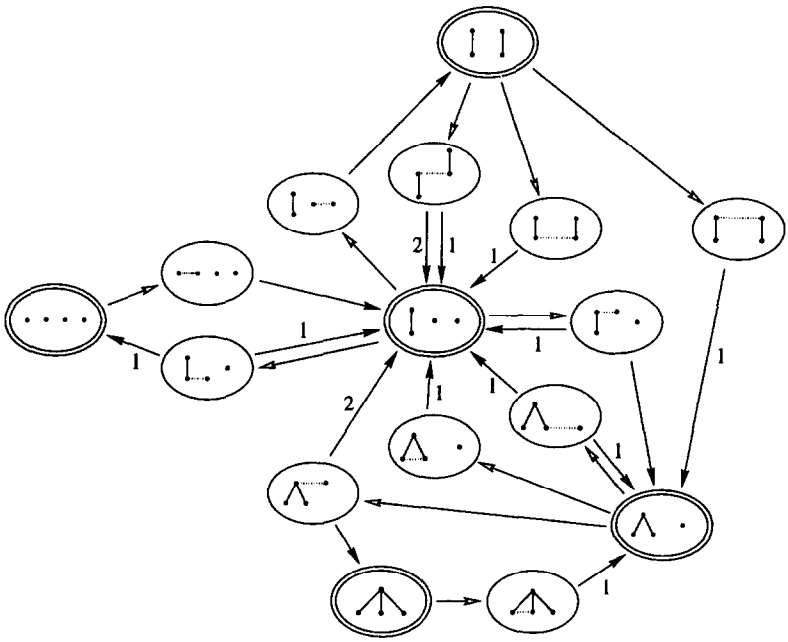

Fig. 4. A mean payoff game that corresponds to the problem of selecting the second largest element with only four storage locations.

but, since none of the standard methods seems to yield a polynomial-time algorithm for them, the study of mean payoff games may require new algorithmic techniques. If such positive approaches are unsuccessful, the example of mean payoff games may help in exploring the structure of NP ∩ co-NP.

Acknowledgements

We would like to thank Sergiu Hart, Ehud Lehrer, Nimrod Megiddo, Moni Naor, Noam Nisan and Avi Wigderson for helpful discussions and suggestions, Alexander Karzanov for pointing out a flaw in the previous version of Theorem 3.1, and Thomas McCormick for bringing references [11] and [13] to our attention.

References

[1] S. Alpern, Cycles in extensive form perfect information games, *J. Math. Anal. Appl.* **159** (1991) 1–17.

[2] H.L. Bodlaender, Complexity of path-forming games, *Theoret Comput Sci.* **110** (1993) 215–245.

[3] A. Borodin, N. Linial and M.E. Saks. An optimal on-line algorithm for metrical task system *J. ACM* **39** (1992) 745–763.

[4] P. Butkovic and R.A. Cuninghame-Green, An $O(n^2)$ algorithm for the maximum cycle mean of an $n \times n$ bivalent matrix, *Discrete Appl. Math.* **35** (1992) 157–162.

[5] R. Cole, R. Hariharan, M. Paterson and U. Zwick, Tighter lower bounds on the exact complexity of string matching, *SIAM J. Comput.* 24 (1995) 30–45.
[6] A. Condon, The complexity of stochastic games, *Inform. Comput.* **96** (1992) 203–224.
[7] T.H. Cormen, C.E. Leiserson and R.L. Rivest, *Introduction to Algorithms* (The MIT Press, Cambridge, (1990).
[8] A. Ehrenfeucht and J. Mycielski, Positional strategies for mean payoff games, *Int. J. Game Theory* **8** (1979) 109–113.
[9] A.S. Fraenkel, E.R. Scheinerman and D. Ullman, Undirected edge geography, *Theoret. Comput. Sci.* **112** (1993) 371–381.
[10] A.S. Fraenkel and S. Simonson, Geography, *Theoret Comput. Sci.* **110** (1993) 197–213.
[11] V.A. Gurvich, A.V. Karzanov and L.G. Khachiyan, Cyclic games and an algorithm to find minimax cycle means in directed graphs, *USSR Comput. Math. Math. Phys.* **28** (1988) 85–91.
[12] R.M. Karp, A characterization of the minimum cycle mean in a digraph, *Discrete Math.* **23** (1978) 309–311.
[13] A.V. Karzanov and V.N. Lebedev, Cyclical games with prohibitions, *Math. Programming* **60** (1993) 277–293.
[14] D.D. Lozovanu, Algorithms to solve some classes of network minimax problems and their applications, *Cybernetics* **29** (1991) 93–100.
[15] D.D. Lozovanu, Strongly polynomial algorithms for finding minimax paths in networks and solution of cyclic games, *Cybernetics Systems Anal.* **29** (1993) 754–759.
[16] W. Ludwig, A subexponential randomized algorithm for the simple stochastic game problem, *Inform. Comput.* **117** (1995) 151–155.
[17] M. Melekopoglou and A. Condon, On the complexity of the policy iteration algorithm for stochastic games, Technical Report TR 941, University of Wisconsin, June 1990.
[18] H. Moulin, Prolongement des jeux a deux joueurs de somme nulle, *Bull. Soc. Math. France, Memoires* **45** (1976).
[19] J.I. Munro and M.S. Paterson, Selection and sorting with limited storage, *Theoret. Comput. Sci.* **12** (1980) 315–323.
[20] J.B. Orlin and R.K. Ahuja, New scaling algorithms for the assignment and minimum mean cycle problems, *Math. Programming* **54** (1992) 41–56.
[21] H.J.M. Peters and O.J. Vrieze, Surveys in game theory and related topics, CWI Tract 39, Centrum voor Wiskunde en Informatica, Amsterdam, 1987.
[22] V.R. Pratt, Every prime has a succinct certificate, *SIAM J. Comput.* **4** (1975) 214–220.
[23] L.S. Shapley, Stochastic games, *Proc. Nat. Acad. Sci. USA*, **39** (1953) 1095–1100.
[24] N.E. Young, R.E. Tarjan and J.B. Orlin, Faster parametric shortest path and minimum-balance algorithms, *Networks* **21** (1991) 205–221.