RUTCOR
RESEARCH
REPORT

A potential reduction algorithm for two-person zero-sum limiting average payoff stochastic games

Endre Boros[a] Khaled Elbassioni[b]
Vladimir Gurvich[c] Kazuhisa Makino[d]

RRR 13-2012, Ferbruary 2012

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003;
(boros@rutcor.rutgers.edu)

[b]Max-Planck-Institute for Informatics; Stuhlsatzenhausweg 85, 66123, Saarbruecken, Germany (elbassio@mpi-sb.mpg.de)

[c]RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003;
(gurvich@rutcor.rutgers.edu)

[d]Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 113-8656, Japan; (makino@mist.i.u-tokyo.ac.jp)

Rutcor Research Report
RRR 13-2012, Ferbruary 2012

A potential reduction algorithm for two-person zero-sum limiting average payoff stochastic games

Abstract. We suggest a new algorithm for two-person zero-sum undiscounted stochastic games focusing on stationary strategies. Given a positive real ϵ, let us call a stochastic game ϵ-ergodic, if its values from any two initial points differ by at most ϵ. The proposed new algorithm outputs for every $\epsilon > 0$ in finite time either a pair of stationary strategies for the two players guaranteeing that the values from any initial points are within an ϵ range, or identifies two initial points u and v and corresponding stationary strategies for the players proving that the game values starting from u and v are at least $\epsilon/24$ apart. In particular, the above result shows that if a stochastic game is ϵ-ergodic, then there are stationary strategies for the players proving 24ϵ-ergodicity. This result strengthens a result by Vrieze (1980) claiming that if a stochastic game is 0-ergodic, then there are ϵ-optimal stationary strategies for every $\epsilon > 0$.
The suggested algorithm extends the approach recently introduced for stochastic games with perfect information, and is based on the classical potential transformation technique that changes the range of local values at all positions without changing the normal form of the game.

Acknowledgements: This research was partially supported by DIMACS, Center for Discrete Mathematics and Theoretical Computer Science, Rutgers University, and by the Scientific Grant-in-Aid from Ministry of Education, Science, Sports and Culture of Japan. Part of this research was done at the Mathematisches Forschungsinstitut Oberwolfach during a stay within the Research in Pairs Program from March 7 to March 20, 2010. The first author also acknowledges the partial support of NSF Grant IIS-0803444 and NSF Grant CMMI-0856663.

1 Introduction

1.1 Basic Concepts and Notation

Stochastic games were introduced in 1953 by Shapley [Sha53] for the discounted case, and extended to the undiscounted case by Gillette [Gil57]. Each such game $\Gamma = (p^{vu}_{k\ell},\ r^{vu}_{k\ell} \ \mid \ k \in K^v,\ \ell \in L^v,\ u, v \in V)$ is played by two players on a finite set of vertices (states) V; K^v and L^v for $v \in V$ are finite sets of actions (pure strategies) of the players, $r^{vu}_{k\ell}$ is the reward player 1 (WHITE) receives from player 2 (BLACK) if k and ℓ are the chosen actions and the game moves from state v to state u, and $p^{vu}_{k\ell}$ is the transition probability from state v to state u if players chose actions $k \in K^v$ and $\ell \in L^v$ at state $v \in V$. We assume w.l.o.g. that the game is non-stopping[1], that is

$$\sum_{u\in V} p^{vu}_{k\ell} = 1 \tag{1}$$

for all states $v \in V$ and for all choices of actions $k \in K^v$ and $\ell \in L^v$. To simplify later expressions, let us denote by $P^{vu} \in \mathbb{R}^{K^v\times L^v}$ the transition matrix, the elements of which are the $p^{vu}_{k\ell}$ probabilities. We associate in Γ a *reward matrix* A^v to every state $v \in V$ defined by

$$(A^v)_{k\ell} \ = \ \sum_{u\in V} p^{vu}_{k\ell} r^{vu}_{k\ell}. \tag{2}$$

When the game Γ is not clear from the context, we shall write $r^{vu}_{k\ell}(\Gamma)$, $p^{vu}_{k\ell}(\Gamma)$, $A^v(\Gamma)$, etc.

In the game Γ, players first agree on an initial vertex $w \in V$ to start. Then, in a general step $j = 0, 1, \ldots$, when the game arrives to state $v \in V$ they choose mixed strategies $\alpha \in \Delta(K^v) := \{\alpha \in \mathbb{R}^{K^v} \mid \sum_{i\in K^v} \alpha_i = 1,\ \alpha_i \geq 0 \ \text{ for } \ i \in K^v\}$ and $\beta \in \Delta(L^v)$, player 1 receives the amount of

$$b_j = \alpha A^v \beta$$

from player 2, and the game moves to the next state u chosen according to the transition probabilities

$$p^{vu}_{\alpha,\beta} \ = \ \alpha P^{vu} \beta. \tag{3}$$

The *undiscounted limiting average effective payoff* (for player 1) is the Cesaro average

$$g^w(\Gamma) \ = \ \liminf_{N\to\infty} \frac{1}{N+1} \sum_{j=0}^{N} \mathbb{E}[b_j],$$

where the expectation is taken over the random choices made (according to mixed strategies and transition probabilities) up to step j of the play. The purpose of player 1 is to maximize $g^w(\Gamma)$, while player 2 would like to minimize it.

[1] Shapley's original stochastic games were assumed to have positive *stopping probabilities*, i.e., at each state v, $\sum_{u\in V} p^{vu}_{k\ell} < 1$, and with probability $1 - \sum_{u\in V} p^{vu}_{k\ell}$, the game stops at state v if actions k and ℓ are selected by the players.

In 1958 Gallai [Gal58] suggested the following simple transformation. Let $x : V \to \mathbb{R}$ be a mapping that assigns to each state $v \in V$ a real number x^v called the *potential* of v. For every transition (v, u) and pair of actions $k \in K^v$ and $\ell \in L^v$ let us transform the payoff $r^{vu}_{k\ell}$ as follows:

$$r^{vu}_{k\ell}(x) = r^{vu}_{k\ell} + x^v - x^u.$$

Then the one step payoff amount changes to $b_j(x) = b_j + x^{v_j} - x^{v_{j+1}}$, where $v_j \in V$ is the position at step j of the play. However, the limiting average payoff remains the same for all finite potentials:

$$g^{v_0}(\Gamma(x)) = g^{v_0}(\Gamma) + \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[x^{v_0} - x^{v_N}] = g^{v_0}(\Gamma).$$

Thus, the transformed game remains equivalent with the original one.

In this paper we shall restrict ourselves (and the players) by the so-called *stationary* strategies, that is, the mixed strategy chosen in a position $v \in V$ can depend only on v but not on the preceding positions or moves. We will denote by $\mathcal{K}(\Gamma)$ and $\mathcal{L}(\Gamma)$ the sets of stationary strategies of White and Black, respectively, that is

$$\mathcal{K}(\Gamma) = \bigotimes_{v\in V} \Delta(K^v) \quad \text{and} \quad \mathcal{L}(\Gamma) = \bigotimes_{v\in V} \Delta(L^v).$$

1.2 Local and Global Values and Concepts of Ergodicity

Consider an arbitrary potential $x \in \mathbb{R}^V$, and let us define the *local value* $m^v(x)$ at position $v \in V$ as the value of the $|K^v| \times |L^v|$ local reward matrix game $A^v(x)$ with entries

$$a^v_{k\ell}(x) = \sum_{u\in V} p^{vu}_{k\ell}(r^{vu}_{k\ell} + x^v - x^u), \quad \text{for all } k \in K^v, \ell \in L^v, \tag{4}$$

that is,

$$m^v(x) \;=\; \text{Val}\,(A^v(x)) := \max_{\alpha^v\in\Delta(K^v)} \; \min_{\beta^v\in\Delta(L^v)} \alpha^v A^v(x)\beta^v = \min_{\beta^v\in\Delta(L^v)} \; \max_{\alpha^v\in\Delta(K^v)} \alpha^v A^v(x)\beta^v.$$

To a pair of stationary strategies $\alpha = (\alpha^v|v \in V) \in \mathcal{K}(\Gamma)$ and $\beta = (\beta^v|v \in V) \in \mathcal{L}(\Gamma)$ we associate a Markov chain $\mathcal{M}_{\alpha,\beta}(\Gamma)$ on states in V, defined by the transition probabilities $p^{vu}_{\alpha,\beta} = \alpha^v P^{vu}\beta^v$. Then, this Markov chain has unique limiting probability distributions $(q^{vu}_{\alpha,\beta}|\ u \in V)$, where $q^{vu}_{\alpha,\beta}$ is the probability of staying in state $u \in V$ when the initial vertex is $v \in V$. Then the reward of player 1 starting from vertex $v \in V$ can be computed as

$$g^v(\alpha, \beta) \;=\; \sum_{u\in V} q^{vu}_{\alpha,\beta}\, (\alpha^u A^u \beta^u)\,. \tag{5}$$

The game is said to be to be solvable in uniformly optimal stationary strategies, if there exist stationary strategies $\bar{\alpha} \in \mathcal{K}(\Gamma)$ and $\bar{\beta} \in \mathcal{L}(\Gamma)$, such that for all initial states $v \in V$

$$g^v(\bar{\alpha}, \bar{\beta}) \;=\; \max_{\alpha \in \mathcal{K}(\Gamma)} g^v(\alpha, \bar{\beta}) \;=\; \min_{\beta \in \mathcal{L}(\Gamma)} g^v(\bar{\alpha}, \beta). \tag{6}$$

This common quantity, if exists, is the value of the game with initial position $v \in V$, and will be simply be denoted by g^v.

Let us remark that so far we defined the game values in terms of stationary strategies. In 1981, Mertens and Neymann in their seminal paper [MN81] proved that every stochastic game has value from any initial point in terms of history dependent strategies. An example (the so-called Big Match) showing that the same does not hold when restricted to stationary strategies was given in 1957 in Gillette's paper [Gil57]; see also [BF68].

For $\epsilon > 0$ let us call a stochastic game Γ ϵ-*ergodic* if the game values from any two initial points differ by at most ϵ.

Let us note that using potential transformations we may be able to obtain a proof for ergodicity/non-ergodicity. To make this more precise, let us further call a stochastic game *locally* ϵ-*ergodic* if there is a potential transformation $x \in \mathbb{R}^V$ such that $|m^v(x) - m^u(x)| \leq \epsilon$ for any two states $u, v \in V$. Let us add, justifying the above definition, that if for some potential transformation we have $m^v(x) \in [a, b]$ for all $v \in V$, for some $a < b$, then the game's values also belong to $[a, b]$ from any initial point.

We will give a constructive proof of the following result.

Theorem 1 *If a stochastic game is locally ϵ-ergodic then it is also ϵ-ergodic. Conversely, if it is ϵ-ergodic, then it is also locally 24ϵ-ergodic.*

This follows from our main result, Theorem 2. Let us also add that local ϵ-ergodicity in this paper will always be guaranteed by stationary strategies.

Remark 1 *The above definition of ergodicity follows Moulin's concept of the ergodic extension of a matrix game [Mou76] (which is a very special example of a stochastic game with perfect information). Let us note that slightly different terminology is used in the Markov chain theory; see, for example, [KS63].*

The following three algorithms for undiscounted stochastic games are based on stronger "ergodicity type" conditions. The strategy iteration algorithm by Hoffman and Karp [HK66] requires that for any pair of stationary strategies of two players the obtained Markov chain has to be irreducible. Two value iteration algorithms by Federgruen are based on similar but slightly weaker requirements; see [Fed80] for the definitions and more details.

1.3 Main Results

Given an undiscounted zero-sum game, we try to reduce the range of its local values by a potential transformation $x \in \mathbb{R}^V$. If they are equalized by some potential x, that is, $m_v(x) = m$ is a constant for all $v \in V$, we say that the game is brought to its ergodic canonical form. In this case the values g^v exist and are equal to m for all initial positions $v \in V$, and furthermore, locally optimal strategies are globally optimal. Thus, the game is solved in uniformly optimal strategies. However, typically we are not that lucky.

To state our main theorem, we need more notation. Let us denote introduce $W > 0$ such that for all positions $v, u \in V$ and actions $k \in K^v$ and $\ell \in L^v$ we have either $p_{k\ell}^{vu} = 0$ or $p_{k\ell}^{vu} \geq 1/W$. Let us further define R as the smallest real such that

$$0 \leq r_{k\ell}^{vu} \leq R \tag{7}$$

for all positions $v, u \in V$ and actions $k \in K^v$ and $\ell \in L^v$ (where we assume w.l.o.g. that all rewards are non-negative), set $N^v = \max\{|K^v|, |L^v|\}$ for all $v \in V$, and set $N = \max_{v \in V} N^v$. Finally set $n = |V|$, and let η be the total bit length for the rational rewards and transition probabilities.

Theorem 2 *For every stochastic game and $\epsilon > 0$ we can find in $O\left(\frac{(NW)^n (nR)^{2^n}}{3\epsilon^{n+1}}\right)\operatorname{poly}(\eta)$ time either a potential vector $x \in \mathbb{R}^V$ proving that the game is locally (24ϵ)-ergodic, or stationary strategies for the players proving that it is not ϵ-ergodic.*

We prove this theorem by an algorithm that extends the approach recently obtained for stochastic games with perfect information [BEGM10]. It is also somewhat similar to the first of two value iteration algorithms suggested by Federgruen in [Fed80], though our approach has some distinct characteristics: It is assumed in [Fed80] that the values g^v exist and are equal for all v; in particular, this assumption implies the ϵ-ergodicity for every $\epsilon > 0$. For our approach we do not need such an assumption. We can verify ϵ-ergodicity for an arbitrary given $\epsilon > 0$, or provide a proof for non-ergodicity (with a small gap) in a finite time.

The approach of [Fed80] is shown to converge, while we provide a bound in terms of the input parameters for the number of steps. Although this bound is doubly exponential in general, it becomes exponential in case of the so-called *switching-controller* games and pseudo-polynomial for the subcase when all but a fixed number of positions are fully controlled by only one of the two players player.

Vrieze [Vri80, Theorem 8.3.2] showed that if a stochastic game has a value $g^v = m$, which is constant independent of the initial state $v \in V$, then it has a value in ϵ-optimal stationary strategies for any $\epsilon > 0$. The above Theorem 2 generalizes this result, and provides a constructive proof.

Remark 2 *Interestingly, potentials appear in [Fed80] implicitly, as the differences of local values of positions, as well as in [HK66], as the dual variables to linear programs corresponding to the controlled Markov processes, which appear when a player optimizes his strategy against a given strategy of the opponent. Yet, the potential transformation is not considered explicitly in these papers.*

Related work. Several other algorithms for solving undiscounted zero-sum stochastic games in stationary strategies are surveyed by Raghavan and Filar; see Sections 4 (B) and 5 in [RF91]. The only algorithmic results that we are aware of that provide bounds on the running time for approximating the value of general (undiscounted) stochastic games are those given in [CMH08, HKL$^+$11]: in [CMH08], the authors provide an algorithm that approximates, within any factor of $\epsilon > 0$, the value of any stochastic game (in history dependent strategies) in time $(nN)^{nN}\,\text{poly}(\eta, \log\frac{1}{\epsilon})$. In [HKL$^+$11], the authors give algorithms for discounted and recursive stochastic games that run in time $2^{N^{O(N^2)}}\,\text{poly}(\eta, \log(\frac{1}{\epsilon}))$, and claim also that similar bounds can be obtained for general stochastic games, by reducing them to the discounted version using a discount factor of $\delta = \epsilon^{\eta N^{O(n^2)}}$ (and this bound on δ is almost tight [Mil11]). Thus, while Theorem 2 provides only an approximation scheme (the running time is polynomial in $\frac{1}{\epsilon}$) rather than $\log\frac{1}{\epsilon}$), it depends only polynomially on N (similar to [HKL$^+$11], but unlike [CMH08]). Furthermore, the algorithm in Theorem 2 exhibits the additional feature that it either provides a solution in stationary strategies in the ergodic case, if one exists, or produces a pair of stationary strategies that witness the non-ergodicity.

2 Pumping Algorithm

Given a subset $S \subseteq V$, let us denote by $e_S \in \{0,1\}^V$ the characteristic vector of S.

Let us further assume that $m^v(x)$ for $v \in V$ are functions depending on *potentials* $x \in \mathbb{R}^n$ (where $n = |V|$) and satisfying the following properties for all subsets $S \subseteq V$ and reals $\delta \geq 0$:

(i) $m^v(x - \delta e_S)$ is a monotone decreasing function of δ if $v \in S$;

(ii) $m^v(x - \delta e_S)$ is a monotone increasing function of δ if $v \notin S$;

(iii) $|m^v(x) - m^v(x - \delta e_S)| \leq \delta$ for all $v \in V$.

We show in this section that under the above conditions we can change iteratively the potentials to some x' such that either all values $m^v(x')$, $v \in V$ are very close to one another or we can find a decomposition of the states V into disjoint subsets proving that such convergence of the values is not possible.

Algorithm 1 PUMP(x, S)

Input: a stochastic game Γ a subset S of states.

Output: a potential $x \in \mathbb{R}^S$.

1: Initialize $\tau := 0$, and $x_\tau := x$.

2: Set $m^+ := \max_{v \in S} m^v(x_\tau)$, $m^- := \min_{v \in S} m^v(x_\tau)$, and $\delta := (m^+ - m^-)/4$.

3: Define

$$\begin{aligned} T_\tau &:= \{v \in S \mid m^v(x_\tau) \geq m^- + 3\delta\} \\ B_\tau &:= \{v \in S \mid m^v(x_\tau) \leq m^- + \delta\} \\ M_\tau &:= S \setminus (T_\tau \cup B_\tau). \end{aligned}$$

4: **if** $T_\tau = \emptyset$ or $B_\tau = \emptyset$ **then**

5: **return** x_τ

6: **end if**

7: Otherwise, set $P_\tau := \{v \in S \mid m^v(x_\tau) \geq m^- + 2\delta\}$ and update

$$x^v_{\tau+1} := \begin{cases} x^v_\tau - \delta & \text{if } v \in P_\tau \\ x^v_\tau & \text{otherwise.} \end{cases}$$

8: Set $\tau := \tau + 1$ and Goto step 3.

We can show next that properties (i), (ii) and (iii) above guarantee some simple properties for the above procedure.

Lemma 1 *We have $T_{\tau+1} \subseteq T_\tau$, $B_{\tau+1} \subseteq B_\tau$ and $M_{\tau+1} \supseteq M_\tau$ for all iterations $\tau = 0, 1, \ldots$*

Proof Indeed, by (i) and (iii) we can conclude that $m^v(x_\tau) \geq m^- + \delta$ holds for all $v \in P_\tau$. Analogously, by (ii) and (iii) $m^v(x_\tau) < m^- + 3\delta$ follows for all $v \notin P_\tau$. □

Lemma 2 *Either $T_\tau = \emptyset$ or $B_\tau = \emptyset$ for some finite τ, or there are nonempty disjoint subsets $I, F \subseteq S$, threshold τ_0, such that for every real Δ there exists a finite index $\tau(\Delta) \geq \tau_0$ such that*

(a) $m^v(x_\tau) \geq m^- + 2\delta$ for all $v \in I$ and $m^v(x_\tau) < m^- + 2\delta$ for all $v \in F$, and for all $\tau \geq \tau_0$;

(b) $x^u_\tau - x^v_\tau \geq \Delta$ for all $v \in I$ and $u \notin I$, and for all $\tau \geq \tau(\Delta)$;

(c) $x^v_\tau - x^u_\tau \geq \Delta$ for all $v \in F$ and $u \notin F$, and for all $\tau \geq \tau(\Delta)$.

Proof By Lemma 1 sets T_τ and B_τ can change only monotonically, and hence only at most $|S|$ times. Thus, if PUMP(x, S) does not stop in a finite number of iterations, then after a finite number of iterations the sets T_τ and B_τ will never change and all positions in T_τ remain always pumped, while all positions in B_τ will be never pumped again.

Assuming now that the pumping algorithm PUMP(x, S) does not terminate, let us define the subset $I \subseteq S$ as the set of all those positions which are always pumped with the exception of a finite number of iterations. Analogously, let F be the subset of all those positions that are never pumped with the exception of a finite number of the iterations. Since I and F are finite sets, there must exist a finite τ_0 such that for all $\tau \geq \tau_0$ we have $I \subseteq P_\tau$ and $F \cap P_\tau = \emptyset$, implying (a).

Let us next observe that all positions not in $I \cup F$ are both pumped and not pumped infinitely many times. Thus, since δ is a fixed constant, for every Δ there must exist an iteration $\tau(\Delta) \geq \tau_0$ such that all positions not in I are not pumped by at least Δ/δ many more times than those in I, and all positions not in F are pumped by at least Δ/δ many more times than those in F, implying (b) and (c). □

Let us next describe the use of PUMP(x, S) for repeatedly shrinking the range of the m^v values, or to produce some evidence that it is not possible. A simplest version is the following:

Algorithm 2 REPEATEDPUMPING(ϵ)

1: Initialize $h := 0$, and $x_h := 0 \in \mathbb{R}^V$.
2: Set $m^+(h) := \max_{v \in V} m^v(x_h)$ and $m^-(h) := \min_{v \in V} m^v(x_h)$.
3: If $m^+(h) - m^-(h) \leq \epsilon$ then STOP.
4: x^{h+1} :=PUMP(x_h, V); $h := h + 1$.
5: Goto step 2.

Note that by our above analysis, REPEATEDPUMPING either returns a potential transformation for which all m^v, $v \in V$ values are within an ϵ band, or returns the sets I and F as in Lemma 2 with arbitrary large potential differences from the other positions. In the next section we use a modification of these procedures for stochastic games, and show that those large potential differences can be used to prove that the game is not ϵ-ergodic.

3 Application of Pumping for Stochastic Games

As we argued in Section 1.1 when the game is 0-ergodic, there exists a potential transformation after which all local matrix game values at the positions will be equal to the game's value, and the locally optimal strategies of players 1 and 2 will form globally optimal strategies for them. A similar statement can be made for the existence of ϵ-ergodic solutions. We show in this section how to use REPEATEDPUMPING to find such optimal potential transformations, or to prove that the game is ϵ-ergodic.

The rest of this section is dedicated to the proof of Theorem 2. We shall give some necessary and sufficient conditions for ϵ-non-ergodicity, and consider a modified version of the pumping algorithm described in the previous section which will provide a constructive proof for the above theorem.

Let us first observe that the local value function of stochastic games satisfies the properties required to run the pumping algorithm described in the previous section.

Lemma 3 *For every subset $S \subseteq V$ and $\delta \geq 0$ and for all $v \in V$ we have*

$$\begin{array}{lllll} m^v(x) & \geq & m^v(x - \delta e_S) & \geq & m^v(x) - \delta \max_{k,\ell} \sum_{u \notin S} p^{vu}_{k\ell} \quad \text{if } v \in S, \\ m^v(x) & \leq & m^v(x - \delta e_S) & \leq & m^v(x) + \delta_{k,\ell} \sum_{u \in S} p^{vu}_{k\ell} \quad \text{if } v \notin S. \end{array}$$

Furthermore, the value functions $m^v(x)$ for $v \in V$ satisfy properties (i), (ii) and (iii).

Proof According to (4) we must have for all $\delta \geq 0$ that $A^v(x) \geq A^v(x - \delta e_S)$ for all $v \in S$ and $A^v(x) \leq A^v(x - \delta e_S)$ for all $v \notin S$ proving properties (i) and (ii). Property (iii) follows directly from the inequalities in the statement. □

The above lemma implies that procedures PUMP and REPEATEDPUMPING could, in principle, be used to find a potential transformation yielding an ϵ-ergodic solution. It does not offer however a way to discover ϵ-non-ergodicity. Towards this end, we need to find some sufficient and algorithmically achievable conditions for ϵ-non-ergodicity.

Let us first analyze 0-non-ergodicity of stochastic games (in stationary strategies).

Lemma 4 *A stochastic game is 0-non-ergodic if and only if it is ϵ-non-ergodic for some positive ϵ.*

Proof A stochastic game is 0-non-ergodic by definition if there exists a threshold σ, positions $v, u \in V$, and stationary strategies α and β for the players, such no matter what other strategy β' player 2 chooses the Markov chain resulting by fixing (α, β') has a value $> \sigma$ when using initial position $v_0 = v$ (guaranteeing for player 1 more than σ from v), and the Markov chain obtained by fixing (α', β) has a value $< \sigma$ when using initial position $v_0 = u$ (guaranteeing for player 2 less than σ from u). Since strategies α' and β' are chosen from a compact space, the above implies that there are $\sigma' > \sigma > \sigma''$ such that α guarantees for player 1 at least σ' from the initial position v, and β guarantees for player 2 at most σ'' from initial position u. Hence the game is ϵ-non-ergodic for any $\epsilon < \sigma' - \sigma''$. □

Lemma 5 *A stochastic game is ϵ-non-ergodic if there exist disjoint subsets of the positions $I, F \subseteq V$, reals a, b with $b - a \geq \epsilon$, and stationary strategies $\alpha = (\alpha^v \mid v \in I)$ for player 1 and $\beta = (\beta^u \mid u \in F)$ for player 2 such that*

(N1) $\alpha^v_k p^{vu}_{k\ell} = 0$ *for all* $v \in I$, $u \notin I$, $k \in K^v$ *and* $\ell \in L^v$,

(N2) $\beta^u_\ell p^{uw}_{k\ell} = 0$ *for all* $u \in F$, $w \notin F$, $\ell \in L(u)$ *and* $k \in K(u)$, *and*

(N3) *there exists a vector of potentials* $x \in \mathbb{R}^V$, *such that*

$$\min_{\beta'} (\alpha^v)^T A^v(x) \beta' \;>\; b \quad \text{and} \quad \max_{\alpha'} (\alpha')^T A^u(x) \beta^u \;<\; a$$

for all $v \in I$ *and* $u \in F$.

Proof Let us note that (N1) and (N3) imply that for all strategies β' of player 2 strategies (α, β') result in a Markov chain in which subset I induces one or more absorbing sets (that is no arc is leaving I with a positive probability), and in which all positions have values $> b$. Analogously, (N2) and (N3) imply that F will always induce an absorbing set with values $< a$. Hence choosing any positions $v \in I$ and $u \in F$ and strategies α and β prove ϵ-nonergodicity. □

Let us introduce notation for denoting upper bounds on the entries of the matrices, more precisely on the part of these entries which do not depend on negative potential differences. Specifically, define

$$\begin{aligned} \widetilde{a}^v_{k\ell}(x) &= \sum_{u \in V} p^{vu}_{k\ell} r^{vu}_{k\ell} \;+ \sum_{u \in V, x^u \leq x^v} p^{vu}_{k\ell}(x^v - x^u) \\ \widetilde{b}^v_{k\ell}(x) &= m^+ - \sum_{u \in V} p^{vu}_{k\ell} r^{vu}_{k\ell} \;- \sum_{u \in V, x^u \geq x^v} p^{vu}_{k\ell}(x^v - x^u) \end{aligned} \tag{8}$$

and define

$$\begin{aligned} R^v(x) &= \max_{k \in K^v, \ell \in L^v} (\widetilde{a}^v_{k\ell}(x)) && \text{if } m^v(x) \geq \tfrac{m^+ + m^-}{2}, \\ R^v(x) &= \max_{k \in K^v, \ell \in L^v} \left(\widetilde{b}^v_{k\ell}(x)\right) && \text{otherwise.} \end{aligned} \tag{9}$$

Note that

$$m^+ - \widetilde{b}^v_{k\ell}(x) \leq a^v_{k\ell}(x) \leq \widetilde{a}^v_{k\ell}(x) \quad \text{for all } v \in V,\ k \in K^v,\ \ell \in L^v \text{ and } x \in \mathbb{R}^V.$$

With this notation we can state a more constructive version of Lemma 5.

Lemma 6 *A stochastic game is ϵ-non-ergodic if there exist disjoint subsets $I, F \subseteq V$, reals a', b' with $b' - a' \geq 3\epsilon$, $a' < \frac{m^+ + m^-}{2}$, $b' \geq \frac{m^+ + m^-}{2}$, and potentials $x \in \mathbb{R}^V$ such that*

(N4) $m^v(x) > b'$ *for all* $v \in I$, *and* $m^u(x) < a'$ *for all* $u \in F$;

(N5) $x^u - x^v \geq L^v W R^v(x)^2/\epsilon$ *for all* $u \notin I$, *and* $v \in I$;

(N6) $x^u - x^v \geq K^v W R^v(x)^2/\epsilon$ *for all* $u \in F$, *and* $v \notin F$.

Proof We first show that (N4)-(N5) imply the existence of strategies α^v, $v \in I$ satisfying (N1) and (N3). We shall then observe that a similar argument can be applied to (N4) and (N6) to show the existence of strategies β^u, $u \in F$ such that those satisfy (N2) and (N3). Consequently, our claim will follow by Lemma 5.

Let us now fix a position $v \in I$ and denote respectively by $\bar{\alpha}$ and $\bar{\beta}$ the optimal strategies of players with respect to the payoff matrix $A^v(x)$. Denote further by $\widehat{\beta} = (1/L^v, 1/L^v, ..., 1/L^v)$ the uniform strategy for player 2, and set $K = \{k \in K^v \mid \sum_{u \notin I} \sum_{\ell \in L^v} p^{vu}_{k\ell} = 0\}$.

Let us then note that we have

$$\left(A^v(x)\widehat{\beta}\right)_k \leq \begin{cases} R^v(x) & \text{if } k \in K, \\ R^v(x) - \frac{R^v(x)^2}{\epsilon} & \text{otherwise,} \end{cases}$$

since the very negative entries of (N5) have at least $1/L^v W$ coefficient in rows which are not in K.

Since we had nonnegative rewards originally by (7), we must have $b > 0$. Thus by the optimality of $\bar{\alpha}$ and by the above inequalities we have

$$0 < b < m^v(x) \leq \bar{\alpha} A^v(x) \widehat{\beta} \leq R^v(x) - \left(\sum_{k \notin K} \bar{\alpha}_k\right) \frac{R^v(x)^2}{\epsilon}$$

implying that

$$\sum_{k \notin K} \bar{\alpha}_k < \frac{\epsilon}{R^v(x)}.$$

Since by (N4) we have $0 < a$, inequalities $\epsilon < m^v(x) \leq R^v(x)$ follow, and hence $\epsilon/R^v(x) < 1$ must hold, implying that the set K is not empty. Let us then denote by $\widetilde{\alpha}$ the truncated strategy defined by

$$\widetilde{\alpha}_k = \begin{cases} \dfrac{\bar{\alpha}_k}{\sum_{k \in K} \bar{\alpha}_k} & \text{if} \quad k \in K, \\ 0 & \text{if} \quad k \notin K. \end{cases}$$

With this we have for any $\beta' \in \mathcal{L}(\Gamma)$

$$\begin{aligned} b' < m^v(x) &\leq (\bar{\alpha} A^v(x)\beta' \\ &= (\widetilde{\alpha} A^v(x)\beta')\left(\sum_{k\in K}\bar{\alpha}_k\right) + \sum_{k\notin K}\bar{\alpha}_k\left(\sum_{\ell\in L^v} a^v_{k\ell}(x)\beta'_\ell\right) \\ &\leq (\widetilde{\alpha} A^v(x)\beta') + \left(\sum_{k\notin K}\bar{\alpha}_k\right) R^v(x) \\ &< (\widetilde{\alpha} A^v(x)\beta') + \epsilon. \end{aligned}$$

Let us then define $\alpha^v = \widetilde{\alpha}$ and repeat the same for all $v \in I$. Then, these strategies satisfy (N1) and (N3) with $b = b' - \epsilon$.

Let us next note that by adding a constant to a matrix game it changes its value with exactly the same constant. Furthermore, multiplying all entries by -1 and transposing it, changes its value by a factor of -1, interchanges the roles of row and column players, but leaves otherwise optimal strategies still optimal. Thus, we can repeat the above arguments for the matrices $B^u(x) = m^+E - A_v(x)^T$, where E is the matrix of all ones, and obtain the same way strategies β^u, $u \in F$ satisfying (N2) and (N3) with $a = a' + \epsilon$. This completes the proof of the lemma. $\square$

To create a finite algorithm to find sets I and F and potentials satisfying (N4)-(N6) we need to do some modifications in our procedures.

Let us replace in procedure PUMP, line7 by the following lines, where $\epsilon > 0$ is a prespecified parameter, and call the new procedure with the above modifications MODIFIED-PUMP(ϵ, x, S):

7a: Otherwise set $P_\tau := \{v \in S \mid m^v(x_\tau) \geq m^- + 2\delta\}$ and compute

$$\begin{aligned} R^v_\tau &:= \max_{k\in K^v, \ell\in L^v} (\widetilde{a}^v_{k\ell}(x_\tau)) && \text{if} \quad v \in P_\tau, \\ R^v_\tau &:= \max_{k\in K^v, \ell\in L^v} \left(\widetilde{b}^v_{k\ell}(x_\tau)\right) && \text{if} \quad v \notin P_\tau, \end{aligned}$$

where $\widetilde{a}$ and $\widetilde{b}$ are defined by (8).

7b: Create an auxiliary directed graph $G = (V, E)$ on vertex set V such that $(v, u) \in E$ iff

$$\begin{aligned} x^u_\tau - x^v_\tau &< \tfrac{L^v W (R^v_\tau)^2}{\epsilon} && \text{if} \quad v \in P_\tau, \\ x^v_\tau - x^u_\tau &< \tfrac{K^v W (R^v_\tau)^2}{\epsilon} && \text{if} \quad v \notin P_\tau. \end{aligned}$$

7c: Find subsets I_τ and F_τ of V such that $I_\tau \subseteq P_\tau$, $F_\tau \cap P_\tau = \emptyset$, and no arcs are leaving these sets in G.

7d: if such sets are found STOP and output these sets, otherwise continue with step 8.

Before starting to analyze this modified pumping algorithm, let us observe that we have for all iterations

$$m^- < m^- + \frac{\epsilon}{2} < \frac{m^- + m^+}{2} \le m^v(x_\tau) \le R^v_\tau \quad \text{for all} \quad v \in P_\tau \tag{10}$$

as long as $m^+ - m^- > \epsilon$.

Lemma 7 *Procedure* MODIFIEDPUMP(ϵ, x, S) *terminates in a finite number of steps.*

Proof Let us observe that by Lemma 2 procedure PUMP would either terminate with $T_\tau = B_\tau = \emptyset$ for some finite τ, or would construct the sets I and F after infinitely many steps. Note that $R^v_\tau \le R^v_{\tau_0}$ for all $v \in I \cup F$, and hence we have $I_\tau = I$ and $F_\tau = F$ for all iterations $\tau \ge \tau(NWQ^2/\epsilon) \ge \tau_0$ where N is the maximum of L^vs and K^vs, and Q is the maximum of R^vs over $v \in I \cup F$. Thus, MODIFIEDPUMP will indeed find some sets I_τ and F_τ and terminate for some finite τ. □

Lemma 8 *Procedure* MODIFIEDPUMP(ϵ, x, V) *either shrinks the* m*-range by a factor of* $3/4$ *or outputs potentials* $x = x_\tau$ *and sets* $I = I_\tau$ *and* $F = F_\tau$ *which satisfy conditions (N4)-(N6) with* $a = b$.

Proof When the procedure terminates without shrinking the m-range, then it outputs sets $I = I_\tau$ and $F = F_\tau$ such that in the auxiliary graph G there are no arcs leaving these sets. Since $I \subseteq P_\tau$ and $F \subseteq V \setminus P_\tau$, condition (N4) holds with $a = b = (m^+ + m^-)/2$. Furthermore, the lack of leaving arcs in G implies that for all (v, u), $v \in I$ and $u \notin I$ and also for all (u, v) with $u \in F$ and $v \notin F$ we must have the reverse inequalities in (7b), implying that conditions (N5) and (N6) hold. □

Let us observe that the bounds and strategies obtained by Lemmas 7 and 8 do not necessarily imply the non-ergodicity of the game since those value ranges of I_τ and F_τ may overlap around $\frac{m^- + m^+}{2}$. To fix this we need to make one more use of the pumping algorithm, as described in the MODIFIEDREPEATEDPUMPING procedure below.

Theorem 3 MODIFIEDREPEATEDPUMPING(ϵ) *terminates in a finite number* $h \le \log \frac{R}{24\epsilon} / \log \frac{8}{7}$, *of iterations, and either provides a potential transformation proving that the game is* 24ϵ*-ergodic, where or outputs two nonempty subsets* $I = I$ *and* F *and strategies* α^v, $v \in I$ *for player 1 and* β^v, $v \in F$ *for player 2 such conditions (N4), (N5) and (N6) hold.*

Algorithm 3 MODIFIEDREPEATEDPUMPING(ϵ)

```
1: Initialize h := 0, and x_h := 0 ∈ ℝ^V.
2: Set m^+(h) := max_{v∈V} m^v(x_h) and m^-(h) := min_{v∈V} m^v(x_h).
3: if m^+(h) − m^-(h) ≤ 24ε then
4:     return x_h.
5: end if
6: x_{h+1} :=ModifiedPump(ε, x_h, V) and let F_τ, I_τ, T_τ, B_τ, P_τ be the sets obtained from
   ModifiedPump.
7: if T_τ = ∅ or B_τ = ∅ then
8:     Set h := h + 1 and Goto step 2
9: end if
10: Otherwise set F = F_τ and I = I_τ.
11: x_{h+1} :=ModifiedPump(ε, x_h, P_τ).
12: if T_τ = ∅ then
13:     Set h := h + 1 and Goto step 2.
14: end if
15: if B_τ = ∅ then
16:     Goto step 19
17: end if
18: Otherwise, update I := I_τ.
19: return x^{h+1} and the sets I and F.
```

Proof Let us note that if $T_\tau = \emptyset$ after the second MODIFIEDPUMP call, then the range of the values has shrunk by a factor of $\frac{7}{8}$ (at least), while if this happens in the first stage the m-range has shrunk by a factor of 3/4.

On the other hand if the m-range is not shrinking, and we have $B_\tau = \emptyset$ after the the second call of MODIFIEDPUMP, then we have $m^v(x_\tau) \geq \frac{5}{8}m^+ + \frac{3}{8}m^- = b$ for all $v \in I$, while $m^u(x_\tau) \leq (m^+ + m^-)/2 = a$ for all $u \in F$, and hence (N4)-(N6) hold with these a and b values. Since the m-range has not shrunk, we must have $m^+ - m^- > 24\epsilon$, and hence $b - a > \epsilon$ follows.

Finally, if the m-range is not shrinking, and the second call returns a new set I_τ, then all values of this set are above $\frac{3}{4}m^+ + \frac{1}{4}m^- > \frac{5}{8}m^+ + \frac{3}{8}m^- = b$, and with the same set F we can conclude again that conditions (N4)-(N6) hold. □

To complete the proof of Theorem 2, we need to analyze the time complexity of the above procedure, in particular, bounding the number of pumping steps performed in MODIFIED-PUMP.

Let us note that as long as $m^+ - m^- > 24\epsilon$ we pump the upper half P_τ by at least $\delta \geq 6\epsilon$.

Let us next sort the positions $v \in V$ such that we have

$$x_\tau^{v_1} \leq x_\tau^{v_2} \leq \cdots \leq x_\tau^{v_n},$$

and denote by $\Delta_j = x_\tau^{v_{j+1}} - x_\tau^{v_j}$ for $j = 1, 2, ..., n-1$.

Then, by (4) we have for $j = 0, 1, 2, ..., n-1$ that

$$R_\tau^{v_{j+1}} \leq R + \sum_{i=1}^{j} \Delta_i, \tag{11}$$

where the empty sum is zero by definition.

Assume now that the pumping procedure terminates with outputting subsets I_τ and F_τ, as above. Let us then denote by i the largest index such that $v_i \in I_\tau$. We know that for all $u \notin I_\tau$ we must have $x_\tau^u - x_\tau^{v_i} > \frac{L_{v_i} W(R^{v_i}(x_\tau))^2}{\epsilon}$, and thus $i < n$, and we must have

$$\Delta_i > \frac{L_{v_i} W\left(R_{v_i}(x_\tau)\right)^2}{\epsilon}.$$

Thus, as long as we have

$$\Delta_i \leq \frac{L_{v_i} W\left(R_{v_i}(x_\tau)\right)^2}{\epsilon}$$

for all $i = 1, 2, ..., n-1$ pumping cannot terminate with producing the sets I_τ and F_τ. We can provide an upper bound on the number of iterations one can do and still satisfy the above inequalities by solving the recursion

$$\Delta_i \leq \frac{L_{v_i} W\left(R_{v_i}(x_\tau)\right)^2}{\epsilon} \leq \frac{NW(R + \sum_{j=1}^{i-1} \Delta_j)^2}{\epsilon} \tag{12}$$

from which

$$\tau \le \frac{x_\tau^{v_n} - x_\tau^{v_1}}{6\epsilon} \le \frac{(NW)^n(nR)^{2^n}}{6\epsilon^{n+1}}.$$

Since the same bound can be developed from the other side for the set F, we can conclude that ModifiedRepeatedPumping(ϵ) must terminate in at most $\frac{(NW)^n(nR)^{2^n}}{3\epsilon^{n+1}}$ iterations, either producing $m^+ - m^- \le 24\epsilon$ or outputting the subsets I_τ and F_τ proving ϵ-non-ergodicity. This completes the proof of Theorem 2. □

4 Closing Remarks

The so-called *switching controller* stochastic games were introduced by Filar [Fil81]. In this special case, the set of positions is partitioned into two: $V = V_1 \cup V_2$, where it is assumed that at each position only one of the two players controls the transition probabilities, i.e., if $V \in V_1$ (resp., $V \in V_2$), then $p_{k\ell}^{vu} = p_{vu}^k$ (resp., $p_{k\ell}^{vu} = p_{vu}^\ell$). Interestingly, in this case, it was shown in [Fil81] that a saddle point exists in stationary strategies, and that it is satisfies the *order filed property*, i.e., there exist rational equilibrium strategies and values, if all the rewards and transition probabilities are rational. Let W' denote the least common denominator of the rational probabilities $p_{k\ell}^{vu}$, and $k = |S|$ be the number of positions $v \in S \subseteq V$ for which both $|L^v| > 1$ and $|K^v| > 1$ (i.e., positions controlled by both players). Using techniques similar to the ones in [BEGM10], we can show that the running time of the algorithm of Theorem 2 can be improved to $\left(\frac{NW'nR}{\epsilon}\right)^{O(k)}$ poly(η), which is pseudo polynomial for constant k. Furthermore, with an accuracy of $\epsilon \approx \left(\frac{1}{nNW'}\right)^{O(kh)}$, where $h = \sum_{v\in S} |L^v||K^v|$, we can get an exact solution in time $(NW'nR)^{O(kh)}$ poly(η).

References

[BEGM10] E. Boros, K. Elbassioni, V. Gurvich, and K. Makino. A pumping algorithm for ergodic stochastic mean payoff games with perfect information. In *Proc. 14th IPCO*, pages 341–354, 2010.

[BF68] D. Blackwell and T.S. Ferguson. The big match. *Ann. Math. Statist.*, 39(1):159–163, 1968.

[CMH08] K. Chatterjee, R. Majumdar, and T. A. Henzinger. Stochastic limit-average games are in exptime. *Int. J. Game Theory*, 37:219–234, 2008.

[Fed80] A. Federgruen. Successive approximation methods in undiscounted stochastic games. *Operations Research*, 1:794–810, 1980.

[Fil81] J. A. Filar. Ordered field property for stochastic games when the player who controls transitions changes from state to state. *J. of Optimization Theory and Applications*, 34(4):503–515, 1981.

[Gal58] T. Gallai. Maximum-minimum Sätze über Graphen. *Acta Mathematica Academiae Scientiarum Hungaricae*, 9:395–434, 1958.

[Gil57] D. Gillette. Stochastic games with zero stop probabilities. In M. Dresher, A. W. Tucker, and P. Wolfe, editors, *Contribution to the Theory of Games III*, volume 39 of *Annals of Mathematics Studies*, pages 179–187. Princeton University Press, 1957.

[HK66] A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Science, Series A*, 12(5):359–370, 1966.

[HKL+11] K. A. Hansen, M. Koucky, N. Lauritzen, P. B. Miltersen, and E. P. Tsigaridas. Exact algorithms for solving stochastic games: extended abstract. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 205–214, New York, NY, USA, 2011. ACM.

[KS63] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. Springer, 1963.

[Mil11] P. B. Miltersen. Discounted stochastic games poorly approximate undiscounted ones, manuscript. Technical report, 2011.

[MN81] J.F. Mertens and A. Neyman. Stochastic games. *Int. J. Game Theory*, 10:5366, 1981.

[Mou76] H. Moulin. Prolongement des jeux à deux joueurs de somme nulle. *Bull. Soc. Math. France, Memoire*, 45, 1976.

[RF91] T. E. S. Raghavan and J. A. Filar. Algorithms for stochastic games a survey. *Mathematical Methods of Operations Research*, 35(6):437–472, 1991.

[Sha53] L. S. Shapley. Stochastic games. *Proc. Nat. Acad. Science, USA*, 39:1095–1100, 1953.

[Vri80] O. J. Vrieze. *Stochastic games with finite state and action spaces*. PhD thesis, Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands, 1980.