Monomial Strategies for Concurrent Reachability Games and Other Stochastic Games*

Søren Kristoffer Stiil Frederiksen and Peter Bro Miltersen

Aarhus University

Abstract. We consider two-player zero-sum finite (but infinite-horizon) stochastic games with limiting average payoffs. We define a family of stationary strategies for Player I parameterized by $\varepsilon > 0$ to be *monomial*, if for each state k and each action j of Player I in state k except possibly one action, we have that the probability of playing j in k is given by an expression of the form $c\varepsilon^d$ for some non-negative real number c and some non-negative integer d. We show that for all games, there is a monomial family of stationary strategies that are ε-optimal among stationary strategies. A corollary is that all concurrent reachability games have a monomial family of ε-optimal strategies. This generalizes a classical result of de Alfaro, Henzinger and Kupferman who showed that this is the case for concurrent reachability games where all states have value 0 or 1.

1 Introduction

We consider two-player zero-sum finite (but infinite-horizon) stochastic games G with state set $\{1, 2, \ldots, N\}$ and set of actions $\{1, 2, \ldots, m\}$ available to each of the two players in each state. The reward to Player I when Player I plays i and Player II plays j in state k is denoted a^k_{ij}. Transition probabilites are denoted p^{kl}_{ij}. We assume stopping probabilitites are 0, i.e., for all k, i, j we have $\sum_l p^{kl}_{ij} = 1$. We are interested in games with limiting average (undiscounted) payoffs [8,12], i.e, payoff $\liminf_{T\to\infty}(\sum_{i=0}^{t-1} r_t)/T$ to Player I, where r_t is the reward collected by Player I at stage t. A *stationary strategy* x for a player in a stochastic game is a fixed (time independent) assignment of probabilities to his actions, for each of the states of the game. We let x^k_j denote the probability of playing action j in state k according to stationary strategy x. We denote the set of stationary strategies for Player I (II) by S_{I} (S_{II}). For a state k, the *lower value in stationary strategies* of k, denoted $\underline{v}_k$, is defined as $\sup_{x\in S_{\text{I}}} \inf_{y\in S_{\text{II}}} u_k(x, y)$, where $u_k(x, y)$ is the expected limiting average payoff when stationary strategy x of Player I

* The authors acknowledge support from The Danish National Research Foundation and The National Science Foundation of China (under the grant 61061130540) for the Sino-Danish Center for the Theory of Interactive Computation and from the Center for re- search in the Foundations of Electronic Markets (CFEM), supported by the Danish Strategic Research Council.

P.A. Abdulla and I. Potapov (Eds.): RP 2013, LNCS 8169, pp. 122–134, 2013.

is played against stationary strategy y of Player II and play starts in state k. Given $\varepsilon > 0$, a stationary strategy x^* for Player I is called *ε-optimal among stationary strategies* if for all states k, we have $\inf_{y \in S_{\mathrm{II}}} u_k(x^*, y) \geq \underline{v}_k - \varepsilon$. Notice that when Player I has fixed his stationary strategy, Player II is just playing a Markov decision process, so he has an optimal positional reponse.

The main purpose of the present paper is to prove that all stochastic games have a family of ε-optimal strategies among stationary strategies of a particular regular kind. We introduce the following definition.

Definition 1. A family of stationary strategies $(x_\varepsilon)_{0<\varepsilon\leq\varepsilon_0}$ for Player I in a stochastic game is called *monomial* if for all states k, and all actions j available to Player I in state k *except possibly one action*, we have that $x^k_{\varepsilon,j}$ is given by a monomial in ε, i.e., an expression of the form $c^k_j \varepsilon^{d^k_j}$, where d^k_j is a non-negative integer and c^k_j is a non-negative real number.

The exception made in the definition for some single action in each state is natural and necessary: The sum of probabilities assigned to the actions in each state must be 1, so without this exception, it is easy to see that a monomial family would have $d^k_j = 0$ for all j, k, i.e., it would be a single strategy rather than a family. Also note that when we specify a monomial family of strategies, we do not have to specify the probability assigned to the "special" action in each state, as it is simply the result of subtracting the sum of the probabilities assigned to the remaining actions from one. We can now state our main theorem:

Theorem 1. *For any game G, there is an $\varepsilon_0 > 0$ and a monomial family of stationary strategies $(x_\varepsilon)_{0<\varepsilon\leq\varepsilon_0}$ for Player I, so that for each $\varepsilon \in (0, \varepsilon_0]$, we have that x_ε is ε-optimal among stationary strategies.*

Discussion of the Main Theorem. A monomial family of strategies can be naturally interpreted as a parameterized strategy where probabilities have well-defined "orders of magnitude", given by the degrees d^k_j. Our main theorem informally states that such "clean" strategies are sufficient for playing stochastic games well, at least if one is restricted to the use of stationary strategies. Our main motivation for the theorem is computational: A monomial family of strategies is a *finite* object, and our theorem makes it possible to *ask* the question of whether a family of ε-optimal strategies parameterized by ε can be efficiently computed for a given game, as the result makes this question well-defined. The existence proof of the present paper is essentially non-constructive and provides no efficient algorithm (although it is possible to derive an inefficient algorithm using standard techniques), so we do not answer the question in this paper. It should also be noted that it is easy to give examples of games with rational rewards and transition probabilties where the coefficients c^k_j cannot be rational numbers, so one has to worry about how to represent those. Fortunately, a straightforward application of the Tarski transfer principle yields that algebraic coefficients suffice, and such a number has a finite representation in the form of a univariate polynomial with rational coefficients and an isolating interval within which the number is the only root of the polynomial.

Our main theorem is particularly natural for classes of stochastic games that are guaranteed to have a *value* in stationary strategies, that is, games for which the lower value $\sup_{x \in S_{\mathrm{I}}} \inf_{y \in S_{\mathrm{II}}} u_k(x, y)$ and the *upper* value $\inf_{y \in S_{\mathrm{II}}} \sup_{x \in S_{\mathrm{I}}} u_k(x, y)$ coincide. A natural subclass of stochastic games with this property is Everett's *recursive* games [6]. In a recursive game, all non-zero rewards occur at absorbing states: states k with only one action "1" available to each player and $p_{1,1}^{kk} = 1$ ("terminal states"). Everett presents several examples of families of ε-optimal strategies for natural recursive games and upon inspection, we note that they are monomial. An interesting subclass of recursive games widely studied in the computer science literature [5,3,11,9] is the class of *concurrent reachability games*. In a concurrent reachability game, Player I is trying to reach a distinguished "goal" state and Player II is trying to prevent him from reaching this state. To view such a game as a recursive game, we simply interpret the goal state as an absorbing state g with reward $r_{1,1}^{g} = 1$. Then, the (lower) value $\underline{v}_k$ of a state k is naturally interpreted as the optimal probability of reaching the goal state from k. De Alfaro, Henzinger and Kupferman [5] presented a polynomial time algorithm for deciding which states in a concurrent reachability game have value 1. Inspecting their proof of correctness, we see that it yields an explicit construction of a monomial family of ε-optimal strategies for Player I if the concurrent reachability games satisfy the (very restrictive) property that each state has value either 0 or 1. Note that even this case requires non-trivial strategies for near optimal play [11]. Also, their polynomial time algorithm can easily be adapted to output this strategy. It is interesting to note that in the computed strategy, all coefficients c_j^k are either 0 or 1.

Discussion of the Proof. Our proof relies heavily on semi-algebraic geometry. In this respect, the proof technique is much in line with classical works on stochastic games, in particular the work of Bewley and Kohlberg [1], and semi-algebraic geometry has seen several uses in stochastic games, see for example [13,4,15,10]. Our proof can be outlined as follows. First, we show that it is possible in first order logic over the reals to uniquely define a particular distinguished ε-optimal strategy among stationary strategies, with ε being a free variable in this definition. Then, standard theorems of semi-algebraic geometry imply that there is a family of ε-optimal strategies the probabilities of which can be described as Pusieux series in the parameter $\varepsilon > 0$. We then "round" these series to their most significant terms and finally massage them into monomials. To argue that ε-optimality is not lost in the process, we appeal to theorems upper bounding the sensitivity of the limiting average values of Markov chains to perturbations of their transition probabilities. These sensitivity theorems are due to Solan [14], building on work on Freidlin and Wentzell [7]. As our main theorem is very simply stated, one might speculate that it has an elementary proof, avoiding the use of semi-algebraic geometry. However, we are not aware of any such proof, even for the case of concurrent reachability games. It should be noted that the proof by De Alfaro, Henzinger and Kupferman is combinatorial in nature, and does not rely on semi-algebraic geometry, so at least for the simpler case considered by them, elementary arguments do exist.

Organization of Paper. In section 2 we will introduce the definitions, lemmas and previous results necessary for the proof. In section 3 we prove a version of the main theorem with monomials replacing Puiseux series. In section 4 we prove the actual main theorem.

2 Preliminaries

For $n \in \mathbb{N}$, let $[n]$ denote $\{1, \ldots, n\}$. A *Puiseux series* p over some indeterminate T and field $\mathbb{F}$ is an expression of the form $p = \sum_{i=K}^{\infty} a_i T^{\frac{i}{M}}$ where $K \in \mathbb{Z}, M \in \mathbb{N}$, and for all i, $a_i \in \mathbb{F}$, with the expression satisfying that if $p \neq 0$ then there $\exists i \in \mathbb{Z} : a_i \neq 0 \wedge \gcd(i, M) = 1$. Similarly, a function $p : \mathbb{R} \to \mathbb{R}$ is a *Puiseux function* on an interval I, if there exists $K \in \mathbb{Z}, M \in \mathbb{N}, a_i \in \mathbb{R}$ such that $p(\epsilon) = \sum_{i=K}^{\infty} a_i \epsilon^{\frac{i}{M}}$ for all $\epsilon \in I$. In the context of this paper we will only look at Puiseux functions, and we will often call the function $p(\epsilon)$ a Puiseux series. The *order* of a Puiseux series $p = \sum_{i=K}^{\infty} a_i T^{\frac{i}{M}}$ is the smallest integer i such that $a_i \neq 0$, and we will write $\mathrm{ord}(p) = i$. If $p = 0$ then the order is defined to be ∞. The proofs of the following elementary lemmas on Puiseux series are easy and we omit them.

Lemma 1. *if $q(\epsilon) = \sum_{i=K}^{\infty} c_i \epsilon^{\frac{i}{M}}$ is a Puiseux series that is convergent and bounded on some $(0, \epsilon_0)$, then $c_i = 0$ for all $i < 0$. In other words, the order of q is greater than or equal to 0.*

Lemma 2. *For any Puiseux series $q(\epsilon) = \sum_{i=K}^{\infty} c_i \epsilon^{\frac{i}{M}}$ with $ord(q) = K \geq 0$ there exists an ϵ_0 such that $sign(q(\epsilon)) = sign(c_K)$ for all $\epsilon \in (0, \epsilon_0)$.*

A *semi-algebraic set* is a subset of real Euclidean space defined by a finite set of polynomial equalities and inequalities. The well-known *Tarsi-Seidenberg theorem* states that any set that can be defined in the language of first order arithmetic is semi-algebraic. We will use this theorem throughout this paper to establish that sets are semi-algebraic. A *semi-algebraic function* is a real-valued function whose graph is a semi-algebraic set. We shall use the following lemma, establishing a close relationship between semi-algebraic functions and Puiseux functions.

Lemma 3. *[13, lemma 6.2] Let $a > 0$, if $f : (0, a) \to \mathbb{R}$ is a semi-algebraic function, then there exists an $0 < \epsilon' < a$ such that f is a Puiseux function on $(0, \epsilon')$.*

For stochastic games, we use the notation introduced in the introduction. We shall use the following theorem, due to Solan, as an important lemma. The theorem applies to *1-player* stochastic games (a.k.a., Markov decision processes). In a 1-player stochastic game, Player 2 has only a single action in each state. We therefore write p_i^{kl} rather than p_{ij}^{kl} for the transition probabilities.

Theorem 2. *[14, theorem 6] Let G and $\tilde{G}$ be 1-player stochastic games with identical state set $\{1, 2, \ldots, N\}$, transition probabilities p_i^{kl}, $\tilde{p}_i^{kl}$ and identical*

rewards. Let c be an upper bound on the absolute value of all rewards. Let $\underline{v}$, $\underline{\tilde{v}}$ be the lower value in stationary strategies in each of the games. Let $\delta \in (0, \frac{1}{2N})$ satisfy $\max_{i,k,l}(\frac{p_i^{kl}}{\tilde{p}_i^{kl}}, \frac{\tilde{p}_i^{kl}}{p_i^{kl}}) - 1 \leq \delta$, where $\frac{x}{0} := \infty, \frac{0}{0} := 1$. Then, $|\underline{v} - \underline{\tilde{v}}| \leq 4cN\delta$.

3 Puiseux Family of Strategies

Lemma 4. *For any game G there exists an ϵ_0 and a family of stationary strategies $(x_\epsilon)_{0<\epsilon\leq\epsilon_0}$ that are ϵ-optimal among stationary strategies, where for all states k and all actions j, $x^k_{\epsilon,j}$ is given by a Puiseux series in ϵ, that is, there is an expression $q^k_j(\epsilon) = \sum_{i=K^k_j}^{\infty} c^k_{i,j}\epsilon^{\frac{i}{M^k_j}}$ such that $x^k_{\epsilon,j} = q^k_j(\epsilon)$ for $\epsilon \in (0, \epsilon_0]$.*

Proof. We want to create a first-order formula $\Phi^k_j(x, \epsilon)$ for every state k and every action j, which is true if and only if x is the probability that Player I should play action j in state k in a specific strategy that is ϵ-optimal among stationary strategies. Then, since we have described the function by a first-order formula, it is semi-algebraic, and by Lemma 3 we get that there exists a Puiseux series that is equal to the function, thus completing the proof. We are going to use several smaller first-order formulas to describe the formulas $\Phi^k_j(x, \epsilon)$.

To ease notation, during the proof k, l will only be refering to states in the game, so they will be numbers $k, l \in [N]$. i, j will be refering to actions in a given state, so they will be numbers $i, j \in [m]$. We will also use the following vectors

$$\boldsymbol{x} := (x_i^k)_{i\in[m]}^{k\in[N]} \quad , \quad \boldsymbol{y} := (y_i^k)_{i\in[m]}^{k\in[N]} \quad , \quad \boldsymbol{v} := (v^k)^{k\in[N]} \quad , \quad \boldsymbol{\nu} := (\nu^k)^{k\in[N]}$$

$\boldsymbol{x}$ and $\boldsymbol{y}$ will represent the strategies of Player I and Player II respectively, while $\boldsymbol{v}$ and $\boldsymbol{\nu}$ will be used to represent different values of stationary strategies of the game starting in each position.

The first two formulas $\Delta_\alpha(\boldsymbol{x})$, $\Delta_\beta(\boldsymbol{y})$ describe that $\boldsymbol{x}$ is a stationary strategy and $\boldsymbol{y}$ is a stationary strategy respectively.

$$\Delta_\alpha(\boldsymbol{x}) := \bigwedge_{k\in[N],i\in[m]} \left[x_i^k \geq 0\right] \wedge \bigwedge_{k\in[m]} \left[\sum_{i\in[m]} x_i^k = 1\right]$$

$$\Delta_\beta(\boldsymbol{y}) := \bigwedge_{k\in[N],i\in[m]} \left[y_i^k \geq 0\right] \wedge \bigwedge_{k\in[N]} \left[\sum_{i\in[m]} y_i^k = 1\right]$$

Next we want to create a first-order formula $\Psi(\boldsymbol{v})$ which expresses that v^k is the lower value in stationary strategies when the game starts in state k, that is, the quantity:

$$\sup_{\boldsymbol{x}\in S_I} \inf_{\boldsymbol{y}\in S_{II}} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \liminf_{T\to\infty} \sum_{t=0}^{T-1} \frac{r_t}{T}$$

We can rewrite this quantity by using the following equations proved in [2, Theorem 5.2]

$$\inf_{\boldsymbol{y}\in S_{II}} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \liminf_{T\to\infty} \sum_{t=0}^{T-1} \frac{r_t}{T} = \inf_{\boldsymbol{y}\in S_{II}} \liminf_{\lambda\to 0} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \frac{\lambda}{1+\lambda} \sum_{t=0}^{\infty} \frac{1}{(1+\lambda)^t} r_t \quad , \forall \boldsymbol{x} \in S_I$$

So the suprema over the two sets are the same, and we can express the value by creating a formula which express that

$$v^k = \sup_{\boldsymbol{x}\in S_I} \inf_{\boldsymbol{y}\in S_{II}} \liminf_{\lambda\to 0} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \frac{\lambda}{1+\lambda} \sum_{t=0}^{\infty} \frac{1}{(1+\lambda)^t} r_t \quad \forall k \in [N]$$

A common way of rewriting these value equations is by expanding the expectations for one state and substituting v^l into the equations

$$v^k = \sup_{\boldsymbol{x}\in S_I} \inf_{\boldsymbol{y}\in S_{II}} \liminf_{\lambda\to 0} \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}} \frac{\lambda}{1+\lambda} \sum_{t=0}^{\infty} \frac{1}{(1+\lambda)^t} r_t \quad \forall k \in [N]$$

$$\Leftrightarrow v^k = \sup_{\boldsymbol{x}\in S_I} \inf_{\boldsymbol{y}\in S_{II}} \liminf_{\lambda\to 0} \frac{\lambda}{1+\lambda} \sum_{i,j\in[m]} x_i^k y_j^k \left(a_{ij}^k + \sum_{l\in[N]} p_{ij}^{kl} \frac{1}{\lambda} v^l \right) \quad \forall k \in [N]$$

First notice that for any semi-algebraic sets A and B, and any function $f : A \to B$ where there is a formula $\Pi(a, b)$ that is true if and only is $f(a) = b$, we can express the supremum $\sup_{a\in A} f(a)$ in the following way

$$\begin{aligned} \Pi_{sup}(s) := &[\forall a \in A \exists b \in B : \Pi(a, b) \wedge s \geq b] \\ &\wedge [\forall \epsilon > 0 \exists a \in A \exists b \in B : \Pi(a, b) \wedge s < b + \epsilon] \end{aligned}$$

And similar formulas can be created for the infimum and the limit, and since $\liminf_{\lambda\to 0} f(\lambda)$ is $\lim_{\lambda'\to 0} \inf_{0<\lambda<\lambda'} f(\lambda)$, we only need to create a formula for the inner part:

$$\frac{\lambda}{1+\lambda} \sum_{i,j\in[m]} x_i^k y_j^k \left(a_{ij}^k + \sum_{l\in[N]} p_{ij}^{kl} \frac{1}{\lambda} v^l \right)$$

We then create the formula

$$\Pi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}, \lambda) := \bigwedge_{k\in[N]} \left[\nu^k = \frac{\lambda}{1+\lambda} \sum_{i,j\in[m]} x_i^k y_j^k \left(a_{ij}^k + \sum_{l\in[N]} p_{ij}^{kl} \frac{1}{\lambda} \nu^l \right) \right]$$

Since $S_I = \{\boldsymbol{x} \in \mathbb{R}^{Nm} | \Delta_\alpha(\boldsymbol{x})\}$, we have that S_I, S_{II} are semi-algebraic. Then from the previous argument we can create a formula $\Pi_{sup}(\boldsymbol{v})$ for the lower value in stationary strategies. Also, by not removing the last supremum, we can create a formula $\Xi(\boldsymbol{x}, \boldsymbol{v})$ that is true if the value of Player I playing strategy $\boldsymbol{x}$ is $\boldsymbol{v}$.

It is now straightforward to create a formula $\Upsilon(\boldsymbol{x}, \epsilon)$ that is true if and only if $\boldsymbol{x}$ is a stationary strategy that is ϵ-optimal among stationary strategies.

$$\Upsilon(\boldsymbol{x},\epsilon) := \exists \boldsymbol{v} \in \mathbb{R}^N \exists \boldsymbol{\nu} \in \mathbb{R}^N : \Lambda_\alpha(\boldsymbol{x}) \wedge (0 < \epsilon < 1) \\ \wedge \Pi_{sup}(\boldsymbol{v}) \wedge \Xi(\boldsymbol{x},\boldsymbol{\nu}) \bigwedge_{k\in[N]} \left[\nu^k \geq v^k - \epsilon\right]$$

Now to create $\Phi_j^k(x,\epsilon)$, we need to select a unique strategy from the set of stationary strategies that are ϵ-optimal among stationary strategies. Let $\varphi : [N]\times[m] \to [Nm]$ be some bijection, which we will use to get an ordering on the pairs consisting of an action i and a state k. Using this we can write a strategy as $(x_\iota)_{\iota\in[Nm]}$. We define formulas $P_\iota(x_1,\ldots,x_\iota,\epsilon)$ for $\iota \in [Nm]$ which are true if there exists a strategy that is ϵ-optimal among stationary strategies and the first ι entries are $(x_1,\ldots,x_\iota)$.

$$P_\iota(x_1,\ldots,x_\iota,\epsilon) := \exists x_{\iota+1},\ldots,x_{Nm} \in \mathbb{R} : \Upsilon(x_1,\ldots,x_\iota,x_{\iota+1},\ldots,x_{Nm},\epsilon)$$

Notice that for each $\iota \in [Nm]$, if we assume that we have chosen $x_1,\ldots,x_{\iota-1}$ such that $P_{\iota-1}(x_1,\ldots,x_{\iota-1},\epsilon)$ is true, then the set $\{x \in \mathbb{R} | P_\iota(x_1,\ldots,x_{\iota-1},x,\epsilon)\}$ is non-empty. From the Tarski-Seidenberg theorem the set is semi-algebraic, so it is defined by a finite set of polynomial equalities and inequalities. This implies that the set must consist of a finite set of intervals[1], so we can choose a unique strategy by the middle of the interval which lower endpoint is closest to 0. Using this observation, we can now create a new series of formulas $\Psi_\iota(x_1,\ldots,x_{\iota-1},x,\epsilon)$ for $\iota \in [Nm]$ which given that $P_{\iota-1}(x_1,\ldots,x_{\iota-1},\epsilon)$ is true, x is the middlepoint of the interval with the lower endpoint closest to 0 among the intervals in the set $\{x \in \mathbb{R} | P_\iota(x_1,\ldots,x_{\iota-1},x,\epsilon)\}$.

$$\begin{aligned}\Psi_\iota(x_1,\ldots,x_{\iota-1},x,\epsilon) := \exists x_{\iota+1},\ldots,x_{Nm},a,b \in \mathbb{R} : a \leq b \wedge x = \frac{a+b}{2} : \\ \Upsilon(x_1,\ldots,x_{\iota-1},x,x_{\iota+1},\ldots,x_{Nm},\epsilon) \\ \wedge [P_\iota(x_1,\ldots,x_{\iota-1},a,\epsilon) \vee (a < b \wedge \forall y \in (a,b) : P_\iota(x_1,\ldots,x_{\iota-1},y,\epsilon))] \\ \wedge [\forall y < a : \neg P_\iota(x_1,\ldots,x_{\iota-1},y,\epsilon)] \\ \wedge [\exists \epsilon > 0 \forall y \in (b,b+\epsilon) : \neg P_\iota(x_1,\ldots,x_{\iota-1},y,\epsilon)]\end{aligned}$$

Now to select our unique strategy we will do the following: For each ϵ, pick x_1 to be the middlepoint of the interval with the lower endpoint closest to 0 among the intervals in the set $\{x \in \mathbb{R} | P_\iota(x,\epsilon)\}$, next we pick x_2 to be the middlepoint of the interval with the lower endpoint closest to 0 among the intervals in the set $\{x \in \mathbb{R} | P_\iota(x_1,x,\epsilon)\}$, and so on. We can then recursively define new formulas $\Omega_\iota(x_1,\ldots,x_\iota,\epsilon)$ for $\iota \in [Nm]$ that are true if and only if the unique choice of the first ι indices described by the above procedure is exactly $x_1,\ldots,x_\iota$.

$$\Omega_1(x,\epsilon) := \Psi_1(x,\epsilon) \ , \ \Omega_\iota(x_1,\ldots,x_\iota,\epsilon) := \Omega_{\iota-1}(x_1,\ldots,x_{\iota-1},\epsilon) \wedge \Psi_\iota(x_1,\ldots,x_\iota,\epsilon)$$

Using this we can now immediately create the formulas $\Phi_\iota(x,\epsilon)$ for $\iota \in [Nm]$ in the following way:

$$\Phi_\iota(x,\epsilon) := \exists x_1,\ldots,x_{Nm} \in \mathbb{R} : \Omega_{Nm}(x_1,\ldots,x_{Nm},\epsilon) \wedge x = x_\iota$$

[1] In this terminology we allow for the interval $[a,a]$ and identify it with the point $\{a\}$.

Now we have obtained that each formula $\Phi_\iota(x,\epsilon)$ implicitly defines a semi-algebraic function $x_\iota(\epsilon)$ and due to Lemma 3 we have that there exists Puiseux series $q_\iota(\epsilon)$ and numbers ϵ_ι such that $x_\iota(\epsilon) = q_\iota(\epsilon)$ for $\epsilon \in (0,\epsilon_\iota)$. Now take $\epsilon_0 = \min_{\iota\in[Nm]} \epsilon_\iota$ and we have the lemma.

4 Proof of Main Theorem

The proof will be carried out in two steps. First we will use the family of strategies obtained from Lemma 4 to create a family of strategies only consisting of the first term of the Puiseux series of the original family. Then by using Theorem 2, we prove their value can not be much worse. Then finally we transform this family into a monomial family of strategies that are $\epsilon-$optimal among stationary strategies.

Proof (of Theorem 1). From Lemma 4 we know that there exists an ϵ_1 and a family of stationary strategies $(x_\epsilon)_{0<\epsilon\leq\epsilon_1}$ that are ϵ-optimal among stationary strategies such that $x^k_{\epsilon,j} = q^k_j(\epsilon) = \sum_{i=K^k_j}^{\infty} c^k_{i,j}\epsilon^{\frac{i}{M^k_j}}$ for $\epsilon \in (0,\epsilon_1]$ and for all states k and actions j. Assume without loss of generality that $K^k_j = \mathrm{ord}(q^k_j)$, and observe that K^k_j can be ∞ if the Puiseux series is identically 0. Also observe that since each $x^k_{\epsilon,j}$ is a probability, it is positive and bounded, so by Lemma 1 we know that all $K^k_j \geq 0$.

Now for each k, look at the set of Puiseux series $\{q^k_j(\epsilon)\}_{j\in[m]}$ and let j_k be an index so $q^k_{j_k}(\epsilon)$ is one of the Puiseux series in the set which has minimal order. Observe that $q^k_{j_k}(\epsilon)$ has order 0. To see this, assume for contradiction that $\mathrm{ord}(q^k_j) > 0$ for all actions j, then all of them behave as power series around 0, thus $q^k_j(\epsilon) \to 0$ for $\epsilon \to 0$ so the sum $\sum_{j\in[N]} q^k_j(\epsilon) \to 0$ for $\epsilon \to 0$, which contradicts that $\sum_{j\in[N]} q^k_j(\epsilon) = 1$ for all $\epsilon \in (0,\epsilon_1]$.

Now look at any k again. We want to approximate the family of strategies defined by $q^k_j(\epsilon)$ by a new family of strategies defined by finite Puiseux series $\rho^k_j(\epsilon)$ for $\epsilon \in (0,\epsilon_2]$, where ϵ_2 will be defined later. We define $\rho^k_j(\epsilon)$ as a conditional function on the following sets

$$
\begin{aligned}
S_1 &= \{(k,j) \in [N]\times[m] \mid \mathrm{ord}(q^k_j) = \infty\}\\
S_2 &= \{(k,j) \in [N]\times[m] \mid j \neq j_k \wedge \mathrm{ord}(q^k_j) \neq \infty\}\\
S_3 &= \{(k,j) \in [N]\times[m] \mid j = j_k\}
\end{aligned}
$$

Then $\rho^k_j(\epsilon)$ is defined as follows

$$
\rho^k_j(\epsilon) = \begin{cases} 0 & \text{if } (k,j) \in S_1 \\ c^k_{K^k_j,j}\epsilon^{\frac{K^k_j}{M^k_j}} & \text{if } (k,j) \in S_2 \\ 1 - \sum_{j\in S_2} c^k_{K^k_j,j}\epsilon^{\frac{K^k_j}{M^k_j}} & \text{if } (k,j) \in S_3 \end{cases}
$$

So $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ is the derived family of strategies from $q_j^k(\epsilon)$, defined by $\rho_j^k(\epsilon) \equiv 0$ when $q_j^k(\epsilon) \equiv 0$, and otherwise equal to the first term in $q_j^k(\epsilon)$ except for one action, $q_{j_k}^k(\epsilon)$ which is 1 minus the sum of the other probabilities, to ensure $\rho_j^k(\epsilon)$ is a probability distribution. Since $q_{j_k}^k(\epsilon)$ is a probability, then it is positive, so from Lemma 2 we have that for $(k,j) \in S_2$ the constant is positive. But then we can choose ϵ_2 to be small enough so that for all $(k,j) \in S_2$, $\rho_j^k(\epsilon) \leq 1$. So for each $k \in [N]$, $(\rho_j^k(\epsilon))_{j\in[m]}$ becomes a probability distribution.

We will use Theorem 2 to prove that the value of the game where Player I fixes his strategy to $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$, is not much different than the value of the game where Player I fixes his strategy to $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$. To do this, we must show that for all states k and all actions j, $\rho_j^k(\epsilon)$ is multiplicatively close to $q_j^k(\epsilon)$ in the sense of Theorem 2. We look at the three cases where a pair (k,j) is either in S_1,S_2 and S_3.

For the case $(k,j) \in S_1$, $q_{j_k}^k(\epsilon) = 0 = \rho_j^k(\epsilon)$, so they are trivially close.

Now we look at an arbritrary $(k,j) \in S_2$. To simplify notation we omit the k,j in the notation, and hence $\rho_j^k(\epsilon)$ becomes $\rho(\epsilon) = c_K\epsilon^{\frac{K}{M}}$ and $q_j^k(\epsilon)$ becomes $q(\epsilon) = \sum_{i=K}^{\infty} c_K\epsilon^{\frac{i}{M}}$. We want to show that there exists an ϵ_j^k for this $(k,j) \in S_2$ such that for all $\epsilon \in (0,\epsilon_j^k)$ we have

$$q(\epsilon)\left(1-\epsilon^{\frac{1}{M}}\frac{1+|c_{K+1}|}{c_K}\right) \leq \rho(\epsilon) \leq q(\epsilon)\left(1+\epsilon^{\frac{1}{M}}\frac{1+|c_{K+1}|}{c_K}\right)$$

To see this holds, we look at the difference between the two numbers

$$\begin{aligned} q(\epsilon)\left(1+\epsilon^{\frac{1}{M}}\frac{1+|c_{K+1}|}{c_K}\right) - \rho(\epsilon) &= \sum_{i=K+1}^{\infty} c_i\epsilon^{\frac{i}{M}} + \epsilon^{\frac{1}{M}}\frac{1+|c_{K+1}|}{c_K}\sum_{i=K}^{\infty} c_i\epsilon^{\frac{i}{M}} \\ &= \epsilon^{\frac{K+1}{M}}(c_{K+1} + c_K\frac{1+|c_{K+1}|}{c_K}) + \ldots \end{aligned}$$

So the first term is positive, and Lemma 2 gives us that the series is positive on some area $(0,\epsilon')$. Similarly we can show that $q(\epsilon)\left(1-\epsilon^{\frac{1}{M}}\frac{1+|c_{K+1}|}{c_K}\right) - \rho(\epsilon)$ is negative on some area $(0,\epsilon'')$, so by letting $\epsilon_j^k = \min(\epsilon',\epsilon'')$ we get the desired inequalities. Since this works for an arbritary state k and action j where $(k,j) \in S_2$, we can create similar inequalities that work for all the states and actions in S_2 by defining

$$C := \max_{(k,j)\in S_2} \frac{1+|c^k_{K_j^k+1,j}|}{c^k_{K_j^k,j}} \quad , \quad Q := \min_{(k,j)\in S_2} \frac{1}{M_j^k} \quad , \quad \epsilon_3 := \min_{(k,j)\in S_2} \epsilon_j^k$$

This immediately implies that for all $(k,j) \in S_2$ we get the following multiplicative relation between $q_j^k(\epsilon)$ and $\rho_j^k(\epsilon)$

$$q_j^k(\epsilon)\left(1-\epsilon^Q C\right) \leq \rho_j^k(\epsilon) \leq q_j^k(\epsilon)\left(1+\epsilon^Q C\right) \quad \forall \epsilon \in (0,\epsilon_3)$$

Now we look at $(k,j) \in S_3$. From the observations on S_2 we have that for all $(l,i) \in S_2$, that $\rho_i^l(\epsilon) \geq q_i^l(\epsilon)\left(1-\epsilon^Q C\right)$ for $\epsilon \in (0,\epsilon_3)$. Furthermore since we know that $\sum_{i\in[m]} q_i^k(\epsilon) = 1$, it holds that $q_j^k(\epsilon) = 1 - \sum_{i\in S_2} q_i^k(\epsilon)$. We use these observations to compute the following

$$\begin{aligned}
\rho_j^k(\epsilon) &= 1 - \sum_{i\in S_2} \rho_i^k(\epsilon) \leq 1 - (1-\epsilon^Q C)\sum_{i\in S_2} q_i^k(\epsilon) \\
&= \epsilon^Q C + (1-\epsilon^Q C) - (1-\epsilon^Q C)\sum_{i\in S_2} q_i^k(\epsilon) \\
&= \epsilon^Q C + (1-\epsilon^Q C)(1-\sum_{i\in S_2} q_i^k(\epsilon)) = \epsilon^Q C + (1-\epsilon^Q C) q_j^k(\epsilon) \\
&= q_j^k(\epsilon)\left(\frac{\epsilon^Q C}{q_j^k(\epsilon)} + 1 - \epsilon^Q C\right) \leq q_j^k(\epsilon)\left(\frac{2\epsilon^Q C}{c_{0,j}^k} + 1 - \epsilon^Q C\right)
\end{aligned}$$

The last inequality is conditioned on ϵ being small enough. To see how small ϵ must be, consider the Puiseux series $q_j^k(\epsilon)$. First recall that for $(i,l) \in S_3$, $q_i^l(\epsilon)$ has order 0, so the initial term is just a constant $c_{0,j}^k$, and from Lemma 2 we know that the constant is positive. Now look at the the tail $\sum_{i=1}^{\infty} c_{i,j}^k \epsilon^{\frac{i}{M_j^k}}$ without the first term. The tail is just a fractional power series, so it tends to 0 for $\epsilon \to 0$. This means that for any constant κ, then there exists an ϵ' such that for all $\epsilon < \epsilon'$ the tail is smaller than κ. By using the constant $\frac{c_{0,j}^k}{2}$, we get that $\rho_j^k(\epsilon)$ must be larger than $\frac{c_{0,j}^k}{2}$ when $\epsilon \in (0,\epsilon')$, giving us the inequality for $\epsilon \in (0,\epsilon')$. If we then chose $\epsilon'' = \min(\epsilon', \epsilon_3)$, then all the inequalities of the above computation hold. In the same way, we get that there exists an ϵ''' such that

$$\rho_j^k(\epsilon) \geq q_j^k(\epsilon)\left(\frac{-2\epsilon^Q C}{c_{0,j}^k} + 1 + \epsilon^Q C\right) \quad \forall \epsilon \in (0,\epsilon''')$$

Now let $\epsilon_j^k = \min(\epsilon'', \epsilon''')$, and let $\epsilon_4 = \min_{(j,k)\in S_3} \epsilon_j^k$. We now get that both inequalities hold for all $(k,j) \in S_3$

$$q_j^k(\epsilon)\left(\frac{-2\epsilon^Q C}{c_{0,j}^k} + 1 + \epsilon^Q C\right) \leq \rho_j^k(\epsilon) \leq q_j^k(\epsilon)\left(\frac{2\epsilon^Q C}{c_{0,j}^k} + 1 - \epsilon^Q C\right)$$

Next by defining $c = \min_{(j,k)\in S_3} c_{0,j}^k$, and inverting the signs of $\epsilon^Q C$ in the above inequalities, the bound also covers $(k,j) \in S_2$ as well. But then we have that for all $\epsilon \in (0,\epsilon_4)$ and all $(k,j) \in S_1 \cup S_2 \cup S_3$ that

$$q_j^k(\epsilon)\left(\frac{-2\epsilon^Q C}{c} + 1 - \epsilon^Q C\right) \leq \rho_j^k(\epsilon) \leq q_j^k(\epsilon)\left(\frac{2\epsilon^Q C}{c} + 1 + \epsilon^Q C\right)$$

Notice that $\frac{2\epsilon^Q C}{c} + 1 + \epsilon^Q C = 1 + \epsilon^Q \frac{2C+cC}{c}$. To ease the notation of the upcoming calculations we define

$$lw(\epsilon) := 1 - \epsilon^Q \frac{2C+cC}{c} \quad , \quad up(\epsilon) := 1 + \epsilon^Q \frac{2C+cC}{c}$$

Now we are ready to use Theorem 2 to bound the difference in the value of the two Markov Decision processes that appear when we fix the strategy of Player I to be $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ and $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$.

Since the strategy $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ is $\epsilon-$optimal among stationary strategies, then when Player I fixes its strategy to $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$, Player II can not gain more than ϵ more than $\underline{v}_k + \epsilon$. Similarly we can look at the game where Player I fixes his strategy to $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$. If we can prove that Player II can not gain more than $\underline{v}_k+\gamma$ in this game, then we get that the strategy is $\gamma-$optimal among stationary strategies.

Let $(p_j^{kl}(\epsilon))_{j\in[m]}^{k,l\in[N]}$ be the transition probabilities of the Markov Decision process where we fix the strategy of Player I to be $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$. Similarly, let $(\tilde{p}_j^{kl}(\epsilon))_{j\in[m]}^{k,l\in[N]}$ be the transition probabilities when we fix Player I's strategy to be $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$. Then, we get:

$$\frac{\tilde{p}_j^{kl}(\epsilon)}{p_j^{kl}(\epsilon)} = \frac{\sum_{j\in\{1,\ldots,m\}} \rho_i^k(\epsilon) p_{ij}^{kl}}{\sum_{j\in\{1,\ldots,m\}} q_i^k(\epsilon) p_{ij}^{kl}} \Rightarrow lw(\epsilon) \leq \frac{\tilde{p}_j^{kl}(\epsilon)}{p_j^{kl}(\epsilon)} \leq up(\epsilon)$$

So we have an upper bound on the fraction $\frac{\tilde{p}_j^{kl}(\epsilon)}{p_j^{kl}(\epsilon)}$. To upper bound the fraction $\frac{p_j^{kl}(\epsilon)}{\tilde{p}_j^{kl}(\epsilon)}$, observe that when ϵ is smaller than some ϵ', then $lw(\epsilon), up(\epsilon) > 0$ and we get the following upper bound

$$lw(\epsilon) \leq \frac{\tilde{p}_j^{kl}(\epsilon)}{p_j^{kl}(\epsilon)} \Rightarrow \frac{p_j^{kl}(\epsilon)}{\tilde{p}_j^{kl}(\epsilon)} \leq \frac{1}{lw(\epsilon)}$$

Also, since $lw(\epsilon) \cdot up(\epsilon) \leq 1$, then $\frac{1}{lw(\epsilon)} \geq up(\epsilon)$, so the fraction $\frac{\tilde{p}_j^{kl}(\epsilon)}{p_j^{kl}(\epsilon)}$ is also upper bounded by $\frac{1}{lw(\epsilon)}$.

We now use Theorem 2 with $\delta := \frac{1}{lw(\epsilon)} - 1$, and a as a an upper bound on the absolute value of the rewards. Now look at any state k, and let $\gamma, \tilde{\gamma} > 0$ be the numbers such that $\underline{v}_k + \gamma$ and $\underline{v}_k + \tilde{\gamma}$ are the values for Player II of the games where Player I has fixed his strategy to $(q_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ and $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ respectively. Then from Theorem 2 we get

$$\begin{aligned} &\underline{v}_k + \gamma - (\underline{v}_k + \tilde{\gamma}) = \gamma - \tilde{\gamma} \geq -4N\delta a \\ \Rightarrow &\tilde{\gamma} \leq 4N \left(\frac{1}{1 - \epsilon^Q \frac{2C+cC}{c}} - 1 \right) a + \gamma \leq 4N \frac{\epsilon^Q \frac{2C+cC}{c}}{1 - \epsilon^Q \frac{2C+cC}{c}} a + \epsilon \end{aligned}$$

Since the denominator $1 - \epsilon^Q \frac{2C+cC}{c}$ tends to 1 for $\epsilon \to 0$, then for ϵ smaller than some ϵ'''', the denominator is always larger than $\frac{1}{2}$. So by letting $\epsilon_0 := \min(\epsilon'''', \epsilon_4)$ we get that $\tilde{\gamma} \leq \frac{8Na(2C+cC)}{c}\epsilon^Q + \epsilon$. This implies that $(\rho_j^k(\epsilon))_{j\in[m]}^{k\in[N]}$ is a $\left(\frac{8Na(2C+cC)}{c}\epsilon^Q + \epsilon\right)$-optimal strategy among stationary strategies. Now consider

the strategy defined by $\varphi_j^k := \rho_j^k\left(\left(\frac{c}{8Na(2C+cC)}\epsilon\right)^{\frac{1}{Q}}\right)$, which is then $(\epsilon^Q+\epsilon)$-optimal among stationary strategies. The strategy $\left(\varphi_j^k(\frac{\epsilon}{2})\right)_{j\in[m]}^{k\in[N]}$ is then an ϵ-optimal strategy, since $\left(\frac{\epsilon}{2}\right)^Q + \frac{\epsilon}{2} \leq \epsilon$.

Finally notice that the strategy $\left(\varphi_j^k(\frac{\epsilon}{2})\right)_{j\in[m]}^{k\in[N]}$ is not a monomial family of strategies, since it could have fractional exponents. To fix this, we define

$$M := \mathrm{lcm}_{j\in\{1,\ldots,m\},k\in\{1,\ldots,N\}} M_j^k,$$

and let $x_{\epsilon,j}^k := \rho_j^k\left(\left(\left(\frac{\epsilon}{2}\right)^Q\right)^M\right)$. Then $(\boldsymbol{x}_\epsilon)_{0<\epsilon\leq\epsilon_0}$ is a monomial family of strategies, which is also $\epsilon-$optimal amough stationary strategies, because $\left(\frac{\epsilon}{2}\right)^{QM} \leq \frac{\epsilon}{2}$, hence adding the exponent QM only improves the approximation of the value.

References

1. Bewley, T., Kohlberg, E.: The asymptotic theory of stochastic games. Mathematics of Operations Research 1(3), 197–208 (1976)
2. Bewley, T., Kohlberg, E.: On stochastic games with stationary optimal strategies. Mathematics of Operations Research 3(2), 104–125 (1978)
3. Chatterjee, K., de Alfaro, L., Henzinger, T.A.: Strategy improvement for concurrent reachability games. In: Third International Conference on the Quantitative Evaluation of Systems, QEST 2006, pp. 291–300. IEEE Computer Society (2006)
4. Chatterjee, K., Majumdar, R., Henzinger, T.A.: Stochastic limit-average games are in EXPTIME. International Journal of Game Theory 37(2), 219–234 (2008)
5. de Alfaro, L., Henzinger, T.A., Kupferman, O.: Concurrent reachability games. Theor. Comput. Sci. 386(3), 188–217 (2007)
6. Everett, H.: Recursive games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games III. Annals of Mathematical Studies, vol. 39, Princeton University Press (1957)
7. Freidlin, M., Wentzell, A.: Random Perturbations of Dynamical Systems. Springer (1984)
8. Gillette, D.: Stochastic games with zero stop probabilities. In: Contributions to the Theory of Games III. Ann. Math. Studies, vol. 39, pp. 179–187. Princeton University Press (1957)
9. Hansen, K.A., Ibsen-Jensen, R., Miltersen, P.B.: The complexity of solving reachability games using value and strategy iteration. In: Kulikov, A., Vereshchagin, N. (eds.) CSR 2011. LNCS, vol. 6651, pp. 77–90. Springer, Heidelberg (2011)
10. Hansen, K.A., Koucky, M., Lauritzen, N., Miltersen, P.B., Tsigaridas, E.P.: Exact algorithms for solving stochastic games. In: Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, pp. 205–214. ACM (2011)
11. Hansen, K.A., Koucky, M., Miltersen, P.B.: Winning concurrent reachability games requires doubly exponential patience. In: 24th Annual IEEE Symposium on Logic in Computer Science (LICS 2009), pp. 332–341. IEEE (2009)

12. Mertens, J.F., Neyman, A.: Stochastic games. International Journal of Game Theory 10, 53–66 (1981)
13. Milman, E.: The semi-algebraic theory of stochastic games. Mathematics of Operations Research 27(2), 401–418 (2002)
14. Solan, E.: Continuity of the value of competitive Markov decision processes. Journal of Theoretical Probability 16, 831–845 (2003)
15. Solan, E., Vieille, N.: Computing uniformly optimal strategies in two-player stochastic games. Economic Theory 42, 237–253 (2010)