Games and Economic Behavior 108 (2018) 523–540

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Acceptable strategy profiles in stochastic games ☆

Eilon Solan

The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997800, Israel

ARTICLE INFO

Article history:
Received 15 August 2016
Available online 8 February 2017

JEL classification:
C72
C73

Keywords:
Stochastic games
Acceptable strategy profiles
Automata

ABSTRACT

This paper presents a new solution concept for multiplayer stochastic games, namely, acceptable strategy profiles. For each player i and state s in a stochastic game, let $w_i(s)$ be a real number. A strategy profile is *w-acceptable*, where $w = (w_i(s))$, if the discounted payoff to each player i at every initial state s is at least $w_i(s)$, provided the discount factor of the players is sufficiently close to 1. Our goal is to provide simple strategy profiles that are w-acceptable for payoff vectors w in which all coordinates are high.

1. Introduction

Shapley (1953) presented the model of *stochastic games*, which are dynamic games in which the state variable changes from stage to stage as a function of the current state and the actions taken by the players. Shapley (1953) proved that the discounted value exists in two-player zero-sum stochastic games, and provided an equation that the discounted value satisfies.

This seminal work led to an extensive research in several directions (see the surveys by, e.g., Neyman and Sorin, 2003; Mertens et al., 2015; Solan and Vieille, 2015; Solan and Ziliotto, 2016; and Jaśkiewicz and Nowak, 2016, 2017), including the study of the discounted value in games with general state and action sets, the study of discounted equilibria in multiplayer stochastic games, and the study of the robustness of equilibria.

A commonly studied robustness concept is that of uniform equilibrium. A strategy profile is a *uniform ε-equilibrium* for $\varepsilon \geq 0$ if it is an ε-equilibrium in (a) the discounted game, provided the discount factor is sufficiently close to 1, namely, the players are sufficiently patient, and (b) the finite horizon game, provided the horizon is sufficiently long.

Progress in the study of the uniform equilibrium turned out to be slow, existence of such a strategy profile was proven only in special cases (see, e.g., Mertens and Neyman, 1981; Solan, 1999; Vieille, 2000a, 2000b; Solan and Vieille, 2001, Simon, 2007, 2012, 2016; and Flesch et al., 1997, 2008, 2009), and the strategy profiles that are uniform ε-equilibria are usually quite complex.

The present paper proposes a new solution concept for stochastic games that combines simplicity in behavior with relatively high payoffs. Let $w = (w_i(s))$ be a vector, where i ranges over all players and s ranges over all states. A strategy profile in a stochastic game is *w-acceptable* if when the players follow it, for every discount factor sufficiently close to 1, the discounted payoff of each player i is at least $w_i(s)$ when the initial state is s. Thus, when the players follow such a

☆ The author thanks Eitan Altman for helping in identifying relevant references, and acknowledges the support of the Israel Science Foundation, Grant #323/13.

E-mail address: eilons@post.tau.ac.il.

http://dx.doi.org/10.1016/j.geb.2017.01.011
0899-8256/

strategy profile, they forgo the option to profit by deviation in order to guarantee a reasonably high payoff for each player. A strategy profile is *min–max ε-acceptable* (resp. *max–min ε-acceptable*) if it is w-acceptable for the vector $w = (w_i(s))$ that is defined by $w_i(s) := \overline{v}_i^1(s) - \varepsilon$ (resp. $w_i(s) := \underline{v}_i^1(s) - \varepsilon$), where $\overline{v}_i^1(s)$ (resp. $\underline{v}_i^1(s)$) is the uniform min–max value (resp. uniform max–min value) of player i at the initial state s. By Neyman (2003), $\overline{v}_i^1(s)$ is the amount that player i can uniformly guarantee when the other players cooperate to lower his payoff, and $\underline{v}_i^1(s)$ is the same amount when the other players can correlate their actions.

In their study of correlated equilibrium, Solan and Vieille (2002) constructed a min–max ε-acceptable strategy profile in every multiplayer stochastic game and for every $\varepsilon > 0$. Their construction uses the technique of Mertens and Neyman (1981) for designing a uniform ε-optimal strategy in two-player zero-sum stochastic games, and in particular is history dependent.

Our goal in this paper is the construction of simple strategy profiles that are min–max ε-acceptable or max–min ε-acceptable, where simplicity is measured by the size of the automata that are needed to implement the individual strategies of the players.

A naïve suggestion for a stationary min–max ε-acceptable strategy profile is a stationary discounted equilibrium, for some discount factor sufficiently close to 1. As we now explain, this approach is bound to fail. The discounted payoff that corresponds to a stationary strategy profile is the weighted average of the payoffs that are received in the various states, where the weight of a state is equal to the discounted time that the play spends in that state. A discounted equilibrium yields a high discounted payoff to all players, which implies that this weighted average is high. It might happen that while the average payoff of all players is high, some players get high payoff in some states, while other players get high payoff in other states. When we fix a λ-discounted equilibrium and we calculate the payoff according to a discount factor λ' that goes to 1, the weights of the various states change, and there is no guarantee that the weighted average payoff of all players remains high. This phenomenon in fact happens, as can be seen in Example 2.6 below.

We prove the existence of a max–min ε-acceptable strategy profile, in which the strategy of each player can be implemented by an automaton whose number of states is at most the number of states in the stochastic game times the number of players.[1] We also prove the existence of a min–max ε-acceptable stationary correlated strategy, which can be implemented by an automaton whose number of states is the number of states in the stochastic game. The proofs are constructive and identifies (at least) one such strategy profile.

Another view on the concept of w-acceptability stems from the folk theorem. The folk theorem for repeated games states that under proper technical conditions, every feasible and individually rational payoff vector is an equilibrium payoff. Solan (2001) extended this result to stochastic games when considering extensive-form correlated equilibria rather than Nash equilibria. The identification of the set of feasible and individually rational payoffs in multiplayer stochastic games is open. A strategy profile is min–max (resp. max–min) ε-acceptable if it generates a feasible and ε-individually rational payoff vector when punishment is given by the uniform min–max (resp. max–min) value.

Identifying individually rational (when punishment is given by the min–max value) correlated strategy profiles in the discrete-time game is useful for continuous-time stochastic games. Indeed, an ε-individually rational correlated strategy in the discrete-time game can be transformed into an ε-equilibrium in the continuous-time game, see Neyman (2012).

The paper is organized as follows. The model of stochastic games, the concept of acceptable strategy profiles, the main results, a discussion, and open problems appear in Section 2. The proof of the main results and additional discussion appear in Section 3.

2. Model and main results

2.1. The model of stochastic games

A multiplayer *stochastic game* is a vector $\Gamma = (I, S, (A_i)_{i\in I}, (u_i)_{i\in I}, q)$ where

- I is a finite set of players.
- S is a finite set of states.
- A_i is a finite set of actions available to player i at each state.[2] Denote by $A := \times_{i\in I} A_i$ the set of all action profiles.
- $u_i : S \times A \to \mathbf{R}$ is player i's payoff function. We assume w.l.o.g. that the payoffs are bounded between 0 and 1.
- $q : S \times A \to \Delta(S)$ is a transition function, where $\Delta(X)$ is the set of probability distributions over X, for every nonempty finite set X.

The game is played as follows. The initial state $s^0 \in S$ is given. At each stage $n \geq 0$, the current state s^n is announced to the players. Each player i chooses an action $a_i^n \in A_i$; the action profile $a^n = (a_i^n)_{i\in I}$ is publicly announced, the new state s^{n+1} is drawn according to the probability distribution $q(\cdot \mid s^n, a^n)$, and the game proceeds to stage $n+1$.

[1] We consider automata in which the output function depends only on the automaton's state, and not on the input (Moore machine). If the output function can depend both on the automaton's state and on the input (Mealy machine), then the number of required automaton's states is at most the number of players.

[2] We could have assumed that the action set of a player depends on the current state. This would have complicated the definition of an automaton that implements a strategy, hence we prefer to assume that the action set is independent of the state.

A *correlated mixed action* is an element of $\Delta(A)$. We extend the domain of q and $(u_i)_{i\in I}$ to correlated mixed actions in a multilinear fashion: for every state $s \in S$ and every correlated mixed action $\alpha \in \Delta(A)$ we define

$$q(s,\alpha) := \sum_{a\in A} \alpha[a] q(s,a),$$

and

$$u_i(s,\alpha) := \sum_{a\in A} \alpha[a] u_i(s,a), \quad \forall i \in I.$$

Let $H := \cup_{n\geq 0} \big((S\times A)^n \times S\big)$ be the set of *finite histories*[3] and $H^\infty := (S\times A)^\infty$ be the set of *plays*. We assume perfect recall. Accordingly, a (behavior) *strategy* of player i is a function $\sigma_i : H \to \Delta(A_i)$. A strategy σ_i of player i is *pure* if for every finite history $h^n \in H$, the support of the mixed action $\sigma_i(h^n)$ contains one action. We note that the superscript n of a history h^n always denotes its length, and the last state of a finite history h^n is always denoted by s^n. Denote by Σ_i the set of all strategies of player i, by $\Sigma := \times_{i\in I}\Sigma_i$ the set of all strategy profiles, and by $\Sigma_{-i} := \times_{j\neq i}\Sigma_j$ the set of all strategy profiles of all players except player i.

A *correlated strategy* is a function $\tau : H \to \Delta(A)$. The set of all correlated strategies is denoted Σ_{corr}. We note that every strategy profile induces a correlated strategy. Below we will provide definitions for correlated strategies that apply also to strategy profiles.

A class of simple strategies is the class of *stationary strategies*. Those are strategies in which the choice of the player at each stage depends only on the current state, and not on previously visited states or on past choices of the players. A stationary strategy of player i can be identified with an element of $(\Delta(A_i))^{|S|} \subset \mathbf{R}^{|S|\times|A_i|}$, and will be denoted $x_i = (x_i(s))_{s\in S}$. A strategy profile $\sigma = (\sigma_i)_{i\in I}$ is *stationary* if for every player $i \in I$ the strategy σ_i is stationary. A stationary correlated strategy is identified with an element of $(\Delta(A))^{|S|}$.

We will endow H^∞ with the σ-algebra generated by finite cylinders. Every initial state $s^0 \in S$ and every correlated strategy $\tau \in \Sigma_{\text{corr}}$ induce a probability distribution $\mathbf{P}_{s^0,\tau}$ over the set of plays H^∞. Denote the corresponding expectation operator by $\mathbf{E}_{s^0,\tau}$.

2.2. Acceptable strategy profiles

For every initial state $s^0 \in S$, every correlated strategy $\tau \in \Sigma_{\text{corr}}$, every player $i \in I$, and every discount factor $\lambda \in [0,1)$ the *λ-discounted payoff* of player i is

$$\gamma_i^\lambda(s^0;\tau) := \mathbf{E}_{s^0,\tau}\left[(1-\lambda)\sum_{n=0}^{\infty}\lambda^n u_i(s^n,a^n)\right].$$

The main concept that we study in this paper is the concept of acceptable strategy profiles.

Definition 2.1. Let $w \in \mathbf{R}^{|S|\times|I|}$. A correlated strategy τ is *w-acceptable at the initial state s^0* if there exists a real number $\lambda_0 \in [0,1)$ such that for every player $i \in I$ and every $\lambda \in [\lambda_0, 1)$,

$$\gamma_i^\lambda(s^0;\tau) \geq w_i(s^0), \quad \forall i \in I.$$

The correlated strategy is *w-acceptable* if it is w-acceptable at all initial states. In this case we say that the vector w is *acceptable*.

In words, a correlated strategy τ is w-acceptable if whenever the players are sufficiently patient it yields each player i at least $w_i(s^0)$, for every initial state s^0.

A natural question that arises is which vectors w are acceptable. A vector w is a *uniform equilibrium payoff*[4] if for every $\varepsilon > 0$ there exists a real number $\lambda_0 \in [0,1)$ and a strategy profile σ^ε such that for every initial state $s^0 \in S$, every player $i \in I$, and every discount factor $\lambda \in [\lambda_0, 1)$ we have $|\gamma_i^\lambda(s^0;\sigma^\varepsilon) - w_i(s^0)| < \varepsilon$ and

$$\gamma_i^\lambda(s^0;\sigma_i,\sigma_{-i}^\varepsilon) \leq \gamma_i^\lambda(s^0;\sigma^\varepsilon) + \varepsilon, \quad \forall \sigma_i \in \Sigma_i.$$

If w is a uniform equilibrium payoff, then for every $\varepsilon > 0$ the vector $w - \varepsilon := (w_i(s) - \varepsilon)_{i\in I, s\in S}$ is acceptable. To date it is not known whether every multiplayer stochastic game admits a uniform equilibrium payoff.

[3] By convention, the set $(S\times A)^0$ contains only the empty history.

[4] The concept that we define here refers to uniformity in the discount factor only. A stronger notion is defined in Mertens and Neyman (1981). We refer to this stronger notion in Section 2.6 below.

The λ-*discounted min–max value of player i at the initial state* s^0 is given by

$$\overline{v}_i^{\lambda}(s^0) := \min_{\sigma_{-i}\in\Sigma_{-i}} \max_{\sigma_i\in\Sigma_i} \gamma_i^{\lambda}(s^0;\sigma_i,\sigma_{-i}). \tag{1}$$

The λ-*discounted max–max value of player i at the initial state* s^0 is given by

$$\underline{v}_i^{\lambda}(s^0) := \max_{\sigma_i\in\Sigma_i} \min_{\sigma_{-i}\in\Sigma_{-i}} \gamma_i^{\lambda}(s^0;\sigma_i,\sigma_{-i}). \tag{2}$$

The interpretation of these two quantities is that player i can guarantee to himself a payoff at least his max–min value, while the other players can ensure that player i's payoff will not exceed his min–max value. Because for every fixed discount factor $\lambda\in[0,1)$ the λ-discounted payoff is a continuous function of the strategies of the players, the maxima and minima in Eqs. (1) and (2) are attained. A strategy σ_i that attains the maximum in Eq. (2) is called a λ-*discounted max–min strategy*.

It is well known (see Neyman, 2003) that the limits

$$\overline{v}_i^{1}(s^0) := \lim_{\lambda\to 1} \overline{v}_i^{\lambda}(s^0), \quad \underline{v}_i^{1}(s^0) := \lim_{\lambda\to 1} \underline{v}_i^{\lambda}(s^0),$$

exist for every player $i\in I$ and every initial state $s^0\in S$. The quantities $\overline{v}_i^{1}(s^0)$ and $\underline{v}_i^{1}(s^0)$ are called the *uniform min–max value* and the *uniform max–min value* of player i at state s^0, respectively.

Neyman (2003) proved the existence of a uniform ε-max–min strategy for each player i and every $\varepsilon>0$, that is, a strategy $\sigma_i\in\Sigma_i$ and $\lambda_0\in[0,1)$ that satisfy

$$\gamma_i^{\lambda}(s^0;\sigma_i,\sigma_{-i}) \geq \underline{v}_i^{1}(s^0)-\varepsilon, \quad \forall\lambda\in[\lambda_0,1), \forall\sigma_{-i}\in\Sigma_{-i}.$$

When each player follows a uniform ε-max–min strategy, all players receive at least their uniform max–min value minus ε. In particular, there is a strategy profile that is w-acceptable for the vector $w=(w_i(s))_{i\in I,s\in S}$ defined by $w_i(s):=\underline{v}_i^{1}(s^0)-\varepsilon$ for every player $i\in I$ and every state $s\in S$.

Definition 2.2. Let $\varepsilon\geq 0$. A correlated strategy τ is *max–min ε-acceptable* if for every player $i\in I$, every initial state $s^0\in S$, and every discount factor λ sufficiently close to 1, we have $\gamma_i^{\lambda}(s^0;\tau)\geq \underline{v}_i^{1}(s^0)-\varepsilon$.

By Neyman (2003), for every $\varepsilon>0$, every initial state $s^0\in S$, and every strategy profile σ_{-i} of the other players, there exists $\lambda_0\in[0,1)$ and a strategy σ_i of player i such that

$$\gamma_i^{\lambda}(s^0;\sigma_i,\sigma_{-i}) \geq \overline{v}_i^{1}(s^0)-\varepsilon, \quad \forall\lambda\in[\lambda_0,1).$$

It is therefore natural to ask whether there are strategy profiles that ensure that all players receive at least their uniform min–max values up to ε for all discount factors sufficiently close to 1. Such a strategy profile will guarantees for all players the minimal amount that they would agree to receive in an equilibrium.

Definition 2.3. Let $\varepsilon\geq 0$. A correlated strategy τ is *min–max ε-acceptable* if for every player $i\in I$, every initial state $s^0\in S$, and every discount factor λ sufficiently close to 1, we have $\gamma_i^{\lambda}(s^0;\tau)\geq \overline{v}_i^{1}(s^0)-\varepsilon$.

A by-product of the study of Solan and Vieille (2002) on extensive-form correlated equilibria in stochastic games is that there always exists a min–max ε-acceptable strategy profile. The constructions of Neyman (2003) and of Solan and Vieille (2002) use the technique of Mertens and Neyman (1981), and therefore the max–min ε-acceptable strategy profile and min–max ε-acceptable strategy profile that are known to exist are complex and history dependent. In this paper we ask whether there are *simple* max–min and min–max ε-acceptable strategy profile.

We first identify two classes of stochastic games, namely, Markov decision processes and absorbing games, in which there are stationary min–max ε-acceptable strategy profiles. We do not know whether stationary min–max ε-acceptable strategy profiles exist in every multiplayer stochastic game.

Blackwell (1962) proved that in stochastic games with a single player ($|I|=1$) there is a pure stationary strategy σ_1 and $\lambda_0\in[0,1)$ that satisfy

$$\gamma_i^{\lambda}(s^0;\sigma_1) \geq \overline{v}_i^{1}(s^0)-\varepsilon, \quad \forall\lambda\in[\lambda_0,1), \forall s^0\in S.$$

It follows that for every stochastic game with a single player there is a pure stationary min–max ε-acceptable strategy, for every $\varepsilon>0$.

A state $s\in S$ is *absorbing* if $q(s\mid s,a)=1$ for every action profile $a\in A$. An *absorbing game* is a stochastic game with a single nonabsorbing state. By Solan (1999, Theorem 4.5) it follows that for every absorbing game there is a stationary min–max ε-acceptable strategy profile, for every $\varepsilon>0$.

2.3. Automata and strategies implemented by automata

A common way to model a decision maker with bounded computational capacity is by an automaton, which is a finite state machine whose output depends on its current state, and whose evolution depends on the current state and on its input (see, e.g., Neyman, 1985 and Rubinstein, 1986). Formally, an *automaton* is given by (1) a finite set Q of states, (2) a finite set In of inputs, (3) a set Out of outputs, (4) an output function $f : Q \to Out$, (5) a transition function $g : Q \times In \to Q$, and (6) an initial state $q^* \in Q$.

Denote by q^n the automaton's state at stage n. The automaton starts in state $q^0 = q^*$, and at every stage $n \geq 0$, as a function of the current state q^n and the current input i^n, the output of the automaton $o^n = f(q^n)$ is determined, and the automaton moves to a new state $q^{n+1} = g(q^n, i^n)$.

The *size* of an automaton is the size of its set of states Q. Below we will use strategies that can be implemented by automata; in this case the size of the automaton measures the complexity of the strategy.

Consider a stochastic game and fix a player $i \in I$. An automaton whose set of inputs is the Cartesian product of the set of action profiles and the set of states of the stochastic game, and the set of outputs is the set of mixed actions of player i, that is, $In = A \times S$ and $Out = \Delta(A_i)$, can implement a behavior strategy of player i. Indeed, at every stage n, the strategy plays the mixed action $f(q^n)$, and the new state of the automaton $q^{n+1} = g(q^n, a^n, s^{n+1})$ depends on its current state q^n, the action profile a^n played at stage n, and the new state of the game s^{n+1}.

Similarly, an automaton can implement a correlated strategy; In this case the set of outputs of the automaton is the set of correlated mixed actions: $Out = \Delta(A)$.

To distinguish between the state of the game and the state of the automaton we refer to the latter as *automaton-states*.

2.4. The main results

We can now present our two main results. The first identifies an upper bound to the size of the smallest automaton that implements a max–min ε-acceptable strategy profile.

Theorem 2.4. *For every stochastic game and every $\varepsilon > 0$ there exists a max–min ε-acceptable strategy profile such that each of the strategies composing the profile can be implemented by an automaton with size $|S| \times |I|$.*

Our second main result states that there exists a stationary min–max ε-acceptable correlated strategy. Such a strategy can be implemented by an automaton of size $|S|$.

Theorem 2.5. *For every stochastic game and every $\varepsilon > 0$ there exists a stationary min–max ε-acceptable correlated strategy.*

The existence of a min–max ε-acceptable correlated strategy in discrete-time stochastic games (Solan and Vieille, 2002) was used by Neyman (2012) to show the existence of a Nash uniform equilibrium in stochastic games in continuous time. If the min–max ε-acceptable correlated strategy is stationary (rather than history dependent), the construction of Neyman (2012) becomes somewhat simpler. Theorem 2.5 therefore simplifies the construction in Neyman (2012).

In our formulation of an automaton, the output function f does not depend on the input, hence it corresponds to a Moore machine. If one allows the output function to depend on the input, one obtains a Mealy machine. With this new formulation, the number of automaton-states required to implement the strategies that compose a max–min ε-acceptable strategy profile is $|I|$, and the number of automaton-states that are required to implement a stationary min–max ε-acceptable correlated strategy is 1.

2.5. Discounted equilibrium and acceptable strategy profiles

A strategy profile σ^λ is a λ-*discounted equilibrium* if for every initial state $s \in S$ and every player $i \in I$ we have

$$\gamma_i^\lambda(s; \sigma^\lambda) \geq \gamma_i^\lambda(s; \sigma_i, \sigma_{-i}^\lambda), \quad \forall \sigma_i \in \Sigma_i.$$

It is well known (see Fink, 1964 or Takahashi, 1964) that a λ-discounted equilibrium in stationary strategies exists in every stochastic game, though it usually depends on the discount factor. As the following example shows, a strategy profile that is a λ-discounted equilibrium for a specific λ may yield some players low payoff when λ changes. This example shows in particular that a λ-discounted equilibrium and a limit of λ-discounted equilibria as λ goes to 1 need not be min–max ε-acceptable.

Example 2.6. Consider the two-player absorbing game that appear in Fig. 1. In the initial state s_0, which is nonabsorbing, each player has two actions. In each entry of the matrix in the figure, the stage payoff appears in the middle and the transition appears on the top-right corner: s_0 means that with probability 1 the play stays in state s_0, while $*$ means that with probability 1 the play continues to an absorbing state, where the payoff vector is the vector written in the entry.

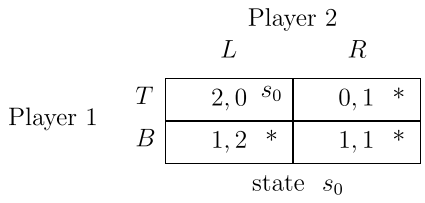

Fig. 1. The absorbing game in Example 2.6.

The uniform min–max value of both players is 1. In the unique λ-discounted equilibrium Player 1 plays the stationary strategy $x_1(\lambda) = [\frac{1}{2-\lambda}(T), \frac{1-\lambda}{2-\lambda}(B)]$ and Player 2 plays the stationary strategy $x_2(\lambda) = [\frac{1}{2-\lambda}(L), \frac{1-\lambda}{2-\lambda}(R)]$. The limit of the equilibrium strategy profiles as λ goes to 1 is for Player 1 to play T and for Player 2 to play L, which yields Player 2 a payoff of 0 that is lower than his uniform min–max value. Similarly, the equilibrium strategy pair for a given discount factor $x(\lambda) := (x_1(\lambda), x_2(\lambda))$ may yield low payoff for discount factors different than λ, because $\lim_{\lambda\to 1}\lim_{\lambda'\to 1}\gamma_2^{\lambda'}(x(\lambda)) = (\frac{1}{2}, \frac{3}{2})$.

2.6. Finite horizon acceptability and limit of the averages acceptability

We defined the concept of acceptability using the discounted evaluation. One could alternatively define this concept using finite horizon games or the infinite game. That is, for every state $s^0 \in S$, every player $i \in I$, every correlated strategy τ, and every $k \in \mathbf{N}$ the *k-stage payoff* is given by

$$\gamma_i^k(s^0;\tau) := \mathbf{E}_{s^0,\tau}\left[\frac{1}{k}\sum_{n=0}^{k-1} u_i(s^n,a^n)\right].$$

Let $w \in \mathbf{R}^{|S|\times|I|}$, and call a correlated strategy τ *average w-acceptable* if for every k sufficiently large

$$\gamma_i^k(s^0;\tau) \geq w_i(s^0), \quad \forall i \in I, \forall s^0 \in S.$$

Call the correlated strategy τ *limit w-acceptable* if the limit $\lim_{k\to\infty}\frac{1}{k}\sum_{n=0}^{k-1} u_i(s^n,a^n)$ exists $\mathbf{P}_{s^0,\tau}$-a.s. and

$$\mathbf{E}_{s^0,\tau}\left[\lim_{k\to\infty}\frac{1}{k}\sum_{n=0}^{k-1} u_i(s^n,a^n)\right] \geq w_i(s^0), \quad \forall i \in I, \forall s^0 \in S.$$

One could define a stronger concept of acceptability that is inspired by the notion of uniform equilibrium: the correlated strategy τ is *uniform w-acceptable* if it is both discounted w-acceptable, average w-acceptable, and limit w-acceptable. The implications of Blackwell (1962), Solan (1999), and Solan and Vieille (2002) for acceptable strategy profiles are valid with the stronger notion of uniform acceptability. Moreover, every correlated strategy that can be implemented by an automaton and is w-acceptable according to the discounted, average, or limit notion, is uniform $(w-\varepsilon)$-acceptable, for every $\varepsilon > 0$.

2.7. Open problems

The introduction of the concept of acceptable strategy profiles raises several open questions. These questions include the following:

- Whether there exists a *stationary* min–max ε-acceptable strategy profile for every $\varepsilon > 0$. If the answer to the above question is negative, then it will be interesting to know the size of the smallest automaton that is needed to implement a min–max ε-acceptable strategy profile.
- The characterization of the set of payoff vectors w for which there exists stationary w-acceptable strategy profiles.
- More generally, one can study the set $X(\widetilde{\Sigma})$ of payoff vectors w for which there exists w-acceptable strategy profiles in some prespecified set $\widetilde{\Sigma}$ of simple strategy profiles, like the set of strategy profiles that can be implemented by automata with at most K states, and determine the dependency of $X(\widetilde{\Sigma})$ on $\widetilde{\Sigma}$.

3. Proof of the main results

We will start by proving Theorem 2.4. Consider the strategy profile in which each player plays a uniform ε-max–min strategy. By definition, the discounted payoff of each player i at least $\underline{v}_i^1(s^0) - \varepsilon$, provided the discount factor is sufficiently close to 1. Since uniform ε-max–min strategies are history dependent, our goal will be to convert this strategy profile into a strategy profile that can be implemented by small automata and in which the limit discounted payoff of the players is not lowered. The steps of the construction can be summarized as follows.

1. We will define a concept of *communicating sets* of states, in which the uniform max–min value of the players is independent of the state and is called the *uniform max–min value in the set*. We will identify communicating sets of two types, A and B.
2. In communicating sets of type A there is a strategy profile that remains in the set and yields to each player i is at least his uniform max–min value in the set minus ε.
3. In communicating sets of type B there is a strategy profile that ensures that the play leaves the set and the expected continuation uniform max–min value of each player is at least his uniform max–min value in the set.
4. We will show that the strategy profiles mentioned in Points 2 and 3 can be chosen to be simple, that is, the individual strategies of the players can be implemented by small automata. Unfortunately, we do not know whether these strategy profiles can be chosen to be stationary.
5. We will show that there is a stationary strategy profile having the following properties: (a) with probability 1 the play reaches a communicating set, and (b) the expected uniform max–min value of each player in the communicating set that is reached is at least $\underline{v}_i^1(s^0)$, for every initial state s^0.
6. We will partition the set of states into maximal communicating sets (w.r.t. set inclusion) and the set of states that do not belong to any communicating set. Combining the stationary strategy profile of Point 5 with the simple strategy profiles of Points 2 and 3 yields a strategy profile that (a) is simple, (b) ensures that the play reaches a communicating set of type A, and (c) yields to all players a limit discounted payoff at least $\underline{v}_i^1(s^0) - \varepsilon$, for every initial state s^0.

The proof of Theorem 2.5 uses the same steps as outlined above but is more involved. To obtain a history dependent strategy profile that yields to all players a discounted payoff at least his min–max value we use the strategy profile of Solan and Vieille (2002), and convert it to a simple strategy profile without lowering the limit discounted payoff.

The proof becomes more complicated because there is a conceptual difference between the uniform ε-max–min strategies as devised by Neyman (2003) and the strategy profile of Solan and Vieille (2002). While to guarantee his uniform max–min value a player need not take into account the play of the other players, to obtain his uniform min–max value the player needs to adapt his play to the play of the other players. Consequently, the uniform ε-max–min strategy of a player is based on a one-parameter family of stationary strategies, namely, the function that assigns a discounted max–min stationary strategy to each discount factor. On the other hand, the strategy profile that ensures to each player his uniform min–max value is based on an $|I|$-parameter family of stationary strategies, namely, the function that assigns for every state s and each vector $\vec{\lambda} = (\lambda_i)_{i\in I}$ of discount factors, one for each player, a $\vec{\lambda}$-discounted equilibrium in the one-shot game that is played at state s and in which the continuation payoff of each player i is given by the λ_i-discounted min–max value. In particular, while a uniform ε-max–min strategy is a perturbation of a stationary strategy, this is not the case with a uniform ε-min–max strategy. This has the consequence that the definition of a communicating set is more intricate.

Once communicating sets are properly defined, one can define communicating sets of types A and B, and show that there are stationary correlated strategies that satisfy the properties mentioned in Points 2 and 3. The rest of the proof is similar to the proof of Theorem 2.4.

3.1. Communicating sets

Let $x \in X$ be a stationary strategy profile. A nonempty set $D \subseteq S$ is *closed* under x if under x the play never leaves D once it enters this set: $q(D \mid s, x(s)) = 1$ for every state $s \in D$. A closed set is *irreducible* if it does not contain any other closed set. Denote by $\mathcal{I}(x)$ the collection of all irreducible sets w.r.t. x, and for every set of states C denote by $\mathcal{I}_C(x)$ the set of all irreducible sets w.r.t. x that are subsets of C. We note that whether or not a set D is an irreducible set under a stationary strategy x is determined by the collection of supports $(\mathrm{supp}(x_i(s)))_{i\in I, s\in D}$.

For every set $C \subseteq S$ denote by ν_C the first arrival time to C:

$$\nu_C := \min\{n > 0 \colon s^n \in C\}.$$

By convention, the minimum of an empty set is $+\infty$. When $D \in \mathcal{I}(x)$ consider the Markov chain over D induced by the stationary strategy profile x; that is, the Markov chain with transition $(q(s' \mid s, x(s)))_{s,s'\in D}$. For every state $s \in D$ denote the long-run frequency of visits to state s by

$$\rho_{D,x}(s) := \frac{1}{\mathbf{E}_{s,x}[\nu_{\{s\}}]}. \tag{3}$$

Since D is irreducible under x, $\rho_{D,x}$ is an invariant distribution of the Markov chain and $\rho_{D,x}(s) > 0$ for every state $s \in D$.

For every irreducible set $D \in \mathcal{I}(x)$ denote the limit discounted payoff under x by

$$\gamma(D; x) := \lim_{\lambda\to 1} \gamma^\lambda(s^0; x) = \sum_{s\in D} \rho_{D,x}(s) u(s, x(s)).$$

We note that the limit discounted payoff $\gamma(D; x)$ is independent of the initial state $s^0 \in D$.

The stationary strategy profile $y \in X$ is an *enlargement* of the stationary strategy profile x if $\mathrm{supp}(x_i(s)) \subseteq \mathrm{supp}(y_i(s))$ for every player $i \in I$ and every state $s \in S$.

We now provide a definition of communicating sets, which is a variant of the common definition of this concept (see Ross and Varadarajan, 1991 for the analogous definition in Markov decision problems, or Solan and Vieille, 2002).

Definition 3.1. A set $C \subseteq S$ is *communicating under* x if

(C.1) The uniform max–min value is constant over C, that is, $\underline{v}^1(s) = \underline{v}^1(s')$ for every pair of states $s, s' \in C$. Denote by $\underline{v}^1(C)$ the common uniform max–min value of the states in C,

(C.2) For every state $s \in C$ there exists an enlargement $y_{\{s\},C}$ of x such that
  (i) the set C is closed under $y_{\{s\},C}$, and
  (ii) under y the play reaches s a.s.:

$$\mathbf{P}_{s^0, y_{\{s\},C}}(\nu_{\{s\}} < \infty) = 1, \quad \forall s^0 \in C.$$

Condition (C.1) distinguishes our concept of communicating set from the standard notion of communication.

We denote by $\mathcal{C}(x)$ the collection of all the sets that communicate under the stationary strategy profile x. If $C_1, C_2 \in \mathcal{C}(x)$ have nonempty intersection, then $C_1 \cup C_2 \in \mathcal{C}(x)$. It follows that the collection $\mathcal{C}_{\max}(x)$ of maximal communicating sets w.r.t. set inclusion contains disjoint sets. Denote by $C^*(x) := \cup_{C \in \mathcal{C}_{\max}(x)} C$ the union of all maximal communicating sets under x.

When C is a communicating set under the stationary strategy profile x and $D \subset C$, there is an enlargement of x that ensures that the play reaches D without leaving C, provided the initial state is in $C \setminus D$. We denote such an enlargement by $y_{D;C}$.

Let C be a communicating set under x. Our construction of a max–min ε-acceptable strategy profile involves visiting different irreducible sets in $\mathcal{I}_C(x)$, and playing the stationary strategy profile x in each one of them for a certain length of time. To simultaneously switch from one irreducible set to the next without counting the number of stages that the play spent in each irreducible set, we need to find an event that can be used as a synchronization device among the players. One way to do that is to identify a player who can play an action that is not in the support of x and that keeps the play in C, and have him play this action. Such a player is called a *signaller*, and is formally defined as follows.

Definition 3.2. Let x be a stationary strategy profile, let $C \subseteq S$ be a set of states, and let $D \in \mathcal{I}_C(x)$. Player i is a *signaller at* D *w.r.t.* C *and* x if there exists a state $s \in D$ and an action $a_i \in A_i \setminus \mathrm{supp}(x_i(s))$ such that $q(C \mid s, a_i, x_{-i}(s)) = 1$.

The following result asserts that for every communicating set C under x and every irreducible set D under x which is a strict subset of C there is a signaller at D w.r.t. C and x.

Lemma 3.3. *Let x be a stationary strategy profile, let $C \in \mathcal{C}(x)$ be a communicating set under x, and let $D \in \mathcal{I}_C(x)$ satisfy $D \subset C$. There is a player who is a signaller at D w.r.t. C and x.*

Proof. Fix a state $s \in C \setminus D$. By definition, there is an enlargement y of x that satisfies that C is closed under y and that under y the play reaches s a.s. when the initial state is in D. Since D is closed under x, there are a player $i \in I$ and a state $s' \in D$ such that $\mathrm{supp}(y_i(s')) \supset \mathrm{supp}(x_i(s'))$. In particular, player i is a signaller at D w.r.t. C and x. □

In the sequel we will construct strategy profiles that satisfy various desirable properties. It will be convenient to define the strategy profiles separately on each maximal communicating set C. We will therefore consider strategy profiles that are defined only for finite histories that remain in some set of states C, that is, for finite histories $h \in H_C := \cup_{n \geq 0}\big((C \times A)^n \times C\big)$.

The following result states that for every probability distribution β over the set of irreducible sets $\mathcal{I}_C(x)$ there is a simple strategy profile that remains in C and according to which the limit discounted payoff converges to the long-run average payoff indicated by β.

Proposition 3.4. *Let x be a stationary strategy profile and let C be a communicating set under x. Let $D^{(1)}, \cdots, D^{(L)}$ be L irreducible sets under x that are subsets of C, and let $\beta = (\beta^{(l)})_{l=1}^L$ be a probability distribution over $\{1, 2, \cdots, L\}$. For every $\varepsilon > 0$ there exists a strategy profile σ^ε that is defined as long as the play remains in C and satisfies the following properties:*

- *Under σ the play does not leave C.*
- *The strategy profile σ^ε can be implemented by automata with size $|C| \times L$.*
- *For every initial state $s \in C$, the limit discounted payoff under σ^ε satisfies*

$$\left| \lim_{\lambda \to 1} \gamma_i^\lambda(s; \sigma^\varepsilon) - \sum_{l=1}^{L} \beta^{(l)} \gamma_i(D^{(l)}; x) \right| \leq \varepsilon, \quad \forall s \in C, \forall i \in N. \tag{4}$$

Proof. According to the strategy profile σ^ε that we will construct the players will play in blocks of random size. In block k they will ensure that the play reaches the set $D^{(l)}$, where $l = k \mod L$, and once the play reaches this set they will follow the stationary strategy profile x. The expected length of the block will be proportional to $\beta^{(l)}$, so as to guarantee that the

limit discounted payoff will satisfy Eq. (4). Let $i^{(l)}$ be the signaller at $D^{(l)}$, and let $(s^{(l)}, a^{(l)}_{i^{(l)}}) \in D^{(l)} \times (A_i(s^{(l)}) \setminus \mathrm{supp}(x_{i^{(l)}}(s^{(l)})))$ be the state and the action that allow player $i^{(l)}$ to signal at $D^{(l)}$, that is, they satisfy $q(C \mid s^{(l)}, a^{(l)}_{i^{(l)}}, x_{-i^{(l)}}(s^{(l)})) = 1$. Player $i^{(l)}$ will indicate when the block ends: whenever the play visits state $s^{(l)}$ he will play the mixed action $(1-\eta^{(l)})x_{i^{(l)}}(s^{(l)}) + \eta^{(l)} a^{(l)}_{i^{(l)}}$, and in case his realized action is $a^{(l)}_{i^{(l)}}$ the block will end. The constants $(\eta^{(l)})_{l=1}^{L}$ will be chosen so that the expected length of each block is proportional to $\beta^{(l)}$.

We now turn to the formal proof. Assume w.l.o.g. that $\beta^{(l)} > 0$ for every $l \in \{1, 2, \cdots, L\}$. For every $l \in \{1, 2, \cdots, L\}$ consider the Markov chain over $D^{(l)}$ induced by the stationary strategy profile x, and denote by $\rho^{(l)}$ the invariant distribution given in Eq. (3). As mentioned before, $\rho^{(l)}(s) > 0$ for every state $s \in D^{(l)}$.

For every $\eta \in (0, \min_{l=1,2,\cdots,L} \beta^{(l)} \rho^{(l)}(s^{(l)}))$ define the following strategy profile $\widetilde{\sigma}^{\eta}$, which is defined only for histories that remain in C:

1. Set $l = 1$.
2. As long as the play is in $C \setminus D^{(l)}$, the players follow the stationary strategy profile $y_{D^{(l)};C}$ that leads the play to the set $D^{(l)}$.
3. As long as the play is in $D^{(l)} \setminus s^{(l)}$, the players play the mixed action profile $x(s)$, where s is the current state.
4. Once the play is in state $s^{(l)}$, the players play the mixed action profile $z^{(l)}$, where

$$z^{(l)}_i := \begin{cases} x_i(s^{(l)}), & \text{if } i \neq i^{(l)}, \\ (1-\eta^{(l)})x_i(s^{(l)}) + \eta^{(l)} a^{(l)}_{i^{(l)}}, & \text{if } i = i^{(l)}, \end{cases}$$

where $\eta^{(l)} = \frac{\eta}{\beta^{(l)}\rho^{(l)}(s^{(l)})}$. If the action $a^{(l)}_{i^{(l)}}$ has been played by player $i^{(l)}$ at state $s^{(l)}$, the index l is increased by 1 modulo L and the players continue to Step 2. Otherwise the players remain in Steps 3 and 4.

Since player $i^{(l)}$ is a signaller at $D^{(l)}$ w.r.t. C and x, the play under $\widetilde{\sigma}^{\eta}$ never leaves the set C. The reader can check that the strategy profile $\widetilde{\sigma}^{\eta}$ can be implemented by an automaton with size $|C| \times L$.

We finally verify that the limit discounted payoff of each player i under $\widetilde{\sigma}^{\eta}$ is close to $\sum_{l=1}^{L} \beta^{(l)} \gamma_i(D^{(l)}; x)$. Denote by $N^{(l)}_{\eta} := \mathbf{E}_{s^{(l)}, \widetilde{\sigma}^{\eta}}[\nu_{C \setminus D^{(l)}}]$ the expected number of stages the play stays in $D^{(l)}$ after it first arrived to state $s^{(l)}$. We note that $\lim_{\eta \to 0} N^{(l)}_{\eta} = \infty$, while the expected number of stages to reach state $s^{(l)}$ in Steps 2 and 3 is uniformly bounded. Consequently,

$$1 = \lim_{\eta \to 0} N^{(l)}_{\eta} \eta^{(l)} \rho^{(l)}(s^{(l)}) = \lim_{\eta \to 0} \eta \frac{N^{(l)}_{\eta}}{\beta^{(l)}},$$

and the result follows by setting σ^{ε} to coincide with $\widetilde{\sigma}^{\eta}$, for η sufficiently small. □

3.2. Exits from communicating sets

In this section we recall the notion of exit from a communicating set, which was used in Solan (1999), Vieille (2000a, 2000b), and Solan and Vieille (2002), and show that players can control the way in which the play leaves a communicating set.

Definition 3.5. Let C be a communicating set under the stationary strategy profile x. An *exit* from C w.r.t. x is a triplet (s, J, a_J) of a state $s \in C$, a set of players $J \subseteq I$, and an action profile $a_J \in \times_{i \in J} A_i(s)$ such that the following two conditions hold:

(E.1) If at state s the players in J play a_J while all other players play x_{-J}, the play leaves C with positive probability: $q(C \mid s, a_J, x_{-J}(s)) < 1$.
(E.2) J is a minimal set of players that has the property spelled out in (E.1): for every strict subset J' of J we have $q(C \mid s, a_{J'}, x_{-J'}) = 1$.

The set of all exits from a communicating set C under x is denoted $\mathrm{Exit}(C, x)$.

For every state $s' \in S$ and every mixed action profile $y(s') \in \times_{i \in I} \Delta(A_i(s'))$, the probability that under $y(s')$ an exit from C w.r.t. x is played when the play visits s' is

$$\sum_{\{(s,J,a_J) \in \mathrm{Exit}(C,x):\ s = s'\}} \left(\prod_{i \in J} y_i(a_i \mid s') \cdot \prod_{i \notin J} y_i(\mathrm{supp}(x_i(s')) \mid s') \right),$$

where $y_i(a_i \mid s')$ is the probability that action a_i is played under $y_i(s')$ and $y_i(\text{supp}(x_i(s') \mid s') := \sum_{a_i \in \text{supp}(x_i(s))} y_i(a_i \mid s')$. There may also be action profiles at s' that are not exits and lead the play outside C. Those are action profiles that are in $\{a_J\} \times \prod_{i \notin J} \text{supp}(x_i(s'))$ for some triplet (s', J, a_J) that satisfy (E.1) and not (E.2). The definition of exits implies that if y is sufficiently close to x in the maximum norm, then the per-stage probability that under $y(s')$ an exit from C w.r.t. x is played, given that an action profile that is in $\{a_J\} \times \prod_{i \notin J} \text{supp}(x_i(s'))$ for some triplet (s', J, a_J) that satisfies (E.1) is played, is high. This observation is summarized in the following lemma.

Lemma 3.6. *For every $\varepsilon > 0$ there is $\delta > 0$ such that the following condition holds: for every stationary strategy profile x, every communicating set C under x, every state $s' \in C$, and every mixed action $y(s') \in \times_{i \in I} \Delta(A_i(s'))$ that satisfies $\|x(s') - y(s')\|_\infty < \delta$, we have*

$$\frac{\sum_{\{(s,J,a_J) \in \text{Exit}(C,x):\ s=s'\}} \left(\prod_{i \in J} y_i(a_i \mid s') \cdot \prod_{i \notin J} y_i(\text{supp}(x_i(s')) \mid s') \right)}{\sum_{\{(s',J,a_J):\ q(C \mid s', a_J, x_{-J}(s')) < 1\}} \left(\prod_{i \in J} y_i(a_i \mid s') \cdot \prod_{i \notin J} y_i(\text{supp}(x_i(s')) \mid s') \right)} \geq 1 - \varepsilon.$$

Denote by ν_C^* the first time in which an exit from C is played:

$$\nu_C^* := \min\left\{ n \geq 0 \colon (s^n, J, a_J^n) \in \text{Exit}(C,x) \text{ and } a_i^n \in \text{supp}(x_i(s^n)) \ \ \forall i \notin J, \text{ for some } J \subseteq I \right\},$$

where $a_J^n := (a_i^n)_{i \in J}$. Note that if $\nu_C^* < \infty$ then $\nu_C^* < \nu_{C^c}$, provided the initial state is in C, where C^c is the complement of C.

Let $C \subset S$ be a communicating set under x, let $(s, J, a_J) \in \text{Exit}(C,x)$ be an exit from C w.r.t. x, let $s^0 \in C$ be the initial state, and let σ be a strategy profile. The probability that (s, J, a_J) is the first exit from C w.r.t. x that is played is given by

$$\mu(s^0, \sigma, C; s, J, a_J) := \mathbf{P}_{s^0,\sigma}\left(s^{\nu_C^*} = s, a_J^{\nu_C^*} = a_J, a_i^{\nu_C^*} \in \text{supp}(x_i(s)) \ \ \forall i \notin J \right).$$

By Lemma 3.6, when $\sigma(h^n)$ is close to $x(s^n)$ for every finite history $h^n \in H$, the sum $\sum_{(s,J,a_J) \in \text{Exit}(C,x)} \mu(s^0, \sigma, C; s, J, a_J)$ is close to $\mathbf{P}_{s^0,\sigma}(\nu_{C^c}^* < \infty)$.

The next result asserts that there is a simple strategy profile that ensures that the play leaves a communicating set according to any distribution over the exits.

Proposition 3.7. *Let C be a communicating set under the stationary strategy profile x and let β be a probability distribution over the set of exits* $\text{Exit}(C,x)$. *There is a strategy profile $\sigma = (\sigma_i)_{i \in I}$ that is defined as long as the play remains in C and satisfies the following properties:*

1. *For each player $i \in I$ the strategy σ_i can be implemented by automata with size $|C| \times |\text{supp}(\beta)|$.*
2. *For every initial state $s^0 \in C$, under σ the play leaves C with probability 1, that is, $\mathbf{P}_{s^0,\sigma}(\nu_{C^c} < \infty) = 1$.*
3. *For every initial state $s^0 \in C$, the distribution of the first exit that is played coincides with β, that is, $\mu(s^0, \sigma, C; s, J, a_J) = \beta(s, J, a_J)$ for every exit $(s, J, a_J) \in \text{Exit}(C,x)$.*

Proof. The idea of the proof is as follows: for each exit (s, J, a_J) the players will play the stationary strategy profile $y_{\{s\};C}$ until the play reaches state s, and at s they will play once the action profile $z = z(s, J, a_J)$ defined by

$$z_i := \begin{cases} x_i(s), & \text{if } i \notin J, \\ (1-\eta)x_i(s) + \eta a_i, & \text{if } i \in J, \end{cases}$$

before continuing with the next exit. The constants η will differ among the various exits, and will be chosen in such a way that the total probability that the play leaves the set C through each exit (s, J, a_J) is $\beta(s, J, a_J)$.

We now turn to the formal proof. Denote by $L := |\text{supp}(\beta)|$ and $\text{supp}(\beta) = \left\{ (s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}}), l \in \{1, 2, \cdots, L\} \right\}$. For every $\eta \in [0,1]$ and every $l \in \{1, 2, \cdots, L\}$ define

$$\eta^{(l)} := \left(\frac{\eta \beta(s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}})}{1 - \sum_{l' < l} \eta \beta(s^{(l')}, J^{(l')}, a^{(l')}_{J^{(l')}})} \right)^{1/|J^{(l)}|} \in [0,1]. \tag{5}$$

Note that for $l = 1$ the denominator in Eq. (5) is equal to 1. For every $l \in \{1, 2, \cdots, L\}$ let $z^{(l)}(\eta)$ be the mixed-action profile at state $s^{(l)}$ defined by

$$z_i^{(l)}(\eta) := \begin{cases} x_i(s^{(l)}), & \text{if } i \notin J^{(l)}, \\ (1-\eta^{(l)})x_i(s^{(l)}) + \eta^{(l)} a_i^{(l)}, & \text{if } i \in J^{(l)}. \end{cases}$$

Let $\sigma(\eta)$ be the strategy profile that plays in blocks of random length and is defined as long as the play remains in C, as follows.

1. Set $l := 1$.
2. Play the stationary strategy profile $y_{\{s^{(l)}\};C}$ until the play reaches the state $s^{(l)}$.
3. At state $s^{(l)}$ play the mixed action profile $z^{(l)}(\eta)$ and the current block ends.
4. If the realized action profile of the players in $J^{(l)}$ is $a^{(l)}_{J^{(l)}}$ and the play did not leave C, we go to Step 1.
5. If the realized action profile of the players in $J^{(l)}$ is not $a^{(l)}_{J^{(l)}}$, the index l is increased by 1 modulo L, and we go to Step 2.

The strategy profile $\sigma(\eta)$ can be implemented by automata with size $|C| \times L$. As soon as $\sum_{l=1}^{L} \eta^{(l)} > 0$ and $s^0 \in C$, the play leaves C with probability 1, that is, $\mathbf{P}_{s^0,\sigma(\eta)}(\nu_{C^c} < \infty) = 1$. Moreover, under $\sigma(\eta)$ with probability 1 the play leaves C through one of the exits in $\mathrm{supp}(\beta)$.

We finally argue that $\mu(s^0, \sigma, C; s, J, a_J) = \beta(s, J, a_J)$ for every exit $(s, J, a_J) \in \mathrm{Exit}(C, x)$. Eq. (5) implies that in every cycle of L blocks the probability that the exit $(s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}})$ is played is $\eta\beta(s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}})$, and once an exit is played, past play is forgotten. It follows that the total probability that the first exit from C that is played is $(s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}})$ is $\beta(s^{(l)}, J^{(l)}, a^{(l)}_{J^{(l)}})$, as desired. □

3.3. Perturbations of stationary strategies

For every player $i \in I$ let $\lambda \mapsto \underline{x}_i^{\lambda}$ be a semi-algebraic function that assigns a λ-discounted max–min strategy of player i to every discount factor λ. From now on we fix this function and we denote the limit stationary strategy by

$$\underline{x}_i^1 := \lim_{\lambda \to 1} \underline{x}_i^{\lambda}, \quad \forall i \in I,$$

and the limit stationary strategy profile by $\underline{x}^1 := (\underline{x}_i^1)_{i \in I}$.

For every player $i \in I$, the λ-discounted max–min strategy $\underline{x}_i^{\lambda}$ satisfies

$$\underline{v}_i^{\lambda}(s) \le (1-\lambda) u_i(s, \underline{x}_i^{\lambda}(s), x_{-i}(s)) + \lambda \varphi_i(s, \underline{x}_i^{\lambda}(s), x_{-i}(s); \underline{v}_i^{\lambda}), \quad \forall s \in S, \forall x_{-i}(s) \in \times_{j \ne i} \Delta(A_j).$$

By taking λ to 1 and setting $x_{-i}(s) = \underline{x}_{-i}^{\lambda}(s)$ we deduce that

$$\underline{v}_i^1(s) \le \varphi_i(s, \underline{x}^1(s); \underline{v}^1), \quad \forall s \in S, \forall i \in I. \tag{6}$$

Eq. (6) implies that every irreducible set under $\underline{x}^1$ is a communicating set. Consequently, when the initial state is outside $C^*(\underline{x}^1)$ and the players follow the stationary strategy $\underline{x}^1$, the play reaches a set in $\mathcal{C}_{\max}(\underline{x}^1)$ a.s. We call the states in $S \setminus C^*(\underline{x}^1)$ *transient states* w.r.t. $\underline{x}^1$.

Definition 3.8. Let $\lambda_0 > 0$. A strategy $\sigma_i \in \Sigma_i$ of player i is called $(\underline{x}_i, \lambda_0)$-*perturbation* if there is a function $\lambda_i : H \to [\lambda_0, 1)$ that satisfies the following properties:

- $\sigma_i(h^n) = \underline{x}_i^{\lambda_i(h^n)}(s^n)$ for every finite history $h^n \in H$.
- For every strategy profile $\sigma_{-i} \in \Sigma_{-i}$ and every initial state $s^0 \in S$ we have $\mathbf{P}_{s^0,\sigma_i,\sigma_{-i}}(\lim_{n\to\infty} \lambda_i(h^n) = 1) = 1$.

In words, a strategy σ_i is $(\underline{x}_i, \lambda_0)$-perturbation if it plays only mixed action in the range of $\underline{x}_i$ around $\lambda = 1$, and if $\lambda_i(h^n)$ converges to 1.

Let $\sigma = (\sigma_i)_{i \in I}$ be a strategy profile in which σ_i is $(\underline{x}_i, \lambda_0)$-perturbation for every player $i \in I$. If λ_0 is close to 1, then for every $n \ge 0$ the quantity $\lambda(h^n)$ is close to 1, hence the mixed action $\sigma(h^n)$ is close to the mixed action $\underline{x}^1(s^n)$. It follows that the play under σ remains for long periods in irreducible sets in $\mathcal{I}(\underline{x}^1)$. Moreover, during each visit to an irreducible set $D \in \mathcal{I}(\underline{x}^1)$ the play resembles the play under $\underline{x}^1$, in the sense that the frequency in which each state in D is visited is close to the invariant distribution under $\underline{x}^1$ over D defined in Eq. (3). This observation is summarized by the following lemma, in which we denote by σ_{h_n} the strategy profile σ conditioned on the history h^n, that is, $\sigma_{h_n} = (\sigma_{i,h_n})_{i \in I}$, where the strategy σ_{i,h_n} is defined by

$$\sigma_{i,h_n}(\widehat{h}^m) = \sigma_i(s^0, a^0, \cdots, s^{n-1}, a^{n-1}, \widehat{s}^0, \widehat{a}^0, \widehat{s}^1, \widehat{a}^1, \cdots, \widehat{s}^m), \quad \forall \widehat{h}^m = (\widehat{s}^0, \widehat{a}^0, \cdots, \widehat{s}^m) \in H.$$

Lemma 3.9. *for every $\varepsilon > 0$ there are $N \in \mathbf{N}$ and $\lambda_* < 1$ sufficiently close to 1, such that for every $\lambda_0 \in [\lambda_*, 1)$, every strategy profile σ in which each player plays an $(\underline{x}_i, \lambda_0)$-perturbation for every player $i \in I$, and for every finite history $h^n \in H$ in which $s^n \in D$, we have*

$$\mathbf{P}_{s^n,\sigma_{h^n}}(\nu_{D^c} > n + N) \ge 1 - \varepsilon \tag{7}$$

and

$$\left| \frac{1}{N} \sum_{m=1}^{N} \mathbf{P}_{s^n,\sigma_{h^n}}(s^{n+m} = s \mid \nu_{D^c} > n+N) - \rho_{D,\underline{x}^1}(s) \right| \leq \varepsilon, \quad \forall s \in D. \tag{8}$$

3.4. Uniform ε-max–min strategies

Mertens and Neyman (1981) proved that each player in a two-player zero-sum stochastic game has a uniform ε-max–min strategy, namely, a strategy that guarantees the uniform value up to ε in every discounted game, provided the discount factor is sufficiently close to 1. As mentioned above, Neyman (2003) extended the result to multiplayer stochastic games. We here present the part of his result that we need in this paper.

A strategy profile of player i is subgame-perfect uniform ε-max–min if it is λ-discounted ε-optimal after every history, provided the discount factor is sufficiently close to 1. Formally,

Definition 3.10. Let $\varepsilon > 0$ and let $i \in I$ be a player. A strategy $\sigma_i \in \Sigma_i$ is *subgame-perfect uniform ε-max–min* if for every $n \geq 0$ there is $\lambda^{(n)} \in [0,1)$ such that for every $\lambda \in [\lambda^{(n)}, 1)$, every finite history $h^n \in H$ of length n, and every strategy profile $\sigma_{-i} \in \Sigma_{-i}$ we have

$$\gamma_i^{\lambda}(s^n; \sigma_{i,h^n}, \sigma_{-i,h^n}) \geq \underline{v}_i^1(s^n) - \varepsilon,$$

where $\sigma_{-i,h^n} = (\sigma_{j,h^n})_{j \neq i}$.

Note that the threshold $\lambda^{(n)}$ depends on the length of the history. When σ_i is a subgame-perfect uniform ε-max–min strategy, for every finite history $h^n \in H$, every bounded stopping time $\nu > n$, and every strategy profile $\sigma_{-i} \in \Sigma_{-i}$ we have

$$\underline{v}_i^1(s^n) \leq \mathbf{E}_{s^n,\sigma_{i,h^n},\sigma_{-i,h^n}}[\underline{v}_i^1(s^{\nu})] + 2\varepsilon. \tag{9}$$

Proposition 3.11 *(Neyman, 2003). For every $\varepsilon > 0$, every λ_0 sufficiently close to 1, and every player $i \in I$ there exists a subgame-perfect uniform ε-max–min strategy $\widehat{\sigma}_i^{\varepsilon,\lambda_0}$ that is $(\underline{x}_i, \lambda_0)$-perturbation.*

We will be interested in communicating sets under $\underline{x}^1$, and will identify two types of such communicating sets. Roughly, the type of a communicating set C under $\underline{x}^1$ is A if under some strategy profile that is composed of subgame-perfect uniform ε-max–min strategies of the players that are $(\underline{x}_i, \lambda_0)$-perturbations, with positive probability the play never leaves C (after some finite history), and the type is B if under some such strategy profile the play is bound to leave C with arbitrarily high probability (after some finite history).

Definition 3.12. A communicating set C under $\underline{x}^1$ has *type A* if there exists a positive number $\zeta > 0$ and for every $\varepsilon > 0$ and every $\lambda_0 > 0$ there exists a finite history $h^n_{\varepsilon,\lambda_0} \in H$ with $s^n_{\varepsilon,\lambda_0} \in C$, where $s^n_{\varepsilon,\lambda_0}$ is the last state of $h^n_{\varepsilon,\lambda_0}$, and for each player $i \in I$ there exists a subgame-perfect uniform ε-max–min strategy $\widehat{\sigma}_i^{\varepsilon,\lambda_0}$ that is $(\underline{x}_i, \lambda_0)$-perturbation, such that

$$\mathbf{P}_{s^n_{\varepsilon,\lambda_0},(\widehat{\sigma}^{\varepsilon,\lambda_0}_{i,h^n_{\varepsilon,\lambda_0}})_{i\in I}}(\nu_{C^c} = \infty) \geq \zeta.$$

A communicating set C under $\underline{x}^1$ has *type B* if for every $\varepsilon > 0$ and every $\lambda_0 \in [0,1)$ there exists a finite history $h^n_{\varepsilon,\lambda_0} \in H$ with $s^n_{\varepsilon,\lambda_0} \in C$, and for each player $i \in I$ there exists a subgame-perfect uniform ε-max–min strategy $\widehat{\sigma}_i^{\varepsilon,\lambda_0}$ that is $(\underline{x}_i, \lambda_0)$-perturbation, such that

$$\mathbf{P}_{s^n_{\varepsilon,\lambda_0},(\widehat{\sigma}^{\varepsilon,\lambda_0}_{i,h^n_{\varepsilon,\lambda_0}})_{i\in I}}(\nu_{C^c} = \infty) \leq \varepsilon.$$

Because a subgame-perfect uniform ε-max–min strategy that is $(\underline{x}_i, \lambda_0)$-perturbation is also a subgame-perfect uniform ε'-max–min strategy that is $(\underline{x}_i, \lambda_0')$-perturbation whenever $\varepsilon' \geq \varepsilon$ and $\lambda_0' \geq \lambda_0$, it follows that any communicating set has at least one type. The definition does not rule out the possibility that a communicating set has both types. For a communicating set C of either type we say that the histories $(h^n_{\varepsilon,\lambda_0})_{\varepsilon>0,\lambda_0\in[0,1)}$ and the strategy profiles $(\widehat{\sigma}^{\varepsilon,\lambda_0})_{\varepsilon>0,\lambda_0\in[0,1)}$ *support* the type of C.

In communicating sets of type A the play may stay in C ad infinitum when all players play a subgame-perfect uniform ε-max–min strategy that is $(\underline{x}_i, \lambda_0)$-perturbation. Since the strategies that the players play are subgame-perfect uniform ε-max–min, this implies that the long-run average payoff for all players on the event that the play stays in the set is high. We will use this property to prove that in this case there is a simple strategy profile that remains in C and yields to all players a high payoff. This is done in Section 3.5. In communicating sets of type B the play may leave C with arbitrarily high probability. We will show that in this case there is a simple strategy profile under which the play leaves C with probability 1 and the expected continuation uniform max–min value is high. This is done in Section 3.6.

3.5. Communicating sets of type A

The following result, together with Proposition 3.4, implies that if C is a communicating set under $\underline{x}^1$ of type A, then for every $\varepsilon > 0$ there is a simple ε-acceptable max–min strategy profile at every initial state that lies in C.

Proposition 3.13. *If C is a communicating set under $\underline{x}^1$ of type A, then there exist $L \leq \min\{|C|, |I|\}$, irreducible sets $D^{(1)}, \cdots, D^{(L)} \in \mathcal{I}_C(\underline{x}^1)$, and a probability distribution $\beta = (\beta^{(l)})_{l=1}^L \in \Delta(\{1, 2, \cdots, L\})$, such that $\sum_{l=1}^L \beta^{(l)} \gamma_i(D^{(l)}; \underline{x}^1) \geq \underline{v}_i^1(C)$.*

Proof. Fix $\varepsilon > 0$ and let λ_0 be sufficiently close to 1 such that Eqs. (7) and (8) holds for every strategy profile that is composed of λ_0-perturbations. Let $(h^n_{\varepsilon,\lambda_0})_{\varepsilon>0}$ and $(\sigma^{\varepsilon,\lambda_0})_{\varepsilon>0}$ be the finite histories and strategy profiles that support the fact that C has type A. By assumption,

$$\mathbf{P}_{s^n_{\varepsilon,\lambda_0}, \sigma^{\varepsilon,\lambda_0}_{h^n_{\varepsilon,\lambda_0}}}(\nu_{C^c} = \infty) \geq \zeta.$$

This implies that there are $n' = n'(\varepsilon, \lambda_0) \in \mathbf{N}$ and a history $h^{n'} = h^{n'}(\varepsilon, \lambda_0) \in H$ that extends the history $h^n_{\varepsilon,\lambda_0}$ such that $s^{n'} \in C$ and

$$\mathbf{P}_{s^{n'}, \sigma^{\varepsilon,\lambda_0}_{h^{n'}}}(\nu_{C^c} = \infty) > 1 - \varepsilon.$$

Since the strategies $(\sigma_i^{\varepsilon,\lambda_0})_{i\in I}$ are subgame-perfect uniform ε-max–min,

$$\lim_{\lambda\to 1} \gamma_i^\lambda(s^{n'}; \sigma^{\varepsilon,\lambda_0}_{h^{n'}}) \geq \underline{v}_i^1(C) - \varepsilon, \quad \forall i \in I. \tag{10}$$

Eq. (8) implies that

$$d_\infty\left(\lim_{\lambda\to 1} \gamma_i^\lambda(s^{n'}; \sigma^{\varepsilon,\lambda_0}_{h^{n'}}), \operatorname{conv}(\{\gamma(D; \underline{x}^1), D \in \mathcal{I}_C(\underline{x}^1)\})\right) \leq \varepsilon, \tag{11}$$

where $d_\infty(z, B) = \sup_{b\in B} d_\infty(z, b)$ for every $z \in \mathbf{R}^d$ and every $B \subseteq \mathbf{R}^d$. By Eqs. (10) and (11), there is a probability distribution β_ε over the set $\mathcal{I}_C(\underline{x}^1)$ that satisfies

$$\sum_{D\in\mathcal{I}_C(\underline{x}^1)} \beta_\varepsilon(D)\gamma_i(D; \underline{x}^1) \geq \underline{v}_i^1(C) - 2\varepsilon, \quad \forall i \in I. \tag{12}$$

The number of elements in $\mathcal{I}_C(\underline{x}^1)$ is at most $|C|$, and by Carathéodory's Theorem it is sufficient to consider probability distributions β_ε whose support is at most $|I|$. Since Eq. (12) holds for every $\varepsilon > 0$, and since the space $\Delta(\mathcal{I}_C(\underline{x}^1))$ is compact, the result follows. □

3.6. Communicating sets of type B

In this section we construct a simple strategy profile that ensures that the play leaves a communicating set and the expected continuation uniform max–min value of all players is high.

Proposition 3.14. *For every communicating set C w.r.t. $\underline{x}^1$ of type B there exists a probability distribution β over the set of exits $\operatorname{Exit}(C, \underline{x}^1)$ that satisfies the following two conditions:*

- *The support of β contains at most $|I|$ exits.*
- *The expected continuation uniform max–min value under β is high:*

$$\sum_{(s,J,a_J)\in\operatorname{Exit}(C,\underline{x}^1)} \beta(s, J, a_J)\varphi_i(s, a_J, \underline{x}^1_{-J}(s); \underline{v}_i^1) \geq \underline{v}_i^1(C).$$

Proof. Let $(h^n_{\varepsilon,\lambda_0})_{\varepsilon>0,\lambda_0\in[0,1)}$ and $(\widehat{\sigma}^{\varepsilon,\lambda_0})_{\varepsilon>0,\lambda_0\in[0,1)}$ be the histories and strategy profiles that support the type of C, and fix $\varepsilon > 0$. Since for every player $i \in I$ the strategy $\widehat{\sigma}_i^{\varepsilon,\lambda_0}$ is an $(\underline{x}_i, \lambda_0)$-perturbation, by Lemma 3.6

$$\sum_{(s,J,a_J)\in\operatorname{Exit}(C,\underline{x}^1)} \mu(s^0, \widehat{\sigma}^{\varepsilon,\lambda_0}, C; s, J, a_J) \geq (1 - \varepsilon)^2 > 1 - 2\varepsilon,$$

provided λ_0 is sufficiently small. By Eq. (9),

$$\sum_{(s,J,a_J)\in \mathrm{Exit}(C,\underline{x}^1)} \mu(s^0,\widehat{\sigma}^{\varepsilon,\lambda_0},C;s,J,a_J)\varphi_i(s,a_J,\underline{x}_{-J};\underline{v}_i^1) > \underline{v}_i^1(C)-4\varepsilon. \tag{13}$$

By letting ε go to 0 in Eq. (13) we deduce that there is a probability distribution β on the set of exits that satisfies

$$\sum_{(s,J,a_J)\in \mathrm{Exit}(C,\underline{x}^1)} \beta(s,J,a_J)\varphi_i(s,a_J,\underline{x}_{-J};\underline{v}_i^1) \geq \underline{v}_i^1(C), \quad \forall i\in I. \tag{14}$$

Since the condition in Eq. (14) involves $|I|$ coordinates, by Carathéodory's Theorem we can assume w.l.o.g. that the support of β contains at most $|I|$ exits, as desired. □

3.7. The construction of a max–min ε-acceptable strategy profile

In this section we complete the proof of Theorem 2.4. Fix $\varepsilon>0$. We will define a max–min ε-acceptable strategy profile $\sigma^{*,\varepsilon}$. This strategy profile will follow the stationary strategy profile $\underline{x}^1$ in states that are transient w.r.t $\underline{x}^1$, thereby ensuring that the play reaches a maximal communicating set w.r.t. $\underline{x}^1$. Moreover, for every maximal communicating set C w.r.t. $\underline{x}^1$, whenever the play enters C the strategy profile $\sigma^{*,\varepsilon}$ will coincide with the strategy profile given by Propositions 3.4 and 3.13 (if C has type A) or Propositions 3.7 and 3.14 (if C has type B).

Define a sequence $(k_n)_{n\geq 0}$ of stopping times that indicates when the play enters a maximal communicating set or visits a transient state w.r.t. $\underline{x}^1$:

$$k_0 := 0,$$

and for every $n\geq 0$,

$$k_{n+1} := \min\{m>n\colon s^m\notin C^*(\underline{x}^1),\ \text{or } s^m\in C\in\mathcal{C}_{\max}(\underline{x}^1) \text{ and } s^n\notin C\}.$$

We now turn to the formal definition of $\sigma^{*,\varepsilon}$. For every $n\geq 0$, define the strategy $\sigma^{*,\varepsilon}$ between stages k_n (including) and $k_{n+1}-1$ (excluding) as follows:

- If $s^{k_n}\notin C^*(\underline{x}^1)$, at stage k_n the strategy profile $\sigma^{*,\varepsilon}$ coincides with $\underline{x}^1$, that is, $\sigma^{*,\varepsilon}(h^{k_n}) := \underline{x}^1(s^{k_n})$.
- Suppose that $s^{k_n}\in C\in\mathcal{C}_{\max}(\underline{x}^1)$ and C is a maximal communicating set of type A. By Propositions 3.4 and 3.13 there is a strategy profile $\sigma^{(1)}$ that is defined as long as the play remains in C, can be implemented by automata with size $|C|\times|I|$, under which the play does not leave C, and the limit discounted payoff of each player is at least $\underline{v}_i^1(C)-\varepsilon$. The conditional strategy profile $\sigma^{*,\varepsilon}_{h^{k_n}}$ coincides with the strategy profile $\sigma^{(1)}$. Note that in this case the play under $\sigma^{*,\varepsilon}_{h^{k_n}}$ never leaves C, that is, $k_{n+1}=\infty$.
- Suppose that $s^{k_n}\in C\in\mathcal{C}_{\max}(\underline{x}^1)$ and C is a maximal communicating set of type B. By Propositions 3.7 and 3.14 there is a strategy profile $\sigma^{(2)}$ that is defined as long as the play remains in C, can be implemented by automata with size $|C|\times|I|$, under which the play leaves C with probability 1, and the expected continuation max–min value of each player i is at least $\underline{v}_i^1(C)$. The conditional strategy profile $\sigma^{*,\varepsilon}_{h^{k_n}}$ coincides with the strategy profile $\sigma^{(2)}$ until the play leaves C.

The reader can verify that the individual strategy of each player can be implemented by an automaton with size $|S|\times|I|$.

Lemma 3.15. *Under the strategy profile $\sigma^{*,\varepsilon}$, with probability 1 the play reaches a maximal communicating set of type A.*

Proof. Assume to the contrary that the claim does not hold. It follows that there is a closed subset of transient states and maximal communicating sets of type B; that is, there is a collection $\{C_1,C_2,\cdots,C_L\}$ of maximal communicating sets w.r.t. $\underline{x}^1$ of type B and a subset $T\subseteq S\setminus C^*(\underline{x}^1)$ of transient states, such that

- $q\left(\left(\cup_{l=1}^L C_l\right)\cup T \mid s,\underline{x}^1(s)\right)=1$ for every state $s\in T$.
- For every $l=1,2,\cdots,L$, every state $s'\in C_l$, and every exit $(s,J,a_J)\in\mathrm{Exit}(C,\underline{x}^1)$ that satisfies $\mu(s',\sigma^{*,\varepsilon},C;s,J,a_J)>0$ we have

$$q\left(\left(\cup_{l=1}^L C_l\right)\cup T \mid s,a_J,\underline{x}^1_{-J}(s)\right)=1.$$

This implies that either there exists an irreducible set w.r.t. $\underline{x}^1$ which is a subset of T, or there exists a communicating set under $\underline{x}^1$ that strictly contains one of the sets $C_1,C_2,\cdots,C_L$. The first alternative contradicts the fact that $C^*(\underline{x}^1)$ contains all maximal communicating sets, while the second alternative contradicts the fact that $C_1,\cdots,C_L$ are maximal communicating sets. □

Define the stopping time N as the minimal integer n such that s^{k_n} belongs to a maximal communicating set of type A:

$$N := \min\{n \geq 0 \colon s^{k_n} \in C \in \mathcal{C}_{\max}(\underline{x}^1), C \text{ has type A}\}.$$

The definition of the strategy profile $\sigma^{*,\varepsilon}$ in transient states and on maximal communicating sets of type B (see Proposition 3.14) imply that the value process is a submartingale, that is, for every player $i \in I$, the sequence $(\underline{v}_i^1(s^{k_n}))_{n=1}^N$ is a submartingale under $\sigma^{*,\varepsilon}$.

Together with Propositions 3.4 and 3.13 we now deduce that the strategy profile $\sigma^{*,\varepsilon}$ is max–min 2ε-acceptable.

3.8. The construction of a stationary correlated min–max ε-acceptable strategy

In this section we prove Theorem 2.5 using the ideas presented in the proof of Theorem 2.4 and few additional ideas.

For every state $s \in S$ and every $\lambda \in [0,1]$ let $G^\lambda(s)$ be the normal-form game with (i) player set I, (ii) the action set of each player $i \in I$ is A_i, and (iii) the payoff function of each player $i \in I$ is

$$U_i^\lambda(s;a) := (1-\lambda)u_i(s,a) + \lambda \sum_{s' \in S} q(s' \mid s,a)\overline{v}_i^\lambda(s'), \quad \forall a \in A.$$

This is the one-shot game played at state s in which the continuation payoff of each player is given by his expected discounted min–max value at tomorrow's state.

For every state $s \in S$ and every $\lambda \in [0,1)$ denote by $E^\lambda(s) \subseteq \times_{i \in I}\Delta(A_i)$ the set of Nash equilibria of the game $G^\lambda(s)$, and let $E^\lambda := \times_{s \in S} E^\lambda(s) \subseteq (\times_{i \in I}\Delta(A_i))^{|S|}$ be the set of stationary strategy profiles composed of equilibria of the games $(G^\lambda(s))_{s \in S}$. Note that for every mixed action profile $x(s) \in E^\lambda(s)$, the payoff to each player $i \in I$ in $G^\lambda(s)$ is at least $\overline{v}_i^\lambda(s)$:

$$\overline{v}_i^\lambda(s) \leq U_i^\lambda(s, x(s)) = (1-\lambda)u_i(s,x(s)) + \lambda \sum_{s' \in S} q(s' \mid s, x(s))\overline{v}_i^\lambda(s'), \tag{15}$$

where $U_i^\lambda(s;x(s))$ is the multilinear extension of $U_i^\lambda(s;\cdot)$ to $\times_{j \in I}\Delta(A_j)$, for each player $i \in I$.

Denote the set of all accumulation points of the sets $(E^\lambda(s))_{\lambda \in [0,1)}$ as λ goes to 1 by

$$E^1(s) := \limsup_{\lambda \to 1} E^\lambda(s).$$

That is, $E^1(s)$ is the set of all accumulation points of sequences $(x^{(k)})_{k \geq 0}$, where $x^{(k)}(s) \in E^{\lambda^{(k)}}(s)$ for every state $s \in S$ and every $k \geq 0$, for some sequence $(\lambda^{(k)})_{k \geq 0}$ that converges to 1. Every point $x \in E^1(s)$ is a Nash equilibrium of the game $G^1(s)$, yet there may be Nash equilibria of $G^1(s)$ that are not in $E^1(s)$. By taking the limit of Eq. (15) as λ goes to 1 we deduce that

$$\overline{v}_i^1(s) \leq \sum_{s' \in S} q(s' \mid s, x(s))\overline{v}_i^1(s') = \varphi(s, x(s); \overline{v}^1), \quad \forall s \in S, \forall x \in E^1(s), \forall i \in I. \tag{16}$$

We will use the following variation of communicating set, which is analogous to Definition 3.1.

Definition 3.16. A set of states $C \subseteq S$ is *communicating under* E^1 if the following conditions hold:

(C′.1) $\overline{v}^1(s) = \overline{v}^1(s')$ for every two states $s, s' \in C$. Denote by $\overline{v}^1(C)$ the common uniform max–min value in states in C.

(C′.2) For every state $s \in C$ there is a stationary strategy profile $x \in E^1$ and an enlargement $y_{\{s\},C}$ of x such that

(i) the set C is closed under $y_{\{s\},C}$, and

(ii) under $y_{\{s\},C}$ the play reaches state s with probability 1, provided the initial state is in C:

$$\mathbf{P}_{s^0, y_{\{s\},C}}(\nu_{\{s\}} < \infty) = 1, \quad \forall s^0 \in C.$$

Denote by $\mathcal{C}_{\max}(E^1)$ the set of all maximal communicating set under E^1 w.r.t. set inclusion and by $C^*(E^1) := \cup_{C \in \mathcal{C}_{\max}(E^1)} C$ the union of all maximal communicating sets under E^1. The reader can verify that if C_1 and C_2 are two communicating sets under E^1 with nonempty intersection, then $C_1 \cup C_2$ is also a communicating set under E^1. Consequently, maximal communicating sets are disjoint.

By Eq. (16), for every stationary strategy profile $x \in E^1$, every irreducible set $D \in \mathcal{I}(x)$ is a communicating set under E^1. This implies that whenever the initial state is not in $C^*(E^1)$ and the players follow a stationary strategy $x \in E^1$, the play reaches some communicating set with probability 1.

The *state-action frequency vector* of a correlated strategy at a given initial state is the long-run average frequency in which each action profile is played at each state.

Definition 3.17. Let τ be a correlated strategy. The *state-action frequency vector* of τ at the initial state $s^0 \in S$ is the probability distribution $\rho_{s^0,\tau}$ over $S \times A$ that is defined as follows:

$$\rho_{s^0,\tau}(s,a) := \lim_{N\to\infty} \frac{1}{N} \mathbf{E}_{s^0,\tau}\left[\sum_{n=1}^{N} \mathbf{1}_{\{s^n=s, a^n=a\}}\right], \quad \forall (s,a) \in S \times A. \tag{17}$$

The state-action frequency vector is well defined only if the $|S| \times |A|$ limits defined in Eq. (17) exist. The *state frequency* of state s under the correlated strategy τ at the initial state s^0 is

$$\rho_{s^0,\tau}(s) := \sum_{a\in A} \rho_{s^0,\tau}(s,a).$$

We will consider below only correlated strategies for which the state-action frequency vector exists, hence issues of nonexistence of the state-action frequency vector and of the state frequency vector will not arise.

The state-action frequency vector can be related to the limit discounted payoff as follows:

$$\lim_{\lambda\to 1} \gamma_i^\lambda(s^0;\tau) = \sum_{s\in S, a\in A(s)} \rho_{s^0,\tau}(s,a) u(s,a).$$

For every communicating set C under E^1 denote by $\Sigma_{\text{corr}}(C)$ (resp. $\Sigma^{\text{stat}}_{\text{corr}}(C)$) the set of all correlated profiles (respectively stationary correlated profiles) under which the play never leaves C, provided the initial state is in C. For every initial state $s^0 \in C$ denote the set of all state-action frequency vectors of correlated strategies that remain in C by

$$\Pi_{\text{corr}}(s^0;C) := \{\rho_{s^0,\tau} : \tau \in \Sigma_{\text{corr}}(C)\} \subset \mathbf{R}^{|S|\times|A|},$$

and the set of all state-action frequency vectors of correlated stationary strategies that remain in C by

$$\Pi^{\text{stat}}_{\text{corr}}(s^0;C) := \{\rho_{s^0,\tau} : \tau \in \Sigma^{\text{stat}}_{\text{corr}}(C)\} \subset \mathbf{R}^{|S|\times|A|}.$$

The following result, which states that the set of state-action frequency vectors of correlated strategies coincides with the closure of the set of state-action frequency vectors of correlated stationary strategies, follows from Altman (1999, Theorem 11.1), Rosenberg et al. (2004), or Mannor and Tsitsiklis (2005).

Theorem 3.18. *For every communicating set C under E^1 and every initial state $s^0 \in C$ we have $\Pi_{\text{corr}}(s^0;C) = \text{closure}(\Pi^{\text{stat}}_{\text{corr}}(s^0;C))$, where for every set X in a Euclidean space, $\text{closure}(X)$ is the closure of X.*

For every communicating set C under E^1, define the set of exits from C as the set of all pairs of a state s in C and an action profile a at s that lead the play outside C with positive probability:

$$\text{Exit}(C) := \{(s,a) : s \in S, a \in A(s), q(C \mid s,a) < 1\}.$$

As before we denote by ν^*_C the first stage in which an exit from C is played. For every correlated strategy τ, the probability that the first exit from C that is used is (s,a) is given by

$$\mu(s^0,\tau,C;s,a) := \mathbf{P}_{s^0,\tau}\left(s^{\nu^*_C} = s, a^{\nu^*_C} = a\right).$$

The analog of Theorem 3.11 that we will use in the proof of Theorem 2.5 is the following result, which is a consequence of the study of Solan and Vieille (2002) on correlated equilibrium in stochastic games.

Theorem 3.19 *(Solan and Vieille, 2002). For every maximal communicating set C under E^1 at least one of the following conditions hold.*

(SV.1) *For every $\varepsilon > 0$ there exists a strategy profile σ^ε that is defined for finite histories that remain in C, under which the play never leaves C, provided the initial state is in C, and satisfies*

$$\lim_{\lambda\to 1} \gamma_i^\lambda(s^0;\sigma^\varepsilon) \geq \overline{v}^1_i(C) - \varepsilon, \quad \forall s^0 \in C.$$

(SV.2) *There is a strategy profile σ that is defined for finite histories that remain in C that satisfies the following conditions:*

- *Under σ the play leaves C with probability 1, that is, $\mathbf{P}_{s^0,\sigma}(\nu_{C^c} < \infty) = 1$, provided $s^0 \in C$.*
- *The expected uniform min–max value upon leaving C is at least the uniform min–max value in C, that is,*

$$\sum_{(s,a)\in \text{Exit}(C)} \mu(s^0,\sigma,C;s,a)\varphi_i(s,a;\overline{v}^1_i) \geq \overline{v}_i(C), \quad \forall i \in I.$$

Communicating sets under E^1 that satisfy Condition (SV.1) (resp. Condition (SV.2)) of Theorem 3.19 correspond to communicating sets under x of type A (resp. type B). Fix $\varepsilon > 0$. By Theorem 3.18, in the former case there is a stationary correlated strategy x_C that satisfies

$$\lim_{\lambda \to 1} \gamma_i^\lambda(s^0; x_C) \geq \overline{v}_i^1(C) - 2\varepsilon. \tag{18}$$

Similarly, in the latter case, there is a stationary correlated strategy x_C that satisfies $\mathbf{P}_{s^0, x_C}(\nu_{C^c} < \infty) = 1$, provided $s^0 \in C$, and

$$\sum_{(s,a) \in \mathrm{Exit}(C)} \mu(s^0, x_C, C; s, a) \varphi_i(s, a; \overline{v}_i^1) \geq \overline{v}_i(C). \tag{19}$$

Let $x \in E^1$ be any stationary strategy profile. We are now ready to define a stationary correlated strategy x^* that is min–max ε-acceptable. For every state $s \in S$:

- If $s \notin C^*(E^1)$, set $x^*(s) := x(s)$.
- If s is in some maximal communicating set $C \in \mathcal{C}_{\max}(E^1)$, set $x^*(s) := x_C(s)$.

Under the stationary correlated strategy x^*, with probability 1 the play reaches a communicating set that satisfies Condition (SV.1). Moreover, denoting by N the first stage in which such a communicating set is reached, by Eqs. (16) and (19) we have

$$\overline{v}_i^1(s^0) \leq \mathbf{E}_{s^0, x^*}[\overline{v}_i^1(s^N)], \quad \forall i \in I, \forall s^0 \in S,$$

that is, for every player $i \in I$ the process $(\overline{v}_i^1(s^{k_n}))_{n \geq 0}$ is a submartingale, where $(k_n)_{n \in \geq 0}$ are the stages in which the play enters a maximal communicating set under E^1 or visits a transient state under E^1, that is, a state in $S \setminus C^*(E^1)$. Together with Eq. (18) this implies that

$$\lim_{\lambda \to 1} \gamma_i^\lambda(s^0; x^*) \geq \overline{v}_i^1(s^0) - 3\varepsilon, \forall i \in I, \forall s^0 \in S,$$

so that x^* is indeed max–min ε-acceptable, as desired.

3.9. Subgame perfectness

The notion of acceptability that we defined is not subgame perfect. That is, even if τ is a w-acceptable correlated strategy, there may be a finite history $h^n \in H$ such that $\limsup_{\lambda \to 1} \gamma_i^\lambda(s^n; \tau_{h^n}) < w_i(s^n)$ for some player $i \in I$. The following definition incorporates subgame perfectness into the definition of acceptability.

Definition 3.20. Let $\varepsilon \geq 0$. A correlated strategy τ is *subgame-perfect w-acceptable* if for every player $i \in I$, every finite history $h^n \in H$, and every discount factor λ sufficiently close to 1, we have $\gamma_i^\lambda(s^n; \tau_{h^n}) \geq w_i(s^n)$.

A stationary w-acceptable correlated strategy is in particular subgame perfect. In particular, the stationary correlated min–max ε-acceptable strategy that was constructed in the proof of Theorem 2.5 is subgame perfect. The reader can verify that the strategy profile that was constructed in the proof of Theorem 2.4 is subgame perfect as well.

3.10. Complexity issues

Our proof allows one to construct a max–min ε-acceptable strategy profile and a min–max ε-acceptable correlated stationary strategy. One necessary step in the construction is the calculation of the uniform max–min value and the min–max value of all players in all states. Arnsfelt Hansen et al. (2012) provided a polynomial-time algorithm to calculate the uniform value of a two-player stochastic game. It may be hoped that the algorithm can be extended to multiplayer games (recall that the technique of Chatterjee et al. (2008) implies that this problem is in the complexity class EXPTIME).

The max–min ε-acceptable strategy profile (resp. the min–max ε-acceptable correlated strategy) that we constructed may use irrational probabilities. Since the strategy profile can be implemented by an automaton, and since the invariant distribution of a Markov chain is continuous in the transition function, any approximation of the strategy profile by a strategy profile that uses only rational probabilities is max–min 2ε-acceptable (resp. min–max 2ε-acceptable). Thus, Theorem 2.4 (resp. Theorem 2.5) can be strengthened to require that the max–min ε-acceptable strategy profile (resp. the min–max ε-acceptable correlated strategy) that can be implemented by small automata uses only rational probabilities.

References

Altman, E., 1999. Constrained Markov Decision Processes. Chapman and Hall/CRC.

Arnsfelt Hansen, K., Koucký, M., Lauritzen, N., Bro Miltersen, P., Tsigaridas, E.P., 2012. Exact algorithms for solving stochastic games. Preprint. Extended abstract published at STOC 2011, pp. 205–214.

Blackwell, D., 1962. Discrete dynamic programming. Ann. Math. Stat. 33, 719–726.

Chatterjee, K., Majumdar, R., Henzinger, T.A., 2008. Stochastic limit-average games are in EXPTIME. Int. J. Game Theory 37, 219–234.

Fink, A.M., 1964. Equilibrium in a stochastic n-person game. J. Sci. Hiroshima Univ., Ser. A–I Math. 28, 89–93.

Flesch, J., Thuijsman, F., Vrieze, O.J., 1997. Stochastic games with additive transitions. Eur. J. Oper. Res. 179, 483–497.

Flesch, J., Schoenmakers, G., Vrieze, K., 2008. Stochastic games on a product state space. Math. Oper. Res. 33, 403–420.

Flesch, J., Schoenmakers, G., Vrieze, K., 2009. Stochastic games on a product state space: the periodic case. Int. J. Game Theory 38, 263–289.

Jaśkiewicz, A., Nowak, A.S., 2016. Zero-sum stochastic games. In: Başar, T., Zaccour, G. (Eds.), Handbook of Dynamic Game Theory. Springer International Publishing AG.

Jaśkiewicz, A., Nowak, A.S., 2017. Non-zero-sum stochastic games. In: Başar, T., Zaccour, G. (Eds.), Handbook of Dynamic Game Theory. Springer International Publishing AG.

Mannor, S., Tsitsiklis, J., 2005. On the empirical state-action frequencies in Markov decision processes under general policies. Math. Oper. Res. 30, 545–561.

Mertens, J.-F., Neyman, A., 1981. Stochastic games. Int. J. Game Theory 10, 53–66.

Mertens, J.-F., Sorin, S., Zamir, S., 2015. Repeated Games. Cambridge University Press.

Neyman, A., 1985. Bounded complexity justifies cooperation in the finitely-repeated prisoners' dilemma. Econ. Letters 19, 227–229.

Neyman, A., 2003. Real Algebraic tools in stochastic games. In: Neyman, A., Sorin, S. (Eds.), Stochastic Games and Applications. Kluwer Academic Publishers, pp. 57–75.

Neyman, A., 2012. Continuous-time stochastic games. Discussion Paper #616, Center for the Study of Rationality, Hebrew University of Jerusalem.

Neyman, A., Sorin, S., 2003. Stochastic Games and Applications, vol. 570. Springer Science & Business Media.

Rosenberg, D., Solan, E., Vieille, N., 2004. Approximating a sequence of observations by a simple process. Ann. Stat. 32, 2742–2775.

Ross, K.W., Varadarajan, R., 1991. Multichain Markov decision processes with a sample path constraint: a decomposition approach. Math. Oper. Res. 16, 195–207.

Rubinstein, A., 1986. Finite automata play the repeated prisoner's dilemma. J. Econ. Theory 39, 83–96.

Shapley, L.S., 1953. Stochastic games. Proc. Natl. Acad. Sci. USA 39, 1095–1100.

Simon, R.S., 2007. The structure of non-zero-sum stochastic games. Adv. Appl. Math. 38, 1–26.

Simon, R.S., 2012. A topological approach to quitting games. Math. Oper. Res. 37, 180–195.

Simon, R.S., 2016. The challenge of non-zero-sum stochastic games. Int. J. Game Theory 45, 191–204.

Solan, E., 1999. Three-player absorbing games. Math. Oper. Res. 24, 669–698.

Solan, E., 2001. Characterization of correlated equilibria in stochastic games. Int. J. Game Theory 30, 259–277.

Solan, E., Vieille, N., 2001. Quitting games. Math. Oper. Res. 26, 265–285.

Solan, E., Vieille, N., 2002. Correlated equilibrium in stochastic games. Games Econ. Behav. 38, 362–399.

Solan, E., Vieille, N., 2015. Stochastic games: a perspective. Proc. Natl. Acad. Sci. USA 112 (45), 13743–13746.

Solan, E., Ziliotto, B., 2016. Stochastic games with signals. Adv. Dynam. Evolutionary Games 14, 77–94.

Takahashi, M., 1964. Equilibrium points of stochastic non-cooperative n-person games. J. Sci. Hiroshima Univ., Ser. A–I Math. 28, 95–99.

Vieille, N., 2000a. Two-player stochastic games I: a reduction. Isr. J. Math. 119, 55–91.

Vieille, N., 2000b. Two-player stochastic games II: the case of recursive games. Isr. J. Math. 119, 93–126.