Subgame Maxmin Strategies in Zero-Sum Stochastic Games with Tolerance Levels

János Flesch[1] · P. Jean-Jacques Herings[2] · Jasmine Maes[2] · Arkadi Predtetchinski[2]

Abstract

We study subgame ϕ-maxmin strategies in two-player zero-sum stochastic games with a countable state space, finite action spaces, and a bounded and universally measurable payoff function. Here, ϕ denotes the tolerance function that assigns a nonnegative tolerated error level to every subgame. Subgame ϕ-maxmin strategies are strategies of the maximizing player that guarantee the lower value in every subgame within the subgame-dependent tolerance level as given by ϕ. First, we provide necessary and sufficient conditions for a strategy to be a subgame ϕ-maxmin strategy. As a special case, we obtain a characterization for subgame maxmin strategies, i.e., strategies that exactly guarantee the lower value at every subgame. Secondly, we present sufficient conditions for the existence of a subgame ϕ-maxmin strategy. Finally, we show the possibly surprising result that each game admits a strictly positive tolerance function ϕ^* with the following property: if a player has a subgame ϕ^*-maxmin strategy, then he has a subgame maxmin strategy too. As a consequence, the existence of a subgame ϕ-maxmin strategy for every positive tolerance function ϕ is equivalent to the existence of a subgame maxmin strategy.

Keywords Stochastic games · Zero-sum games · Subgame ϕ-maxmin strategies

JEL Classification C72 · C73

✉ P. Jean-Jacques Herings
     P.Herings@maastrichtuniversity.nl

     János Flesch
     J.Flesch@maastrichtuniversity.nl

     Jasmine Maes
     Jasmine.Maes@maastrichtuniversity.nl

     Arkadi Predtetchinski
     A.Predtetchinski@maastrichtuniversity.nl

[1]   Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

[2]   Department of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

1 Introduction

We consider two-player zero-sum stochastic games with a countable state space, finite action spaces, and a bounded and universally measurable payoff function. Such a game models a repeated interaction between two players with opposite objectives. The environment in which the interaction takes place is fully characterized by the state variable. The transition from one state variable to the next one is influenced by both players as well as an element of chance. Throughout the paper, we take the perspective of the maximizing player. We are interested in strategies of the maximizing player that guarantee the lower value at every subgame and call such strategies subgame maxmin strategies. Under the assumptions as made in the paper, the value may not exist, which explains why we consider the lower value instead.

As the name subgame maxmin strategy suggests, this concept is closely related to the concept of a subgame perfect equilibrium as defined in Selten ([19]). In two-player zero-sum games where the value exists, for conditions see Maitra and Sudderth ([13]) and Martin ([14]), the notions of a subgame maxmin strategy and a subgame minmax strategy coincide with the notion of a subgame optimal strategy. Moreover, in such games a strategy profile is a subgame perfect equilibrium if and only if it consists of a pair of subgame optimal strategies.

As illustrated by the Big Match, a game introduced in Gillette ([9]) and analyzed in Blackwell and Ferguson ([2]), even if the value exists, it is not guaranteed that optimal strategies exist, so a fortiori, subgame optimal strategies and subgame perfect equilibria may not exist. A large part of the literature therefore focuses on so-called subgame perfect ϵ-equilibria as defined in Radner ([17]). This equilibrium concept is more permissive than a subgame perfect equilibrium and consists of a strategy pair such that every player obtains the value at each history up to a fixed error term of $\epsilon/2$.

Instead of having a fixed error term at each subgame, we allow the error term to vary across different subgames. This error term is expressed as a function ϕ of the histories and is called the tolerance function. The central topic of this paper is the concept of a subgame ϕ-maxmin strategy. This is a strategy of the maximizing player that guarantees the lower value at every subgame within the allowed tolerance level. Intuitively, a subgame ϕ-maxmin strategy performs sufficiently well across all subgames. This type of strategy is related to the concept of ϕ-tolerance equilibrium as defined in Flesch and Predtetchinski ([6]). Indeed, if the value exists, then a strategy profile in which both players use a subgame $(\phi/2)$-optimal strategy is a ϕ-tolerance equilibrium.

One motivation for letting the tolerance level vary across subgames is given by Mailath, Postlewaite, and Samuelson ([11]) when introducing the concept of a contemporaneous perfect ϵ-equilibrium. The authors focus on games in which the payoff function of the players is given by the discounted sum of periodic rewards. Due to this discounting, there exists a period after which the maximal total discounted reward a player can receive is smaller than ϵ. If the allowed tolerance level ϵ is fixed across all subgames, any strategy will be an ϵ-maxmin strategy of a subgame in such a period. Therefore, the concept of subgame ϵ-maxmin strategy does not impose any restrictions on the actions chosen at a very distant future. The issue here is that it would be more intuitive to discount not only the reward but also the allowed tolerance level.

Additional motivation for letting the tolerance level vary across subgames stems from the fact that the notion of what is considered sufficiently good might be relative. For instance, Tversky and Kahneman [21] observe that people evaluate decisions with respect to a reference level. They find that significantly more people were willing to exert extra effort to save $5 on a $15 purchase than to save $5 on a $125 purchase. To apply this to the context of zero-sum

games, consider the following game to which we will refer as the high stakes-low stakes game. In the first stage of the game, a chance move determines whether the player will engage in the high stakes or the low stakes variant of this game. The high stakes and the low stakes games are identical in terms of possible strategies. The only difference is that the payoffs in the high stakes game are a thousand fold the payoffs in the low stakes game. Furthermore, assume that in the high stakes subgame, the payoff of player 1 ranges between 0 and 2000 and the value is 1000, while in the low stakes subgame the payoff of player 1 ranges between 0 and 2 and the value is 1. Players that evaluate decisions with respect to a reference level may label a strategy which guarantees a payoff of 999 in the high stakes game as sufficiently good. However, in the low stakes game, 0 corresponds to the minimum payoff. Allowing the tolerance level to vary across subgames can therefore be used to accommodate such behavioral effects into the model of zero-sum stochastic games.

Another case where history-dependent tolerance levels are natural is the following. In situations that commonly occur, a player may use a familiar strategy irrespective of the scale of the payoffs. To understand this better, imagine a player who is highly experienced in playing a certain zero-sum game. He has a trusted strategy which guarantees him the value of this game within some error. Now, consider the high stakes–low stakes game again. The player might well use the trusted strategy in both the low stakes and the high stakes subgame. Therefore, the error related to this strategy will be relative with respect to the value of the respective subgame.

Finally, in the class of stochastic games as identified in Flesch, Thuijsman, and Vrieze ([8]), the only way to obtain ϵ-optimality is to use strategies that are called improving. Improving strategies are closely related to subgame ϕ-maxmin strategies such that the tolerance level in some subgames is smaller than the tolerance level at the root.

With respect to the concept of subgame ϕ-maxmin strategies, this paper attempts to provide answers to the following two fundamental questions:

1. For positive tolerance functions ϕ, when does a subgame ϕ-maxmin strategy exist?
2. How is the existence of a subgame ϕ-maxmin strategy related to the existence of a subgame maxmin strategy?

To answer these questions we will first derive necessary and sufficient conditions for a strategy to be a subgame ϕ-maxmin strategy. This is done in Sect. 4. As a special case of these necessary and sufficient conditions, we obtain a characterization of subgame maxmin strategies. For the special class of positive and negative stochastic games, a related characterization of subgame maxmin strategies was obtained by Flesch, Predtetchinski, and Sudderth ([7]).

The necessary and sufficient conditions for strategies to be subgame ϕ-maxmin can be split into a local condition and an equalizing condition. Informally, the local condition states that the lower value one expects to get in the next subgame should always be at least the lower value of the current subgame. The equalizing condition requires that for every strategy of the other player, a subgame ϕ-maxmin strategy almost surely results in a play with an eventually good enough payoff, where eventually good enough means being at least the lower value in very deep subgames up to the allowed tolerance level.

In Sect. 5, we consider the question of existence of subgame ϕ-maxmin strategies for positive tolerance functions ϕ. We prove that for a positive tolerance function ϕ, a subgame ϕ-maxmin strategy exists if every play is either a point of upper semicontinuity of the payoff function or if the sequence of tolerance levels which occur along the play has a positive lower bound. The novel aspect of the theorem is the way in which it encompasses a number of important special cases, some of which have been studied extensively in the literature. One special case of interest is when the payoff function is upper semicontinuous everywhere.

In this case, the existence of subgame maxmin strategies follows by a result of Laraki, Maitra, and Sudderth ([10]). Another is the case when along each play tolerance levels remain bounded away from zero. This latter case subsumes the familiar class of all positive constant tolerance functions. When the tolerance level is equal to a constant ϵ at each history, subgame ϕ-maxmin strategies are exactly the so-called subgame ϵ-optimal strategies. Existence of subgame ϵ-optimal strategies for each $\epsilon > 0$ is established in Mashiah–Yaakovi ([15]).

Our technique relies on switching between strategies: one starts by following a $\frac{\phi(x_0)}{2}$-optimal strategy until a history (say h) is reached where it ceases to be $\phi(h)$-optimal. At that moment, one discards the original strategy and switches instead to a $\frac{\phi(h)}{2}$-optimal strategy, and so on. This technique is well-attested in the literature, and in particular is used in Mashiah–Yaakovi ([15]). We refine the method to accommodate both types of plays: those where the tolerance levels remain bounded away from zero, and those where they converge to zero (the latter are, by assumption, the plays where the payoff function is upper semicontinuous).

In Sect. 6, we study the relation between the existence of subgame ϕ-maxmin strategies and subgame maxmin strategies. Our main result states that the existence of a particular subgame ϕ^*-maxmin strategy, with $\phi^* > 0$, is equivalent to the existence of a subgame maxmin strategy. Consequently, the existence of subgame ϕ-maxmin strategies for all positive tolerance functions ϕ is equivalent to the existence of a subgame maxmin strategy. In the special case of an upper semi-continuous payoff function, the results of Sects. 5 and 6 imply the existence of a subgame maxmin strategy. As noted above, this result is not new, and is a consequence of Laraki, Maitra, and Sudderth ([10]).

The connection between existence of subgame ϕ-maxmin strategies for every positive tolerance function ϕ and the existence of subgame maxmin strategies is not only useful to further understand the results obtained by Laraki, Maitra, and Sudderth ([10]) but also highlights an important and surprising difference between subgame ϕ-maxmin strategies and the closely related concept of subgame ϵ-maxmin strategies. Indeed, the existence of a subgame ϵ-maxmin strategy for every $\epsilon > 0$ does not imply the existence of a subgame maxmin strategy.

The rest of the paper is structured as follows. In Sect. 2, we formulate the model setup, and in Sect. 3, we formally define the main concepts. Then in Sect. 4, we discuss the necessary and sufficient conditions for a strategy to be a subgame ϕ-maxmin strategy and give a characterization for subgame maxmin strategies. We continue in Sect. 5 by providing sufficient conditions to guarantee the existence of a subgame ϕ-maxmin strategy. Then in Sect. 6, we explain why the existence of subgame ϕ-maxmin strategies for every positive tolerance function ϕ is equivalent to existence of a subgame maxmin strategy. Finally, in Sect. 7, we discuss the importance of the assumptions we made and whether they might be relaxed.

2 Two-Player Zero-Sum Stochastic Games

We consider a two-player zero-sum stochastic game with finitely many actions and countably many states. The payoff function is required to be bounded and universally measurable. The model encompasses all two-player zero-sum games with perfect information and stochastic moves.

Actions, states, and histories The action sets of players 1 and 2 are given by the finite sets $\mathcal{A}$ and $\mathcal{B}$, respectively. The state space is given by the countable set $\mathcal{X}$. Let x_0 denote the initial state and define the set $\mathcal{Z} = \mathcal{A} \times \mathcal{B} \times \mathcal{X}$. The game consists of an infinite sequence of one-shot games. At the initial state x_0, the one-shot game $G(x_0)$ is played as follows:

Players 1 and 2 simultaneously select an action from their respective action sets, denoted by a_1 and b_1, respectively. Then, the next state x_1 is selected according to the transition probability $q(\cdot|x_0, a_1, b_1)$. At the new state x_1, this process repeats itself and both players play the one-shot game $G(x_1)$ by selecting actions a_2 and b_2 from their respective action sets. The next state x_2 is selected according to the transition probability $q(\cdot|x_0, a_1, b_1, x_1, a_2, b_2)$. This goes on indefinitely and creates a play $p = (x_0, a_1, b_1, x_1, a_2, b_2, \ldots)$. Note that the transition probability can depend on the entire history.

For every $t \in \mathbb{N} = \{0, 1, 2, \ldots\}$, let $\mathcal{H}^t = \{x_0\} \times \mathcal{Z}^t$ denote the set of all histories that are generated after t one-shot games. The set $\mathcal{H}^0$ consists of the single element x_0. For $t \geq 1$, elements of $\mathcal{H}^t$ are of the form $(x_0, a_1, b_1, \ldots, a_t, b_t, x_t)$. Let $\mathcal{H} = \cup_{t \in \mathbb{N}} \mathcal{H}^t$ denote the set of all histories. For all $h \in \mathcal{H}$, let $\|h\| = \|(x_0, a_1, b_1, \ldots, a_t, b_t, x_t)\| = t$ denote the number of one-shot games that occurred during the history h. We refer to $\|h\|$ as the length of the history h. For all $t \leq \|h\|$, the restriction of the history h to the first t one-shot games is denoted by $h_{|t} = (x_0, a_1, b_1, \ldots, a_t, b_t, x_t)$. We write $h' \preceq h$ if there exists $t \leq \|h\|$ such that $h_{|t} = h'$, so if the history h extends the history h'.

The space of plays Define $\mathcal{P} = \{x_0\} \times \mathcal{Z}^{\mathbb{N}}$ to be the set of plays. Elements of $\mathcal{P}$ are of the form $p = (x_0, a_1, b_1, x_1, a_2, b_2, \ldots)$. For $t \in \mathbb{N}$, let $p_{|t}$ denote the prefix of p of length t: that is $p_{|0} = x_0$ and $p_{|t} = (x_0, a_1, b_1, \ldots, a_t, b_t, x_t)$ for $t \geq 1$. We write $h \prec p$ if a history h is a prefix of p. For every $h \in \mathcal{H}$, let $\mathcal{P}(h) = \{p \in \mathcal{P} | h \prec p\}$ denote the cylinder set consisting of all plays which extend history h.

We endow $\mathcal{Z}$ with the discrete topology and $\mathcal{P}$ with the product topology. The collection of all cylinder sets is a basis for the product topology on $\mathcal{P}$.

For $t \in \mathbb{N}$, let $\mathcal{F}^t$ be the sigma-algebra generated by the collection of cylinder sets $\{\mathcal{P}(h) \mid h \in \mathcal{H}^t\}$. Each set in $\mathcal{F}^t$ can be written as the union of sets in $\{\mathcal{P}(h) \mid h \in \mathcal{H}^t\}$. Let $\mathcal{F}^{\infty}$ be the sigma-algebra generated by $\cup_{t \in \mathbb{N}} \mathcal{F}^t$. This is exactly the Borel sigma-algebra generated by the product topology on $\mathcal{P}$. The sigma-algebra of universally measurable subsets of $\mathcal{P}$ is denoted by $\mathcal{F}$. Elements of $\mathcal{F}$ are sets that belong to the completion of each Borel probability measure on $\mathcal{P}$. For details of the definition of the sigma-algebra $\mathcal{F}$, the reader is referred to Appendix A. It holds that $\mathcal{F}^t \subseteq \mathcal{F}^{t+1} \subseteq \cdots \subseteq \mathcal{F}^{\infty} \subseteq \mathcal{F}$. The set of plays $\mathcal{P}$ together with the universally measurable sigma-algebra $\mathcal{F}$ determines a measurable space $(\mathcal{P}, \mathcal{F})$. A stochastic variable is a universally measurable function from $\mathcal{P}$ to $\mathbb{R}$.

Strategies Let $\Delta(\mathcal{A})$ denote the set of probability measures over the action set of player 1 and $\Delta(\mathcal{B})$ the set of probability measures over the action set of player 2. A behavioral strategy for player 1 is a function $\sigma : \mathcal{H} \to \Delta(\mathcal{A})$. Hence, at each history player 1 chooses a mixed action. A pure strategy for player 1 is a function that assigns to every history an action, with a minor abuse of notation, $\sigma : \mathcal{H} \to \mathcal{A}$. Similarly, one can define a behavioral and a pure strategy τ for player 2. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ denote the sets of behavioral strategies of players 1 and 2, respectively.

It follows from the Ionescu Tulcea extension theorem that every history $h \in \mathcal{H}$ and strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$ determine a probability measure $\mathbb{P}_{h,\sigma,\tau}$ on the measurable space $(\mathcal{P}(h), \mathcal{F}^{\infty}_{\mathcal{P}(h)})$, where $\mathcal{F}^{\infty}_{\mathcal{P}(h)}$ denotes the Borel sigma-algebra over the set of plays extending the history h. The measure $\mathbb{P}_{h,\sigma,\tau}$ can be extended to the measurable space $(\mathcal{P}, \mathcal{F}^{\infty})$ in the obvious way. Taking the completion of the probability space $(\mathcal{P}, \mathcal{F}^{\infty}, \mathbb{P}_{h,\sigma,\tau})$ yields the probability space $(\mathcal{P}, \mathcal{F}, \mathbb{P}^c_{h,\sigma,\tau})$. With a minor abuse of notation, we will omit the superscript c and write $\mathbb{P}_{h,\sigma,\tau}$ in the remainder of this paper.

Payoff function We assume that the payoff function $u : \mathcal{P} \to \mathbb{R}$ of player 1 is bounded, where we define $M = \sup_{p \in \mathcal{P}} |u(p)|$, and universally measurable. We also assume the game to be zero-sum. The payoff function of player 2 is therefore given by $-u$. We denote the game as described above by $\Gamma_{x_0}(u)$. Throughout the paper, we take the point of view of player 1. This

is without loss of generality, since the situation of player 2 in the game $\Gamma_{x_0}(u)$ is identical to that of player 1 in the game $\Gamma_{x_0}(-u)$.

The expected payoff of player 1 corresponding to strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$ at history $h \in \mathcal{H}$ is given by $\mathbb{E}_{h,\sigma,\tau}[u]$, where the expectation is taken with respect to the probability measure $\mathbb{P}_{h,\sigma,\tau}$. The expected payoff of player 1 at the history x_0 is denoted by $\mathbb{E}_{\sigma,\tau}[u]$.

We imposed the condition above that the payoff function u is universally measurable. As Borel measurable functions are also universally measurable, this condition is weaker than requiring that u be Borel measurable. The main difference between these two conditions is the following. When the payoff function u is Borel measurable, the game is known to admit a value (cf. Sect. 3 for the formal definition). However, it is not provable from the standard ZFC axioms of set theory[1] that our games with a universally measurable payoff function also admit a value. Therefore, instead of the value, in the remaining part of the paper we will use the lower value (cf. Sect. 3 for the formal definition). We would like to emphasize that except for the difference between the concepts of the value and the lower value, assuming Borel measurable payoffs would not lead to any simplification in the proofs. So the reader unfamiliar with universally measurable functions can read the entire paper imagining Borel measurable payoffs.

3 Subgame ϕ-Maxmin Strategies

We start this section by defining subgame ϕ-maxmin strategies. Then, we discuss an illustrative example. Finally, we define and examine guarantee levels of strategies.

3.1 Definition of Subgame ϕ-Maxmin Strategies

For every game $\Gamma_{x_0}(u)$, for every history $h \in \mathcal{H}$, we define the lower value $\underline{v}(h)$ and the upper value $\overline{v}(h)$ by

$$\underline{v}(h) = \sup_{\sigma \in \mathcal{S}_1} \inf_{\tau \in \mathcal{S}_2} \mathbb{E}_{h,\sigma,\tau}[u], \tag{3.1}$$

$$\overline{v}(h) = \inf_{\tau \in \mathcal{S}_2} \sup_{\sigma \in \mathcal{S}_1} \mathbb{E}_{h,\sigma,\tau}[u]. \tag{3.2}$$

Because the payoff function u is assumed to be bounded, we have that $\underline{v}(h), \overline{v}(h) \in \mathbb{R}$. Therefore, the lower and upper value exist in every subgame of $\Gamma_{x_0}(u)$. Furthermore, we have that $\underline{v}(h) \leq \overline{v}(h)$. Whenever $\underline{v}(h) = \overline{v}(h)$ we say that the subgame at history h has a value and we denote it by $v(h)$. The lower value, the upper value, and the value at the initial state x_0 are denoted by $\underline{v}$, $\overline{v}$, and v, respectively. If u is Borel measurable, then the value exists by Maitra and Sudderth ([13]) and Martin ([14]). Since we do not assume u to be Borel measurable, we present our results in terms of the lower value.

Even if the value exists, player 1 may not have a strategy that guarantees the value in each subgame and the literature has therefore studied subgame ϵ-optimal strategies. These are strategies which guarantee the value in each subgame up to an allowed error term of ϵ. If the payoff function is bounded and Borel measurable, it has been shown by Mashiah–Yaakovi ([15]) that for each $\epsilon > 0$ player 1 has a subgame ϵ-optimal strategy. The concept

---

[1] See the discussion on page 1566 in Martin ([14]), for the specific case of perfect information games in which the payoff function is the characteristic function of an analytic set.
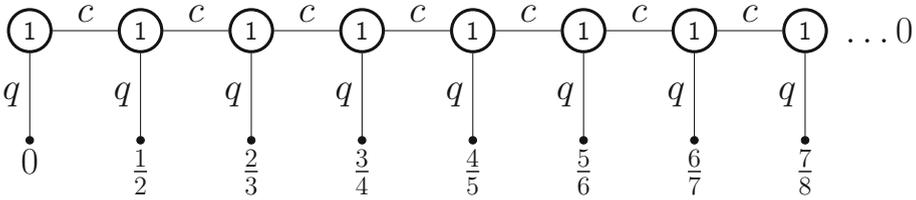
Fig. 1 Characterization of ϕ such that pure subgame ϕ-maxmin strategies exist

of a subgame ϵ-optimal strategy has a constant error term ϵ across all subgames. However, as argued in the introduction, there are instances in which it is more natural to consider a variable error term. Therefore, instead of considering a constant error term, we follow Flesch and Predtetchinski ([6]), who allow the error term to vary across histories in their investigation of the ϕ-tolerance equilibrium. This leads us to the study of subgame ϕ-maxmin strategies, where $\phi : \mathcal{H} \to [0, \infty)$ is a tolerance function assigning an allowed tolerance level to each history.

Definition 3.1 Let $\phi : \mathcal{H} \to [0, \infty)$ be a tolerance function. A strategy $\sigma \in \mathcal{S}_1$ is a *subgame ϕ-maxmin strategy* in the game $\Gamma_{x_0}(u)$ if for every history $h \in \mathcal{H}$ it holds that

$$\forall \tau \in \mathcal{S}_2, \; \mathbb{E}_{h,\sigma,\tau}[u] \geq \underline{v}(h) - \phi(h). \tag{3.3}$$

In case ϕ is identically equal to zero, we omit it from the notation, and simply refer to a subgame maxmin strategy.

A subgame ϕ-maxmin strategy guarantees at each history h of the game the lower value up to the tolerance level $\phi(h)$. If the tolerance function is such that for some $\epsilon \geq 0$, for every $h \in \mathcal{H}$, $\phi(h) = \epsilon$, then we refer to the strategy as a subgame ϵ-maxmin strategy. If the value exists, then the notion of subgame ϵ-maxmin strategy coincides with the notion of subgame ϵ-optimal strategy.

The following example illustrates that even for a strictly positive tolerance function ϕ player 1 may not have subgame ϕ-maxmin strategies. Interestingly, however, player 1 has a subgame ϵ-maxmin strategy for every positive $\epsilon > 0$, which in fact is pure. We also give an example of a tolerance function ϕ for which player 1 has a subgame ϕ-maxmin strategy, but not a pure such strategy.

Example 3.2 The decision problem depicted in Fig. 1 corresponds to a two-player zero-sum stochastic game in which the state space is trivial and the second player is a dummy player. Whenever the state space or action sets are degenerate, the corresponding states and actions are omitted from the notation in examples. The set of actions of player 1 is $\mathcal{A} = \{c, q\}$, where c stands for continue and q for quit. The game stops as soon as player 1 chooses to quit. If player 1 decides to quit at period t, then his payoff is $t/(t+1)$. If player 1 never quits, his payoff is 0.

In this game, player 1 has a subgame ϵ-maxmin strategy for every positive $\epsilon > 0$. For example, the strategy which quits whenever quitting leads to a payoff of at least $1 - \epsilon$.

We now turn to the existence of a subgame ϕ-maxmin strategy. Clearly, there exist no subgame maxmin strategy. As we will see later, Theorem 6.1 then implies that there is some strictly positive tolerance function ϕ for which there does not exist a subgame ϕ-maxmin strategy. Intuitively, such a tolerance function forces player 1 to continue with such a large probability that the total probability of never quitting becomes nearly one.

Interestingly, there is a tolerance function ϕ for which player 1 has a subgame ϕ-maxmin strategy, but for which he has no pure strategy that is subgame ϕ-maxmin. To see this, define the tolerance function ϕ by $\phi(c^t) = 1 - \frac{1}{2} \cdot \frac{t}{t+1} - \frac{1}{2} \cdot \frac{t+1}{t+2}$ for each $t \in \mathbb{N}$. Now consider the strategy σ which always chooses action c with probability 1/2 and action q with probability 1/2. In the subgame at any history c^t, it quits immediately with probability 1/2 and quits at a later period with probability 1/2. In the former case, it gives payoff $\frac{t}{t+1}$, whereas in the latter case, it gives an expected payoff of at least $\frac{t+1}{t+2}$. Hence, the strategy σ is a subgame ϕ-maxmin strategy. Yet, there is no pure strategy that is subgame ϕ-maxmin. Indeed, the tolerance function does not allow a pure strategy to quit at any period, whereas a pure strategy that always continues is not subgame ϕ-maxmin either. As we show in the working paper version of our paper, Flesch, Herings, Maes, and Predtetchinski ([5]), a pure subgame ϕ-maxmin strategy exists if and only if the following two conditions hold: (1) for every $t \in \mathbb{N}$, $\phi(c^t) > 0$, and (2) for infinitely many $t \in \mathbb{N}$, $\phi'(c^t) = \min_{n \leq t} \phi(c^n) \geq \frac{1}{t+1}$. $\qquad\square$

3.2 Guarantee Levels

To identify subgame ϕ-maxmin strategies, it is useful to define the function $\underline{u} : \mathcal{S}_1 \times \mathcal{H} \to \mathbb{R}$ by

$$\underline{u}(\sigma, h) = \inf_{\tau \in \mathcal{S}_2} \mathbb{E}_{h,\sigma,\tau}[u]. \qquad (3.4)$$

The payoff $\underline{u}(\sigma, h)$ corresponds to the guarantee level that player 1 is expected to receive at history h when playing the strategy σ. A strategy $\sigma \in \mathcal{S}_1$ is called a $\phi(h)$-maxmin strategy for the subgame at history h if $\underline{u}(\sigma, h) \geq \underline{v}(h) - \phi(h)$.

For every strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$, for every $t \in \mathbb{N}$, define the stochastic variables $U_{\sigma,\tau}^t$, $\underline{U}_\sigma^t$, and $\underline{V}^t$ by letting $U_{\sigma,\tau}^t(p) = \mathbb{E}_{p|t,\sigma,\tau}[u]$, $\underline{U}_\sigma^t(p) = \underline{u}(\sigma, p_{|t})$, and $\underline{V}^t(p) = \underline{v}(p_{|t})$, respectively, for each play $p \in \mathcal{P}$. All three stochastic variables are $\mathcal{F}^t$-measurable. We have $\underline{U}_\sigma^t \leq U_{\sigma,\tau}^t$ and $\underline{U}_\sigma^t \leq \underline{V}^t$ everywhere on $\mathcal{P}$.

The next lemma states the submartingale property of guarantee levels. It says that the guarantee level that player 1 can expect to receive increases over time.

Lemma 3.3 (Submartingale property of guarantee levels) *Let a strategy profile* $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$, $t \in \mathbb{N}$, *and a history* $h \in \mathcal{H}^t$ *of length t be given. The process* $(\underline{U}_\sigma^{t+n})_{n \in \mathbb{N}}$ *is a* $\mathbb{P}_{h,\sigma,\tau}$*-submartingale. In particular, it holds that* $\underline{u}(\sigma, h) \leq \mathbb{E}_{h,\sigma,\tau}[\underline{U}_\sigma^{t+1}]$.

Proof We first prove the second claim. Take $\delta > 0$. Let $\tau' \in \mathcal{S}_2$ be such that $\tau'(h) = \tau(h)$ and for each $(a, b, x) \in \mathcal{Z}$ it holds that

$$\mathbb{E}_{(h,a,b,x),\sigma,\tau'}[u] \leq \underline{u}(\sigma, (h, a, b, x)) + \delta.$$

We have that

$$\underline{u}(\sigma, h) \leq \mathbb{E}_{h,\sigma,\tau'}[u]$$

$$= \sum_{(a,b,x) \in \mathcal{Z}} \sigma(h)(a) \cdot \tau(h)(b) \cdot q(x|h, a, b) \cdot \mathbb{E}_{(h,a,b,x),\sigma,\tau'}[u]$$

$$\leq \sum_{(a,b,x) \in \mathcal{Z}} \sigma(h)(a) \cdot \tau(h)(b) \cdot q(x|h, a, b) \cdot (\underline{u}(\sigma, (h, a, b, x)) + \delta)$$

$$= \mathbb{E}_{h,\sigma,\tau}[\underline{U}_\sigma^{t+1}] + \delta.$$

It follows that $\underline{u}(\sigma, h) \leq \mathbb{E}_{h,\sigma,\tau}[\underline{U}_\sigma^{t+1}]$ since $\delta > 0$ is arbitrary.

To prove that the process $(U_\sigma^{t+n})_{n\in\mathbb{N}}$ is a $\mathbb{P}_{h,\sigma,\tau}$-submartingale, note that the conditional expectation $\mathbb{E}_{h,\sigma,\tau}(U_\sigma^{t+n+1} \mid \mathcal{F}^{t+n})$, when evaluated at a play $p \in \mathcal{P}(h)$, equals $\mathbb{E}_{p|_{t+n},\sigma,\tau}(U_\sigma^{t+n+1})$, which, by the claim already proven, is at least $\underline{u}(\sigma, p|_{t+n}) = \underline{U}_{t+n}(\sigma)(p)$. $\qquad\square$

4 Conditions for Strategies to be Subgame ϕ-Maxmin

4.1 *n*-Day Maxmin Strategies and Equalizing Strategies

The goal of this section is to provide necessary and sufficient conditions for a strategy to be subgame ϕ-maxmin and to provide a characterization of subgame maxmin strategies.

Definition 4.1 A strategy $\sigma \in \mathcal{S}_1$ is an *n-day ϕ-maxmin strategy* in the game $\Gamma_{x_0}(u)$ if for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$ of length t, and for every strategy $\tau \in \mathcal{S}_2$,

$$\mathbb{E}_{h,\sigma,\tau}[\underline{V}^{t+n}] \geq \underline{v}(h) - \phi(h). \tag{4.1}$$

Definition 4.2 A strategy $\sigma \in \mathcal{S}_1$ is *ϕ-equalizing* in the game $\Gamma_{x_0}(u)$ if for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$ of length t, and for every strategy $\tau \in \mathcal{S}_2$,

$$u \geq \limsup_{t\to\infty} \underline{V}^t - \phi(h), \quad \mathbb{P}_{h,\sigma,\tau}\text{-almost surely.} \tag{4.2}$$

When $\phi = 0$, we use the terms *n-day maxmin* and *equalizing* to mean *n*-day 0-maxmin and 0-equalizing, respectively.

The first definition is very intuitive. It says that player 1 should play in such a way that, on average, the lower value increases over time. The notion of 1-day maxmin strategies is particularly well-known in dynamic programming and stochastic games, see Puterman ([16]). A simple characterization of 1-day maxmin strategies is provided in following theorem.

Theorem 4.3 *Consider a strategy $\sigma \in \mathcal{S}_1$ in the game $\Gamma_{x_0}(u)$. The following three conditions are equivalent:*

1. *For each $n \in \mathbb{N}$, σ is an n-day maxmin strategy.*
2. *σ is a 1-day maxmin strategy.*
3. *For each history $h \in \mathcal{H}^t$ of length t and each strategy $\tau \in \mathcal{S}_2$, the process $(\underline{V}^{t+n})_{n\in\mathbb{N}}$ is a $\mathbb{P}_{h,\sigma,\tau}$-submartingale.*

Proof That condition 1 implies condition 2 is obvious. That condition 2 implies condition 3 follows by the law of iterated expectations. Finally, that condition 3 implies condition 1 follows from the properties of a submartingale. $\qquad\square$

Note that the equivalence in Theorem 4.3 is no longer true for nonzero tolerance functions. For example, if $\phi > 0$, then a 1-day ϕ-maxmin strategy is generally not 2-day ϕ-maxmin.

A strategy is ϕ-equalizing if, roughly speaking, it almost surely results in a play with an eventually good enough payoff, where eventually good enough means being about as large as the lower value in very deep subgames.

The following example illustrates both the notion of a 1-day maxmin strategy and an equalizing strategy.

Example 4.4 Consider the centipede game shown in Fig. 2. At every history, the active player can choose to continue (c) or to quit (q). As soon as a player decides to quit the game ends
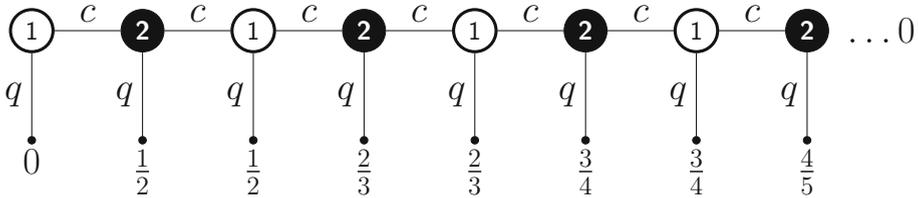
Fig. 2 Strategies that are n-day maxmin may not be subgame maxmin

and in that case the payoff is as given in Fig. 2. If the game continues indefinitely then the payoff is 0. It is easily verified that for every $t \in \mathbb{N}$, $\underline{v}(c^{2t}) = \underline{v}(c^{2t+1}) = (t+1)/(t+2)$. A pure strategy $\sigma \in \mathcal{S}_1$ is then easily seen to be a 1-day maxmin strategy if and only if $\sigma(c^{2t}) = c$ for every $t \in \mathbb{N}$.

On the other hand, a pure strategy $\sigma \in \mathcal{S}_1$ is equalizing if and only if for infinitely many $t \in \mathbb{N}$ it holds that $\sigma(c^{2t}) = q$. Indeed, for a strategy σ such that q is chosen for infinitely many $t \in \mathbb{N}$ the generated play ends with q after each history $h \in \mathcal{H}$ irrespective of $\tau \in \mathcal{S}_2$, which shows that σ is equalizing. On the other hand, if σ is such that q is chosen for finitely many $t \in \mathbb{N}$, then there is a history $h \in \mathcal{H}$ such that player 1 always plays continue after this history. If $\tau \in \mathcal{S}_2$ is such that player 2 always chooses continue, then the play c^{∞} with $u(c^{\infty}) = 0$ results, which is strictly less than $\limsup_{t \to \infty} \underline{v}(c^t) = 1$, so such a σ is not equalizing.

It follows that player 1 does not have a pure strategy which is both 1-day maxmin and equalizing. $\qquad \square$

The following theorem states sufficient conditions under which a strategy σ of player 1 is a subgame ϕ-maxmin strategy.

Theorem 4.5 (Sufficient condition) *Let $\phi : \mathcal{H} \to [0, \infty)$ be a tolerance function. The strategy $\sigma \in \mathcal{S}_1$ is a subgame ϕ-maxmin strategy in the game $\Gamma_{x_0}(u)$ if there exist tolerance functions $\phi_1 : \mathcal{H} \to [0, \infty)$ and $\phi_2 : \mathcal{H} \to [0, \infty)$ such that $\phi_1 + \phi_2 \leq \phi$ and*

1. *for every $n \in \mathbb{N}$, σ is n-day ϕ_1-maxmin,*
2. *σ is ϕ_2-equalizing.*

Proof Let ϕ_1, ϕ_2, and σ be such that the conditions in the theorem are satisfied. We show that σ is a subgame ϕ-maxmin strategy.

Fix a history $h \in \mathcal{H}^t$ and a strategy $\tau \in \mathcal{S}_2$. Then, we have that

$$
\begin{aligned}
\underline{v}(h) - \phi(h) &\leq \underline{v}(h) - \phi_1(h) - \phi_2(h) \\
&\leq \limsup_{n \to \infty} \mathbb{E}_{h,\sigma,\tau}[\underline{V}^{t+n}] - \phi_2(h) \\
&\leq \mathbb{E}_{h,\sigma,\tau}[\limsup_{n \to \infty} \underline{V}^{t+n}] - \phi_2(h) \\
&\leq \mathbb{E}_{h,\sigma,\tau}[u],
\end{aligned}
$$

where the second inequality holds since σ is an n-day ϕ_1-maxmin strategy, the third inequality is by Fatou lemma, and the last inequality holds since σ is ϕ_2-equalizing. $\qquad \square$

According to Theorem 4.5, to conclude that σ is a subgame ϕ-maxmin strategy, it is enough to find tolerance functions ϕ_1 and ϕ_2 such that at each history their sum does not exceed ϕ and the strategy σ is both n-day ϕ_1-maxmin and ϕ_2-equalizing.

Particularly natural are situations where the tolerance level does not increase as time progresses. More formally, the tolerance function ϕ is said to be non-increasing if $\phi(h) \geq \phi(h')$ whenever $h \prec h'$. The following result states necessary conditions for a strategy to be subgame ϕ-maxmin.

Theorem 4.6 (Necessary condition) *Let* $\sigma \in S_1$ *be a subgame ϕ-maxmin strategy in the game* $\Gamma_{x_0}(u)$. *Then, it holds that:*

1. *For every* $n \in \mathbb{N}$, σ *is an n-day ϕ-maxmin strategy.*
2. *If the tolerance function ϕ is non-increasing, then σ is ϕ-equalizing.*

Proof Let $\sigma \in S_1$ be a subgame ϕ-maxmin strategy in the game $\Gamma_{x_0}(u)$. Take a history $h \in \mathcal{H}^t$ of length t and a strategy $\tau \in S_2$.

We prove condition 1. We have

$$\mathbb{E}_{h,\sigma,\tau}[\underline{V}^{t+n}] \geq \mathbb{E}_{h,\sigma,\tau}[\underline{U}^{t+n}_\sigma] \geq \mathbb{E}_{h,\sigma,\tau}[\underline{U}^t_\sigma] = \underline{u}(\sigma, h) \geq \underline{v}(h) - \phi(h),$$

where the first inequality holds since for each play $p \in \mathcal{P}(h)$ we have $\underline{V}^{t+n}(p) = \underline{v}(p_{|t+n}) \geq \underline{u}(\sigma, p_{|t+n}) = \underline{U}^{t+n}_\sigma(p)$. The second inequality holds by Lemma 3.3. The next equation holds since t is the length of history h, and the final inequality holds since σ is a subgame ϕ-maxmin strategy.

We prove condition 2. We have, $\mathbb{P}_{h,\sigma,\tau}$-almost surely,

$$u(p) = \lim_{t \to \infty} \mathbb{E}_{p_{|t},\sigma,\tau}[u] \geq \limsup_{t \to \infty}(\underline{v}(p_{|t}) - \phi(p_{|t})) \geq \limsup_{t \to \infty} \underline{V}^t(p) - \phi(h),$$

where the equality is by Levy's zero-one law (Lemma A.1 in Appendix A), the first inequality follows since σ is a subgame ϕ-maxmin strategy, and the second inequality holds since ϕ is non-increasing. ☐

Notice that in the case of a nonzero tolerance function, the necessary and sufficient conditions do not coincide, and we do not obtain a characterization of subgame ϕ-maxmin strategies. We now turn to the case where the tolerance function ϕ is identically equal to 0.

Corollary 4.7 *A strategy $\sigma \in S_1$ is a subgame maxmin strategy in the game $\Gamma_{x_0}(u)$ if and only if it is 1-day maxmin and equalizing.*

When we compare the sufficient conditions of Theorem 4.5 to the sufficient conditions of Corollary 4.7, we notice that in the case of a non-zero tolerance function we require the strategy to be n-day ϕ-maxmin. In the case of a zero tolerance function, the corresponding sufficient conditions only require the strategy to be 1-day maxmin. The reason for this difference is that we should avoid that strategies accumulate the allowed tolerance levels, causing them to become too permissive over time. The following example illustrates this issue.

Example 4.8 Consider the decision problem depicted in Fig. 3. Each period the decision maker can choose to continue (c) or to quit (q). Notice that $v(c^t) = 1/(t+1)$ and hence $\lim_{t \to \infty} v(c^t) = 0$. Any strategy of player 1 is therefore equalizing. Furthermore, it is clear that in this decision problem the subgame maxmin strategy is unique and requires the decision maker to quit at each time. Now suppose the decision maker has the following tolerance function:

$$\phi(c^t) = v(c^t) - v(c^{t+1}) = \frac{1}{t+1} - \frac{1}{t+2}, \quad t \in \mathbb{N}.$$

Consider the strategy σ where the decision maker always chooses to continue. From the definition of the tolerance function, it follows that the strategy σ is a 1-day ϕ-maxmin
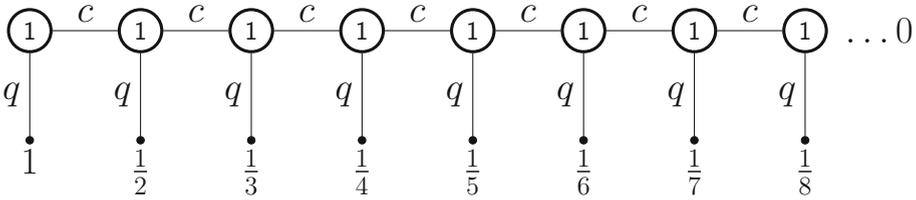
Fig. 3 A 1-day ϕ-maxmin strategy may not be n-day ϕ-maxmin

strategy. Indeed, it holds that $v(c^{t+1}) \geq v(c^t) - \phi(c^t)$. It is also easily seen that the strategy σ is equalizing. Nevertheless, it is clear that σ is not a subgame ϕ-maxmin strategy.

The underlying problem with the strategy σ is that every time the decision maker chooses to continue this causes an acceptable loss in the value, but over time these losses add up to an unacceptable loss. The requirement that for every $n \in \mathbb{N}$ a strategy is n-day ϕ-maxmin guarantees that the accumulated losses over any finite period of time remain acceptable.

If we require a strategy to be subgame maxmin, then we do not tolerate any losses. Therefore, the accumulation problem mentioned above will never occur, and it will be sufficient to only require that the strategy is 1-day maxmin.   ◇

Example 4.9 is such that player 1 has a maxmin strategy in every subgame, but has no subgame maxmin strategy. From Corollary 4.7, it follows that any subgame maxmin strategy needs to be both 1-day maxmin and equalizing. The example presents a game where all the strategies are 1-day maxmin but none of them is equalizing and therefore subgame maxmin strategies do not exist.

Example 4.9 Consider the following perfect information game. Both players have two actions, left (L) and right (R), so $\mathcal{A} = \mathcal{B} = \{L, R\}$. The players take turns playing an action, which generates a play $p \in \{L, R\}^{\mathbb{N}}$. Let $r_1(p)$ and $r_2(p)$ denote the number of times that player 1 and player 2, respectively, play action R during the play p and define $r(p) = \min\{r_1(p), r_2(p)\}$. Player 1 obtains a payoff of 0 if both players choose R infinitely often. When at least one of them chooses R only a finite number of times, then player 1 receives a payoff of $r(p)/(r(p) + 1)$. The payoff function u is therefore obtained by defining, for $p \in \{L, R\}^{\mathbb{N}}$,

$$u(p) = \begin{cases} \frac{r(p)}{r(p)+1}, & \text{if } r_1(p) \neq \infty \text{ or } r_2(p) \neq \infty, \\ 0, & \text{if } r_1(p) = r_2(p) = \infty. \end{cases} \tag{4.3}$$

At each history $h \in \mathcal{H}$, the value of the game exists and is given by $v(h) = r_2(h)/(r_2(h) + 1)$, where $r_2(h)$ denotes the number of times player 2 has chosen R in the history h. Indeed, player 1 can guarantee this payoff by choosing the action R $\max\{r_2(h) - r_1(h), 0\}$ times after history h. Player 2 can guarantee to lose at most this amount by playing only left after history h. Hence at every history $h \in \mathcal{H}$, player 1 has a maxmin strategy.

For every history $h \in \mathcal{H}$ where player 1 takes an action and for every action $a \in \mathcal{A}$, we have that $r_2(ha) = r_2(h)$ and $v(ha) = v(h)$. Therefore, all strategies of player 1 are 1-day maxmin.

On the other hand, no equalizing strategy exists for player 1. To see this, take any strategy $\sigma \in \mathcal{S}_1$ for player 1 and consider the strategy $\tau \in \mathcal{S}_2$ in which player 2 always chooses R. Let p be the play generated by the strategy profile (σ, τ). Then, we have that $u(p) < 1$ and $\lim_{t \to \infty} v(p_{|t}) = 1$. It follows that σ is not equalizing. Using Corollary 4.7, we conclude that player 1 does not have a subgame maxmin strategy.   □

4.2 The case of an upper semi-continuous payoff function

The remainder of this section is devoted to the case where the payoff function is upper semi-continuous. We argue that in this case, any strategy of player 1 is equalizing. Because all upper semi-continuous functions are Borel measurable and because we assumed finite action sets and a countable state space, the value exists, see Maitra and Sudderth ([13]) and Martin ([14]), and the lower value equals the value.

The function u is upper semi-continuous at a play $p \in \mathcal{P}$ if for every sequence $\{p_t\}_{t \in \mathbb{N}}$ of plays converging to p it holds that

$$\limsup_{t \to \infty} u(p_t) \leq u(p).$$

Lemma 4.10 *Let the payoff function u be upper semi-continuous at the play p. Then, we have that*

$$u(p) \geq \limsup_{t \to \infty} \underline{v}(p_{|t}). \tag{4.4}$$

Proof Fix $\epsilon > 0$. For every $t \in \mathbb{N}$, define $h_t = p_{|t}$ and let $p_t \in \mathcal{P}(h_t)$ be such that $u(p_t) \geq \underline{v}(h_t) - \epsilon$. Such a play p_t exists as player 1 can guarantee a payoff of at least $\underline{v}(h_t) - \epsilon$ at history h_t. Since the sequence $\{p_t\}_{t \in \mathbb{N}}$ converges to p, we have

$$u(p) \geq \limsup_{t \to \infty} u(p_t) \geq \limsup_{t \to \infty} \underline{v}(h_t) - \epsilon.$$

Since this holds for every $\epsilon > 0$, the lemma follows.   □

In view of Lemma 4.10, we obtain the following corollary to Theorems 4.5, 4.6, and 4.7.

Corollary 4.11 *Let $\Gamma_{x_0}(u)$ be such that u is upper semi-continuous. Then, each strategy of player 1 is equalizing. The strategy $\sigma \in \mathcal{S}_1$ is a subgame ϕ-maxmin strategy if and only if for every $n \in \mathbb{N}$ it is an n-day ϕ-maxmin strategy. The strategy $\sigma \in \mathcal{S}_1$ is a subgame maxmin strategy if and only if it is a 1-day maxmin strategy.*

Example 4.12 (**Staying in the set game**) Let some subset $\mathcal{X}^*$ of $\mathcal{X}$ be given. For a play $p = (x_0, a_1, b_1, x_1, a_2, b_2 \ldots)$, we define $u(p)$ to be 1 if $x_t \in \mathcal{X}^*$ for every $t \in \mathbb{N}$ and to be 0 otherwise. Maitra and Sudderth ([12]) refer to such a payoff function as "staying in a set" and in the computer science literature it goes under the name of "safety objective," see Bruyère ([4]). Since u is upper semi-continuous, any strategy $\sigma \in \mathcal{S}_1$ is equalizing.   □

Example 4.13 Consider again the centipede game shown in Figure 2, but with one slight modification. If the game continues indefinitely, then player 1 receives a payoff of 2 instead of 0. The payoff function is now upper semi-continuous. As argued in Example 4.4, the strategy σ in which player 1 continues at each history is 1-day maxmin. Because of Corollary 4.11, we can conclude that the strategy σ is a subgame maxmin strategy.   □

Note that the limit average payoff is generally not upper semi-continuous. Hence, in games where the payoff function is given by the limit average payoff, there may exist strategies of the maximizing player which are not equalizing.

5 Existence of subgame ϕ-maxmin strategies

The goal of this section is to give sufficient conditions for the existence of a subgame ϕ-maxmin strategy if the tolerance function ϕ is positive. We will prove the following theorem.

Theorem *Let a game $\Gamma_{x_0}(u)$ and a tolerance function $\phi > 0$ be given such that for every $p \in \mathcal{P}$ at least one of the following two conditions holds:*

1. *(point of upper semicontinuity) The function u is upper semi-continuous at p.*
2. *(positive limit inferior) $\liminf_{t \to \infty} \phi(p_{|t}) > 0$.*

Then, there exists a subgame ϕ-maxmin strategy in the game $\Gamma_{x_0}(u)$.

Two special cases of the above theorem are worth noting. One is when the payoff function is everywhere upper semicontinuous. In this special case, the existence of a ϕ-maxmin strategy (in fact, of a maxmin strategy) follows by a result of Laraki, Maitra, and Sudderth ([10]). We note that the latter result is not entirely subsumed by the theorem above, since the authors allow for a general Borel state space. Another special case of interest is when along each play, the tolerance remains bounded away from zero (i.e., if each play satisfies condition 2 of the theorem). This special case, to the best of our knowledge, has not been explicitly considered in the literature. However, it does encompass the familiar situation of a constant tolerance function: $\phi(h) \equiv \epsilon > 0$. For such tolerance functions, ϕ-maxmin strategies are exactly subgame ϵ-optimal strategies, and existence follows by Proposition 11 in Mashiah–Yaakovi ([15]). Note that the case when all plays satisfy condition 2 of the theorem is more general than the case of a constant tolerance function, since no uniform lower bound on the tolerance is imposed. A novel feature of the above theorem is the way it combines these various special cases. We note that the theorem remains original even if we restrict our attention to Borel measurable payoff functions.

The main technique used in the proof of this theorem are switching strategies. The idea of switching between progressively "more optimal" strategies is certainly not new. A similar construction is used for example in Rosenberg, Solan, and Vieille ([18]), Solan and Vieille ([20]), and Mashiah–Yaakovi ([15]). However, the method needs to be fine-tuned, in order for it to accommodate both the plays where the tolerance remains bounded away from zero (condition 2), as well as those where the payoff function is upper semicontinuous (condition 1). It is not straightforward to use this proof technique when we have a general Borel measurable state space as the selection of switching strategies would need to satisfy measurability conditions. It is therefore an interesting open question whether the result can be extended to that case.

The switching strategy σ^ϕ is formally defined in Sect. 5.1. To analyze its properties, we define in Sect. 5.2, for every $k \in \mathbb{N}$, a strategy σ^k that coincides with σ^ϕ as long as at most k switches have been made and it stops switching thereafter. We also present three technical lemmas for finite switching strategies in this subsection, the most important one showing that the guaranteed expected payoff of a strategy σ^k is increasing in k. Sect. 5.3 uses the lemmas of Sect. 5.2 to show the desired existence result in Theorem 5.7.

5.1 Definition of the Switching Strategy σ^ϕ

Fix a tolerance function $\phi > 0$. For every history $h \in \mathcal{H}$, player 1 has a $(\phi(h)/2)$-maxmin strategy for the subgame at history h, denoted by σ^h. The function $\psi : \mathcal{H} \to \mathcal{H}$ maps histories into histories and is such that player 1 is going to use strategy $\sigma^{\psi(h)}$ at subgame $h \in \mathcal{H}$. The function ψ is used to describe when player 1 switches strategies and is recursively defined by setting $\psi(x_0) = x_0$ and, for every $h \in \mathcal{H}$, for every $z \in \mathcal{Z}$,

$$\psi(hz) = \begin{cases} \psi(h), & \text{if } \underline{u}(\sigma^{\psi(h)}, hz) \geq \underline{v}(hz) - \phi(hz), \\ hz, & \text{otherwise.} \end{cases}$$

The condition $\underline{u}(\sigma^{\psi(h)}, hz) \geq \underline{v}(hz) - \phi(hz)$ verifies whether the strategy to which player 1 switched last, $\sigma^{\psi(h)}$, is a $\phi(hz)$-maxmin strategy for the subgame at history hz. If this is the case, then there is no need to switch and $\psi(hz) = \psi(h)$. Otherwise, player 1 switches to σ^{hz}, which is achieved by setting $\psi(hz) = hz$. Formally, we define the switching strategy $\sigma^{\phi} : \mathcal{H} \to \Delta(\mathcal{A})$ by

$$\sigma^{\phi}(h) = \sigma^{\psi(h)}(h), \quad h \in \mathcal{H}. \tag{5.1}$$

The following example illustrates the switching strategy σ^{ϕ} and shows that it may not be subgame ϕ-maxmin.

Example 5.1 Consider again the centipede game depicted in Fig. 2. We recall that, for every $t \in \mathbb{N}$, $\underline{v}(c^{2t}) = \underline{v}(c^{2t+1}) = (t+1)/(t+2)$. Take a tolerance function ϕ with the property that, for every $t \in \mathbb{N}$, $\phi(c^{2t}) < 1/((t+1)(t+2))$.

Let $h \in \mathcal{H}$ be the active history and let $k \in \mathbb{N}$ be such that $h = c^{2k}$ or $h = c^{2k+1}$. The strategy σ^h in which player 1 chooses continue at periods $0, 2, \ldots, 2k$ and chooses quit at every later period, i.e.,

$$\sigma^h(c^{2t}) = \begin{cases} c, & \text{if } 2t \leq 2k, \\ q, & \text{otherwise,} \end{cases}$$

is a maxmin strategy at subgame h.

We now consider the switching strategy σ^{ϕ}. We show by induction that for every $h \in H$, for every $z \in Z$, it holds that $\psi(hz) = hz$ if hz is an active history of player 1 and $\psi(hz) = h$ if hz is an active history of player 2. The statement trivially holds for the initial history. Let h be an active history of player 2 and let $t \in \mathbb{N}$ be such that $h = c^{2t+1}$. Since $\sigma^h = \sigma^{c^{2t}}$ is a maxmin strategy at subgame h, it holds that $\psi(c^{2t+1}) = c^{2t}$. Let h be an active history of player 1 and let $t \in \mathbb{N} \setminus \{0\}$ be such that $h = c^{2t}$. We have that

$$\underline{u}(\sigma^{\psi(c^{2t-1})}, c^{2t}) = \underline{u}(\sigma^{c^{2t-2}}, c^{2t}) = \tfrac{t}{t+1} = \underline{v}(c^{2t}) - \tfrac{1}{(t+1)(t+2)} < \underline{v}(c^{2t}) - \phi(c^{2t}),$$

so $\psi(c^{2t}) = c^{2t}$. Since the tolerance function ϕ is so stringent, it forces player 1 to switch at each of his active histories. For every $t \in \mathbb{N}$, it holds that $\sigma^{\phi}(c^{2t}) = \sigma^{c^{2t}}(c^{2t}) = c$, so under σ^{ϕ} player 1 chooses c at each of his active histories. The strategy σ^{ϕ} is not a subgame ϕ-maxmin strategy as it fails to be ϕ-equalizing, see Example 4.4. $\square$

5.2 Definition and Analysis of Finite Switching Strategies

Given the switching strategy σ^{ϕ}, for every $k \in \mathbb{N}$, we define the strategy $\sigma^k : \mathcal{H} \to \Delta(\mathcal{A})$ such that it coincides with σ^{ϕ} as long as at most k switches have been made and it stops switching thereafter. Formally, we recursively define the function $\kappa : \mathcal{H} \to \mathbb{N}$ which counts the number of switches along a history h by setting $\kappa(x_0) = 0$ and, for all histories $h, hz \in \mathcal{H}$,

$$\kappa(hz) = \begin{cases} \kappa(h), & \text{if } \underline{u}(\sigma^{\psi(h)}, hz) \geq \underline{v}(hz) - \phi(hz), \\ \kappa(h) + 1, & \text{otherwise.} \end{cases}$$

For every $k \in \mathbb{N}$, we define the stopping time $T_k : \mathcal{P} \to \mathbb{N} \cup \{\infty\}$ by

$$T_k(p) = \inf\{t \in \mathbb{N} | \kappa(p_{|t}) = k\}, \quad p \in \mathcal{P}. \tag{5.2}$$

The stopping time T_k indicates the time at which switch k occurred. Since, for $t < \infty$, the expression $T_k(p) \leq t$ only depends on the history up to period t, it holds that the set $\{p \in \mathcal{P} | T_k(p) \leq t\}$ is $\mathcal{F}^t$-measurable and T_k is a stopping time indeed.

We now formally define $\sigma^k : \mathcal{H} \rightarrow \Delta(\mathcal{A})$. Take any $p \in \mathcal{P}(h)$ and let

$$\sigma^k(h) = \begin{cases} \sigma^\phi(h), & \text{if } \kappa(h) \leq k, \\ \sigma^{h|_{T_k(p)}}(h), & \text{otherwise.} \end{cases} \tag{5.3}$$

If $\kappa(h) > k$, then the time at which switch k has occurred is the same for every $p \in \mathcal{P}(h)$, so it holds that σ^k is well-defined.

For every $k \in \mathbb{N}$, let $\mathcal{R}_k \subseteq \mathcal{P}$ be the set of plays along which at least k switches occur,

$$\mathcal{R}_k = \{p \in \mathcal{P} | T_k(p) < \infty\}, \tag{5.4}$$

so $\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \mathcal{R}_3 \supseteq \cdots$. Furthermore, let $\mathcal{R}_\infty \subseteq \mathcal{P}$ denote the set of plays along which infinitely many switches occur,

$$\mathcal{R}_\infty = \cap_{k=1}^\infty \mathcal{R}_k. \tag{5.5}$$

The remainder of the section is devoted to the proof of three technical lemmas concerning the finite switching strategies. These lemmas are needed in the analysis of σ^ϕ.

Consider the strategies $\sigma^k, \sigma^{k+1}, \ldots$. All these strategies agree with σ^ϕ for as long as σ^ϕ does not require more than k switches. Consequently, the measures that these strategies induce on $\mathcal{P}$ assign the same probability to any event that is "determined" before switch $k+1$ occurs, i.e., to any event in the sigma-algebra $\mathcal{F}^{T_{k+1}}$.

Lemma 5.2 *Let a strategy $\tau \in \mathcal{S}_2$, a history $h \in \mathcal{H}$, and some $k \in \mathbb{N}$ be given. For $\sigma = \sigma^k, \sigma^{k+1}, \ldots, \sigma^\phi$, the probability measures $\mathbb{P}_{h,\sigma,\tau}$ all coincide on the sigma-algebra $\mathcal{F}^{T_{k+1}}$. Furthermore, these probability measures all agree on each universally measurable subset of $\mathcal{P} \setminus \mathcal{R}_{k+1}$.*

Proof A set A of the universally measurable sigma-algebra $\mathcal{F}$ is called *agreeable* if for $\sigma = \sigma^k, \sigma^{k+1}, \ldots, \sigma^\phi$ the measures $\mathbb{P}_{h,\sigma,\tau}$ all assign the same probability to A.

We argue first that each cylinder set in $\mathcal{F}^{T_{k+1}}$ is agreeable. A cylinder set $\mathcal{P}(h)$ is a member of $\mathcal{F}^{T_{k+1}}$ if and only if $\kappa(h) \leq k+1$. Let a cylinder set $\mathcal{P}(h')$ in $\mathcal{F}^{T_{k+1}}$ be given. Since $\kappa(h') \leq k+1$, we know that $\kappa(h'') \leq k$ for each history h'' preceding h'. It follows that $\sigma^k(h'') = \sigma^{k+1}(h'') = \cdots = \sigma^\phi(h'')$. Since this applies to each history h'' that precedes h', the set $\mathcal{P}(h')$ is agreeable.

Now take any $E \in \mathcal{F}^{T_{k+1}}$. For $t \in \mathbb{N}$, let $E_t = E \cap \{p \in \mathcal{P} \mid T_{k+1}(p) = t\}$ and let $E_\infty = E \cap \{p \in \mathcal{P} \mid T_{k+1}(p) = \infty\}$. To show that E is agreeable, it suffices to show that the sets E_t and E_∞ are.

Let some $t \in \mathbb{N}$ be given. We know that E_t is a member of $\mathcal{F}^t$. Consequently, E_t can be written as a disjoint union of cylinder sets in $\mathcal{F}^t$, say $E_t = \cup \{C_n \mid n \in \mathbb{N}\}$, with each C_n a member of $\mathcal{F}^t$. Since each C_n is a subset of the set $\{p \in \mathcal{P} \mid T_{k+1}(p) = t\}$, it is a member of $\mathcal{F}^{T_{k+1}}$, so is agreeable by the result of the second paragraph in the proof. It now follows that E_t is agreeable.

To show that E_∞ is agreeable, we make use of the fact that E_∞ is a Borel set and of the regularity of σ on the Borel sigma-algebra. Let O be any open subset of $\mathcal{P}$ containing E_∞. The set O can be written as a disjoint union of cylinder sets, say $O = \cup \{\mathcal{P}(h_n) \mid n \in \mathbb{N}\}$. Without loss of generality it holds that for every $n \in \mathbb{N}$ the set $\mathcal{P}(h_n)$ has a point in common with E_∞. Thus in particular, there is $p \in \mathcal{P}(h_n)$ with $T_{k+1}(p) = \infty$. This implies that $\kappa(h_n) \leq k$ and hence that $\mathcal{P}(h_n)$ is a member of $\mathcal{F}^{T_{k+1}}$. We conclude that each $\mathcal{P}(h_n)$ is agreeable by the result of the second paragraph in the proof. It follows that O is an agreeable set.

To prove the second claim, we notice that all Borel subsets of $\mathcal{P} \setminus \mathcal{R}_{k+1} = \{p \in \mathcal{P} \mid T_{k+1}(p) = \infty\}$ are members of $\mathcal{F}^{T_{k+1}}$, so are agreeable. The result for universally measurable

subsets of $\{p \in \mathcal{P} \mid T_{k+1}(p) = \infty\}$ follows since each such set differs from a Borel set by a negligible set.   □

Let T be a stopping time. We define the stochastic variable $\underline{U}_\sigma^T$ by letting it agree with $\underline{U}_\sigma^t$ on the set $\{p \in \mathcal{P} : T(p) = t\}$ for each $t \in \mathbb{N}$ and by letting it agree with $\underline{U}_\sigma^\infty$ on the set $\{p \in \mathcal{P} : T(p) = \infty\}$, where $\underline{U}_\sigma^\infty$ is some suitably chosen $\mathcal{F}^\infty$-measurable stochastic variable. The next lemma relates the guaranteed expected payoffs of strategies σ^k and σ^{k+1} at the moment of switch $k + 1$. We write Φ^t to denote the $\mathcal{F}^t$-measurable stochastic variable given by $\Phi^t(p) = \phi(p_{|t})$.

Lemma 5.3 *Let $k \in \mathbb{N}$ and $\mathcal{F}^\infty$-measurable stochastic variables $\underline{U}_{\sigma^k}^\infty, \underline{U}_{\sigma^{k+1}}^\infty$ such that $\underline{U}_{\sigma^k}^\infty = \underline{U}_{\sigma^{k+1}}^\infty$ be given. Then, it holds that*

$$\underline{U}_{\sigma^k}^{T_{k+1}} \leq \underline{U}_{\sigma^{k+1}}^{T_{k+1}} - \tfrac{1}{2}\Phi^{T_{k+1}} \cdot I(T_{k+1} < \infty). \tag{5.6}$$

Proof Let some $p \in \mathcal{P}$ be given. We distinguish the following two cases.

CASE 1: $T_{k+1}(p) < \infty$. In this case at least $k + 1$ switches occur along the play p. For $h = p_{|T_{k+1}(p)}$, we have the following inequalities

$$\underline{u}(\sigma^k, h) < \underline{v}(h) - \phi(h) = \underline{v}(h) - \tfrac{1}{2}\phi(h) - \tfrac{1}{2}\phi(h) \leq \underline{u}(\sigma^{k+1}, h) - \tfrac{1}{2}\phi(h),$$

where the first inequality holds since σ^k is not a $\phi(h)$-maxmin strategy for the subgame at history h and the second inequality holds because the strategy σ^{k+1} is a $(\phi(h)/2)$-maxmin strategy for the subgame at history h. Since $I(T_{k+1}(p) < \infty) = 1$, (5.6) holds.

CASE 2: $T_{k+1}(p) = \infty$. In this case, we have

$$\underline{U}_{\sigma^k}^{T_{k+1}}(p) = \underline{U}_{\sigma^k}^\infty(p) = \underline{U}_{\sigma^{k+1}}^\infty(p) = \underline{U}_{\sigma^{k+1}}^{T_{k+1}}(p).$$

Thus, (5.6) holds as an equality.   □

The following lemma states the intuitive property that for histories with less than $k + 1$ switches or histories at which switch $k + 1$ occurs, strategy σ^{k+1} guarantees at least the same payoff to player 1 than strategy σ^k.

Lemma 5.4 *Let $t \in \mathbb{N}$ and a history $h \in \mathcal{H}^t$ of length t be given. Let $k \in \mathbb{N}$ be such that $T_{k+1}(p) \geq t$ for every $p \in \mathcal{P}(h)$. Then, it holds that $\underline{u}(\sigma^k, h) \leq \underline{u}(\sigma^{k+1}, h)$.*

Proof Fix strategy $\tau \in \mathcal{S}_2$ of player 2.

We first define $\underline{U}_{\sigma^k}^\infty$. Consider the probability measure $\mathbb{Q}$ on the measurable space $(\mathcal{P}, \mathcal{F})$ given by $\mathbb{Q}(A) = \sum_{k \in \mathbb{N}} 2^{-k-1}\mathbb{P}_{h,\sigma^k,\tau}(A)$ for each $A \in \mathcal{F}$. Let $\bar{u}$ be an $\mathcal{F}^\infty$-measurable stochastic variable with the property that $\bar{u} = u$, $\mathbb{Q}$-almost surely. Since $\mathbb{P}_{h,\sigma^k,\tau}$ is absolutely continuous with respect to $\mathbb{Q}$, it holds that $\bar{u} = u$, $\mathbb{P}_{h,\sigma^k,\tau}$-almost surely, for every $k \in \mathbb{N}$. We define $\underline{U}_{\sigma^k}^\infty$ to be equal to $\bar{u}$, for every $k \in \mathbb{N}$.

We now obtain the following inequalities. First, by Lemma 3.3 and the optional sampling theorem with unbounded stopping times as presented in Theorem 8.16 of Yeh ([22], p. 139), we have[2]

$$\underline{u}(\sigma^k, h) \leq \mathbb{E}_{h,\sigma^k,\tau}[\underline{U}_{\sigma^k}^{T_{k+1}}].$$

---

[2] We verify that the conditions of Theorem 8.16 in Yeh ([22]) applied to the process $(\underline{U}_\sigma^n)_{n \geq t}$ on the measurable space $(\mathcal{P}, \mathcal{F}, \mathbb{P}_{h,\sigma,\tau})$ with filtration $(\mathcal{F}^n)_{n \geq t}$ are satisfied. The process $(\underline{U}_\sigma^n)_{n \geq t}$ is a $\mathbb{P}_{h,\sigma,\tau}$-submartingale with respect to the filtration $(\mathcal{F}^n)_{n \geq t}$ by Lemma 3.3. Take ξ of Theorem 8.16 in Yeh ([22]) equal to u. Since $\underline{U}_\sigma^\infty$ is an $\mathcal{F}^\infty$-measurable stochastic variable that $\mathbb{P}_{h,\sigma,\tau}$-almost surely coincides with u, it is a version of $\mathbb{E}_{h,\sigma,\tau}[u|\mathcal{F}^\infty]$, as is required by the theorem. To verify condition (1) of Theorem 8.16 in Yeh ([22]), notice that, for every $n \geq t$, for every play $p \in \mathcal{P}$, we have $\underline{U}_\sigma^n(p) = \underline{u}(\sigma, p_{|n}) \leq \mathbb{E}_{p_{|n},\sigma,\tau}[u]$. The right-hand side of this inequality is a version of $\mathbb{E}_{h,\sigma,\tau}[u|\mathcal{F}^n]$. Consequently, $\underline{U}_\sigma^n \leq \mathbb{E}_{h,\sigma,\tau}[u|\mathcal{F}^n]$ holds $\mathbb{P}_{h,\sigma,\tau}$-almost surely, as desired.

Secondly, from the fact that $\underline{U}^{T_{k+1}}_{\sigma^k}$ is an $\mathcal{F}^{T_{k+1}}$-measurable stochastic variable, we obtain by Lemma 5.2 that

$$\mathbb{E}_{h,\sigma^k,\tau}[\underline{U}^{T_{k+1}}_{\sigma^k}] = \mathbb{E}_{h,\sigma^{k+1},\tau}[\underline{U}^{T_{k+1}}_{\sigma^k}].$$

Thirdly, we know that

$$\underline{U}^{T_{k+1}}_{\sigma^k} \leq \underline{U}^{T_{k+1}}_{\sigma^{k+1}} \leq \mathbb{E}_{h,\sigma^{k+1},\tau}[u|\mathcal{F}^{T_{k+1}}], \quad \mathbb{P}_{h,\sigma^{k+1},\tau}\text{-almost surely,}$$

where the first of these inequalities follows from Lemma 5.3 and the second one follows by Lemma 3.3 and the optional sampling theorem with unbounded stopping times as presented in Theorem 8.16 of Yeh ([22], p. 139). Taking the expectation of the last array of inequalities with respect to $\mathbb{P}_{h,\sigma^{k+1},\tau}$ and making use of the law of iterated expectation yields

$$\mathbb{E}_{h,\sigma^{k+1},\tau}[\underline{U}^{T_{k+1}}_{\sigma^k}] \leq \mathbb{E}_{h,\sigma^{k+1},\tau}[u].$$

Combining these facts yields $\underline{u}(\sigma^k, h) \leq \mathbb{E}_{h,\sigma^{k+1},\tau}[u]$. Taking the infimum over all strategies τ of player 2 completes the proof.   □

5.3 Properties of the Switching Strategy σ^ϕ

We show first that the switching strategy σ^ϕ is always n-day ϕ-maxmin for every $n \in \mathbb{N}$.

Lemma 5.5 *Let a game $\Gamma_{x_0}(u)$ and a tolerance function $\phi > 0$ be given. For every $n \in \mathbb{N}$, the switching strategy σ^ϕ is n-day ϕ-maxmin.*

Proof Let a history $h \in \mathcal{H}$ with length t be given. Let $k = \kappa(h)$ denote the number of switches that have occurred along the history h. For every $\tau \in \mathcal{S}_2$, we obtain the chain of inequalities

$$\underline{v}(h) - \phi(h) \leq \underline{u}(\sigma^k, h) \leq \underline{u}(\sigma^{t+n}, h) \leq \mathbb{E}_{h,\sigma^{t+n},\tau}[\underline{U}^{t+n}_{\sigma^{t+n}}] \leq \mathbb{E}_{h,\sigma^{t+n},\tau}[\underline{V}^{t+n}]$$
$$= \mathbb{E}_{h,\sigma^\phi,\tau}[\underline{V}^{t+n}],$$

where the first inequality holds since σ^k is a $\phi(h)$-maxmin strategy for the subgame at history h, the second inequality holds by Lemma 5.4 since $k \leq t \leq t+n$, the third inequality holds by Lemma 3.3, and the fourth one follows since $\underline{U}^{t+n}_{\sigma^{t+n}} \leq \underline{V}^{t+n}$. The final equality follows from the fact that the stochastic variable $\underline{V}^{t+n}$ is $\mathcal{F}^{t+n}$-measurable and the fact that the strategy σ^{t+n} coincides with σ^ϕ at least until time $t + n$.   □

The next lemma says that along almost any play $p \in \mathcal{P}$ only finitely many switches occur or the tolerance level goes to zero. This lemma is crucial to show the equalizing property of switching strategies.

Lemma 5.6 *For every history $h \in \mathcal{H}$, for every strategy $\tau \in \mathcal{S}_2$ of player 2, it holds that*

$$\lim_{k \to \infty} \Phi^{T_k} \cdot I(T_k < \infty) = 0, \quad \mathbb{P}_{h,\sigma^\phi,\tau}\text{-almost surely.}$$

Proof Although it is possible that player 1 switches infinitely many times along a play p, and therefore incurs infinitely many increases in his guarantee level along this play, we show first that the total overall increase in his guarantee level is bounded, since the payoff function itself is a bounded function. More formally, let $t \in \mathbb{N}$, a history $h \in \mathcal{H}^t$ of length t, and

a strategy $\tau \in \mathcal{S}_2$ of player 2 be given. Let $\kappa(h)$ denote the number of switches that have occurred along the history h. We show that

$$\sum_{k=\kappa(h)+1}^{\infty} \mathbb{E}_{h,\sigma^\phi,\tau}\left[\tfrac{1}{2}\Phi^{T_k} \cdot I\,(T_k < \infty)\right] \leq 2M, \tag{5.7}$$

where we recall that $M = \sup_{p \in \mathcal{P}} |u(p)|$.

As in the proof of Lemma 5.4, let $\bar{u}$ be any $\mathcal{F}^\infty$-measurable stochastic variable with the property that, for every $k \in \mathbb{N}$, $\bar{u} = u$, $\mathbb{P}_{h,\sigma^k,\tau}$-almost surely. For every $k \in \mathbb{N}$, we define $\underline{U}_{\sigma^k}^\infty = \bar{u}$.

Let some $k > \kappa(h)$ be given. For every $p \in \mathcal{P}(h)$, it holds that $t \leq T_k(p) \leq T_{k+1}(p)$, so Lemma 3.3 and the optional sampling theorem with unbounded stopping times as presented in Theorem 8.16 of Yeh ([22], p. 139) implies

$$\underline{U}_{\sigma^k}^{T_k} \leq \mathbb{E}_{h,\sigma^k,\tau}[\underline{U}_{\sigma^k}^{T_{k+1}}|\mathcal{F}^{T_k}], \quad \mathbb{P}_{h,\sigma^k,\tau}\text{-almost surely}. \tag{5.8}$$

We now have

$$\mathbb{E}_{h,\sigma^\phi,\tau}[\underline{U}_{\sigma^k}^{T_k}] = \mathbb{E}_{h,\sigma^k,\tau}[\underline{U}_{\sigma^k}^{T_k}] \leq \mathbb{E}_{h,\sigma^k,\tau}[\underline{U}_{\sigma^k}^{T_{k+1}}] = \mathbb{E}_{h,\sigma^\phi,\tau}[\underline{U}_{\sigma^k}^{T_{k+1}}],$$

where the two equalities follow from Lemma 5.2 and the fact that both $\underline{U}_{\sigma^k}^{T_k}$ and $\underline{U}_{\sigma^k}^{T_{k+1}}$ are $\mathcal{F}^{T_{k+1}}$-measurable stochastic variables, and the inequality follows by taking the expectation on both sides of inequality (5.8) and the law of iterated expectation.

Using Lemma 5.3, we can conclude that

$$\begin{aligned}
\mathbb{E}_{h,\sigma^\phi,\tau}\left[\tfrac{1}{2}\Phi^{T_{k+1}} \cdot I(T_{k+1} < \infty)\right] &\leq \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^{k+1}}^{T_{k+1}}\right] - \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^k}^{T_{k+1}}\right] \\
&\leq \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^{k+1}}^{T_{k+1}}\right] - \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^k}^{T_k}\right].
\end{aligned}$$

Summing the preceding inequality over $k = \kappa(h) + 1, \dots, K$, we obtain

$$\sum_{k=\kappa(h)+1}^{K} \mathbb{E}_{h,\sigma^\phi,\tau}\left[\tfrac{1}{2}\Phi^{T_k} \cdot I(T_k < \infty)\right] \leq \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^K}^{T_K}\right] - \mathbb{E}_{h,\sigma^\phi,\tau}\left[\underline{U}_{\sigma^{\kappa(h)+1}}^{T_{\kappa(h)+1}}\right] \leq 2M.$$

The result follows by taking the limit as $K \to \infty$.

Let us write $X_k = \Phi^{T_k} \cdot I(T_k < \infty)$ and $k' = \kappa(h) + 1$. Since $X_k \geq 0$, the monotone convergence theorem implies that $\mathbb{E}_{h,\sigma^\phi,\tau}[\sum_{k=k'}^{\infty} X_k] = \sum_{k=k'}^{\infty} \mathbb{E}_{h,\sigma^\phi,\tau}[X_k]$. The latter expression is finite by (5.7). Thus $\sum_{k=k'}^{\infty} X_k$ has a finite mean with respect to the probability measure $\mathbb{P}_{h,\sigma^\phi,\tau}$. Hence $\sum_{k=k'}^{\infty} X_k < \infty$ holds $\mathbb{P}_{h,\sigma^\phi,\tau}$-almost surely. This implies that $X_k \to 0$ holds $\mathbb{P}_{h,\sigma^\phi,\tau}$-almost surely.   □

We are now in a position to state the main result of this section, which gives sufficient conditions for the existence of subgame ϕ-maxmin strategies.

Theorem 5.7 *Let a game $\Gamma_{x_0}(u)$ and a tolerance function $\phi > 0$ be given such that for every $p \in \mathcal{P}$ at least one of the following two conditions holds:*

1. *(point of upper semicontinuity) The function u is upper semi-continuous at p.*
2. *(positive limit inferior) $\liminf_{t\to\infty} \phi(p_{|t}) > 0$.*

Then, there exists a subgame ϕ-maxmin strategy in the game $\Gamma_{x_0}(u)$.

Proof We define the tolerance function ϕ' by

$$\phi'(h) = \tfrac{1}{2} \min\{\phi(h') | h' \preceq h\}, \quad h \in \mathcal{H}.$$

Thus, ϕ' is a non-increasing tolerance function with $\phi' \leq \tfrac{1}{2}\phi$.

We show that $\sigma^{\phi'}$ is ϕ'-equalizing. Since $2\phi' \leq \phi$ and $\sigma^{\phi'}$ is an n-day ϕ'-maxmin strategy for every $n \in \mathbb{N}$ by Lemma 5.5, it then follows from Theorem 4.5 that $\sigma^{\phi'}$ is a subgame ϕ-maxmin strategy.

Let $\mathcal{U}$ denote the set of plays $p \in \mathcal{P}$ at which u is upper semi-continuous and let $\mathcal{I}$ denote the set of plays $p \in \mathcal{P}$ such that $\liminf_{t \to \infty} \phi(p_{|t}) > 0$. By the assumption of the theorem it holds that $\mathcal{P} = \mathcal{U} \cup \mathcal{I}$. By the definition of ϕ', we have $\liminf_{t \to \infty} \phi'(p_{|t}) > 0$ for each $p \in \mathcal{I}$.

Let some $t \in \mathbb{N}$, a history $h \in \mathcal{H}^t$, and a strategy $\tau \in \mathcal{S}_2$ be given.

STEP 1 For every $p \in \mathcal{U}$, $u(p) \geq \limsup_{n \to \infty} \underline{V}^n(p) - \phi'$.

This follows directly from Lemma 4.10.

For every $k \in \mathbb{N}$, we define $\mathcal{J}_k = \{p \in \mathcal{I} \cap (\mathcal{R}_k \setminus \mathcal{R}_{k+1}) \mid \lim_{n \to \infty} U^n_{\sigma^k, \tau}(p) = u(p)\}$ and $\mathcal{J} = \bigcup_{k \in \mathbb{N}} \mathcal{J}_k$, where we remind the reader that $\mathcal{R}_k$ contains the plays along which at least k switches occur.

STEP 2 For every $p \in \mathcal{J}$, $u(p) \geq \limsup_{n \to \infty} \underline{V}^n(p) - \phi'$.

Let $k \in \mathbb{N}$ and $p \in \mathcal{J}_k$ be given. Exactly k switches occur along the play p and the last switch occurs at time $T_k(p)$. By our construction of $\sigma^{\phi'}$, this means that for each time $n > T_k(p)$, the strategy σ^k is a $\phi'(p_{|n})$-maxmin strategy for the subgame at history $p_{|n}$, so $\underline{u}(\sigma^k, p_{|n}) \geq \underline{v}(p_{|n}) - \phi'(p_{|n})$. Since $U^n_{\sigma^k, \tau}(p) \geq \underline{u}(\sigma^k, p_{|n})$ and since for $n \geq t$ we have $\phi'(p_{|n}) \leq \phi'(h)$, we conclude that for $n \geq t$

$$U^n_{\sigma^k, \tau}(p) \geq \underline{V}^n(p) - \phi'(h).$$

Taking the limit as n goes to infinity, and making use of the fact that $p \in \mathcal{J}_k$, we obtain

$$u(p) = \lim_{n \to \infty} U^n_{\sigma^k, \tau}(p) \geq \limsup_{n \to \infty} \underline{V}^n(p) - \phi'(h).$$

STEP 3: $\mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{I} \cap \mathcal{R}_\infty) = 0$.

Recall that by Lemma 5.6, $\Phi^{T_k} \cdot I(T_k < \infty)$ converges to 0 as k goes to infinity, $\mathbb{P}_{h, \sigma^{\phi'}, \tau}$-almost surely. Also recall that $\mathcal{R}_\infty$ is the set of plays where infinitely many switches occur. Thus $I(T_k < \infty)$ is identically equal to 1 on $\mathcal{R}_\infty$. Furthermore, $\liminf_{k \to \infty} \Phi^{T_k} > 0$ everywhere on $\mathcal{I}$. We conclude that $\Phi^{T_k} \cdot I(T_k < \infty)$ does not converge to zero on $\mathcal{I} \cap \mathcal{R}_\infty$ and the result follows.

STEP 4: $\mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{U} \cup \mathcal{J}) = 1$.

For every $k \in \mathbb{N}$, it holds by Levy's zero-one law (Lemma A.1 in Appendix A) that

$$\mathbb{P}_{h, \sigma^k, \tau}(\mathcal{J}_k) = \mathbb{P}_{h, \sigma^k, \tau}(\mathcal{I} \cap (\mathcal{R}_k \setminus \mathcal{R}_{k+1})).$$

Using Lemma 5.2 twice yields

$$\mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{J}_k) = \mathbb{P}_{h, \sigma^k, \tau}(\mathcal{J}_k) = \mathbb{P}_{h, \sigma^k, \tau}(\mathcal{I} \cap (\mathcal{R}_k \setminus \mathcal{R}_{k+1})) = \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{I} \cap (\mathcal{R}_k \setminus \mathcal{R}_{k+1})).$$

We now have

$$\mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{J}) = \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{I} \setminus \mathcal{R}_\infty) = \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{I}),$$

where the last equality follows from Step 3. Finally, we obtain

$$\mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{U} \cup \mathcal{J}) \geq \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{U} \setminus \mathcal{I}) + \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{J}) = \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{U} \setminus \mathcal{I}) + \mathbb{P}_{h, \sigma^{\phi'}, \tau}(\mathcal{I}) = 1,$$

where the last equality follows from the fact that the sets $\mathcal{U}$ and $\mathcal{I}$ cover $\mathcal{P}(h)$.   □

Theorem 5.7 generalizes Proposition 11 in Mashiah–Yaakovi ([15]), where the existence of a subgame ϵ-optimal strategy in a two-player zero-sum stochastic game with Borel measurable payoff functions is shown. The tolerance function ϕ there is given by $\phi(h) = \epsilon$ for every $h \in \mathcal{H}$, so for every play the tolerance function has a positive limit inferior. Theorem 5.7 yields the existence of a subgame ϵ-maxmin strategy. Because the Borel measurability of the payoff function guarantees the existence of the value (Maitra and Sudderth, [13], and Martin, [14]), this is equivalent to proving the existence of a subgame ϵ-optimal strategy.

In Sect. 6, we argue that Theorem 5.7 also provides further insight into the main result of Laraki, Maitra, and Sudderth ([10]). There the authors prove among other things that if the payoff function is bounded and upper semi-continuous, then the first player has a subgame optimal strategy.

6 Subgame maxmin strategies

The goal of this section is to explore the relationship between the concept of a subgame maxmin strategy and that of a subgame ϕ-maxmin strategy for $\phi > 0$. The main result of this section is the following theorem.

Theorem 6.1 *For every game $\Gamma_{x_0}(u)$, there exists a tolerance function $\phi^* > 0$ such that the following statements are equivalent:*

1. *The game $\Gamma_{x_0}(u)$ has a subgame maxmin strategy.*
2. *The game $\Gamma_{x_0}(u)$ has a subgame ϕ^*-maxmin strategy.*

Because every subgame maxmin strategy is a subgame ϕ-maxmin strategy, statement 1 clearly implies statement 2. To prove the converse, we define a tolerance function $\phi^* > 0$; the precise definition is given in Subsection 6.3. Intuitively, we fix a positive sequence $(\delta_t)_{t \in \mathbb{N}}$ with the property that $\sum_{t=0}^{\infty} \delta_t < \infty$. Then, we choose ϕ^* such that (1) $\phi^*(p_{|t}) < \delta_t$ for very play $p \in \mathcal{P}$ and every $t \in \mathbb{N}$ and (2) every one-day ϕ^*-maxmin action is δ_t-close to the region of optimal mixed actions in the one-day game. The first condition implies that tolerance function ϕ^* decreases rapidly to zero, i.e., $\sum_{t=0}^{\infty} \phi^*(p_{|t}) < \infty$. By assumption, player 1 has a subgame ϕ^*-maxmin strategy. We obtain a subgame maxmin strategy as follows: at every history h_t we pick a mixed action which guarantees the lower value $\underline{v}(h_t)$ in the one-day game and is δ_t-close to the mixed action chosen by the subgame ϕ^*-maxmin strategy. This method creates a strategy which is clearly one-day maxmin. The condition that the tolerance function ϕ^* fastly decreases to 0 is then used to show that expected payoffs from a history h_t with length t obtained by the subgame ϕ^*-maxmin strategy and the constructed strategy are within a distance of $M \cdot \sum_{n=t}^{\infty} \delta_n$, where $M = \sup_{p \in \mathcal{P}} |u(p)|$. From this, the equalizing property of the constructed strategy follows. Then, Corollary 4.7 from Sect. 4 lets us conclude that the constructed strategy is indeed a subgame maxmin strategy.

Theorem 6.1 has the following corollary.

Corollary 6.2 *The following statements are equivalent:*

1. *The game $\Gamma_{x_0}(u)$ has a subgame maxmin strategy.*
2. *For every $\phi > 0$, the game $\Gamma_{x_0}(u)$ has a subgame ϕ-maxmin strategy.*

We would like to make two remarks. First, as the payoff function u need not be continuous, one cannot simply use a continuity argument to prove that statement 2 of Corollary 6.2 implies

statement 1. Second, note that the existence of a subgame ϵ-maxmin strategy for every $\epsilon > 0$ is a weaker requirement than statements 1 and 2 of Corollary 6.2. Indeed, as Example 3.2 already illustrated, there exist games which admit a subgame ϵ-maxmin strategy for every $\epsilon > 0$ but do not admit a subgame maxmin strategy.

The special case where the payoff-function is upper semi-continuous deserves some additional attention. From Maitra and Sudderth ([13]) and Martin ([14]), it follows that every two-player zero-sum stochastic game with a countable state space, finite action sets, and a Borel measurable payoff function admits a value. Because every upper semi-continuous payoff function is Borel measurable, the existence of the value is guaranteed. From Theorem 5.7, we obtain the existence of a subgame ϕ-optimal strategy for every $\phi > 0$. Combining this with Corollary 6.2 shows that player 1 has a subgame optimal strategy. Hereby, we obtain a special case of the result by Laraki, Maitra, and Sudderth ([10]), where the authors allow the state space to be a Borel subset of a Polish space and transition probabilities to be determined by a Borel transition function.

When we are interested in pure strategies that guarantee the maxmin levels, we can strengthen the result of Theorem 6.1. In the context of simultaneous move games, pure strategies are of course rather restrictive. Still, there are important classes of games, such as perfect information games, in which they are natural and play a prominent role.

Theorem 6.3 *For every game $\Gamma_{x_0}(u)$, there exists a tolerance function $\phi^* > 0$ such that the following statements are equivalent:*

1. *The pure strategy $\sigma \in \mathcal{S}_1$ is a subgame maxmin strategy.*
2. *The pure strategy $\sigma \in \mathcal{S}_1$ is a subgame ϕ^*-maxmin strategy.*

The tolerance function ϕ^* in Theorem 6.3 will be defined based on the data of the game and the lower values in the subgames.

This section is structured as follows. We start by proving Theorem 6.3, which is easier than Theorem 6.1 and helps us explain the main ideas. Then, we turn to the proof of Theorem 6.1.

6.1 The proof of Theorem 6.3

In this subsection, we prove Theorem 6.3. Since the statement of this theorem is about pure strategies, the proof is less technical and the intuition is more transparent.

We only need to prove that there is $\phi^* > 0$ such that statement 2 of Theorem 6.3 implies statement 1 of Theorem 6.3. From now on, for any $t \in \mathbb{N}$, history $h \in \mathcal{H}^t$ with some final state x, and mixed actions $m_1 \in \Delta(\mathcal{A})$ and $m_2 \in \Delta(\mathcal{B})$, we denote the expectation of the lower value at the next stage by

$$\mathbb{E}_{h,m_1,m_2}\left[\underline{V}^{t+1}\right] = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \sum_{x' \in \mathcal{X}} q(x'|h, a, b) \cdot \underline{v}(h, a, b, x'). \tag{6.1}$$

The proof consists of two steps.

STEP 1. Definition of $\phi^* > 0$.

For every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, there exists a number $d(h) > 0$ such that for every action $a \in \mathcal{A}$ of player 1 either one of the following holds:

- Action a guarantees that the lower value does not drop in expectation: for every $m_2 \in \Delta(\mathcal{B})$, $\mathbb{E}_{h,a,m_2}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h)$.
- There exists a mixed action $m_2 \in \Delta(\mathcal{B})$ for player 2 such that, if player 1 uses action a, the lower value drops in expectation by more than $d(h)$: $\mathbb{E}_{h,a,m_2}\left[\underline{V}^{t+1}\right] < \underline{v}(h) - d(h)$.

This statement is true because the action set $\mathcal{A}$ of player 1 is finite. We define $\phi^* > 0$ as follows. For every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, let $\phi^*(h) = \min\{d(h), 2^{-t}\}$. The term 2^{-t} is included so that the tolerance levels tend to 0 along each play. For every $p \in \mathcal{P}$, it holds that $\lim_{t \to \infty} \phi^*(p_{|t}) = 0$.

STEP 2. If the pure strategy $\sigma \in \mathcal{S}_1$ is a subgame ϕ^*-maxmin strategy, then σ is a subgame maxmin strategy.

Let the pure strategy $\sigma \in \mathcal{S}_1$ be subgame ϕ^*-maxmin. We verify that σ satisfies the conditions of Corollary 4.7.

First, we show that σ is 1-day maxmin. Fix $t \in \mathbb{N}$ and $h \in \mathcal{H}^t$. Let $a = \sigma(h)$ denote the action that σ recommends at h. Then, for every mixed action $m_2 \in \Delta(\mathcal{B})$ and strategy $\tau \in \mathcal{S}_2$ with $m_2 = \tau(h)$, we have

$$\mathbb{E}_{h,a,m_2}\left[\underline{V}^{t+1}\right] = \mathbb{E}_{h,\sigma,\tau}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h) - \phi^*(h) \geq \underline{v}(h) - d(h),$$

where the first inequality follows from the fact that σ is subgame ϕ^*-maxmin and by condition 1 of Theorem 4.6, and the second inequality follows from the definition of $\phi^*(h)$. Therefore, by the choice of $d(h)$ in Step 1, for every $m_2 \in \Delta(\mathcal{B})$, it holds that

$$\mathbb{E}_{h,a,m_2}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h).$$

Hence, for every strategy $\tau \in \mathcal{S}_2$, it holds that

$$\mathbb{E}_{h,\sigma,\tau}\left[\underline{V}^{t+1}\right] = \mathbb{E}_{h,a,\tau(h)}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h).$$

Thus, σ is 1-day maxmin.

We show that σ is equalizing. Take some $\tau \in \mathcal{S}_2$. Because σ is subgame ϕ^*-maxmin, for every $p \in \mathcal{P}(h)$, for every $n \geq t$, we have

$$U_{\sigma,\tau}^n(p) = \mathbb{E}_{p_{|n},\sigma,\tau}[u] \geq \underline{v}(p_{|n}) - \phi^*(p_{|n}) = \underline{V}^n(p) - \Phi^{*n}(p),$$

where $\Phi^{*n}(p) = \phi^*(p_{|n})$. We conclude that

$$\lim_{n \to \infty} U_{\sigma,\tau}^n \geq \limsup_{n \to \infty}\left(\underline{V}^n - \Phi^{*n}\right) = \limsup_{n \to \infty} \underline{V}^n - \lim_{n \to \infty} \Phi^{*n} = \limsup_{n \to \infty} \underline{V}^n,$$

where the last equality from the fact that for all $p \in \mathcal{P}$ we have $\lim_{n \to \infty} \phi^*(p_{|n}) = 0$, so

$$u = \lim_{n \to \infty} U_{\sigma,\tau}^n \geq \limsup_{n \to \infty} \underline{V}^n, \quad \mathbb{P}_{h,\sigma,\tau}\text{-almost surely,}$$

where the equality follows from Lemma A.1 in Appendix A. We have shown that σ is equalizing.

6.2 The One-Shot Game Υ_h

To prove Theorem 6.1, we first analyze a one-shot zero-sum game in this subsection. For each history, the one-shot game is such that the payoff is given by the lower value at the next stage. In the next subsection, we use these one-shot games to construct the tolerance function $\phi^* > 0$.

For some $t \in \mathbb{N}$, let $h \in \mathcal{H}^t$ be a history in the game $\Gamma_{x_0}(u)$. The one-shot zero-sum game Υ_h is played as follows. Player 1 chooses an action $a \in \mathcal{A}$ and player 2 simultaneously chooses an action $b \in \mathcal{B}$. Then, player 1 receives from player 2 the amount $\mathbb{E}_{h,a,b}\left[\underline{V}^{t+1}\right]$. As the action sets $\mathcal{A}$ and $\mathcal{B}$ are finite, the game Υ_h has a value, which we denote by $w(h)$. Furthermore, both players have optimal mixed actions in the game Υ_h.

The following lemma states that the value $w(h)$ of the one-shot game Υ_h equals the lower value $\underline{v}(h)$ at the history h in the original game $\Gamma_{x_0}(u)$. In case the value exists, this is a classic result of dynamic programming.

Lemma 6.4 *For every history $h \in \mathcal{H}$, we have $w(h) = \underline{v}(h)$.*

Proof The proof is by contradiction. Fix $t \in \mathbb{N}$ and a history $h \in \mathcal{H}^t$. Now suppose that $w(h) \neq \underline{v}(h)$. Then, we have either $w(h) > \underline{v}(h)$ or $w(h) < \underline{v}(h)$.
CASE 1: $w(h) > \underline{v}(h)$.
Let $\delta = w(h) - \underline{v}(h)$. We derive a contradiction by showing that, in the subgame of $\Gamma_{x_0}(u)$ at history h, player 1 can guarantee an expected payoff of at least $\underline{v}(h) + \delta/2$.

Let $m_1 \in \Delta(\mathcal{A})$ be an optimal mixed action for player 1 in the one-shot game Υ_h. Let $\sigma \in \mathcal{S}_1$ be such that $\sigma(h) = m_1$ and such that it induces a $(\delta/2)$-maxmin strategy for the subgame at each history in period $t+1$, i.e., for every $h' \in \mathcal{H}^{t+1}$, for every $\tau \in \mathcal{S}_2$,

$$\mathbb{E}_{h',\sigma,\tau}[u] \geq \underline{v}(h') - \frac{\delta}{2}. \tag{6.2}$$

Then, for every $\tau \in \mathcal{S}_2$, it holds that

$$\mathbb{E}_{h,\sigma,\tau}[u] = \mathbb{E}_{h,\sigma,\tau}\left[U_{\sigma,\tau}^t\right] = \mathbb{E}_{h,\sigma,\tau}\left[U_{\sigma,\tau}^{t+1}\right]$$
$$\geq \mathbb{E}_{h,\sigma,\tau}\left[\underline{V}^{t+1}\right] - \frac{\delta}{2} = \mathbb{E}_{h,m_1,\tau(h)}\left[\underline{V}^{t+1}\right] - \frac{\delta}{2} \geq w(h) - \frac{\delta}{2} = \underline{v}(h) + \frac{\delta}{2},$$

where the second equality follows from the fact that $(U_{\sigma,\tau}^n)_{n \geq t}$ is a $\mathbb{P}_{h,\sigma,\tau}$-martingale, the first inequality follows from (6.2), and the second inequality follows from the choice of m_1.
CASE 2: $w(h) < \underline{v}(h)$.[3]
Let $\delta = \underline{v}(h) - w(h)$. We derive a contradiction by showing that, for every strategy of player 1, there is a strategy for player 2 such that the expected payoff is at most $\underline{v}(h) - \delta/2$ in the subgame of $\Gamma_{x_0}(u)$ at history h.

Fix $\sigma \in \mathcal{S}_1$ and let $m_1 = \sigma(h)$. Let $m_2 \in \Delta(\mathcal{B})$ be an optimal mixed action for player 2 in the one-shot game Υ_h. Let $\tau \in \mathcal{S}_2$ be such that $\tau(h) = m_2$ and the expected payoff under (σ, τ) in the subgame at each history h' at period $t+1$ is at most the lower value $\underline{v}(h') + \delta/2$, i.e., for every $h' \in \mathcal{H}^{t+1}$,

$$\mathbb{E}_{h',\sigma,\tau}[u] \leq \underline{v}(h') + \frac{\delta}{2}. \tag{6.3}$$

We have that

$$\underline{v}(h) - \frac{\delta}{2} = w(h) + \frac{\delta}{2} \geq \mathbb{E}_{h,m_1,m_2}\left[\underline{V}^{t+1}\right] + \frac{\delta}{2}$$
$$\geq \mathbb{E}_{h,m_1,m_2}\left[U_{\sigma,\tau}^{t+1}\right] = \mathbb{E}_{h,\sigma,\tau}\left[U_{\sigma,\tau}^{t+1}\right] = \mathbb{E}_{h,\sigma,\tau}\left[U_{\sigma,\tau}^t\right] = \mathbb{E}_{h,\sigma,\tau}[u],$$

where the first inequality follows from the choice of m_2, the second inequality follows from (6.3), and the penultimate equality follows from the fact that $(U_{\sigma,\tau}^n)_{n \geq t}$ is a $\mathbb{P}_{h,\sigma,\tau}$-martingale. Because σ is chosen arbitrarily, we have derived a contradiction with the definition of the lower value $\underline{v}(h)$. □

The total variation distance between two mixed actions $m_1, n_1 \in \Delta(\mathcal{A})$ of player 1 is defined as

$$\|m_1 - n_1\|_{\mathrm{TV}} = \sum_{a \in \mathcal{A}} |m_1(a) - n_1(a)|.$$

---

[3] The proof of this case in not symmetric to the proof of Case 1, because we consider the lower value. Imitating the proof of Case 1 for player 2 would yield results in terms of the upper value.

The total variation distance between two probability measures $\mathbb{P}$ and $\mathbb{P}'$ on $(\mathcal{P}, \mathcal{F})$ is defined as

$$\|\mathbb{P} - \mathbb{P}'\|_{\mathrm{TV}} = \sup\{\sum_{i=1}^{n} |\mathbb{P}(F_i) - \mathbb{P}'(F_i)| : F_1, \ldots, F_n \in \mathcal{F} \text{ and } \{F_1, \ldots, F_n\} \text{ is a partition of } \mathcal{P}\}.$$

Let $t \in \mathbb{N}$ and a history $h \in \mathcal{H}^t$ be given. Let $O_h \subseteq \Delta(\mathcal{A})$ denote the set of optimal mixed actions of player 1 in the one-shot game Υ_h. By Lemma 6.4 it holds that

$$O_h = \left\{ m_1 \in \Delta(\mathcal{A}) \mid \text{for every } m_2 \in \Delta(\mathcal{B}), \ \mathbb{E}_{h,m_1,m_2}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h) \right\}.$$

Note that O_h is a compact subset of $\Delta(\mathcal{A})$. For every $m_1 \in \Delta(\mathcal{A})$, the distance of m_1 to O_h is defined by

$$\|m_1 - O_h\|_{\mathrm{TV}} = \min_{n_1 \in O_h} \|m_1 - n_1\|_{\mathrm{TV}}.$$

Due to the compactness of O_h, the minimum is attained. For $\delta > 0$, let D_h^δ be the set of mixed actions of player 1 which have a distance of at least δ to the set O_h, so

$$D_h^\delta = \{m_1 \in \Delta(\mathcal{A}) \mid \|m_1 - O_h\|_{\mathrm{TV}} \geq \delta\}. \tag{6.4}$$

The mixed actions in D_h^δ are not optimal in the one-shot game Υ_h. The following lemma says that the loss in utility caused by these mixed actions has a positive lower bound.

Lemma 6.5 *Let $h \in \mathcal{H}$ and $\delta > 0$ be given. If D_h^δ is non-empty, then there is $\epsilon > 0$ such that for every $m_1 \in D_h^\delta$ there exists $b \in \mathcal{B}$ such that*

$$\mathbb{E}_{h,m_1,b}\left[\underline{V}^{t+1}\right] \leq \underline{v}(h) - \epsilon.$$

Proof Assume D_h^δ is non-empty. Consider the function $e_h^\delta : D_h^\delta \to \mathbb{R}$ defined by

$$e_h^\delta(m_1) = \underline{v}(h) - \min_{b \in \mathcal{B}} \mathbb{E}_{h,m_1,b}\left[\underline{V}^{t+1}\right], \quad m_1 \in D_h^\delta. \tag{6.5}$$

Since $\mathcal{B}$ is finite, the minimum exists. For every $m_1 \in D_h^\delta$, we have $m_1 \notin O_h$, and therefore there exists $m_2 \in \Delta(\mathcal{B})$ such that $\mathbb{E}_{h,m_1,m_2}\left[\underline{V}^{t+1}\right] < \underline{v}(h)$. The function e_h^δ is therefore a positive and continuous function on a compact set, so has a positive minimum. □

6.3 Definition of the Tolerance Function ϕ^*

In this subsection, we define a positive tolerance function ϕ^*. Fix a positive and decreasing sequence $(\delta_t)_{t \in \mathbb{N}}$ such that $\sum_{t=0}^{\infty} \delta_t < \infty$. For example, $\delta_t = 2^{-t}$ for every $t \in \mathbb{N}$. Note that $\lim_{t \to \infty} \delta_t = 0$.

For every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, we define the constant $c(h)$ as follows. If the set $D_h^{\delta_t}$ is non-empty, then $c(h)$ is equal to the positive number ϵ of Lemma 6.5 and $c(h) = \delta_t$ otherwise. We define

$$\phi^*(h) = \frac{\min\{c(h), \delta_t\}}{2}. \tag{6.6}$$

Notice that $0 < \phi^*(h) < \delta_t$.

We summarize the properties of the tolerance function ϕ^*:

1. For every history $h \in \mathcal{H}$, we have $\phi^*(h) > 0$.
2. For every play $p \in \mathcal{P}$, we have $\sum_{t=0}^{\infty} \phi^*(p_{|t}) \leq \sum_{t=0}^{\infty} \delta_t < \infty$.
3. For every play $p \in \mathcal{P}$, $\lim_{t \to \infty} \phi^*(p_{|t}) = 0$.

4. If the set $D_h^{\delta_t}$ is non-empty, then by the choice of $c(h)$, for every $m_1 \in D_h^{\delta_t}$ there exists $b \in \mathcal{B}$ such that

$$\mathbb{E}_{h,m_1,b}\left[\underline{V}^{t+1}\right] \leq \underline{v}(h) - c(h) < \underline{v}(h) - \phi^*(h).$$

The importance of $\sum_{t=0}^{\infty} \delta_t < \infty$ is underlined by the following lemma. Recall that $M = \sup_{p \in \mathcal{P}} |u(p)|$.

Lemma 6.6 *Let the strategies $\sigma, \sigma' \in \mathcal{S}_1$ be such that, for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, $\|\sigma(h) - \sigma'(h)\|_{\mathrm{TV}} \leq \delta_t$. Then, for every strategy $\tau \in \mathcal{S}_2$, for every $t \in \mathbb{N}$, and for every history $h \in \mathcal{H}^t$,*

$$\left|\mathbb{E}_{h,\sigma,\tau}[u] - \mathbb{E}_{h,\sigma',\tau}[u]\right| \leq M \cdot \sum_{n=t}^{\infty} \delta_n.$$

Proof Let $\tau \in \mathcal{S}_2$ be given. It follows from a more general result in Abate, Redig, and Tkachev ([1], theorem 1) that, for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$,

$$\|\mathbb{P}_{h,\sigma,\tau} - \mathbb{P}_{h,\sigma',\tau}\|_{\mathrm{TV}} \leq \sum_{n=t}^{\infty} \delta_n. \tag{6.7}$$

For completeness, we provide a direct proof of this inequality in Lemma B.1 in Appendix B. The claim of Lemma 6.6 follows directly. □

Lemma 6.6 says the following. Consider two arbitrary strategies $\sigma, \sigma' \in \mathcal{S}_1$ such that the total variation distance between the mixed actions at every history in period t is at most δ_t. Now consider $t \in \mathbb{N}$, a history $h \in \mathcal{H}^t$, and a strategy $\tau \in \mathcal{S}_2$ of player 2. Then, the expected payoffs under (σ, τ) and (σ', τ) in the subgame at h differ at most the constant M times $\sum_{n=t}^{\infty} \delta_n$. Note that this bound does not depend on the strategy τ and it only depends on the history h through its period t. Moreover, these bounds tend to 0 as t goes to infinity.

The importance of property 4 of the tolerance function is shown by the following lemma. It says that if $\sigma \in \mathcal{S}_1$ is a subgame ϕ^*-maxmin strategy, then for every $h \in H$ the mixed action $\sigma(h)$ is close to the set of optimal mixed actions O_h in the one-shot game Υ_h.

Lemma 6.7 *Let $\sigma \in \mathcal{S}_1$ be a subgame ϕ^*-maxmin strategy. Then, for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, we have $\sigma(h) \notin D_h^{\delta_t}$, so $\|\sigma(h) - O_h\|_{\mathrm{TV}} < \delta_t$.*

Proof Because σ is a subgame ϕ^*-maxmin strategy, it follows from condition 1 of Theorem 4.6 that for every mixed action $m_2 \in \Delta(\mathcal{B})$ of player 2

$$\mathbb{E}_{h,\sigma(h),m_2}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h) - \phi^*(h).$$

If $D_h^{\delta_t}$ is empty, then there is nothing to prove. If $D_h^{\delta_t}$ is non-empty, then property 4 of ϕ^* shows that $\sigma(h) \notin D_h^{\delta_t}$. □

6.4 The proof of Theorem 6.1

Proof Let $\Gamma_{x_0}(u)$ be a game and take the tolerance function ϕ^* as defined in Subsection 6.3. We only need to show that statement 2 implies statement 1. Let $\sigma \in \mathcal{S}_1$ be a subgame ϕ^*-maxmin strategy of $\Gamma_{x_0}(u)$.

With the help of σ, we define a strategy $\sigma^* \in \mathcal{S}_1$ in Step 1 of the proof. Then, it is shown that σ^* is a subgame maxmin strategy of $\Gamma_{x_0}(u)$ in Steps 2 and 3 of the proof by verifying that σ^* satisfies the conditions of Corollary 4.7.

STEP 1: Definition of $\sigma^* \in \mathcal{S}_1$.
Take $t \in \mathbb{N}$ and a history $h \in \mathcal{H}^t$. In view of Lemma 6.7, it holds that $\|\sigma(h) - O_h\|_{TV} < \delta_t$. Therefore, there exists $m^*(h) \in O_h$ such that

$$\|\sigma(h) - m^*(h)\|_{TV} < \delta_t. \tag{6.8}$$

Now define $\sigma^*(h) = m^*(h)$.

STEP 2: σ^* is 1-day maxmin.
Consider some $\tau \in \mathcal{S}_2$. For every $t \in \mathbb{N}$, for every $h \in \mathcal{H}^t$, since $\sigma^*(h) \in O_h$ we have that

$$\mathbb{E}_{h,\sigma^*,\tau}\left[\underline{V}^{t+1}\right] \geq \underline{v}(h).$$

STEP 3: σ^* is equalizing.
Consider some $\tau \in \mathcal{S}_2$. In view of (6.8), we can apply Lemma 6.6 to conclude that, for every $t \in \mathbb{N}$, for every $h \in \mathcal{H}^t$,

$$\left|\mathbb{E}_{h,\sigma,\tau}[u] - \mathbb{E}_{h,\sigma^*,\tau}[u]\right| \leq M \cdot \sum_{n=t}^{\infty} \delta_n.$$

Hence, for every $t \in \mathbb{N}$, for every history $h \in \mathcal{H}^t$, and for every $n \geq t$,

$$\left|U_{\sigma,\tau}^n[u] - U_{\sigma^*,\tau}^n[u]\right| \leq M \cdot \sum_{i=n}^{\infty} \delta_i, \quad \mathbb{P}_{h,\sigma^*,\tau}\text{-almost surely}. \tag{6.9}$$

Because σ is subgame ϕ^*-maxmin, for every history $h \in \mathcal{H}$, we have

$$\mathbb{E}_{h,\sigma,\tau}[u] \geq \underline{v}(h) - \phi^*(h). \tag{6.10}$$

Thus, for every history $h \in \mathcal{H}$ it holds that

$$u = \lim_{t \to \infty} U_{\sigma^*,\tau}^t = \lim_{t \to \infty} U_{\sigma,\tau}^t \geq \limsup_{t \to \infty}\left(\underline{V}^t - \Phi^{*t}\right) = \limsup_{t \to \infty} \underline{V}^t, \quad \mathbb{P}_{h,\sigma^*,\tau}\text{-almost surely},$$

where the first equality is due to Lemma A.1 in Appendix A, the second equality follows from (6.9), the inequality is by (6.10), and the last equality is a consequence of property 3 of ϕ^* from Subsection 6.3. □

Note that it is not necessary that the initial δ_t are small, or that for initial histories the tolerance level $\phi^*(h)$ is small. The conditions put on the sequence $(\delta_t)_{t\in\mathbb{N}}$ are there to ensure that the equalizing condition for subgame maxmin strategies is satisfied. The equalizing condition essentially only cares about what happens to the payoff at deep subgames. The requirement that the sequence $(\delta_t)_{t\in\mathbb{N}}$ decreases to zero very fast guarantees that $\lim_{t\to\infty}\|\mathbb{P}_{p_{|t},\sigma,\tau} - \mathbb{P}_{p_{|t},\sigma,\tau}\|_{TV} = 0$. This allows us to use the ϕ^*-equalizing property of the subgame ϕ^*-maxmin strategy to show that the strategy σ^* satisfies the equalizing condition.

7 Discussion

Countable action sets. We have assumed throughout the paper that the action sets are finite. If one player has countably many actions, several of our arguments and results are no longer valid. For example, consider the following game. In the initial state, player 1's action set is

$\mathbb{N}$, player 2 has only one action, and if player 1 chooses action $n \in \mathbb{N}$ then regardless how the play continues, the payoff is $1 - \frac{1}{n+1}$. This is essentially a one-shot game. Clearly, for every tolerance function $\phi > 0$ player 1 has a pure subgame ϕ-maxmin strategy, i.e. he can play action n with n sufficiently large. Yet, player 1 has no subgame maxmin strategy. So the statements of Theorems 6.1 and 6.3 and Corollary 6.2 are not valid for this game.

One-player games. Suppose that player 2 has a trivial action set, i.e. $\mathcal{B}$ is a singleton. Even though this is essentially a game with only one player, as shown in Example 3.2, it may happen that player 1 has a subgame ϕ-maxmin strategy for a certain tolerance function ϕ, but he has no pure strategy that is subgame ϕ-maxmin. Nevertheless, if $\mathcal{B}$ is a singleton and the conditions of Theorem 5.7 are satisfied, then player 1 has a pure subgame ϕ-maxmin strategy. Indeed, player 1 has a pure δ-maxmin strategy for each $\delta > 0$. Hence, the switching strategy used in the proof of Theorem 5.7 will also be pure.

Discounted games. Our model encompasses discounted stochastic games with finite action and countable state spaces, provided that the immediate payoffs are bounded. Since these games admit a value and the discounted payoffs are continuous, it follows by Theorem 5.7 that player 1 has a subgame ϕ-optimal strategy for every tolerance function $\phi > 0$. Therefore, Corollary 6.2 implies the well-known fact that player 1 has a subgame 0-optimal strategy. Our proof does not directly imply that this strategy can be chosen to be stationary.

Appendix A: Universal Measurability

This appendix contains a review of the definitions of universally measurable sets, integrals of universally measurable functions, stopping times, and conditional expectations, as well as a technical lemma that is used in the paper.

Universally measurable sets Let $(\mathcal{P}, \mathcal{F}^\infty)$ be a measurable space, where $\mathcal{F}^\infty$ denotes the Borel sigma-algebra and let $\mathcal{M}$ denote the collection of all probability measures over this measurable space. For each probability measure $\mathbb{P} \in \mathcal{M}$, we can extend the probability space $(\mathcal{P}, \mathcal{F}^\infty, \mathbb{P})$ to a complete probability space $(\mathcal{P}, \mathcal{F}_\mathbb{P}, \mathbb{P}^c)$ by including all $\mathbb{P}$-negligible sets. To be more precise, let

$$\mathcal{F}_\mathbb{P}^0 = \{Q \subset \mathcal{P} \mid \exists Q' \in \mathcal{F}^\infty \text{ such that } \mathbb{P}(Q') = 0 \text{ and } Q \subseteq Q'\}$$

be the set of all subsets of $\mathbb{P}$-negligible sets of $\mathcal{F}^\infty$. We define

$$\mathcal{F}_\mathbb{P} = \{Q \cup Q^0 \subseteq \mathcal{P} \mid Q \in \mathcal{F}^\infty \text{ and } Q^0 \in \mathcal{F}_\mathbb{P}^0\}$$

and we define $\mathbb{P}^c : \mathcal{F}_\mathbb{P} \to [0, 1]$ by

$$\mathbb{P}^c(Q \cup Q^0) = \mathbb{P}(Q), \quad Q \in \mathcal{F}^\infty, \ Q^0 \in \mathcal{F}_\mathbb{P}^0.$$

It can be shown that $(\mathcal{P}, \mathcal{F}_{\mathbb{P}}, \mathbb{P}^c)$ is a probability space. Now define

$$\mathcal{F} = \bigcap_{\mathbb{P} \in \mathcal{M}} \mathcal{F}_{\mathbb{P}}.$$

It can be shown that $\mathcal{F}$ is a sigma-algebra that contains the Borel sigma-algebra $\mathcal{F}^{\infty}$ as a proper subset. The collection $\mathcal{F}$ is the universally measurable sigma-algebra and the elements of $\mathcal{F}$ are called universally measurable sets.

Integrals of universally measurable functions A function $u : \mathcal{P} \to \mathbb{R}$ is called universally measurable if $u^{-1}[a, b] \in \mathcal{F}$ for every $[a, b] \subseteq \mathbb{R}$. The class of universally measurable functions contains the class of Borel measurable functions. Furthermore, for every universally measurable function there exists a Borel measurable function that coincides with it almost everywhere.

A function $g : \mathcal{P} \to \mathbb{R}$ is called a simple universally measurable function if it is of the form $g(p) = \sum_{i=1}^{n} c_i I(p \in Z_i)$, where $\{Z_1, \ldots, Z_n\}$ is a partition of $\mathcal{P}$ with $Z_i \in \mathcal{F}$ for all $i = 1, \ldots, n$. The expected value of a simple universally measurable payoff with respect to a probability measure $\mathbb{P}$ is defined as

$$\int_{p \in \mathcal{P}} g(p) \mathbb{P}(dp) = \sum_{i=1}^{n} c_i \mathbb{P}^c(Z_i). \tag{A.1}$$

Let $\mathcal{G}$ denote the set of simple universally measurable functions. The expected value of a bounded universally measurable function $u : \mathcal{P} \to \mathbb{R}$ is then given by

$$\int_{p \in \mathcal{P}} u(p) \mathbb{P}(dp) = \sup_{\substack{g \in \mathcal{G} \\ g \le u}} \int_{p \in \mathcal{P}} g(p) \mathbb{P}(dp) = \inf_{\substack{g \in \mathcal{G} \\ g \ge u}} \int_{p \in \mathcal{P}} g(p) \mathbb{P}(dp). \tag{A.2}$$

Stopping times A stopping time is a function $T : \mathcal{P} \to \mathbb{N} \cup \{\infty\}$ such that for each $t \in \mathbb{N}$ the set $\{p \in \mathcal{P} \mid T(p) = t\}$ is an element of $\mathcal{F}^t$. Given a stopping time T, let $\mathcal{F}^T$ denote the sigma-algebra of sets $A \in \mathcal{F}^{\infty}$ such that $A \cap \{p \in \mathcal{P} \mid T(p) = t\} \in \mathcal{F}^t$ for each $t \in \mathbb{N}$.

For every $t \in \mathbb{N}$, let X^t be an $\mathcal{F}^t$ measurable stochastic variable, and let X^{∞} be an $\mathcal{F}^{\infty}$ measurable stochastic variable. The stochastic variable X^T is defined by letting it coincide with X^t on $\{p \in \mathcal{P} \mid T(p) = t\}$ for each $t \in \mathbb{N}$ and with X^{∞} on $\{p \in \mathcal{P} \mid T(p) = \infty\}$. Following Yeh ([22], Theorem 3.28), it holds that X^T is $\mathcal{F}^T$ measurable.

Conditional expectations Consider a bounded stochastic variable $F : \mathcal{P} \to \mathbb{R}$, a strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$, and a history $h \in \mathcal{H}^{\ell}$ of length ℓ. Let some $t \ge \ell$ be given. The conditional expectation of F with respect to the sigma-algebra $\mathcal{F}^t$ and the measurable space $(\mathcal{P}, \mathcal{F}, \mathbb{P}_{h,\sigma,\tau})$ is denoted by $\mathbb{E}_{h,\sigma,\tau}[F|\mathcal{F}^t]$. The conditional expectation $\mathbb{E}_{h,\sigma,\tau}[F|\mathcal{F}^t]$ can be identified with the stochastic variable $p \mapsto \mathbb{E}_{p_{|t},\sigma,\tau}[F]$.[4]

Lemma A.1 states a version of Levy's zero-one law for universally measurable functions. It relies on the fact that a universally measurable function can be approximated by a Borel measurable function.

Lemma A.1 (Levy's zero-one law for universally measurable functions) *For every strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$, for every history $h \in \mathcal{H}$, we have*

$$\lim_{t \to \infty} U_{\sigma,\tau}^t = u, \quad \mathbb{P}_{h,\sigma,\tau}\text{-almost surely.} \tag{A.3}$$

---

[4] As usual, a conditional expectation is not defined uniquely, since some histories might not be reached with positive probability under the strategy profile (σ, τ). Our particular choice is both convenient and inconsequential, since any two conditional probability systems coincide $\mathbb{P}_{h,\sigma,\tau}$-almost surely. We refer to Bogachev ([3], p. 350) for a careful discussion of conditional expectations.

Proof Fix a strategy profile $(\sigma, \tau) \in \mathcal{S}_1 \times \mathcal{S}_2$ and a history $h \in \mathcal{H}$. Since u is universally measurable, there exists a Borel measurable function $\bar{u}$ such that $u = \bar{u}$, $\mathbb{P}_{h,\sigma,\tau}$-almost surely. Then,

$$\lim_{t \to \infty} U_{\sigma,\tau}^t = \lim_{t \to \infty} \mathbb{E}_{h,\sigma,\tau}\left[u | \mathcal{F}^t\right] = \lim_{t \to \infty} \mathbb{E}_{h,\sigma,\tau}\left[\bar{u} | \mathcal{F}^t\right] = \mathbb{E}_{h,\sigma,\tau}\left[\bar{u} | \mathcal{F}^\infty\right] = \bar{u},$$

$\mathbb{P}_{h,\sigma,\tau}$-almost surely. In the first equality, we use the definition of the stochastic variable $U_{\sigma,\tau}^t$. The second equality follows from the fact that $u = \bar{u}$ $\mathbb{P}_{h,\sigma,\tau}$-almost surely. The third equality follows from Levy's zero-one law (see e.g., Bogachev, [3], Example 10.3.15). The last equality follows from the fact that $\bar{u}$ is $\mathcal{F}^\infty$ measurable. This is because $\bar{u}$ is Borel measurable and $\mathcal{F}^\infty$ is the Borel sigma-algebra on $\mathcal{P}$. The fact that $u = \bar{u}$ $\mathbb{P}_{h,\sigma,\tau}$-almost surely concludes the proof.   □

Appendix B: The Proof of Inequality (6.7)

The following statement follows from the more general result of Theorem 1 in Abate, Redig, and Tkachev ([1]). For completeness, we provide a direct proof in this section.

Lemma B.1 *Let* $\sigma, \sigma' \in \mathcal{S}_1$ *be such that, for every* $t \in \mathbb{N}$, *for every* $h \in \mathcal{H}^t$, $\|\sigma(h) - \sigma'(h)\|_{\mathrm{TV}} \leq \delta_t$. *Then, for every strategy* $\tau \in \mathcal{S}_2$, *for every* $t \in \mathbb{N}$, *and for every history* $h \in \mathcal{H}^t$,

$$\|\mathbb{P}_{h,\sigma,\tau} - \mathbb{P}_{h,\sigma',\tau}\|_{\mathrm{TV}} \leq \sum_{i=t}^{\infty} \delta_i.$$

The class $\mathcal{V} \subseteq \mathcal{F}$ is called an inner (outer) approximating class for the class $\mathcal{W} \subseteq \mathcal{F}$ if $\mathcal{V}$ is closed under unions (intersections) and if for every $\epsilon > 0$, and for every probability measure $\mathbb{P} \in \mathcal{M}$ and every set $W \in \mathcal{W}$ there exists a set $V \in \mathcal{V}$ such that $V \subseteq W$ ($V \supseteq W$) and $|\mathbb{P}(W) - \mathbb{P}(V)| \leq \epsilon$.

Lemma B.2 *If* $\mathcal{V}$ *is an inner (outer) approximating class for the class* $\mathcal{W}$ *and* $\mathcal{V} \subseteq \mathcal{W}$, *then it holds for every two probability measures* $\mathbb{P} \in \mathcal{M}$ *and* $\mathbb{P}' \in \mathcal{M}$ *that*

$$\sup_{W \in \mathcal{W}} |\mathbb{P}(W) - \mathbb{P}'(W)| = \sup_{V \in \mathcal{V}} |\mathbb{P}(V) - \mathbb{P}'(V)|.$$

Proof Fix $\epsilon > 0$. Assume that the class $\mathcal{V}$ is an inner approximating class for $\mathcal{W}$. The proof for an outer approximating class is similar. Fix two probability measures $\mathbb{P}$ and $\mathbb{P}'$ and a set $W \in \mathcal{W}$. Then there exist sets $V \in \mathcal{V}$ and $V' \in \mathcal{V}$ such that $V \subseteq W$, $V' \subseteq W$, $|\mathbb{P}(W) - \mathbb{P}(V)| \leq \epsilon$, and $|\mathbb{P}'(W) - \mathbb{P}'(V')| \leq \epsilon$. Let $\tilde{V} = V \cup V'$. Because $\mathcal{V}$ is closed under unions we have that $\tilde{V} \in \mathcal{V}$. Furthermore, it follows trivially that $\tilde{V} \subseteq W$, $|\mathbb{P}(W) - \mathbb{P}(\tilde{V})| \leq \epsilon$, and $|\mathbb{P}'(W) - \mathbb{P}'(\tilde{V})| \leq \epsilon$. We find that

$$|\mathbb{P}(W) - \mathbb{P}'(W)| = |\mathbb{P}(W) - \mathbb{P}(\tilde{V}) + \mathbb{P}(\tilde{V}) - \mathbb{P}'(\tilde{V}) + \mathbb{P}'(\tilde{V}) - \mathbb{P}'(W)|$$
$$\leq |\mathbb{P}(W) - \mathbb{P}(\tilde{V})| + |\mathbb{P}(\tilde{V}) - \mathbb{P}'(\tilde{V})| + |\mathbb{P}'(\tilde{V}) - \mathbb{P}'(W)|$$
$$\leq |\mathbb{P}(\tilde{V}) - \mathbb{P}'(\tilde{V})| + 2\epsilon.$$

Because this holds for any $\epsilon > 0$ and any set $W \in \mathcal{W}$, we can conclude that

$$\sup_{W \in \mathcal{W}} |\mathbb{P}(W) - \mathbb{P}'(W)| \leq \sup_{V \in \mathcal{V}} |\mathbb{P}(V) - \mathbb{P}'(V)|.$$

Because $\mathcal{V} \subseteq \mathcal{W}$, it is clear that:

$$\sup_{W \in \mathcal{W}} |\mathbb{P}(W) - \mathbb{P}'(W)| \geq \sup_{V \in \mathcal{V}} |\mathbb{P}(V) - \mathbb{P}'(V)|.$$

$\square$

In the following lemma, we use Lemma B.2 to simplify the computation of the total variation norm. Instead of having to compute the supremum over all sets of the universally measurable sigma-algebra, it will be sufficient to compute the supremum over the subclass of open sets $\mathcal{O}^t$, $t \in \mathbb{N}$, defined by

$$\mathcal{O}^t = \{\cup_{h \in H} \mathcal{P}(h) | H \subseteq \mathcal{H}^t\} \tag{B.4}$$

The set $\mathcal{O}^t$ is the class of open sets such that all the plays sharing a common history at time t are such that either all or none of them are contained in a specific open set.

Lemma B.3 *For every $h \in \mathcal{H}$, it holds that*

$$\|\mathbb{P}_{h,\sigma,\tau} - \mathbb{P}_{h,\sigma',\tau}\|_{\text{TV}} = \sup_{t \in \mathbb{N}, O \in \mathcal{O}^t} |\mathbb{P}_{h,\sigma,\tau}(O) - \mathbb{P}_{h,\sigma',\tau}(O)|.$$

Proof We define $\mathcal{O} = \cup_{t \in \mathbb{N}} \mathcal{O}^t$. Since any set in $\mathcal{O}$ is a union of open sets, it holds that $\mathcal{O}$ is a subset of the class of all open sets, $\mathcal{O}^*$. We now show that $\mathcal{O}$ is an inner approximating class for $\mathcal{O}^*$. Fix an open set $O^* \in \mathcal{O}^*$. For every $t \in \mathbb{N}$, we define $O^t = \{p \in \mathcal{P} | \mathcal{P}(p_{|t}) \subseteq O^*\}$. It is clear that for every $t \in \mathbb{N}$, we have $O^t \in \mathcal{O}$ and $O^t \subseteq O^*$. Furthermore, we have that $O^1 \subseteq O^2 \subseteq \dots$ and $O^* = \cup_{t \in \mathbb{N}} O^t$. For every probability measure $\mathbb{P}$ it therefore holds that $\lim_{t \to \infty} \mathbb{P}(O^t) = \mathbb{P}(O^*)$, so for every $\epsilon > 0$ there exists $t \in \mathbb{N}$ and $O^t \in \mathcal{O}^t$ such that $|\mathbb{P}(O^*) - \mathbb{P}(O^t)| < \epsilon$. Hence, $\mathcal{O}$ is an inner approximating class of $\mathcal{O}^*$.

Because of the outer regularity of Borel measures on metric spaces, we have that the class of open sets is an outer approximating class for the class of Borel sets. In addition, we have that the class of Borel sets is an inner approximating class for the class of universally measurable sets. Indeed, for any probability measure and any universally measurable set, there exists a Borel set which is contained in the universally measurable set and has the same probability. Repeated application of Lemma B.2 concludes the proof. $\square$

Proof of Lemma B.1 Fix $t \in \mathbb{N}$ and a history $h_t \in \mathcal{H}^t$. We prove by induction that for every $n \geq t$, for every $O \in \mathcal{O}^{n+1}$,

$$\left|\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O)\right| \leq \sum_{i=t}^{n} \delta_i. \tag{B.5}$$

Induction basis ($n = t$).
We prove that for every $O \in \mathcal{O}^{t+1}$, $\left|\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O)\right| \leq \delta_t$. Fix a set $O \in \mathcal{O}^{t+1}$. We define

$$\mathcal{Z}_{h_t} = \{(a, b, x) \in \mathcal{A} \times \mathcal{B} \times \mathcal{X} | \mathcal{P}(h_t abx) \subseteq O\}.$$

Let $\mathcal{A}_{h_t} = \{a \in \mathcal{A} | \exists (b, x) \in \mathcal{B} \times \mathcal{X} : (a, b, x) \in \mathcal{Z}_{h_t}\}$ be the projection of the set $\mathcal{Z}_{h_t}$ on the set $\mathcal{A}$. Let x_t denote the state at the history h_t. We have that

$$\mathbb{P}_{h_t,\sigma,\tau}(O) = \sum_{(a,b,x) \in \mathcal{Z}_{h_t}} \sigma(h_t)(a) \cdot \tau(h_t)(b) \cdot q(x|a, b, x_t),$$

$$\mathbb{P}_{h_t,\sigma',\tau}(O) = \sum_{(a,b,x)\in\mathcal{Z}_{h_t}} \sigma'(h_t)(a) \cdot \tau(h_t)(b) \cdot q(x|a,b,x_t).$$

We find that

$$
\begin{aligned}
\left|\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O)\right| &= \left|\sum_{(a,b,x)\in\mathcal{Z}_{h_t}} \big(\sigma(h_t)(a) - \sigma'(h_t)(a)\big) \cdot \tau(h_t)(b) \cdot q(x|a,b,x_t)\right| \\
&\leq \sum_{(a,b,x)\in\mathcal{Z}_{h_t}} \left|\sigma(h_t)(a) - \sigma'(h_t)(a)\right| \cdot \tau(h_t)(b) \cdot q(x|a,b,x_t) \\
&\leq \sum_{(a,b,x)\in\mathcal{A}_{h_t}\times\mathcal{B}\times\mathcal{X}} \left|\sigma(h_t)(a) - \sigma'(h_t)(a)\right| \cdot \tau(h_t)(b) \cdot q(x|a,b,x_t) \\
&= \left[\sum_{a\in\mathcal{A}_{h_t}} \left|\sigma(h_t)(a) - \sigma'(h_t)(a)\right|\right] \cdot \left[\sum_{(b,x)\in\mathcal{B}\times\mathcal{X}} \tau(h_t)(b) \cdot q(x|a,b,x_t)\right] \\
&\leq \sum_{a\in\mathcal{A}} \left|\sigma(h_t)(a) - \sigma'(h_t)(a)\right| = \|\sigma(h_t) - \sigma'(h_t)\|_{\mathrm{TV}} \leq \delta_t.
\end{aligned}
$$

Induction step.
From the induction hypotheses, it follows that for every open n-level set $O^n \in \mathcal{O}^n$ with $n-1 \geq t$,

$$|\mathbb{P}_{h_t,\sigma,\tau}(O^n) - \mathbb{P}_{h_t,\sigma',\tau}(O^n)| \leq \sum_{i=t}^{n-1} \delta_i. \tag{B.6}$$

Fix a set $O \in \mathcal{O}^{n+1}$. We can assume without loss of generality that $\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O) \geq 0$. Define

$$\mathcal{H}_+^n = \{h_n \in \mathcal{H}^n \mid \mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \geq 0\}. \tag{B.7}$$

We have that

$$
\begin{aligned}
&\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O) \\
&= \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma',\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \\
&= \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \\
&\quad + \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) - \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma',\tau}(O|\mathcal{P}(h_n))\mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \\
&= \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n)) \cdot \big(\mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n))\big) \\
&\quad + \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \cdot \big(\mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(O|\mathcal{P}(h_n))\big) \\
&\leq \sum_{h_n\in\mathcal{H}_+^n} \mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n)) \cdot \big(\mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n))\big) \\
&\quad + \sum_{h_n\in\mathcal{H}^n} \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \cdot \left|\big(\mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(O|\mathcal{P}(h_n))\big)\right|
\end{aligned}
$$

$$\leq \sum_{h_n \in \mathcal{H}_+^n} \left(\mathbb{P}_{h_t,\sigma,\tau}(\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \right) + \sum_{h_n \in \mathcal{H}^n} \mathbb{P}_{h_t,\sigma',\tau}(\mathcal{P}(h_n)) \cdot \delta_n$$

$$\leq |\mathbb{P}_{h_t,\sigma,\tau}(\cup_{h_n \in \mathcal{H}_+^n} \mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(\cup_{h_n \in \mathcal{H}_+^n} \mathcal{P}(h_n))| + \delta_n$$

$$\leq \sum_{i=t}^{n-1} \delta_i + \delta_n = \sum_{i=t}^{n} \delta_i,$$

where the fact that $\left| \left(\mathbb{P}_{h_t,\sigma,\tau}(O|\mathcal{P}(h_n)) - \mathbb{P}_{h_t,\sigma',\tau}(O|\mathcal{P}(h_n)) \right) \right| \leq \delta_n$ follows by assumption. Using Lemma B.3, we can conclude that

$$\begin{aligned} \|\mathbb{P}_{h_t,\sigma,\tau} - \mathbb{P}_{h_t,\sigma',\tau}\|_{\mathrm{TV}} &= \sup_{n \in \mathbb{N}, O \in \mathcal{O}^n} |\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O)| \\ &= \sup_{n \geq t, O \in \mathcal{O}^{n+1}} |\mathbb{P}_{h_t,\sigma,\tau}(O) - \mathbb{P}_{h_t,\sigma',\tau}(O)| \\ &\leq \sup_{n \geq t} \sum_{i=t}^{n} \delta_i = \sum_{i=t}^{\infty} \delta_i. \end{aligned}$$

$\square$

References

1. Abate A, Redig F, Tkachev I (2014) On the effect of perturbation of conditional probabilities in total variation. Stat Probab Lett 88:1–8
2. Blackwell D, Ferguson T (1968) The big match. Ann Math Stat 39:159–163
3. Bogachev VI (2007) Measure Theory, vol II. Springer, Berlin
4. Bruyère V (2017) Computer aided synthesis: a game-theoretic approach. In: Charlier E, Leroy J, Rigo M (eds) Developments in Language Theory, vol 10396. Lecture Notes in Computer Science. Springer, pp 3–35
5. Flesch J, Herings PJJ, Maes J, Predtetchinski A (2018) Subgame maxmin strategies in zero-sum stochastic games with tolerance levels. GSBE Research Memorandum 18/20, Maastricht University, Maastricht, pp. 1-37
6. Flesch J, Predtetchinski A (2016) On refinements of subgame perfect ϵ-equilibrium. Int J Game Theory 45:523–542
7. Flesch J, Predtetchinski A, Sudderth W (2018) Characterization and simplification of optimal strategies in positive stochastic games. J Appl Probab 55:728–741
8. Flesch J, Thuijsman F, Vrieze OJ (1998) Improving strategies in stochastic games. In: Proceedings of the 37th IEEE Conference on Decision and Control, Vol 3, IEEE, pp. 2674-2679
9. Gillette D (1957) Stochastic games with zero stop probabilities. In: Dresher M, Tucker AW, Wolfe P (eds) Contributions to the Theory of Games, vol III. Annals of Mathematics Studies, Volume 39. Princeton University Press, Princeton, New Jersey, pp 179–187
10. Laraki R, Maitra A, Sudderth W (2013) Two-person zero-sum stochastic games with semicontinuous payoff. Dyn Games Appl 3:162–171
11. Mailath G, Postlewaite A, Samuelson L (2005) Contemporaneous perfect epsilon-equilibria. Games Econ Behav 53:126–140
12. Maitra A, Sudderth W (1996) Discrete gambling and stochastic games. Springer, New York
13. Maitra A, Sudderth W (1998) Finitely additive stochastic games with Borel measurable payoffs. Int J Game Theory 27:257–267
14. Martin DA (1998) The determinacy of Blackwell games. J Symb Logic 63:1565–1581
15. Mashiah-Yaakovi A (2015) Correlated equilibria in stochastic games with Borel measurable payoffs. Dyn Games Appl 5:120–135
16. Puterman ML (1994) Markov decision processes, discrete stochastic dynamic programming. Wiley, Hoboken
17. Radner R (1980) Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. J Econ Theory 22:136–154

18. Rosenberg D, Solan E, Vieille N (2001) Stopping games with randomized strategies. Probab Theory Related Fields 119:433–451
19. Selten R (1965) Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit: Teil I: Bestimmung des dynamischen Preisgleichgewichts. Zeitschrift für die gesamte Staatswissenschaft 121:301–324
20. Solan E, Vieille N (2002) Uniform value in recursive games. Ann App Probab 12:1185–1201
21. Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458
22. Yeh J (1995) Martingales and stochastic analysis, vol 1. World Scientific, Singapore