Stochastic Equilibria under Imprecise Deviations in Terminal-Reward Concurrent Games*

Patricia Bouyer Nicolas Markey Daniel Stan
LSV, CNRS & ENS Cachan, Université Paris-Saclay, France

We study the existence of mixed-strategy equilibria in concurrent games played on graphs. While existence is guaranteed with safety objectives for each player, Nash equilibria need not exist when players are given arbitrary terminal-reward objectives, and their existence is undecidable with qualitative reachability objectives (and only three players). However, these results rely on the fact that the players can enforce infinite plays while trying to improve their payoffs. In this paper, we introduce a relaxed notion of equilibria, where deviations are imprecise. We prove that contrary to Nash equilibria, such (stationary) equilibria always exist, and we develop a PSPACE algorithm to compute one.

1 Introduction

Games (especially games played on graphs) are a prominent formalism for modelling and reasoning about interactions between components of computerized systems [15, 9]. Until recently, those games have mainly been studied in the special case where only two players are interacting and have opposite objectives. This setting is especially relevant for modelling reactive systems evolving in a presumably hostile environment. Over the last decade, multi-player games with non-zero-sum objectives have come into the picture: they allow for conveniently modelling complex infrastructures where each individual system tries to fulfill its own objectives, while still being subject to interactions with the surrounding systems. As an example, consider (a simplified version of) the team-formation problem [7], as depicted in Fig. 1: several agents are trying to complete tasks; each task requires some resources, which are shared by the players. Completing a task thus requires the formation of a team that has all the required resources for that task: each player selects the task she wants to achieve (and so proposes her resources for achieving that task), and if a task receives enough resources, the associated team receives the corresponding payoff (to be divided among the players in the team). In such a game, there is a need of cooperation (to gather enough resources), and an incentive to selfishness (to maximise the payoff).

*This work is partly supported by ERC project EQualIS (308087) and by FP7 project Cassting (601148).

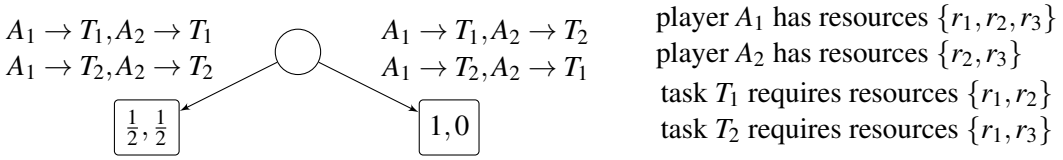


Figure 1: An instance of the team-formation problem [7]. For any deterministic choice of actions, one of the players has an incentive to change her choice: there is no pure Nash equilibrium. However there is one mixed Nash equilibrium, where each player plays T_1 and T_2 uniformly at random.

Submitted to:
GandALF 2016

In that setting, focusing only on optimal strategies for one single agent is not relevant. In game theory, several solution concepts have been defined, which more accurately represents rational behaviours of these multi-player systems; Nash equilibrium [11] is the best-known such concept. A Nash equilibrium is a strategy profile (that is, one strategy to each player) where no player can improve her own payoff by unilaterally changing her strategy. In other terms, in a Nash equilibrium, each individual player has a satisfactory strategy with regards to the other players' strategies. Notice that Nash equilibria need not exist (except for some classes of games) nor be unique, and they are not necessarily "optimal": Nash equilibria where all players lose may coexist with other Nash equilibria with positive payoffs. Many other concepts do exist, which refine the notion of Nash equilibria (like subgame-perfect equilibria [13] or trembling-hand equilibria [14]), or relax the notion (like ε-Nash equilibria [6]). The existence and computation of (constrained) equilibria (for various concepts) are important problems in the area, for which many results have been recently obtained.

In particular, in a recent paper [4], we proved that the existence of Nash equilibria in randomized strategies is undecidable in deterministic concurrent games with terminal-reward (while the problem is decidable for pure strategies [3]). Those games are concurrent games played on graphs, with terminal nodes assigning a reward to every player. The undecidability result holds for three players or more, and the status of two-player games is open: it is not known whether there always exists a Nash equilibrium in two-player concurrent games, even when the terminal rewards are in $\{0,1\}$ (which corresponds to a reachability objective).

In order to circumvent this undecidability result, we consider in this paper a relaxed version of Nash equilibria, with a stronger notion of *profitable deviation*. A deviation is called *really-profitable* only if all the "neighbouring" deviations (with small changes in the probability distribution) remain profitable (in the standard sense). In this paper, we prove that under this restriction, such equilibria always exist, even for concurrent games with stochastic states. We also show that *stationary* equilibria exist, and provide an algorithm to compute one.

To prove the existence result, we show that the notion of imprecise deviations is captured by adding constraints to the set of strategy profiles one can use. This allows to show the convexity of the set of best responses to a given strategy profile, as well as a terminating property (that is, with a lower-bounded positive probability the game progresses toward the terminal states). Then Kakutani's fixed-point theorem [10] can be applied to get the existence result, as is done in many other contexts. Note that the above-mentioned terminating property is a property that one either proves through discounting, like in stay-in-a-set games [12] and for ε-Nash equilibria in reachability games [6], or that one imposes, like in "games that end almost surely" in [1].

Related work. Our notion of equilibria is close to the notion of *trembling-hand perfect equilibria*, which has been proposed in the context of matrix games in [14]; in trembling-hand equilibria, strategy profiles should be robust to small perturbations when playing (or implementing) the strategies while keeping the standard optimality criteria of Nash equilibria. This concept obviously shares conceptual considerations of our notion of equilibria against imprecise deviations; however the point-of-view is somehow dual: the imprecision is in the implementation of the equilibrium in [14], whereas it is in the existence of really-profitable deviations in our work. While the notion of trembling-hand perfection refines that of Nash equilibria (it allows for a selection in the set of Nash equilibria), our notion relaxes that notion. In particular every trembling-hand perfect equilibrium is a Nash equilibriun, and every Nash equilibrium is an equilibrium against imprecise deviations (and the inclusions are strict).

ε-Nash equilibria [6] relax the notion of Nash equilibria as well, but in a different way: deviations are interpreted in a standard way, but single deviations should not increase the payoff by more than ε. This

is another way to introduce imprecision in Nash equilibria, which also ensures the existence of stationary equilibria in stochastic concurrent games with terminal rewards.

2 Definitions

In this paper, we study multiplayer stochastic concurrent games. This section presents a definition of those games, discusses mixed strategy Nash equilibria, and defines the new concept of equilibria under imprecise deviations.

2.1 Concurrent game

In the following, if A is an at most denumerable set, $\mathit{Dist}(A)$ will denote the set of probability distributions over A. If δ is such a distribution, $\mathit{Supp}(\delta)$ denotes the support of δ, that is the subset $\{a \in A \mid \delta(a) > 0\}$. Pointwise addition for distributions will be written $+$, and multiplication by a scalar is written $\cdot$, so that for any two distributions δ and δ' on the same set A, and for any $p \in [0,1]$, $p \cdot \delta + (1-p) \cdot \delta'$ is still a distribution on A.

Definition 1. *A* stochastic concurrent arena $\mathscr{A}$ *is a* 5*-tuple* $\langle \mathsf{States}, \mathsf{Agt}, \mathsf{Act}, (\mathsf{Allow}_i)_{i \in \mathsf{Agt}}, \mathsf{Tab} \rangle$ *where*

- States *is a finite set of states,* Agt *is a finite set of agents (or players),* Act *is a finite set of actions;*
- *for each* $i \in \mathsf{Agt}$, $\mathsf{Allow}_i \colon \mathsf{States} \longrightarrow 2^{\mathsf{Act}} \setminus \{\emptyset\}$ *is a function describing the set of actions available to player* i *from a given state;*
- $\mathsf{Tab} \colon \mathsf{States} \times \mathsf{Act}^{\mathsf{Agt}} \to \mathit{Dist}(\mathsf{States})$ *is the transition function, which assigns to every combined action of the players a distribution on the next states.*

We say that the arena is deterministic *whenever the transition function is deterministic (i.e., only makes use of Dirac distributions).*

We fix a stochastic concurrent arena $\mathscr{A} = \langle \mathsf{States}, \mathsf{Agt}, \mathsf{Act}, (\mathsf{Allow}_i)_{i \in \mathsf{Agt}}, \mathsf{Tab} \rangle$ for the rest of this section. We say a state $s \in \mathsf{States}$ is *final* if $\mathit{Supp}(\mathsf{Tab}(s,A)) = \{s\}$ for all $A \in \mathsf{Act}^{\mathsf{Agt}}$ (that is, s is a sink state). The set of final states is denoted by $\mathtt{F}$. A *history* (resp. *run*) ρ in $\mathscr{A}$ is a finite non-empty (resp. infinite) sequence of states $s_0 s_1 s_2 \cdots \in \mathsf{States}^+$ (resp. $\in \mathsf{States}^\omega$) such that there are actions $A_1, A_2, \cdots \in \mathsf{Act}^{\mathsf{Agt}}$ with $s_i \in \mathsf{Tab}(s_{i-1}, A_i)$ for every $i \geq 1$. We denote by $\mathit{first}(\rho)$ (resp. $\mathit{last}(\rho)$, when relevant) the first (resp. last) state of ρ. We say that ρ is *terminating* whenever it visits a state in $\mathtt{F}$.

A *reward function* is a function that associates with any (infinite) run a real number. This function is *terminal-reward* whenever there exists a function $\nu \colon \mathtt{F} \to \mathbb{R}$ such that:

- any non-terminating run has reward 0;
- if ρ is a terminating run which visits $f \in \mathtt{F}$, then its reward is $\nu(f)$.

In this case, we write ϕ as ϕ_ν.

Definition 2. *A stochastic concurrent game* $\mathscr{G}$ *is a pair* $\langle \mathscr{A}, \phi \rangle$ *where* $\mathscr{A}$ *is a stochastic concurrent arena and* ϕ *associates with each player* $i \in \mathsf{Agt}$ *a reward function* ϕ_i*. The game has terminal-reward payoffs whenever each* ϕ_i *(*$i \in \mathsf{Agt}$*) is terminal-reward.*

2.2 Strategies and outcomes

During a play, players in Agt choose their next (distribution over) moves concurrently and independently of each other, based on the current history h of the play, and what they are allowed to do in the current state $last(h)$. This is given by strategies, that we define now.

Definition 3. *A* mixed strategy *for player $i \in \mathsf{Agt}$ is a mapping $\sigma_i \colon \mathsf{States}^+ \to Dist(\mathsf{Act})$, with the requirement that for all $h \in \mathsf{States}^+$, $Supp(\sigma_i(h)) \subseteq \mathsf{Allow}_i(last(h))$.*

Note that strategies, as defined above, can only observe the sequence of visited states along the history, but they may not depend on the exact distributions chosen by the players along the history, nor on the actual sequence of actions played by the players. Notice that this model is more general than the model where actions are visible, which are sometimes considered in the literature—see for instance [17] and [2, Section 6] or [5] for discussions—and the results presented here are valid when considering visible actions.

In this paper, we consider several subclasses of strategies:

- the set of *mixed strategies* of player i in arena $\mathscr{A}$, denoted $\mathbb{S}_i^{\mathscr{A}}$, is the set containing all the strategies of player i as defined above;
- the set of *pure strategies* of player i, denoted $S_i^{\mathscr{A}}$ contains those strategies in which all probability distributions are Dirac functions (that is, strategies are in some sense deterministic);
- the set of *stationary strategies* of player i, written $\mathbb{M}_i^{\mathscr{A}}$, in which the value of the strategy over history h only depends on $last(h)$;
- the set of *(pure) memoryless strategies*, denoted with $M_i^{\mathscr{A}}$, which contains the strategies that are pure and stationary.

A strategy profile is a tuple $\sigma = (\sigma_i)_{i \in \mathsf{Agt}}$, in which σ_i is a strategy for player i. Following the definitions introduced above, we consider the full class $\mathbb{S}^{\mathscr{A}}$ of mixed strategy profiles, the class $S^{\mathscr{A}}$ of pure strategy profiles, the class $\mathbb{M}^{\mathscr{A}}$ of stationary strategy profiles, and the class $M^{\mathscr{A}}$ of memoryless (that is, pure stationary) strategy profiles. If $\mathscr{A}$ is clear in the context, we will simplify the various notations and skip the superscript $\mathscr{A}$ in the notation.

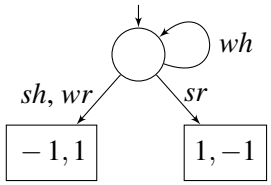
Let σ be a strategy profile. We denote by $\mathbb{P}^\sigma(-)$ the probability measure induced by σ on the infinite runs in States^ω as follows: the probability of cylinder $h\mathsf{States}^\omega$, with $h = s_1 \dots s_p$, is defined as $\mathbb{P}^\sigma(h \cdot \mathsf{States}^\omega) = \prod_{i=1}^{p} \sigma(h_{<i})(s_i)$, where $h_{<i}$ is the prefix of length $i-1$ of h (if $i = 1$, $h_{<i}$ is the empty word); it extends in a unique way to the σ-algebra generated by the above cylinders.

If $h \in \mathsf{States}^+$ is a history such that $\mathbb{P}^\sigma(h \cdot \mathsf{States}^\omega) > 0$, we define the conditional probability measure $\mathbb{P}^\sigma(- \mid h)$ in a natural way: $\mathbb{P}^\sigma(h' \cdot \mathsf{States}^\omega \mid h) = \frac{\mathbb{P}^\sigma(h' \cdot \mathsf{States}^\omega)}{\mathbb{P}^\sigma(h \cdot \mathsf{States}^\omega)}$ if h is a prefix of h' and $\mathbb{P}^\sigma(h' \cdot \mathsf{States}^\omega \mid h) = 0$ otherwise; this extends in a natural way to the generated σ-algebra. For any finite history $h' \in \mathsf{States}^+$, we write $\mathbb{P}^\sigma(h' \mid h)$ as a shorthand for $\mathbb{P}^\sigma(h' \cdot \mathsf{States}^\omega \mid h)$.

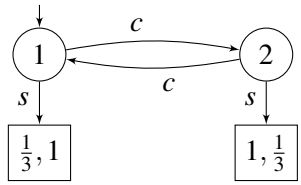
For every $i \in \mathsf{Agt}$, let ϕ_i be a terminal-reward reward function for player i, and define $\phi = (\phi_i)_{i \in \mathsf{Agt}}$. We denote by $\mathbb{E}^\sigma(\phi_i \mid h)$ the expected value of the reward function ϕ_i induced by the probability mesure $\mathbb{P}^\sigma(- \mid h)$. By extension, we write $\mathbb{E}^\sigma(\phi \mid h)$ for the tuple $(\mathbb{E}^\sigma(\phi_i \mid h))_{i \in \mathsf{Agt}}$

2.3 Nash equilibria

We now define the notion of Nash Equilibrium, as introduced by Nash [11].

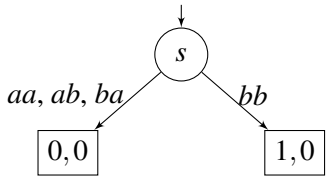


(a) Hide-or-run game



(b) The first player to quit the loop loses

Figure 2: Two examples of games with cycling behaviours



$\sigma_1(a \mid s) = 1; \quad \sigma_2(b \mid s) = \varepsilon$

Figure 3: A simple game

Definition 4. *A* Nash Equilibrium *from state s_0 is a (mixed) strategy profile $\sigma \in \mathbb{S}$ such that:*

$$\forall i \in \mathsf{Agt}\ \forall \sigma_i' \in \mathbb{S}_i \quad \mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s_0) \leq \mathbb{E}^{\sigma}(\phi_i \mid s_0).$$

where $\sigma[i/\sigma_i']$ is the strategy profile obtained from σ by replacing strategy σ_i for player i with σ_i'.

In this definition, strategy σ_i' corresponds to a *deviation* of player i with respect to the profile σ; we will often use this terminology thereafter.

Example 1. *Fig. 2 displays two examples of games that we will describe now. The hide-or-run game (on the left) represents a game where one player has one snowball and wants to shoot the other player; the second player is hiding, and wants to run to the other side of the road. The first player can either wait or shoot, while the second can hide or run. Label "sr" on a transition represents the concurrent action "s (shoot) for the first player and r (run) for the second player". The payoff is $(0,0)$ if the players keep on playing "wh" (loop on the initial state). The first player wins after "sr", and loses after "sh" and "wr" (represented by rewards $(1,-1)$ or $(-1,1)$). One can easily check that this game has no Nash equilibrium: if the probability to jointly take wh (resp. sr) is positive, then the second player can deviate and earn more with action pair wr (resp. sh); if the probability to jointly take sh (resp. wr) is positive, then the first player can deviate and earn more with action pair wh (resp. sr).*

The second game is turn-based, and numbers labelling nodes correspond to the players: in the left-most state, the first player can decide whether to stop (action s) or to continue (action c) playing the game; symmetrically for the second player in the right-most state. Again, the payoff is $(0,0)$ if the play does not reach a terminal state. This game has pure Nash equilibria: for instance, the memoryless strategy profile where player 1 plays c and player 2 plays s is an equilibrium, with payoff $(1,1/3)$. Another solution concept would allow a tradeoff between players who will commit a fixed probability each to exit the game (for example $\varepsilon > 0$). In general, such tradeoff is not a Nash equilibrium as the other player can change his mind (play c).

While one can compute pure (that is, deterministic) Nash equilibria in deterministic terminal-reward games [3], in the general case, computing mixed Nash equilibria in terminal-reward games is undecidable. Even for turn-based games, [17] proved that it is impossible to decide wheter a turn-based game with at least 14 players has a Nash equilibrium where one player wins almost surely (called 0-safe condition). This result was later improved by [8] to 0-safe equibria with finite memory and pure strategies in turn-based games with at least 5 players. In the concurrent setting, [18] showed the existence of a Nash equilibrium is undecidable for 14-player concurrent deterministic games using similar techniques, and when strategies do not observe the actions which are played (as in the current paper), the number of players can even be reduced to 3 ([4]). The 0-safe condition (one player should win) can be omitted in the concurrent setting, thanks to a gadget, composed of a 2-player zero-sum concurrent game having almost-optimal strategies but no optimal strategy, hence no Nash equilibrium (this is the first example mentioned

previously, and depicted on Fig. (2a)). If only non-negative terminal rewards are allowed, these undecidability results still hold in the concurrent setting, but under the additional 0-safe condition (there is no known game with no Nash equilibrium in this setting); indeed, the previous gadgets cannot be adapted as non-negative terminal rewards imply that every game is non-zero sum, then no player has an incentive to make the game cycling, ensuring global payoff 0, instead of reaching a terminal state. We summarize this discussion with the stronger undecidability result which applies to the precise setting of this paper.

Theorem 5 ([4])**.** *The existence problem of a Nash equilibrium in concurrent deterministic games with three players and terminal-reward payoff functions is undecidable.*

On the positive side, [6] showed that the relaxed notion of ε-Nash equilibrium, where deviations may only improve the payoffs by at most ε, always exists and can be computed. However, while the game of Fig. (2b) is very symmetric, there is no (ε-)Nash equilibrium (except the cycling one with payoff 0 for both players) where the two players have close payoffs. This is due to the discontinuity yielded by the pure deviation which consists in cycling; and if this pure strategy is not played precisely, there will actually be no improvement in the payoffs. We will therefore propose a new notion of equilibria where improvements by deviations should not come from a (punctual) discontinuity in the payoff function.

2.4 Equilibria under imprecise deviations

In this paper, we propose a new solution concept, with some *robustness* constraints on possible deviations, which will enjoy rather nice termination and continuity properties.

Definition 6. *An* equilibrium under ε-imprecise deviations *from state s_0 is a strategy profile $\sigma \in \mathbb{S}$ s.t.*

$$\forall i \in \mathsf{Agt}.\ \forall \sigma_i' \in \mathbb{S}_i.\ \exists \sigma_i'' \in \mathbb{S}_i \text{ s.t. } \mathbb{E}^{\sigma[i/\sigma_i'']}(\phi_i \mid s_0) \leq \mathbb{E}^{\sigma}(\phi_i \mid s_0) \text{ and } d(\sigma_i', \sigma_i'') \leq \varepsilon$$

where $d(\sigma, \sigma')$ is the supremum distance between the two distributions:

$$d(\sigma, \sigma') = \sup_{h \in \mathsf{States}^+} d(\sigma(h), \sigma'(h))$$

The intuition behind that definition is that, to have an incentive to deviate, a player should be sure to improve her payoff, even if her deviation is perturbed by ε (this corresponds to some noise the other players can add, or to a lack of precision in playing distributions). Said differently, a deviation is only considered profitable when all the surrounding (up to a distance of ε) strategies are also profitable.

We will prove that this new solution concept enjoys very nice properties: (a) for every $\varepsilon > 0$, equilibria under ε-imprecise deviations always exist, and (b) we can decide (and compute) such equilibria with constraints over the payoffs of the players.

Example 2. *Back to the first game in Example 1 (Fig. 2a). The strategy profile such that the first player plays s with proba 1 and player 2 plays r with probability ε is an equilibrium under ε-imprecise deviations with payoff $(2\varepsilon - 1, 1 - 2\varepsilon)$ (only the second player can deviate and improve, but its deviation will be smaller (w.r.t. the distance) than ε).*

In the second game in Example 1 (Fig. 2b). The strategy profiles where each player plays s with probability ε yields payoffs $1 - 2/(6-3\varepsilon)$ for player 1 and $1 - (2-2\varepsilon)/(6-3\varepsilon)$ for player 2 from the initial state. It is an equilibrium under ε-imprecise deviations. The only way to really improve the payoff for a player is to play with higher probability action c. But with the lack of precision, she might lose some payoff anyway. The payoff values get arbitrarily close to $2/3$ as ε goes to 0. Such an equilibrium is neither a Nash equilibrium, neither a ε-Nash equilibrium, since the pure deviation c allows an improvement of almost $1/3$.

Finally, consider the game of Fig. 3, and the strategy profile (σ_1,σ_2)*: the payoff is then* $(0,0)$*, and player* 1 *can improve her payoff by* ε *by playing action b from s. So* (σ_1,σ_2) *is an* ε*-Nash equilibrium but not an equilibrium under* ε*-imprecise deviations: any strategy at distance* ε *from* σ_1' *strictly improves the payoff of player* 1*. Thus we conclude that the two concepts are incomparable.*

Remark 1. *As we already noticed, equilibria under imprecise deviations are not Nash equilibria in the classical sense, but Nash equilibria are equilibria under imprecise deviations. So our notion relaxes that of Nash equilibria. Finally the concept of trembling-hand equilibria [14], already discussed in the introduction, is an orthogonal notion.*

3 Existence of equilibria under imprecise deviations

In this section, we prove the following existence result:

Theorem 7. *Let* $\mathscr{G}$ *be a stochastic concurrent game with terminal-reward payoffs, and let* s_0 *be a state of* $\mathscr{G}$*. For every* $\varepsilon > 0$*, there always exists an equilibrium under* ε*-imprecise deviations from state* s_0*.*

The proof will rely on an alternative notion of equilibria, where players are enforced to leave cycles of the game. We formalize this now, and we fix for the rest of this section a stochastic concurrent game with terminal-reward payoffs $\mathscr{G} = \langle \mathscr{A}, \phi_v \rangle$, with $\mathscr{A} = \langle \mathsf{States}, \mathsf{Agt}, \mathsf{Act}, (\mathsf{Allow}_i)_{i\in\mathsf{Agt}}, \mathsf{Tab} \rangle$

3.1 Non-cycling games

Definition 8. *A state* s *of* $\mathscr{A}$ *is said* cycling *if there exists a mixed strategy profile* $\sigma \in \mathbb{S}$ *such that no player can enforce (by deviating) reaching a final state, that is:*

$$\forall i \in \mathsf{Agt}\ \forall \sigma_i' \in \mathbb{S}_i,\ \mathbb{P}^{\sigma[i/\sigma_i']}(\mathsf{States}^*\mathtt{F}^\omega \mid s) = 0.$$

The arena $\mathscr{A}$ *(and by extension, the game* $\mathscr{G}$*) is said* cycle-free *if it contains no cycling state.*

We notice first that in the above definition, strategy profiles can be restricted to memoryless profiles ($\sigma \in M$), and deviations can be restricted to stationary deviations ($\sigma_i' \in \mathbb{M}_i$). Furthermore only the supports of these deviations matter.

We further notice that from any cycling state, there is a Nash equilibrium with payoff zero for all the players (playing profile σ from the definition). Those are also equilibria under imprecise deviations (since no payoff can be improved).

They are therefore somehow pathological behaviours, that we will remove. This is formalized as follows:

Proposition 9. *One can construct a cycle-free game* $\widetilde{\mathscr{G}} = \langle \widetilde{\mathscr{A}}, \phi_{\tilde{v}} \rangle$ *which has less Nash equilibria and less equilibria under imprecise deviations (whatever the bound on the imprecision): for every equilibrium (Nash, resp. under imprecise deviations)* $\widetilde{\sigma}$ *in* $\widetilde{\mathscr{G}}$*, one can build an equilibrium (Nash, resp. under imprecise deviations)* σ *with the same payoffs in* $\mathscr{G}$*.*

This proposition allows to prove Theorem 7 by restricting to cycle-free games: if the existence holds for cycle-free games, then it will hold as well for the whole class of stochastic concurrent games with terminal-reward payoffs.

3.2 Strong components and terminating strategy profiles

We will see that equilibria under imprecise deviations with stationary strategies always exist. The main argument of the existence theorem relies on the structure of the strategy profiles, that can be forced to terminate the game, even in the presence of deviations. We describe in this subsection a definition of the constraints we impose on our strategies. These constraints should be tight enough for the game to terminate, later implying the existence theorem of a *stable* profile, but should also be general enough for this same *stable* profile to capture the notion of equilibria under imprecise deviations.

Definition 10. *Let C be a non-empty set of states of $\mathscr{A}$, and $\sigma \in \mathbb{M}$ be a stationary strategy profile. We say that σ stabilizes C if for every $s \in C$, for every $s' \in \mathsf{States}$, $\mathbb{P}^{\sigma}(\mathsf{States}^* \cdot s' \mid s) > 0 \quad \textit{iff} \quad s' \in C$. When such a profile exists for C, we say that C is a* strong component*, and write* SC *the set of strong components.*

Notice that for defining the stabilization property, one could equivalently require the probability be equal to 1. Also notice that every strong component intersecting $\mathtt{F}$ is reduced to a singleton.

Definition 11. *Let $C \in \mathrm{SC}$ be a strong component, and $s \in C$. An action $a \in \mathsf{Act}$ is an* exiting action *from C for a state s and player i if there exists $\sigma \in \mathbb{S}$ which stabilizes C such that:*

$$\mathbb{P}^{\sigma[i/(s \mapsto a)]}(s \cdot (\mathsf{States} \setminus C) \cdot \mathsf{States}^{\omega} \mid s) > 0.$$

We set $\mathrm{Exit}(C) = \{(a,i,s) \mid a \text{ is an exiting action from } C \text{ for a state } s \text{ and player } i\}$.

We then trivially have:

Lemma 12. *If $\mathscr{A}$ is cycle-free, then for any $C \in \mathrm{SC}$, $\mathrm{Exit}(C) \neq \emptyset$.*

For the rest of this subsection, we will systematically assume that $\mathscr{A}$ is cycle-free.

We will now restrict the set of strategy profiles in which we search for equilibria. Under this restriction, each play will eventually reach a final state with probability 1. Nash equilibria restricted to this set of strategies will actually correspond to our modified notion of equilibria, in a sense that we will make precise.

Definition 13. *Let $\varepsilon > 0$ and assume $\mathscr{A}$ is cycle-free. For every strong component $C \in \mathrm{SC}$, we define the set of* (ε, C)-exiting stationary strategy profiles *as follows:*

$$\Delta_{\varepsilon}(C) = \{\sigma \in \mathbb{M} \mid \forall (a,i,s) \in \mathrm{Exit}(C) \sigma_i(s)(a) \geq \varepsilon\}$$

We also let $\Delta_{\varepsilon} = \bigcap_{C \in \mathrm{SC}} \Delta_{\varepsilon}(C)$.

Note that, to be properly defined and non-empty, Δ_{ε} requires the assumption that the game arena is cycle-free.

Lemma 14. *For all $\varepsilon \leq \frac{1}{|\mathsf{Act}|}$ and $\mathscr{A}$ cycle-free, it holds $\Delta_{\varepsilon} \neq \emptyset$.*

Proof. Consider the stationary strategy profile σ_u which makes each player play uniformly at random over the set of allowed actions, at each state.

For any $C \in \mathrm{SC}$, since $\mathrm{Exit}(C)$ is non-empty, this strategy profile is in $\Delta_{\varepsilon}(C)$. Hence $\sigma_u \in \Delta_{\varepsilon}$. □

The strategy profiles in Δ_{ε} enjoy the following property, which establishes some kind of fairness with respect to final states for strategies in Δ_{ε}. This will be useful in the sequel:

Proposition 15. *Fix $0 < \varepsilon \leq \frac{1}{|\mathsf{Act}|}$ and $\mathscr{A}$ cycle-free. There exist $0 < p < 1$ and $k \in \mathbb{N}$ such that for every $\sigma \in \Delta_{\varepsilon}$, for every $s \in \mathsf{States}$, for every $n \geq 0$, $\mathbb{P}^{\sigma}(\mathsf{States}^{k \cdot n} \cdot \mathtt{F}^{\omega} \mid s) \geq 1 - p^n$.*

3.3 Restricting to memoryless deviations

This part is devoted to the proof of the following key lemma:

Lemma 16. *Let s_0 be a state of a stochastic concurrent game $\mathscr{G}$ with terminal-reward payoffs. For any stationary strategy profile $\sigma \in \mathbb{M}$, it holds: σ is an equilibrium under ε-imprecise deviations iff*

$$\forall i \in \mathsf{Agt}.\ \forall \sigma_i' \in M_i.\ \exists \sigma_i'' \in \mathbb{M}_i.\quad d(\sigma_i', \sigma_i'') \leq \varepsilon \wedge \mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s_0) \leq \mathbb{E}^{\sigma}(\phi_i \mid s_0)$$

In other terms, it is sufficient to consider memoryless deviations when checking if a stationary strategy profile is an equilibrium under imprecise deviations.

We prove this lemma by considering an intermediate two-player game to represent deviations of Player i and their counter-deviations at distance ε.

The notion of equilibria under imprecise deviation has been introduced in a very general setting with arbitrarily complex strategies and deviations. An important step when proving existence of stationary equilibria is to check that one can restrict ourselves to deviations that are also stationary. Intuitively, one can even wonder if we can, as in the case of Nash Equilibria, only consider pure memoryless deviations, that will be imprecise up to ε, hence leading to stationary deviations, but in finite number.

Let $\mathscr{G}$ a game, σ a stationary strategy profile and $i \in \mathsf{Agt}$ a player. We write $\mathscr{G}\langle\sigma\rangle_{-i}$ for the 1-player game obtained from $\mathscr{G}$ by assigning to all players, but player i, her strategy in σ. Note that for any $\sigma_i' \in \mathbb{S}_i$, we have $\mathbb{E}_{\mathscr{G}}^{\sigma[i/\sigma_i']}(\phi_i \mid s) = \mathbb{E}_{\mathscr{G}\langle\sigma\rangle_{-i}}^{\sigma_i'}(\phi_i \mid s)$ In the following, we are mainly interested in the possible ε-imprecise deviations of player i alone in this new game.

In order to make the reduction clear, we consider in the following the particular case of games where each player is allowed at most two actions. When exactly two distinct actions are allowed, they will be noted a and b. The general case will be discussed in remark 2.

For a stationary profile σ, we consider the 1-player game $\mathscr{G}\langle\sigma\rangle_{-i}$ as defined above (with Player i alone, all other strategies being fixed) and construct a 2-player turn-based game with an additional antagonistic Player $\hat{i}$, whose role is to "change" the strategy of Player i by a distance at most ε. Formally, for any state s where Player i has two allowed actions a and b (resulting in distributions $\delta(s,a)$ and $\delta(s,b)$, resp.), we modify the game as follows:

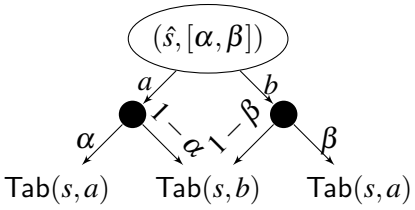
- from s, Player i is given the opportunity to move to one of the following four states: $(s,[0,\varepsilon])$, $(s,[0,2\varepsilon])$, $(s,[1-2\varepsilon,1])$ and $(s,[1-\varepsilon,1])$.
- from each state $(s,[\alpha,\beta])$, Player $\hat{i}$ has two actions, leading to distributions $\alpha\cdot\mathsf{Tab}(s,a)+(1-\alpha)\cdot\mathsf{Tab}(s,b)$ and $\beta\cdot\mathsf{Tab}(s,a)+(1-\beta)\cdot\mathsf{Tab}(s,b)$, respectively. If Player $\hat{i}$ plays action a with probability p, then the final distribution is $[p\alpha+(1-p)\beta]\cdot\mathsf{Tab}(s,a)+[(1-p)(1-\alpha)+p(1-\beta)]\cdot\mathsf{Tab}(s,b)$.

For a 1-player game $\mathscr{G}$ for i, we denote by $\widehat{\mathscr{G}^{\varepsilon}}$ the previous transformation. Our aim is to have a correspondence between (stochastic) moves of Player i from s in $\mathscr{G}$, and her move from the corresponding state $\hat{s}$ in $\widehat{\mathscr{G}^{\varepsilon}}$. Our notion of correspondence is defined as follows:

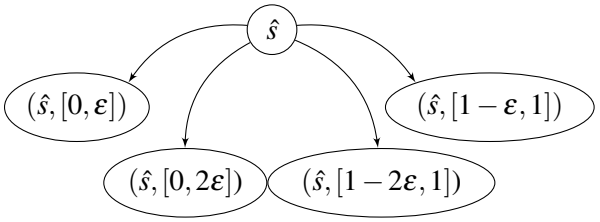
Definition 17. *Let $\sigma_i, \sigma_i' \in \mathbb{S}$ two strategies for the 1-player game $\mathscr{G}$ (played by i) such that $d(\sigma_i,\sigma_i') \leq \varepsilon$, and $\hat{\sigma}$ a strategy profile in $\widehat{\mathscr{G}^{\varepsilon}}$. We say that (σ_i,σ_i') corresponds to $\hat{\sigma}$ if the following holds for any history $\hat{h}$ ending in state s of $\widehat{\mathscr{G}^{\varepsilon}}$:*

$$\mathsf{Tab}(s,\sigma_i'(\pi_{\mathsf{States}}(\hat{h}))) \equiv \widehat{\mathsf{Tab}}(s,\hat{\sigma}(\hat{h}))$$

where $\pi_{\mathsf{States}}(h)$ the projection on the letters corresponding to the original states States.



(a) Simple $(\hat{s},I)$ node with I interval of $[0,1]$ played by $\hat{i}$, ensuring any distribution $p\mathsf{Tab}(s,a)+(1-p)\mathsf{Tab}(s,b)$ for any $p \in I$, hence a *deviation range I*.



(b) Replaced node s, where i can choose between 4 ranges of probabilities to play action a.

Figure 4: Translation of a node s with allowed action a and b to $\widehat{a}$.

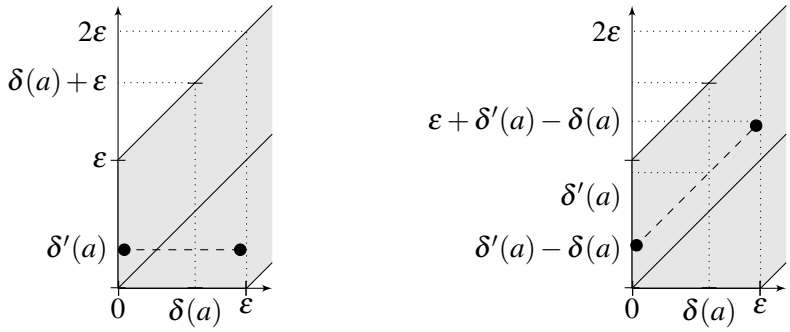


Figure 5: Intuition of the construction for $\delta(a) \leq \varepsilon$: seeing $\delta(a)$ as a convex combination of 0 and ε, we obtain $\delta'(a)$ as the same convex combination of the black dots.

We now explicit explicit the purpose of the construction by establishing a correspondence between strategies in the original game and strategies in our 2-player version.

Lemma 18. *For any σ_i strategy of $\mathscr{G}$, there exists a strategy $\hat{\sigma}_i$ in $\widehat{\mathscr{G}^{\varepsilon}}$ for player i, such that, for any strategy σ_i' of $\mathscr{G}$ such that $d(\sigma_i,\sigma_i') \leq \varepsilon$, there exists $\hat{\sigma}_{\hat{i}}$ such that (σ_i,σ_i') corresponds to $\hat{\sigma}$.*

Moreover, any pure memoryless strategy profile of $\widehat{\mathscr{G}^{\varepsilon}}$ corresponds to some pair of strategies (σ_i,σ_i') in $\mathscr{G}$ where σ_i is pure memoryless and σ_i' is stationary.

The constructed game is a turn-based stochastic game with a quantitative terminal reachability objective, which can be interpreted as a special case of limit-average objective. Hence, thanks to a result of [16], such a game is determined with pure memoryless optimal strategies for both players.

As a consequence of this construction, we can infer two possible characterizations of imprecise deviations in stationary profiles:

Corollary 19. *The value of $\widehat{\mathscr{G}^{\varepsilon}}$ at state $\widehat{s}$ can be expressed as the following quantity on game $\mathscr{G}$:*

$$\sup_{\sigma \in M^{\mathscr{G}}} \inf_{\substack{\sigma' \in \mathbb{M}^{\mathscr{G}} \\ d(\sigma,\sigma') \leq \varepsilon}} \mathbb{E}^{\sigma'}(\phi_i \mid s)$$

Corollary 20. *Let $\sigma \in \mathbb{M}^{\mathscr{G}}$ a stationary strategy profile in $\mathscr{G}$. σ is an equilibrium under ε-imprecise deviations from state s_0, if and only if:*

$$\forall i \in \mathsf{Agt}.\ \forall \sigma_i' \in M_i^{\mathscr{G}}.\ \exists \sigma_i'' \in \mathbb{M}_i^{\mathscr{G}}\ s.t.\ \mathbb{E}^{\sigma[i/\sigma_i'']}(\phi_i \mid s_0) \leq \mathbb{E}^{\sigma}(\phi_i \mid s_0)\ \text{and}\ d(\sigma_i',\sigma_i'') \leq \varepsilon$$

Remark 2. *One can notice the construction of the deviation game and inferred results have been applied to nodes with two allowed actions only. In fact, the same reasoning can be generalized to an arbitrary number of allowed actions at the expense of an exponential blowup: player i has to announce simultaneously, for each allowed action a, if its probability in the expected distribution will be larger than ε and/or smaller than $1-\varepsilon$. Note however that for a given fixed bound on the number of actions, the size of $\widehat{\mathscr{G}}^{\varepsilon}$ is still polynomial.*

3.4 Existence of equilibria under imprecise deviations

We are now ready to prove Theorem 7, that is, for every $\varepsilon > 0$, the existence of a (stationary) equilibrium under ε-imprecise deviations from any state of stochastic concurrent games with terminal-reward payoffs.

Our proof will rely on the following well-known fixed-point theorem, that we will apply to a well-adapted sets of strategy profiles.

Theorem 21 ([10])**.** *Let X be a non-empty, compact and convex subset of some Euclidean space. Let $f\colon X \to 2^X$ be a set-valued function on X with a closed graph and the property that $f(x)$ is non-empty and convex for all $x \in X$. Then f has a fixed point.*

A Nash equilibrium σ can be characterized as containing, for each player i, the best response σ_i to the strategies of the other players. This can be expressed as a fixed point of the *best-response function* ([11]). Nevertheless, over game graphs, continuity of this best-response function is not ensured. More precisely, the graph of the function is not closed. Let us consider for example game of Figure 2b, and write any stationary strategy profile σ in this game as the tuple $(\sigma_1(s \mid 1), \sigma_2(s \mid 2))$. Then, if one player decides to stop the game with any positive probability, the other player has all incentive to purely continue the game, until reaching the terminal state (with probability), hence: $\mathrm{BR}((x,y)) = \{(0,0)\}$ for every $x, y > 0$, where BR denotes the best-response function. However, if the other player purely continues the game, the only way to win some positive payoff $1/3$ is to play the stopping action with positive probability, hence: $\mathrm{BR}((0,0)) = \{(x,y) \mid x, y > 0\}$. We conclude that the graph is not closed, so Theorem 21 cannot apply to the classical BR function. This is not surprising as we know that Nash equilibria need not always exist (recall the example given in Figure 2a). On the other hand, in [6], stationary ε-Nash equilibria are characterized as fixed points of the best-response function.

In the following we will see that the (standard) best-response function will fit well in our setting.

Definition 22. *We consider $T \subseteq \mathbb{M}$ a subset of stationary strategy profiles. Let $\mathrm{BR}_T : T \to 2^T$ with*

$$\mathrm{BR}_T(\sigma) = \left\{ \sigma' \in T \;\middle|\; \forall i \in \mathsf{Agt}.\ \forall s \in \mathsf{States}.\ \sigma'_i \in \mathrm{argmax}_{\sigma''_i \text{ s.t. } \sigma'[i/\sigma''_i] \in T} \mathbb{E}^{\sigma[i/\sigma''_i]}(\phi_i \mid s) \right\}$$

Note that $\mathrm{BR}_{\mathbb{M}}$ is the usual notion of best response function.

Lemma 23. *For every $0 < \varepsilon \leq \frac{1}{|\mathsf{Act}|}$ and $\mathscr{A}$ cycle-free, $\mathrm{BR}_{\Delta_\varepsilon}$ has a fixed point.*

Proof. We apply Theorem 21.

- First notice that $T = \Delta_\varepsilon$ can be viewed as a non-empty compact convex subset of $\mathbb{R}^N$ where $N = \mathsf{Act} \times \mathsf{Agt} \times \mathsf{States}$. Moreover, T can be decomposed in a product of individual strategy sets for each player $T = T_1 \times \ldots T_{|\mathsf{Agt}|}$ where

$$\forall i \in \mathsf{Agt}\ T_i = \{\sigma_i \mid \forall (a,s)\ (a,i,s) \in \mathrm{Exit}(C) \Rightarrow \sigma_i(s)(a) \geq \varepsilon\}$$

  Hence, for every $(\sigma, \sigma') \in T^2$, and $i \in \mathsf{Agt}$, we still have $\sigma[i/\sigma'_i] \in T$.

- Let k and p be the constants appearing in the statement of Proposition 15. For every $n \geq 0$, we define g_n for the function assigning to every pair of strategy profiles $(\sigma, \sigma') \in T^2$ the following vector value in $\mathbb{R}^{\mathsf{Agt} \times \mathsf{States}}$:

$$\left(\sum_{j=0}^{k \cdot n} \sum_{f \in \mathtt{F}} \mathbb{P}^{\sigma[i/\sigma_i']}\big((\mathsf{States} \setminus \mathtt{F})^j \cdot f^\omega \mid s\big) \cdot \nu_i(f) \right)_{i \in \mathsf{Agt}, s \in \mathsf{States}}$$

  Then, we obviously see that for every $(i,s) \in \mathsf{Agt} \times \mathsf{States}$, $\lim_{n\to\infty} g_n(\sigma,\sigma')_{i,s} = \mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s)$. Furthermore, as an application of Proposition 15, we get:

$$|\mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s) - g_n(\sigma,\sigma')_i| \leq K \cdot p^n$$

  where $K = \max_{i \in \mathsf{Agt}, f \in \mathtt{F}} |\nu_i(f)|$. This implies that the above convergence is indeed uniform, and that $g_\infty : (\sigma,\sigma') \mapsto \left(\mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s)\right)_{i,s}$ is therefore continuous on T^2.

- Let us now show that the graph of BR_T is closed. In order to do so, we consider a converging sequence of strategy profiles $(\sigma^k)_{k>0}$ with limit σ^∞ and for each $k > 0$, $\sigma'^k \in \mathrm{BR}_T(\sigma^k)$ converging to σ'^∞. We will prove that $\sigma'^\infty \in \mathrm{BR}_T(\sigma^\infty)$. For a fixed σ', we have $\mathbb{E}^{\sigma^k[i/\sigma_i']}(\phi_i \mid s) \leq \mathbb{E}^{\sigma^k[i/\sigma_i'^k]}(\phi_i \mid s)$, hence by continuity, $\mathbb{E}^{\sigma^\infty[i/\sigma_i']}(\phi_i \mid s) \leq \mathbb{E}^{\sigma^\infty[i/\sigma_i'^\infty]}(\phi_i \mid s)$.

- It remains to show that $\mathrm{BR}_T(\sigma)$ is convex. We fix $i \in \mathsf{Agt}$ and show that $(\mathrm{BR}_T(\sigma))_i$ is convex hence the result. Let $0 < \lambda < 1$ and $\sigma', \sigma'' \in \mathrm{BR}_T(\sigma)$: this means that both vectors $(\mathbb{E}^{\sigma[i/\sigma_i']}(\phi_i \mid s))_s$ and $(\mathbb{E}^{\sigma[i/\sigma_i'']}(\phi_i \mid s))_s$ are maximal, and equal to some vector m_i. Indeed, if two different maximal vectors exists, we take the combined strategy that uses best action in each state, this new strategy is still in T_i.

  By convexity of $T = \Delta_\varepsilon$, $\sigma^\lambda = \sigma[i/\lambda \cdot \sigma_i' + (1-\lambda) \cdot \sigma_i''] \in T$, so $\forall s$, $\mathbb{P}^{\sigma^\lambda}(\mathsf{States}^* \mathtt{F} \mid s) = 1$. This implies that the payoff vector $(\mathbb{E}^{\sigma^\lambda}(\phi_i \mid s))_s$ is the *unique* solution of the equation

$$\begin{cases} \forall f \in \mathtt{F} \quad \mathbb{E}^{\sigma^\lambda}(\phi_i \mid f) = \nu_i(f) \\ \forall s \notin \mathtt{F} \quad \mathbb{E}^{\sigma^\lambda}(\phi_i \mid s) = \sum_{s'} \mathsf{Tab}(s, \sigma^\lambda(s))(s') \mathbb{E}^{\sigma^\lambda}(\phi_i \mid s') \\ \qquad\qquad\qquad\quad = \sum_{s'} \left[\lambda\, \mathsf{Tab}(s, \sigma[i/\sigma_i'](s)) + (1-\lambda)\mathsf{Tab}(s, \sigma[i/\sigma_i''](s))\right](s') \cdot \mathbb{E}^{\sigma^\lambda}(\phi_i \mid s') \end{cases}$$

  On the other hand, m_i satisfies the following equation:

$$\begin{cases} \forall f \in \mathtt{F} \quad m_{i,f} = \nu_i(f) \\ \forall s \notin \mathtt{F} \quad m_{i,s} = \sum_{s'} \mathsf{Tab}(s, \sigma[i/\sigma_i'](s))(s') m_{i,s'} = \sum_{s'} \mathsf{Tab}(s, \sigma[i/\sigma_i''](s))(s') m_{i,s'} \end{cases}$$

  We can check that $(\mathbb{E}^{\sigma^\lambda}(\phi_i \mid s))_{s \in \mathsf{States}} = m_i$ is a valid solution, hence the actual value, so $\sigma_i^\lambda \in \mathrm{BR}_T(\sigma)_i$. □

Thanks to Corollary 20 (stationary deviations), and this fixed-point theorem, we infer the following proposition:

Proposition 24. *If $0 < \varepsilon \leq \frac{1}{|\mathsf{Act}|}$ and $\mathscr{A}$ is cycle-free, then there exists $\sigma \in \Delta_\varepsilon$ fixed point of $\mathrm{BR}_{\Delta_\varepsilon}$ which is an equilibrium under ε-imprecise deviations from every state s of $\mathscr{G}$.*

The general Theorem 7 follows immediately for any $\varepsilon > 0$ and any arena, thanks to Proposition 9.

4 Computing stationary equilibria under imprecise deviations

We describe a polynomial-space algorithm for computing stationary equilibria under imprecise deviations for non-negative terminal reward games. A similar proof for Nash equilibria in turn-based stochastic games is given in [19]. We briefly describe the later proof, which will help understanding our current encoding.

The algorithm proceeds by encoding a Nash Equilibrium as an existential first-order formula over the reals, which satisfiability can be decided in PSPACE. The formula quantifies over all stationary strategy profiles and payoffs at each state, and checks that:

1. the strategy profile σ under consideration is properly defined;
2. the payoff in each state corresponds to the real payoff of the strategy profile;
3. for any i, Player i cannot benefit from deviating in $\mathscr{G}\langle\sigma\rangle_{-i}$.

These properties cannot, *in general*, be expressed locally, but in the setting of [19], one can first, non-deterministically, guess the support of the strategy. On the one hand, this allows us to compute (in linear time) the set of states from which F is never reached. Those states have payoff 0 for all agents, and the payoff in the other states (from which F is reachable with some positive probability) can be expressed as a combination of the payoff values of the successor states and the (local) strategy profile. On the other hand, we can also compute (still in linear time) the set of states that are reachable from s_0. It is easy to see that Player i has an incentive to deviate if, and only if, her payoff can be increased by deviating locally from such a reachable state. Hence we can express stability of the Nash Equilibrium as a (polynomial size) conjunction of inequalities.

Another way of expressing this stability property is by saying that for any Player i, s_0 should yield a payoff in the equilibrium that is larger than the optimal value $v_i(s_0)$ in the Markov decision process representing the possible deviations of Player i, namely $\mathscr{G}\langle\sigma\rangle_{-i}$. Since the initial guess can be done in NPSPACE and the generated formula is of polynomial size, the whole algorithms runs in PSPACE.

In the case of equilibria under ε-imprecise deviations, we apply a similar technique but deviations are now to be considered as strategies for Player i in $\widehat{\mathscr{G}\langle\sigma\rangle_{-i}}^{\varepsilon}$ against the worst strategies of Player $\widehat{i}$. In fact, we want to check that s_0 has a payoff (in the equilibrium) larger for player i than the maximal value she could get by imprecisely deviating. Thanks to corollary 19, this optimal value is the same as in $\mathscr{G}_i = \widehat{\mathscr{G}\langle\sigma\rangle_{-i}}^{\varepsilon}$, denoted by $v_{\varepsilon,i}(s)$. In order to compute these values for each game $\mathscr{G}_i$, we non-deterministically compute optimal strategies for players i and $\hat{i}$. These strategies can be supposed to be pure memoryless. In order to do so, we first guess a strategy for Player i in the game game $\widehat{\mathscr{G}\langle\sigma\rangle_{-i}}^{\varepsilon}$. Without knowing the exact probability values of this game (which depends on σ), we can still derive its structure since the support is known, thus we can compute the set of states for which Player $\hat{i}$ can totally spoil i's payoff, that is, enforce a non-terminating run; such a run has payoff 0, which is optimal for Player $\hat{i}$. We later guess a pure memoryless strategy for Player $\hat{i}$ keeping in mind that $\hat{i}$ has to play such a cycling strategy from any state where she is able to. From the other states, for which Player i can still ensure positive probability to terminate, the value of the game can again be expressed locally as a combination of the guessed strategy profile and the values of the successor states. As for the previous algorithm for Nash Equilibrium in $\mathscr{G}$, the optimality of both strategies can be expressed as stability by local deviations. Finally, stability by imprecise deviations in $\mathscr{G}$ consists in coding the fact that payoff in $\mathscr{G}$ for Player i should be larger than the optimal value $v_{\varepsilon,i}(s_0)$.

We now make precise the result and the algorithm.

Theorem 25. *Let $k > 0$. Let $\mathscr{G} = \langle \mathscr{A}, \phi_\nu \rangle$ be a stochastic concurrent game with non-negative terminal rewards with $|\mathsf{Act}| \leq k$. Let $s_0 \in \mathsf{States}$ and $\varepsilon > 0$. For every $i \in \mathsf{Agt}$, we fix $x_i, y_i \in \mathbb{R}_+$ two real numbers. We can decide in* PSPACE *whether there is a stationary equilibrium under ε-imprecise deviations σ from s_0, such that for every $i \in \mathsf{Agt}$, $x_i \leq \mathbb{E}^\sigma(\phi_i \mid s_0) \leq y_i$.*

Remark 3. *The previous theorem can be applied to compute* some *equilibria in the case of negative payoffs by considering the new payoff function $\nu' = \nu - \min \nu \geq 0$. However, $\phi' = \phi_\nu - \min \nu$ and $\phi_{\nu'}$ coincide only on runs that reach a final state since ϕ' assigns positive value $-\min \nu$ to non-terminating runs. A possible work-around is to first compute the cycle-free arena $\widetilde{\mathscr{A}}$ and exiting conditions Δ_ε, which size is bounded by the number of pairs $(a,i,s) \in \mathsf{Act} \times \mathsf{Agt} \times \mathsf{States}$. Then we can apply the previous theorem on game $\langle \widetilde{\mathscr{A}}, \phi_{\nu'} \rangle$ with the extra formula $\sigma \in \Delta_\varepsilon$. Thanks to this last constraint, we ensure that the run always terminates, thus the payoff functions coincide. Finally we conclude the computation by applying proposition 9 to get back an equilibrium on $\mathscr{G}$.*

References

[1] D. Auger & O. Teyraud (2012): *The Frontier of Decidability in Partially Observable Recursive Games*. Int. Journal of Foundations of Computer Science 23(7), pp. 1439–1450, doi:10.1142/S0129054112400576.

[2] P. Bouyer, R. Brenguier, N. Markey & M. Ummels (2011): *Nash Equilibria in Concurrent Games with Büchi Objectives*. In: Proc. 30th Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'11), LIPIcs 13, Leibniz-Zentrum für Informatik, pp. 375–386, doi:10.4230/LIPIcs.FSTTCS.2011.375.

[3] P. Bouyer, R. Brenguier, N. Markey & M. Ummels (2015): *Pure Nash Equilibria in Concurrent Games*. Logical Methods in Computer Science 11(2:9), doi:10.2168/LMCS-11(2:9)2015.

[4] P. Bouyer, N. Markey & D. Stan (2014): *Mixed Nash Equilibria in Concurrent Games*. In: Proc. 33rd Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'14), LIPIcs 29, Leibniz-Zentrum für Informatik, pp. 351–363, doi:10.4230/LIPIcs.FSTTCS.2014.351.

[5] K. Chatterjee & L. Doyen (2014): *Partial-Observation Stochastic Games: How to Win when Belief Fails*. ACM Transactions on Computational Logic 15(2:16), doi:10.1145/2579821.

[6] K. Chatterjee, M. Jurdziński & R. Majumdar (2004): *On Nash Equilibria in Stochastic Games*. In: Proc. 18th Int. Workshop on Computer Science Logic (CSL'04), LNCS 3210, Springer, pp. 26–40, doi:10.1007/978-3-540-30124-0_6.

[7] T. Chen, M. Kwiatkowska, D. Parker & A. Simaitis (2011): *Verifying Team Formation Protocols with Probabilistic Model Checking*. In: Proc. 12th Int. Workshop on Computational Logic in Multi-Agent Systems (CLIMA'11), LNAI 6814, Springer, pp. 190–207, doi:10.1007/978-3-642-22359-4_14.

[8] A. Das, S. Krishna, L. Manasa, A. Trivedi & D. Wojtczak (2015): *On Pure Nash Equilibria in Stochastic Games*. In: Theory and Applications of Models of Computation, LNCS 9076, Springer, pp. 359–371, doi:10.1007/978-3-319-17142-5_31.

[9] Thomas A. Henzinger (2005): *Games in System Design and Verification*. In: Proceedings of the 10th Conference on Theoretical Aspects of Rationality and Knowledge, TARK '05, National University of Singapore, Singapore, Singapore, pp. 1–4. Available at `http://doi.acm.org/10.1145/1089933.1089935`.

[10] S. Kakutani (1941): *A generalization of Brouwer's fixed point theorem*. Duke Mathemastical Journal 8(3), pp. 457–459, doi:10.1215/S0012-7094-41-00838-4.

[11] J.F. Nash (1950): *Equilibrium Points in n-Person Games*. Proceedings of the National Academy of Sciences of the United States of America 36(1), pp. 48–49, doi:10.1073/pnas.36.1.48.

[12] P. Secchi & W.D. Sudderth (2001): *Stay-in-a-Set Games*. Int. Journal of Game Theory 30, pp. 479–490, doi:10.1007/s001820200092.

[13] R. Selten (1965): *Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit*. Zeitschrift für die gesamte Staatswissenschaft 121(2), pp. 301–324 and 667–689. Available at `http://www.jstor.org/stable/40748884`.

[14] R. Selten (1975): *A reexamination of the perfectness concept for equilibrium points in extensive games*. Int. Journal of Game Theory 4, pp. 25–55, doi:10.1007/BF01766400.

[15] W. Thomas (2002): *Infinite Games and Verification*. In: Proc. 14th Int. Conf. on Computer Aided Verification (CAV'02), LNCS 2404, Springer, pp. 58–64, doi:10.1007/3-540-45657-0_5. Invited Tutorial.

[16] S.A. Lippman T.M. Liggett (1969): *Short Notes: Stochastic Games With Perfect Information and Time Average Payoff*. SIAM Review 11(4), pp. 604–607, doi:10.1137/1011093.

[17] M. Ummels (2008): *The Complexity of Nash Equilibria in Infinite Multiplayer Games*. In: Proc. 11th Int. Conf. on Foundations of Software Science and Computation Structures (FoSSaCS'08), LNCS 4962, Springer, pp. 20–34, doi:10.1007/978-3-540-78499-9_3.

[18] M. Ummels & D. Wojtczak (2011): *The Complexity of Nash Equilibria in Limit-Average Games*. In: Proc. 22nd Int. Conf. on Concurrency Theory (CONCUR'11), LNCS 6901, Springer, pp. 482–496, doi:10.1007/978-3-642-23217-6_32.

[19] M. Ummels & D. Wojtczak (2011): *The Complexity of Nash Equilibria in Stochastic Multiplayer Games*. Logical Methods in Comp. Science 7(3), doi:10.2168/LMCS-7(3:20)2011.