TWO-PLAYER NON-ZERO-SUM GAMES: A REDUCTION

NICOLAS VIEILLE
HEC School of Management
Jouy-en-Josas, France

1. Introduction

In this chapter, we focus on finite two-player non-zero-sum stochastic games. Following the notations used in earlier chapters, we let S be the state space, and A and B be the action sets of players 1 and 2 respectively. All three sets S, A and B are finite. Generic elements of S, A and B will be denoted by z, a and b. We let $p(\cdot|z, a, b)$ be the transition function of the game and $r : S \times A \times B \to \mathbb{R}^2$ be the (stage) payoff function of the game. We deal with games with complete information, perfect recall and perfect monitoring. Thus, at each stage $n \geq 0$, the two players know the play $h_n = (z_1, a_1, b_1, ..., z_n)$ up to that stage, and simultaneously choose actions a_n and b_n. The next state z_{n+1} is drawn according to $p(\cdot|z_n, a_n, b_n)$ and the play proceeds to the next stage.

The goal of this chapter, together with the next one, is to give an overview of the proof of Theorem 1.

Theorem 1 *Any finite two-player stochastic game has a uniform equilibrium payoff.*

The proof partitions the state space S into three kinds of (disjoint) regions: solvable sets, controlled sets and remaining states. We devise an ε-equilibrium (σ, τ) that coincides with a simple (periodic-like) profile (σ^*, τ^*) until at least one player detects a deviation from the equilibrium play. From that stage on, a punishment threat is activated, i.e., each player switches to an ε-optimal strategy in a related zero-sum game.

Under (σ^*, τ^*), the play eventually reaches a solvable set and remains there forever. In each controlled set, (σ^*, τ^*) is of the type discussed in [4], Section 5. Outside of the solvable sets and of the controlled sets, (σ^*, τ^*) coincides with a stationary profile.

We shall only describe the structure of the proof, and provide selected details. For the complete proof, we refer to [8],[9].

We define solvable sets in Section 2. Section 3 explains the division of the proof into two parts, and describes the structure of the first part. The last two sections contain details.

Given $C \subset S$, we let $\theta_C := \inf\{n : z_n \notin C\}$ denote the first exit stage from C. Given a measure μ, expectations with respect to μ are denoted by $\mathbf{E}_\mu$. We abbreviate $\mathbf{E}_{p(\cdot|z,\alpha_z,\beta_z)}[\cdot]$ to $\mathbf{E}[\cdot|z,\alpha_z,\beta_z]$.

2. Solvable Sets

We let Γ be any finite two-player stochastic game. We denote by Γ^i, $i = 1, 2$ the zero-sum game obtained from Γ when the other player minimizes player i's payoff and we let $v^i : S \to \mathbb{R}$ be the uniform value of Γ^i.

2.1. DEFINITION

Let (α, β) be a stationary profile, and R be a recurrent set for the Markov chain over S induced by (α, β). By the ergodic theorem for Markov chains, $\lim_{n\to+\infty} \gamma_n(z, \alpha, \beta)$ exists and is independent of $z \in R$. We simply write $\gamma(R, \alpha, \beta)$. For each $C \subseteq S$, we denote by $\mathcal{R}_{\mathcal{C}}(\alpha, \beta)$ the (possibly empty) collection of sets $R \subseteq C$ such that R is recurrent under (α, β).

Definition 2 *A* solvable set *is a triple* $(C, (\alpha, \beta), \mu)$ *where (i)* (α, β) *is a stationary profile, (ii)* $C \subseteq S$ *is a weak communicating set w.r.t.* (α, β)*, and (iii)* μ *is a probability distribution over* $\mathcal{R}_{\mathcal{C}}(\alpha, \beta)$ *such that, for every* $z \in C$*,* $a \in A$ *and* $b \in B$*,*

$$\begin{array}{l} \sum_{R\in\mathcal{R}_{\mathcal{C}}(\alpha,\beta)} \mu(R)\gamma^1(R, \alpha, \beta) \geq \mathbf{E}\left[v^1|z, a, \beta_z\right] \\ \sum_{R\in\mathcal{R}_{\mathcal{C}}(\alpha,\beta)} \mu(R)\gamma^2(R, \alpha, \beta) \geq \mathbf{E}\left[v^2|z, \alpha_z, b\right]. \end{array}$$

Weak communicating sets are defined in [4], Definition 4.2. The concept of solvable set is slightly more general than the concept of *easy initial states*, introduced by Vrieze and Thuijsman [6]. For simplicity, we abuse definitions and say that $C \subseteq S$ is a solvable set as soon as $(C, (\alpha, \beta), \mu)$ is a solvable set for some (α, β) and μ.

2.2. MAIN PROPERTIES

Solvable sets do exist (see [6] or [7] for two players, and [10] for a generalization to N-player games). Existence for two-player games is also a corollary to Proposition 8 below.

Lemma 3 *Let $(C,(\alpha,\beta),\mu)$ be a solvable set. Then*

$$\sum_{R\in\mathcal{R}_C(\alpha,\beta)}\mu(R)\gamma(R,\alpha,\beta)$$

is a uniform equilibrium payoff, provided the initial state z belongs to C.

The players will play history-dependent perturbations of (α,β) such that the average payoff converges to $\sum_{R\in\mathcal{R}_C(\alpha,\beta)}\mu(R)\gamma(R,\alpha,\beta)$. The properties of a solvable set ensure that a unilateral deviation, followed by indefinite punishment, is not profitable.

Proof. Let $\varepsilon>0$ and $z\in C$ be given. Since C is a weak communicating set w.r.t. (α,β), one can construct a profile (σ^*,τ^*) such that (i) $\|(\sigma^*(h_n),\tau^*(h_n))-(\alpha_{z_n},\beta_{z_n})\|<\varepsilon$ for every history h_n, (ii) $\mathbf{P}_{z,\sigma^*,\tau^*}(\theta_C<+\infty)=0$ and (iii) $\lim_{n\to+\infty}\overline{r}_n=\sum_{R\in\mathcal{R}_C(\alpha,\beta)}\mu(R)\gamma(R,\alpha,\beta), \mathbf{P}_{z,\sigma^*,\tau^*}$-a.s. Adding to (σ^*,τ^*) appropriate punishment threats yields a 2ε-equilibrium. ∎

Remark 4 *Lemma 3 holds as soon as current states and payoffs are observed by the players.*

We now proceed to a first modification of the game. For each solvable set $(C,(\alpha,\beta),\mu)$ and each $z\in C$, turn z into an absorbing state with payoff $\sum_{R\in\mathcal{R}_C(\alpha,\beta)}\mu(R)\gamma(R,\alpha,\beta)$.[1] Denote by Γ' the game obtained this way. It can be checked that: (i) the solvable sets of Γ' coincide with the absorbing states of Γ and (ii) each equilibrium payoff of Γ' is an equilibrium payoff of Γ. The converse to (ii) does not hold.

3. Overview

Let $\Gamma=(S,A,B,r,p)$ be a game, and $\varepsilon>0$. By the previous section, we may assume that the solvable sets of Γ coincide with the absorbing states. W.l.o.g., we assume that $r^1(\cdot)<0<r^2(\cdot)$. We denote by $S^*\subseteq S$ and $S_0=S\backslash S^*$ the subsets of absorbing and of non-absorbing states respectively. A profile (σ,τ) is *ε-absorbing* if $\mathbf{P}_{z,\sigma,\tau}(\theta_{S_0}<+\infty)\geq 1-\varepsilon$ for each $z\in S$. Given such a profile, the play reaches S^* with high probability.

3.1. A FEW DEFINITIONS

Definition 5 *A* controlled set *is a pair (C,Q), where $C\subseteq S_0$, and Q is an exit distribution from C that is controllable for* any *payoff vector $\gamma\geq v$.*

[1]For states z that belong to several solvable sets, choose any of the corresponding payoffs.

Controllable exit distributions are introduced in [4], Definition 5.1. For simplicity, we sometimes refer to the set C itself as a *controlled set*.

Given a controlled set (C, Q), we let Γ_C be the game obtained from Γ by turning the subset C into a dummy state $\{C\}$, where transitions are specified by Q, irrespective of actions being played, and payoffs are arbitrary. Thus, the state space of Γ_C is $S \backslash C \cup \{C\}$.

More generally, let $\mathcal{C}$ be a family of disjoint controlled sets. We let $\Gamma_{\mathcal{C}}$ be the game obtained from Γ by turning each set $C \in \mathcal{C}$ into a dummy state. To avoid confusion, the transition function of $\Gamma_{\mathcal{C}}$ is denoted by $p_{\mathcal{C}}$, and we will add such subscripts whenever useful. For instance, $S_{0,\mathcal{C}}$ is the set of non-absorbing states of $\Gamma_{\mathcal{C}}$. It coincides with $(S_0 \backslash \cup_{C \in \mathcal{C}} C) \cup (\cup_{C \in \mathcal{C}} \{C\})$.

It can be shown that if $\Gamma_{\mathcal{C}}$ has an ε-absorbing ε-equilibrium for each $\varepsilon > 0$, then so does Γ. (Note in particular that this excludes the possibility of "cycles" between the different controlled sets.)

Definition 6 *Let $D \subseteq S_0$ and β be a stationary strategy. The pair (β, D) is a* blocking pair *for player 1 if for each $z \in D, a \in A$,*

$$\left\{ p(D|s, a, \beta_z) < 1 \Rightarrow \mathbf{E}\left[v^1|s, a, \beta_z\right] < \max_D v^1 \right\}.$$

A blocking pair (α, D) for player 2 is defined by exchanging the roles of the two players in Definition 6.

We also extend this definition to games obtained by reducing Γ.

Definition 7 *Let $\mathcal{C}$ be a family of disjoint controlled sets. A pair (β, D), where $D \subseteq S_{0,\mathcal{C}}$, is a* reduced blocking pair *for player 1 if for each $z \in D, a \in A$,*

$$\left\{ p_{\mathcal{C}}(D|s, a, \beta_z) < 1 \Rightarrow \mathbf{E}_{\mathcal{C}}\left[v^1|s, a, \beta_z\right] < \max_{z \in D} v^1 \right\}.$$

We stress the fact that the value v^1 that is used is the value associated with the *original* game, and *not* the value $v^1_{\mathcal{C}}$ of the reduced game.[2] A state $z \in S_{0,\mathcal{C}}$ is either a state of S_0 or a controlled set $C \in \mathcal{C}$. In the latter case, we set $v^i(C) = \sup_{z' \in C} v^i(z')$. There is no relation between reduced blocking pairs and blocking pairs of the reduced game.

The notion of reduced blocking pair is relative to the family $\mathcal{C}$ of controlled sets. No ambiguity should ever arise.

[2]There is no specific relation between v and $v_{\mathcal{C}}$. In particular, $v_{\mathcal{C}}$ may depend on the choice of the payoff in the dummy states which replace the controlled sets.

3.2. STRUCTURE OF THE PROOF

We here list the main steps of the proof, summarized by Propositions 8, 9 and 10 below. The corresponding proofs will be sketched in later sections. The proof is divided into two independent parts.

The first part (see Proposition 8) consists of constructing a (possibly empty) family of disjoint controlled sets $\mathcal{C}$ that, in a sense, exhausts the blocking opportunities for one player, say for player 1. We then turn the corresponding reduced game $\Gamma_{\mathcal{C}}$ into a recursive game Γ_R, by setting the payoff function to zero in each non-absorbing state, and by leaving $\Gamma_{\mathcal{C}}$ unchanged in other respects. It will follow from the properties of $\mathcal{C}$ that any ε-absorbing ε-equilibrium profile of Γ_R is an ε-absorbing ε-equilibrium profile of the reduced game $\Gamma_{\mathcal{C}}$, when supplemented with a punishment threat (see Proposition 9).

The second part of the proof (see Proposition 10) consists of showing that recursive games such as Γ_R do have ε-absorbing ε-equilibrium profiles.

Proposition 8 *There is a collection $\mathcal{C}$ of disjoint controlled sets such that there is no reduced blocking pair for player 2.*

Of course, there is nothing specific about player 2. The same result holds for player 1 as well.

Let Γ_R be the game obtained from $\Gamma_{\mathcal{C}}$ by setting to zero the payoff function in each non-absorbing state.

The game Γ_R is recursive (see [1], [5]). Moreover,

F1 all absorbing payoffs of player 2 (resp. of player 1) are positive (resp. negative).

F2 for every α, there exists β, such that (α, β) is (0)-absorbing.

Proposition 9 *If Γ_R has a uniform equilibrium payoff, the game $\Gamma_{\mathcal{C}}$ has an ε-absorbing ε-equilibrium profile, for every $\varepsilon > 0$.*

Proposition 10 *Every (finite, two-player) recursive game that satisfies* **F1** *and* **F2** *has a uniform equilibrium payoff.*

Section 4 provides insights into the proof of Proposition 8. Proposition 9 is in most respects standard (see Section 5). Proposition 10 is discussed in the next chapter.

4. The Reduction

We here prove Proposition 8. If there is no blocking pair, the statement holds vacuously. We thus deal with games that have at least one blocking pair. In Sections 4.1 and 4.2, we prove that such games have a controlled set (see Proposition 13). This is not sufficient by itself, since the corresponding reduced game may contain additional reduced blocking pairs. In Section 4.3, we explain how to proceed iteratively.

4.1. THE BASIC PRINCIPLE

We start with a basic yet crucial observation. Denote by $\overline{\alpha}^\lambda$ an optimal strategy of player 1 in the λ-discounted version of the zero-sum game Γ^1. Hence,

$$\lambda r^1(z, \overline{\alpha}^\lambda_z, b) + (1-\lambda)\mathbf{E}\left[v^1_\lambda(\cdot)|z, \overline{\alpha}^\lambda_z, b\right] \geq v^1_\lambda(z) \text{ for every } z \in S \text{ and } b \in B. \tag{1}$$

Set $\overline{\alpha} := \lim_{\lambda\to 0} \overline{\alpha}^\lambda$, where the limit is taken up to a subsequence. Letting λ go to zero in (1), one gets

$$\mathbf{E}\left[v^1(\cdot)|z, \overline{\alpha}_z, b\right] \geq v^1(z), \text{ for every } z \in S \text{ and } b \in B.$$

Define also $\overline{\beta}$ as a limit, as λ converges to zero, of optimal stationary strategies of player 2 in the discounted zero-sum game Γ^2.

Lemma 11 *Let (α, D) be a blocking pair for player 2. There exists $\overline{D} \subseteq D$, such that: (i) v^2 is constant on $\overline{D}$; (ii) $\overline{D}$ is a weak communicating set w.r.t. $(\alpha, \overline{\beta})$; (iii) $(\alpha, \overline{D})$ is a blocking pair for player 2.*

Proof. Let $\widetilde{D} = \left\{z \in D : v^2(z) = \max_D v^2\right\}$ contain the states in D where v^2 is highest. Clearly, $(\alpha, \widetilde{D})$ is a blocking pair for player 2. In particular, $\widetilde{D}$ is closed for the Markov chain induced by $(\alpha, \overline{\beta})$. Consider the subsets of $\widetilde{D}$ which are maximal for the (weak) communication property. At least one of them will satisfy (i), (ii) and (iii). ∎

Lemma 11 is the simplest statement of a basic principle that underlies the reduction algorithm. Apply Lemma 11 to any blocking pair (α, D) for player 2.

- If $p(\overline{D}|z, a, \overline{\beta_z}) < 1$ and $\mathbf{E}\left[v^1|z, a, \overline{\beta}_z\right] \geq \max_{\overline{D}} v^1$, for some $z \in \overline{D}$, $a \in A$, then $\overline{D}$ is a controlled set. Indeed, choose, among those pairs, a pair (z^*, a^*) for which $\mathbf{E}\left[v^1|z, a, \overline{\beta}_z\right]$ is highest. It can be checked that the exit distribution $p(\cdot|z^*, a^*, \overline{\beta}_{z^*})$ from $\overline{D}$ is controllable for every continuation payoff vector $\gamma \geq v$.
- Otherwise, $(\overline{\beta}, \overline{D})$ is a blocking pair for player 1. Apply again Lemma 11 to the pair $(\overline{\beta}, \overline{D})$, with the roles of the two players exchanged, and let $\overline{\overline{D}} \subseteq \overline{D}$ be the corresponding subset. Then, as above,
  - either the exit distribution $p(\cdot|z^*, \overline{\alpha}_{z^*}, b^*)$ from $\overline{\overline{D}}$ is controllable, for some pair $(z^*, b^*) \in \overline{\overline{D}} \times B$,

- or $(\overline{\alpha}, \overline{\overline{D}})$ is a blocking pair for player 2.

In the latter case, notice that $(\overline{\alpha}, \overline{\overline{D}})$ is a blocking pair for player 2 *and* $(\overline{\beta}, \overline{\overline{D}})$ is a blocking pair for player 1. Hence we have proven Corollary 12 below.

Corollary 12 *Let (α, D) be a blocking pair for player 2. There exists $\overline{\overline{D}} \subseteq D$, such that either $\overline{\overline{D}}$ is a controlled set or both $(\overline{\alpha}, \overline{\overline{D}})$ and $(\overline{\beta}, \overline{\overline{D}})$ are blocking pairs.*

In the next section, we prove that, in the latter case, there is a controllable exit distribution from $\overline{\overline{D}}$, based on joint perturbations of the two players.

4.2. THE RELATION BETWEEN CONTROLLED AND SOLVABLE SETS

We let $D \subseteq S_0$ be a weak communicating set w.r.t. $(\overline{\alpha}, \overline{\beta})$, such that both $(\overline{\alpha}, D)$ and $(\overline{\beta}, D)$ are blocking pairs. We shall prove that either D is solvable or there is some controllable exit distribution from D. Since, by assumption, S_0 does not contain any solvable set, this yields Proposition 13 below.

Proposition 13 *If Γ has a blocking pair, then Γ has a controlled set.*

The argument in Section 4.1 used only the subharmonic properties of v^1 (resp. of v^2) with respect to the kernel $p(\cdot|z, \overline{\alpha}_z, \beta_z)$ (resp. $p(\cdot|z, \alpha_z, \overline{\beta}_z)$). We here use arguments of a different nature.

For expository purposes, we assume that for each state $z \in D$, and each action pair (a, b), one has $p(D|z, a, b) = 0$ as soon as $p(D|z, a, b) < 1$.

4.2.1. *Reminder about ε-Optimal Strategies*

We first point out a consequence of the proof of the existence of the value in finite zero-sum stochastic games (see [3]).

Let $(\alpha^\lambda)_{\lambda \leq \lambda_0}$ and $(\beta^\mu)_{\mu \leq \mu_0}$ be parametric families of stationary strategies, where $\lambda_0 > 0$ and $\mu_0 > 0$. Assume that, for every $z \in S$, $\lambda \in (0, \lambda_0)$, $\mu \in (0, \mu_0)$, one has

$$\lambda r^1(z, \alpha_z^\lambda, \beta_z^\mu) + (1 - \lambda)\mathbf{E}[v_\lambda^1 | z, \alpha_z^\lambda, \beta_z^\mu] \geq v_\lambda^1(z). \tag{2}$$

Then, close inspection of the proof in [2] shows that for every $\varepsilon > 0$, the following holds under (2). There exists a strategy σ which always plays according to some α^λ, in the sense that, for each finite history h_n, there is $\lambda(h_n) < \lambda_0$ such that $\sigma(h_n) = \alpha_{z_n}^{\lambda(h_n)}$, such that: for each strategy τ which always plays like some β^μ, and each initial state z,

$$\gamma_n^1(z, \sigma, \tau) \geq v^1(z) - \varepsilon, \text{ for } n \text{ large enough.} \tag{3}$$

Moreover, the same result holds in every subgame, in the following sense. Given a finite history h_p, let $(\sigma^{h_p}, \tau^{h_p})$ be the profile induced by (σ, τ) in the subgame initiated at h_p, and let z_p be the initial state of that subgame. Then, for every n large enough, one has $\gamma_n^1(z_p, \sigma^{h_p}, \tau^{h_p}) \geq v^1(z_p) - \varepsilon$.

Observe also that if the inequality (2) holds for each $\lambda \in (0, \lambda_0)$, it holds *a fortiori* for each $\lambda \in (0, \overline{\lambda}_0)$, provided $\overline{\lambda}_0 \leq \lambda_0$. Therefore, one may assume that $\lambda(h_n) < \varepsilon$, for every finite history h_n.

Finally, it can be checked that if some states $z \in S_0$ are replaced by absorbing states with payoff $v(z) + \varepsilon$, then the conclusion (3) will still hold in the modified game.[3]

4.2.2. *Application*

In Mertens and Neyman's original proof, α^λ is taken to be an optimal strategy of player 1 in the λ-discounted zero-sum game. No restriction on the strategies of player 2 is needed.

By contrast, we here define α^λ as follows. By definition of $\overline{\alpha}$, there exists, for each $\lambda > 0$, an optimal stationary strategy $\overline{\alpha}^\lambda$ in the λ-discounted game such that $\lim_{\lambda \to 0} \overline{\alpha}^\lambda = \overline{\alpha}$. For $z \in D$, set $A_z = \{a \in A,\, p(D|z, a, \overline{\beta}_z) = 1\}$, and define α_z^λ to be the conditional distribution of $\overline{\alpha}_z^\lambda$, conditioned on A_z.

Define β^μ accordingly for player 2, by conditioning $\overline{\beta}^\mu$ on the set $B_z = \{b \in B,\, p(D|z, \overline{\alpha}_z, b) = 1\}$. It is straightforward to check that:

- $\lim_{\lambda \to 0} \alpha^\lambda = \overline{\alpha}$, and $\lim_{\mu \to 0} \beta^\mu = \overline{\beta}$.
- For every z, the inequality (2), together with its counterpart for player 2, holds, provided λ and μ are small enough.

It will be convenient to apply the result of Section 4.2.1 to a slightly modified game. Let $\varepsilon > 0$ be given. Turn each state $z \notin D$ into an absorbing state with payoff $(v^1(z) + \varepsilon, v^2(z) + \varepsilon)$. Using the final remark of Section 4.2.1, there exists a profile $(\sigma_\varepsilon, \tau_\varepsilon)$ of strategies, such that:

1. for each history $h_n = (z_1, a_1, b_1, \cdots, z_n)$, there exists λ_n and β_n such that $\sigma_\varepsilon(h_n) = \overline{\alpha}_{z_n}^{\lambda_n}$ and $\tau_\varepsilon(h_n) = \overline{\beta}_{z_n}^{\mu_n}$. Moreover, both $||\sigma_\varepsilon(h_n) - \overline{\alpha}_{z_n}|| < \varepsilon$ and $||\tau_\varepsilon(h_n) - \overline{\beta}_{z_n}|| < \varepsilon$ hold.
2. For n large enough, $\gamma_n(z, \sigma_\varepsilon, \tau_\varepsilon) \geq v(z) - \varepsilon$, and the same holds in any subgame.[4]

Let $p_\varepsilon := \mathbf{P}_{z, \sigma_\varepsilon, \tau_\varepsilon}(\theta_D < +\infty)$ be the probability that the play ever leaves D. We discuss two non-mutually exclusive cases.

[3]Where the left-hand side is the expected average payoff in the new game, while the right-hand side contains the value of the original game. We use here the fact that the value of the new game is at least the value of the original game.

[4]In this inequality, $\gamma_n(z, \sigma_\varepsilon, \tau_\varepsilon)$ stands for the average payoff in the new game, and $v(z)$ for the value in the original game.

Case 1 There is a sequence (ε_n) converging to zero, with $p_{\varepsilon_n} = 1$ for every n.

We here argue that there is a controllable exit distribution from D. Indeed, let ε belong to the sequence (ε_n), and denote by

$$Q_\varepsilon(\cdot) = \mathbf{P}_{z,\sigma_\varepsilon,\tau_\varepsilon}(z_{\theta_D} = \cdot)$$

the law of the exit state from D. Since θ_D is a.s. finite, $\lim_n \gamma_n(z, \sigma_\varepsilon, \tau_\varepsilon) = \mathbf{E}_{Q_\varepsilon}[v] + \varepsilon$. Therefore,

$$\mathbf{E}_{Q_\varepsilon}[v] \geq v(z) - 2\varepsilon. \tag{4}$$

On the other hand, $p(D|z_n, \sigma_\varepsilon(h_n), \overline{\beta}_{z_n}) = 1 = p(D|z_n, \overline{\alpha}_{z_n}, \tau_\varepsilon(h_n))$, for every history h_n. Thus, Q_ε belongs to the set $\mathcal{Q}^2(\overline{\alpha}, \overline{\beta})$ (see [4]), that is, Q_ε is a convex combination of the distributions $p(\cdot|z,a,b)$ where $(z,a,b) \in D \times A \times B$ is any triple such that $p(D|z,a,\overline{\beta}_z) = p(D|z,\overline{\alpha}_z,b) = 1$. Since $\mathcal{Q}^2(\overline{\alpha}, \overline{\beta})$ is compact, and using (4), there is a distribution $Q \in \mathcal{Q}^2(\overline{\alpha}, \overline{\beta})$ with $\mathbf{E}_Q[v] \geq v(z)$.

Since Q involves no unilateral exits, and since both pairs $(\overline{\alpha}, D)$ and $(\overline{\beta}, D)$ are blocking, it is not difficult to conclude that the exit Q is controllable (with respect to any $\gamma \geq v$).

Case 2 There is a sequence (ε_n) converging to zero, with $p_{\varepsilon_n} < 1$, for every n.

We here argue that D is solvable. Indeed, let ε belong to the sequence (ε_n). Let h_p be any history following which the probability $\mathbf{P}_{z,\sigma_\varepsilon,\tau_\varepsilon}(\theta_D < +\infty|h_p)$ of leaving D in finite time is close to 0. Since $p_\varepsilon < 1$, such a history exists. Following previously used notations, let $(\sigma_\varepsilon^{h_p}, \tau_\varepsilon^{h_p})$ be the profile induced by $(\sigma_\varepsilon, \tau_\varepsilon)$ in the subgame initiated at h_p, and let z_p be the initial state of that subgame. Since σ_ε and τ_ε always play approximately like $\overline{\alpha}$ and $\overline{\beta}$, the average payoff $\gamma_n(z_p, \sigma_\varepsilon^{h_p}, \tau_\varepsilon^{h_p})$ is, for n large, close to the convex hull of the payoff vectors $\gamma(R, \overline{\alpha}, \overline{\beta})$, $R \in \mathcal{R}_D(\overline{\alpha}, \overline{\beta})$. Letting ε converge to zero, this shows the existence of $\mu \in \Delta(\mathcal{R}_D(\overline{\alpha}, \overline{\beta}))$ such that $\sum_{R \in \mathcal{R}_D(\overline{\alpha}, \overline{\beta})} \mu_D \gamma(R, \overline{\alpha}, \overline{\beta}) \geq v$.

Since D is a weak communicating set w.r.t. $(\overline{\alpha}, \overline{\beta})$, and since both pairs $(\overline{\alpha}, D)$ and $(\overline{\beta}, D)$ are blocking, the triple $(D, (\overline{\alpha}, \overline{\beta}), \mu)$ is solvable.

4.3. THE ALGORITHM

In the previous section, we showed why the existence of a blocking pair implies the existence of a controlled pair. This is not sufficient by itself, since the corresponding reduced game may contain additional blocking pairs. In that case, the above proof needs to be applied again to the reduced game. This raises a difficulty. Indeed, the proof ultimately relied on the fact that all solvable sets in the game correspond to absorbing states. This need not be the case for the reduced game. The reduced game being only a fiction, such additional solvable sets would be very difficult to interpret in the original game.

It turns out that such troubles can safely be avoided by somehow reversing the argument and by proceeding in only two steps.

Step 1 Let $\mathcal{D}$ be the collection of all sets $D \subseteq S_0$ that satisfy **D1-3** below:

D1 v is constant on D;
D2 D is a weak communicating set w.r.t. $(\overline{\alpha}, \overline{\beta})$;
D3 $(\overline{\beta}, D)$ is a blocking pair for player 1.

Let $\mathcal{D}'$ be the collection that consists of the maximal (w.r.t. inclusion) elements of $\mathcal{D}$. Any two distinct elements of $\mathcal{D}'$ are disjoint subsets of S_0. As shown above, for each $D \in \mathcal{D}'$, there is a controllable exit distribution from D, based on perturbations of $(\overline{\alpha}, \overline{\beta})$.

Lemma 14 *Consider the reduced game $\Gamma_{\mathcal{D}'}$. There is no set $D \subseteq S_{0,\mathcal{D}'}$, such that $(\overline{\beta}, D)$ is a reduced blocking pair for player 1.*

Proof. We argue by contradiction, and assume that there is such a reduced blocking pair $(\overline{\beta}, D)$, where D contains only non-absorbing states of $\Gamma_{\mathcal{D}'}$. It is easy to check that there is a subset $\widetilde{D} \subseteq D$ such that: (i) v^2 is constant on $\widetilde{D}$ and (ii) $(\overline{\beta}, \widetilde{D})$ is a reduced blocking pair for player 1. A variation on Lemma 11 shows that there exists a subset $\overline{D} \subseteq \widetilde{D}$ such that v is constant on $\overline{D}$ and $(\overline{\beta}, \overline{D})$ is a reduced blocking pair. Identifying $\overline{D}$ to a subset of S_0, the pair $(\overline{\beta}, \overline{D})$ is a blocking pair for player 1 in the original game Γ. Note in addition that, for each $D \in \mathcal{D}'$, one either has $D \subseteq \overline{D}$ or $D \cap \overline{D} = \emptyset$. Since elements of $\mathcal{D}'$ are transient states in the reduced game $\Gamma_{\mathcal{D}'}$, the set $\overline{D}$ cannot coincide with an element of $\mathcal{D}'$ −a contradiction to the definition of $\mathcal{D}'$. ∎

Step 2 For given $C \subseteq S_{0,\mathcal{D}'}$, we define the properties **C1** and **C2** as:

C1 v^2 is constant on C;
C2 for some α, C is communicating for $(\alpha, \overline{\beta})$ and (α, C) is a reduced blocking pair for player 2.

Let C_1 be a maximal subset of $S_{0,\mathcal{D}'}$ (if any) that satisfies both **C1** and **C2**. Iteratively, we let C_l be a maximal subset of $S_{0,\mathcal{D}'} \backslash (C_1 \cup ... C_{l-1})$ that satisfies both **C1** and **C2,** if any. Let $\mathcal{E}$ be the collection of subsets of $S^0_{\mathcal{D}'}$ obtained in this way.[5]

[5] It is not uniquely defined.

Let C be an arbitrary element of $\mathcal{E}$. By **Step 1**, the pair $(C, \overline{\beta})$ is *not* a reduced blocking pair for player 1. This can be shown to imply that $(C, \overline{\beta})$ is not a blocking pair for player 1. Therefore, for some $(z, a) \in C \times A$, the pair $(C,\, p(\cdot|z, a, \overline{\beta}_z))$ is a controlled set.

As above, we may view elements of $\mathcal{E}$ as subsets of S_0. We finally let $\mathcal{C} = \mathcal{E} \cup \mathcal{D}_1$, where $\mathcal{D}_1$ is the collection of sets $D \in \mathcal{D}'$ that are disjoint from each $C \in \mathcal{E}$.

The next lemma is a slightly involved variant of Lemma 14. It concludes the sketch of the proof of Proposition 8.

Lemma 15 *Consider the reduced game $\Gamma_{\mathcal{C}}$. There is no reduced blocking pair for player 2.*

5. Reduced Games and Recursive Games

It remains to prove Proposition 9. It is based on the next lemma, which itself relies on the idea that the reduction eliminated all blocking pairs for player 2.

Lemma 16 *Let $\mathcal{C}$ be a collection of disjoint controlled sets such that there is no reduced blocking pair for player 2. In the reduced game $\Gamma_{\mathcal{C}}$, the following holds. For each stationary strategy α, there exists β such that the profile (α, β) is (0)-absorbing and $\gamma^2(z, \alpha, \beta) \geq v^2(z)$.*

We stress once more the fact that v^2 is the value associated to the original game Γ^2.

Proof. Choose β such that, for each z, the support of β_z consists exactly of those $b \in B$ such that

$$\mathbf{E}\left[v^2|z, \alpha_z, b\right] \geq v^2(z),$$

and notice that $\mathbf{E}_{\mathcal{C}}\left[v^2|z, a, b\right] \geq \mathbf{E}\left[v^2|z, a, b\right]$, for every (z, a, b). It can be shown that β satisfies both conclusions. ∎

We conclude by sketching the proof of Proposition 9. Let (σ, τ) be an ε-absorbing ε-equilibrium of Γ_R. For $N \in \mathbf{N}$ large enough, the profile which plays (σ, τ) up to stage N, and punishment strategies afterwards, is an ε'-absorbing ε'-equilibrium, where $\varepsilon' > \varepsilon$ goes to zero with ε.[6]

References

1. Everett, H. (1957) Recursive games, in M. Dresher, A. W. Tucker and P. Wolfe (eds.), *Contributions to the Theory of Games, Vol III,* Annals of Mathematics Studies 39, Princeton University Press, Princeton, NJ, pp. 47–78.

[6]This sketch is misleading. A correct proof needs to connect the value of Γ_R to the value of the reduced game. I will not elaborate on this specific point.

2. Mertens, J.-F. and Neyman, A. (1981) Stochastic games, *International Journal of Game Theory* **10**, 53–66.
3. Neyman, A. (2003) Stochastic games: Existence of the minmax, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 11, pp. 173–193.
4. Solan, E. (2003) Perturbations of Markov chains with applications to stochastic games, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 17, pp. 265–280.
5. Thuijsman, F. (2003) Recursive games, in A. Neyman and S. Sorin (eds.), *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Kluwer Academic Publishers, Dordrecht, Chapter 16, pp. 253–264.
6. Thuijsman, F. and Vrieze, O.J. (1991) Easy initial states in stochastic games, in T. S. Ferguson, T. E. S. Raghavan, T. Parthasarathy and O.J. Vrieze (eds.), *Stochastic Games and Related Topics,* Kluwer Academic Publishers, Dordrecht, pp. 85–100.
7. Vieille, N. (1993) Solvable states in stochastic games, *International Journal of Game Theory* **21**, 395–405.
8. Vieille, N. (2000) Two-player stochastic games I: A reduction, *Israel Journal of Mathematics* **119**, 55–91.
9. Vieille, N. (2000) Two-player stochastic games II: The case of recursive games, *Israel Journal of Mathematics* **119**, 93–126.
10. Vieille, N. (2000) Solvable states in N-player stochastic games, *SIAM Journal of Control and Optimization* **38**, 1794–1804.