Continuous-time Stochastic Games

Abraham Neyman*

September 23, 2016

Abstract

We study continuous-time stochastic games (with finitely many states and actions), with a focus on the existence of their equilibria that are insensitive to a small imprecision in the specification of players' evaluations of streams of payoffs.

We show that the stationary, namely, time-independent, discounting game has a stationary equilibrium, that the (not necessarily stationary) discounting game and the more general game with time separable payoffs (where the discounting or time-separable payoffs can vary among the players) has an epsilon equilibrium, and that in all the above-mentioned cases there are strategy profiles that are epsilon equilibrium in all the games with a sufficiently small perturbation of the players' valuations.

A limit point of discounting valuations need not be a discounting valuation as some of the "mass" may be pushed to infinity; it is represented by an average of a discounting valuation (which represents the impatient part of the valuation) and a mass at infinity (which represents the patient part of the valuation). We show that for every such limit point there is a strategy profile that is an epsilon equilibrium of all the discounting games with discounting valuations that are sufficiently close to the limit point.

*Institute of Mathematics, and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel. *E-mail: aneyman@math.huji.ac.il*

This research was supported in part by Israel Science Foundation grant 1596/10.

1 Introduction

1.1 Motivating continuous-time stochastic games

A fundamental shortfall of economic forecasting is its inability to make deterministic forecasts. A notable example is the repeated occurrence of financial crises. Even though each financial crisis was forecasted by many economic experts, many other experts claimed just before each occurrence that "this time will be different," as documented in the book "This Time Is Different: Eight Centuries of Financial Folly" by Reinhart and Rogoff. On the other hand, many economic experts have forecasted financial crises that never materialized. Finally, even those who correctly predicted a financial crisis could not predict the time of its occurrence and many of them warned that it was imminent when in fact it occurred many years later.

An analogous observation comes from sports, e.g., soccer. In observing the play of a soccer game, or knowing the qualities of both competing teams, we can recognize a clear advantage of one team over the other. Therefore, a correct forecast is that the stronger team is more likely than the other team to score the next goal. However, it is impossible to be sure which team will be the first to score the next goal, and it is impossible to forecast the time of the scoring.

In both cases, the state of the economy and the score of the game can dramatically change in a split second, and players' actions impact the likelihood of each state change.

A game-theoretic model that accounts for the change of a state between different stages of the interaction, and where such change is impacted by the players' actions, is a stochastic game. However, no single deterministic-time dynamic game can capture the important common feature of these two examples: namely, the probability of a state change in any short time interval can be positive yet arbitrarily small. This feature can be analyzed by studying the asymptotic behavior in a sequence of discrete-time stochastic games, where the individual stage represents short time intervals that converge to zero and the transition probabilities to a new state also converge to zero. The limit of such a sequence of discrete-time stochastic games is a continuous-time stochastic game; see [27, 28].

The analysis of continuous-time stochastic games enables us to model and analyze the important properties that (1) the state of the interaction

can change, (2) the stochastic law of states is impacted by players' actions, and (3) in any short time interval the probability of a discontinuous state change can be positive (depending on the players' actions), but arbitrarily small for sufficiently short time intervals.

Accounting for stochastic state changes — both gradual (continuous) and instantaneous (discontinuous) but with infinitesimal probability in any short time interval — is common in the theory of continuous-time finance; see, e.g., [20]. However, in continuous-time finance theory the stochastic law of states is not impacted by agents' actions. Accounting for (mainly deterministic and more recently also stochastic) continuous state changes that are impacted by agents' actions is common in the theory of differential games; see, e.g., [9].

In the present paper we study the three above-mentioned properties, with a special focus on discontinuous state changes. To do this we study the continuous-time stochastic game with finitely many states, since if there are finitely many states, then any state change is discontinuous.

1.2 Difficulties with continuous-time strategies

A classical game-theoretic analysis of continuous-time games entails a few unavoidable pathologies, as for some naturally defined strategies there is no distribution on the space of plays that is compatible with these strategies, and for other naturally defined strategies there are multiple distributions on the space of plays that are compatible with these strategies. In addition, what defines a strategy is questionable. In spite of these pathologies, we will describe unambiguously equilibrium payoffs and equilibrium strategies in continuous-time stochastic games.

Earlier game-theoretic studies overcome the above-mentioned pathologies by restricting the study either to strategies with inertia, or to Markov (memoryless) strategies that select an action as a function of (only) the current time and state. See, e.g., [1, 33] for the study of continuous-time supergames, [29] for the study of continuous-time bargaining, [9] for the study of differential games, and [44, 4, 5] for the study of continuous-time stochastic games, termed in the above-mentioned literature Markov games, or Markov chain games. A more detailed discussion of the relation between these earlier contributions and the present paper appears in Section 7.

The common characteristic of these earlier studies is that they each con-

sider only a subset of strategies, so that a profile[1] of strategies selected from the restricted class defines a distribution over plays of the games, and thus optimality and equilibrium are well defined, but only within the restricted class of strategies. Therefore, in an equilibrium, there is no beneficial unilateral deviation within only the restricted class of strategies. Therefore, there is neither optimality nor nonexistence of beneficial unilateral deviation claims for general strategies.

The restriction to Markov (memoryless) strategies (see, e.g., [44, 31, 21, 22, 11, 12, 14, 7]) is (essentially) innocuous in the study of continuous-time, discounted or finite-horizon Markov decision processes and two-person zero-sum stochastic games.

The two properties of the discounted or finite-horizon payoffs that make this restriction innocuous are: (1) impatience, namely, the contribution to the payoff of the play in the distant future is negligible, and (2) time-separability of the payoff, namely, the payoff of a play on $[0, \infty)$ is, for every $s > 0$, a sum of a function of the play on the time interval $[0, s)$ and a function of the play on the time interval $[s, \infty)$.

The Markov (memoryless) assumption is restrictive in the Markov decision processes with a payoff that is not time-separable and in the zero-sum stochastic games with non-impatient payoffs.

The Markov (memoryless) assumption is restrictive in the non-zero-sum game-theoretic framework. It turns out that in discounted stochastic games (with finitely many states and actions) a Markov strategy profile that is an equilibrium in the universe of Markov strategies is also an equilibrium in the universe of history-dependent strategies. However, the set of equilibrium strategies and equilibrium payoffs in the universe of Markov strategies is a proper subset of those in the universe of history-dependent strategies.

A fundamental shortcoming of the restriction to memoryless strategies (and to oblivious strategies, which depend only on the state process) arises in the study of approximate equilibria that are robust to small changes of the discounting valuation. For example, there is a continuous-time (and discrete-time, even two-person zero-sum, e.g., the Big Match [2]) stochastic game for which there is no Markov strategy profile that is an approximate equilibrium in all the discounted games with a sufficiently small discounting rate.

[1] I.e., a list, one for each player.

The present paper proves that every continuous-time stochastic game with finitely many states and actions has such strategy profiles.

1.3 Discrete-time stochastic games

We recall the basic properties of the discrete-time stochastic game model; this will enable us to point out its similarities and differences with the continuous-time model to be described later.

In a discrete-time stochastic game, play proceeds in stages and the stage state, the stage action profile, the previous stage, and the next stage are well defined. The stage payoff is a function $g(z, a)$ of the stage state z and the stage action profile a, and the transitions to the next state z' are defined by conditional probabilities $p(z' \mid z, a)$ of the next state z' given the present state z and the stage action profile a. Players' stage-action choices are made simultaneously and are observed by all players following the stage play. Pure, mixed, and behavioral strategies are well and unambiguously defined in the discrete-time model.

If at stage $m = 0, 1, \ldots$, the state is z_m and the action profile played is a_m, the stage payoff is $g_m := g(z_m, a_m)$. A play of the discrete-time stochastic game generates a stream of payoffs $g_0, g_1, \ldots$.

A measure w on the nonnegative integers defines a payoff function u_w on streams of payoffs by $u_w(g_0, g_1, \ldots) = \sum_{m=0}^{\infty} w(m) g_m$. The stationary discounting corresponds to measures w of the form $w(m) = c\lambda^m$ where c is a positive constant and $0 \leq \lambda < 1$. A (not necessarily stationary) discounting corresponds to a measure w with $0 \leq w(m+1) \leq w(m)$.

The common normalization $u_w(1, 1, \ldots) = 1$ corresponds to $\sum_{m=0}^{\infty} w(m) = 1$ and is useful in comparing outcomes of the same game with different discountings.

The measures that define the payoff function of different players need not be identical.

1.4 Continuous-time stochastic games: payoffs

In a continuous-time game, the payoff is usually defined as an accumulation of infinitesimal payoffs. If at time $t \in [s, s + ds)$ the state is z and the action profile played is a the accumulation of the payoff in the time interval $[s, s + ds)$ is $g(z, a)ds$. Therefore, if the function $t \mapsto (z_t, a_t)$, where z_t is

the state at time t and a_t is the action profile at time t, is measurable, then the accumulation of payoffs in the time interval $[0, s)$ is $\int_0^s g_t\,dt$ where $g_t := g(z_t, a_t)$.

The unnormalized, respectively, the normalized, ρ-discounted payoff, $\rho > 0$, is $\int_0^\infty e^{-\rho t} g_t\,dt$, respectively, $\rho \int_0^\infty e^{-\rho t} g_t\,dt$, and the unnormalized, respectively, the normalized, s-horizon payoff, $s > 0$, is $\int_0^s g_t\,dt$, respectively, $\frac{1}{s}\int_0^s g_t\,dt$.

Similarly, the payoffs are also defined on the space of measurable functions $t \mapsto (z_t, x_t)$, where z_t is the state at time t and x_t is the mixed action-profile (i.e., mixtures of pure action profiles) at time t, by setting $g(z, x)$ to be the linear extension of $g(z, a)$ and $g_t = g(z_t, x_t)$.

In general, the payoff in a stochastic game can be an arbitrary function of the function $t \mapsto (z_t, x_t)$.

Consider for example a continuous-time stochastic game that models a single quarter-final or semi-final in the UEFA Champions League. The rule that specifies the winner, which qualifies for the next stage, follows.

A single quarter-final or semi-final is played under the knockout system, on a home-and-away basis (two legs). The team that scores the greater aggregate of goals in the two matches qualifies for the next stage. Otherwise, the team that scores more away goals qualifies for the next stage.

If this procedure does not produce a result, i.e., if both teams score the same number of goals at home and away, a 30-minute period of overtime is played at the end of the second leg. If both teams score the same number of goals during overtime, away goals count double (i.e. a visiting team that scored in the overtime qualifies).

If no goals are scored during overtime, kicks from the penalty mark determine which team qualifies for the next stage. For simplicity, assume that the outcome of the penalty kicks is random, e.g., each team is equally likely to win this phase.

Each match lasts 90 minutes. Therefore, if the state of the stochastic game, z_t, is the cumulative score at the t-th minute, the rules provides us a function of $(z_{90}, z_{180}, z_{210})$ that specifies the probability (1, 0, or 1/2) that team 1 wins.

In the present paper we focus on payoffs of the stochastic game that satisfy several conditions.

First, the payoff to a player, say player i, is a function u of the stream $\mathbf{g} : t \mapsto g_t^i := g^i(z_t, x_t)$ of his "stage payoffs."

Second, we assume that the payoff function u is linear and monotonic and is defined over the ordered linear space of all bounded measurable streams of payoffs $\mathbf{f} : t \mapsto f_t$ (with the order $\mathbf{f} \geq \mathbf{g}$ iff $f_t \geq g_t \ \forall t$, and the natural linear structure $\alpha\mathbf{f} + \beta\mathbf{g} : t \mapsto \alpha f_t + \beta g_t$).

We allow for the payoff functions to differ among players.

The payoff function u is *normalized* if $u(\mathbf{1}) = 1$, where $\mathbf{1}$ is the stream $\mathbf{f}$ with $f_t = 1$ for every $t \geq 0$.

The payoff function u is *impatient* if $u(\mathbf{1}_{\geq s}) \to_{s\to\infty} 0$, where $\mathbf{1}_{\geq s}$ is the stream $\mathbf{f}$ with $f_t = 1$ for every $t \geq s$ and $f_t = 0$ for every $0 \leq t < s$.

A payoff function u that is defined (on the space of bounded measurable streams of payoffs $\mathbf{f}$) by $u(\mathbf{f}) := \int_0^\infty f_t \, dw(t)$, where w is a measure on $[0, \infty)$, is called a *time-separable payoff* and is denoted by u_w.

A time-separable payoff function is linear and monotonic. If, in addition, w is a probability measure, then u_w is normalized.

In the other direction, any linear and monotonic payoff function u that is impatient defines a unique measure w on $[0, \infty)$ such that for any bounded continuous stream of payoffs $\mathbf{f}$ we have $u(\mathbf{f}) = u_w(\mathbf{f}) := \int_0^\infty f_t \, dw(t)$.

An important special class of time-separable payoff functions u, called *impatient valuations*, is defined below.

For a given measurable subset B of $\mathbb{R}_+ = [0, \infty)$, we denote by $\mathbf{1}_B$ the indicator function of B (i.e., the stream of payoffs $\mathbf{f}$ such that $f_t = 1$ if $t \in B$ and $f_t = 0$ if $t \notin B$).

An *impatient valuation* is a time-separable payoff function u that is normalized and for every measurable subset B of $\mathbb{R}_+$ we have $u(\mathbf{1}_B) \geq u(\mathbf{1}_{c+B})$, where $c \geq 0$ and $c + B := \{c + b \mid b \in B\}$. The last condition reflects the preference of advancing the time of positive payoffs.

The time-separable payoff function that is defined by a measure w on $\mathbb{R}_+$ is an impatient valuation if and only if w is a probability measure and $w([b, b + c))$ is, for every fixed $c > 0$, nonincreasing in $b \geq 0$.

An approximate description of a normalized, linear, monotonic, and impatient payoff function u is obtained by specifying, within a small positive error term, its value on finitely many continuous streams of payoffs with bounded support.

Therefore, we say that a sequence u_k of normalized time-separable payoff functions u_k converges if the sequence $u_k(\mathbf{f})$ converges for every continuous function $\mathbf{f}$ with bounded support.

Examples of converging time-separable payoff functions include: the normalized ρ-discount payoff as ρ goes to 0, the normalized s-horizon payoff as s goes to infinity, and the sequence of payoff functions u_k that are defined by the dirac measures on a sequence of points that converge to some fixed point x_0.

Let $u_k = u_{w_k}$ be a converging sequence of normalized time-separable payoff functions. Then, the limit $u(\mathbf{f}) := \lim_{k\to\infty} u_k(\mathbf{f})$ exists for every continuous function $\mathbf{f}$ that is defined on $[0,\infty]$. (Note that a continuous function $\mathbf{f}$ on $[0,\infty]$ is identified with a continuous function $\mathbf{f}$ on $[0,\infty)$ for which the limit of f_t, as t goes to infinity, exists.)

The limit u, which is defined on the space of all continuous functions on $[0,\infty]$, is linear, monotonic, and normalized. However, it need not coincide with a time-separable payoff function u, as some of the corresponding w_k-mass may be pushed to infinity.

The restriction of the limit u to the space of continuous functions $\mathbf{f}$ on $[0,\infty]$ is represented by a (unique) probability measure w on $[0,\infty]$:

$$\lim_{k\to\infty} u_k(\mathbf{f}) = \lim_{k\to\infty} \int_0^\infty f_t \, dw_k(t) = \int_{[0,\infty]} f_t \, dw(t).$$

If u_k are impatient valuations then the limiting probability measure w, which may have an atom at ∞ and at 0, is absolutely continuous[2] on $(0,\infty)$ and $\frac{dw}{dt}$ is nonincreasing on $(0,\infty)$; equivalently, w is a probability measure on $[0,\infty]$ with $w([b,b+c))$ being, for every fixed $c > 0$, nonincreasing in $b \geq 0$.

Therefore, in studying approximate optimization that is insensitive to a small imprecision in the specification of an impatient valuation, it is useful to enlarge the space of measures on $[0,\infty)$ to the space W of all measures on $[0,\infty]$.

[2]A finite measure μ on Borel subsets of the real line is absolutely continuous if for every positive number $\varepsilon > 0$ there is a positive number $\delta > 0$ such that $\mu(A) < \varepsilon$ for all Borel sets A of Lebesgue measure less than δ. If μ is absolutely continuous than there exists a Lebesgue integrable function g, denoted by $\frac{d\mu}{dt}$, on the real line such that $\mu(A) = \int_A g \, dt$ for all Borel subsets A of the real line.

For a given measure $w \in W$, the w-payoff is the integral $\int_{[0,\infty]} g_t \, dw(t)$, where $g_\infty = \lim_{s\to\infty} \frac{1}{s} \int_0^s g_t \, dt$ if the limit exists.

The definition of g_∞ is motivated by the characterization of *valuations* that follows.

A payoff function u satisfies the time value of money principle if

$$u(\mathbf{f}) \geq u(\mathbf{g}) \text{ whenever } f_0 \geq g_0 \text{ and } \int_0^s f_t \, dt \geq \int_0^s g_t \, dt \;\; \forall s > 0.$$

The time value of money principle reflects the preference for expediting the time of positive payoffs: the faster the accumulation of payoffs, the better. The principle states that a unit of payoffs in a given time period is preferable to its being spread over future time periods.

Note that a payoff function that satisfies the time value of money principle is monotonic.

A *valuation* is a normalized linear payoff function u that satisfies the time value of money principle. A *patient valuation* is a valuation u such that $u(\mathbf{1}_{\geq s}) = 1 \; \forall s \geq 0$.

One can show that a payoff function u is a patient valuation if and only if it is a linear function over the bounded stream of payoffs such that

$$\liminf_{s\to\infty} \frac{1}{s} \int_0^s f_t \, dt \leq u(\mathbf{f}) \leq \limsup_{s\to\infty} \frac{1}{s} \int_0^s f_t \, dt,$$

and that any valuation is a mixture of an impatient valuation and a patient one.

1.5 Continuous-time stochastic games: transitions

The transitions of states are described by nonnegative real-valued transition rates $\mu(z', z, a)$, defined for all triples z', z, a, of two distinct states $z' \neq z$ and an action profile a.

The nonnegative real number $\mu(z', z, a)$ describes the rate of transition from state z to state z' when action profile a is played at state z.

If the state at time t is $z_t = z$ and the action profile at all times $t \leq s < t + \delta$ is a, then the probability that the state $z_{t+\delta} = z' \neq z$ is approximately $\delta\mu(z', z, a)$ for small δ; explicitly, for two distinct states z, z' and a time t,

conditional on the history of the play up to time t, $z_t = z$, and $a_s = a$ for all $t \le s < t + \delta$, we have $P(z_{t+\delta} = z')/\delta \to \mu(z', z, a)$ as $\delta \to 0+$.

The actions in a continuous-time game can depend on time. Therefore, one has to define the transitions resulting from time-dependent actions.

Given a measurable time-dependent action profile a_t, $t \ge 0$, and two distinct states $z' \ne z$, the transitions obey

$$P(z_{t+\delta} = z' \mid z_t = z) = \int_t^{t+\delta} \mu(z', z, a_s)\, ds + o(\delta) \text{ as } \delta \to 0 + .$$

1.6 Continuous-time stochastic games: strategies

The classical definition of a pure strategy in a discrete-time game is a local definition. It defines, for each stage m, the selected action at stage m as a function of the information derived from (signals from) players' and nature's moves.

In a stochastic game with observable states and actions, a strategy specifies the action at the discrete stage m as a function of the sequence of past states and actions of other players and of the current state.

An equivalent definition is the global one that specifies the sequence of actions of the player as a function of the entire sequence of states and actions of other players, but with the obvious no-looking-ahead property.

The *no-looking-ahead property*, termed also *no anticipation*, asserts that the specified action at stage t depends only on players' past observable actions and past observable chance moves.

The equivalence of the two definitions follows from the fact that the set of decision times is well ordered; see Section 2.2.1.

The set of times $t \ge 0$ is not well ordered. Therefore, there are naturally defined local strategies that do not integrate into a global one; see Section 2.2.1. Therefore, one reverts to the global form of a strategy.

A *full pure strategy* of player i is a function σ^i from plays – functions $h : t \mapsto (z_t, a_t)$, $t \ge 0$ – to functions $t \mapsto \sigma_t^i(h)$ (with proper measurability assumptions), with the no-anticipation property.

Explicitly, if $h = (z_t, a_t)_{t \ge 0}$ and $h' = (z'_t, b_t)_{t \ge 0}$ are two plays with $(z_t, a_t) = (z'_t, b_t)$ for $t \le s$, then $\sigma_t^i(h) = \sigma_t^i(h')$ on $t \le s$.

Full pure strategies are important in the study of subgame perfect equilibria. Our focus in the present paper is on equilibria. Therefore, we confine

attention to reduced strategies, namely strategies that define the action to be taken only on histories that are compatible with the strategy.

A *reduced pure strategy*, henceforth a *pure strategy*, of player i is a function σ^i from plays – functions $h : t \mapsto (z_t, a_t)$, $t \geq 0$ – to functions $t \mapsto \sigma^i_t(h)$ (with proper measurability assumptions), with the no-anticipation property and independent of its own past actions.

Explicitly, if $h = (z_t, a_t)_{t\geq 0}$ and $h' = (z'_t, b_t)_{t\geq 0}$ are two plays with $(z_t, a_t^{-i}) = (z'_t, b_t^{-i})$ for $t \leq s$, then $\sigma^i_t(h) = \sigma^i_t(h')$ on $t \leq s$.

The assumption that the (reduced pure) strategy choice of player i is independent of player i's past actions is conceptually innocuous, as the strategy itself and other players' past actions define player i's past actions.

However, the independence of $\sigma^i_t(h)$ of at least player i's very recent past own actions is required for technical reasons, as otherwise, even in the one-person game, there are strategies that do not define a play; see Section 2.2.1.

The reversion to global strategies does not phase out all pathologies, even in the special case of continuous-time supergames.

Indeed, there are ("naturally" defined global) pure strategy profiles σ, e.g., in the continuous-time matching pennies game, for which there is no play h such that $\sigma(h) = h$, and there are ("naturally" defined global) pure strategy profiles σ in continuous-time supergames for which there is more than one play h with $\sigma(h) = h$ (see Proposition 1 in Section 2.2.1).

We illustrate the above comment by explicit examples of such strategy profiles. This illustration may serve also, in particular, as examples of general strategies. Assume that each player has two actions, labeled 0 and 1.

Define the global strategy σ^1 of player 1 by $\sigma^1((a^2_t)_{t\geq 0}) = (a^1_t)_{t\geq 0}$ where $a^1_0 = 1$, and for $t > 0$, $a^1_t = 1$ if $\exists t > t_n \downarrow 0$ such that $a^2_{t_n} = 0$, and $a^1_t = 0$ otherwise. Note that there is a sequence $t > t_n \downarrow 0$ such that $a^2_{t_n} = 0$ if and only if there is a sequence $t_n \downarrow 0$ such that $a^2_{t_n} = 0$.

Define the global strategy σ^2 of player 2 by $\sigma^2((a^1_t)_{t\geq 0}) = (a^2_t)_{t\geq 0}$ where $a^2_0 = 1$, and for $t > 0$ $a^2_t = 1$ if $\exists t > t_n \downarrow 0$ such that $a^1_{t_n} = 1$, and $a^2_t = 0$ otherwise.

Assume that $\sigma((a_t)_{t\geq 0}) = (a_t)_{t\geq 0}$, where $\sigma = (\sigma^1, \sigma^2)$. If there is a sequence $t_n \downarrow 0$ such that $a^2_{t_n} = 0$, then, by the definition of σ^1, $a^1_t = 1$ for every $t \geq 0$, and therefore, by the definition of σ^2, $a^2_t = 1$ for every $t \geq 0$, contradicting the existence of a sequence $t_n \downarrow 0$ such that $a^2_{t_n} = 0$. If there is no sequence $t_n \downarrow 0$ such that $a^2_{t_n} = 0$, then, by the definition of σ^1, $a^1_t = 0$

for every $t > 0$, and therefore, by the definition of σ^2, $a_t^2 = 0$ for every $t > 0$, contradicting the nonexistence of a sequence $t_n \downarrow 0$ such that $a_{t_n}^2 = 0$. Therefore, there is no play $((a_t)_{t\geq 0})$ such that $\sigma((a_t)_{t\geq 0}) = (a_t)_{t\geq 0}$.

Define the global strategy τ^2 of player 2 by $\tau^2((a_t^1)_{t\geq 0}) = (a_t^2)_{t\geq 0}$ where $a_0^2 = 1$, and for $t > 0$, $a_t^2 = 0$ if $\exists t > t_n \downarrow 0$ such that $a_{t_n} = 1$, and $a_t^2 = 1$ otherwise. Note that the two plays $(a_t)_{t\geq 0}$ where $a_0 = (1, 1)$ and either $a_t = (1, 0)$ for all $t > 0$ or $a_t = (0, 1)$ for all $t > 0$ are compatible with the strategy pair (σ^1, τ^2).

Therefore, there is no proper strategic normal form defined over all strategies.

However, equilibrium is unambiguously defined as a profile of strategies with no unilateral beneficial deviation, as formally defined below.

1.7 Continuous-time stochastic games: equilibria

A strategy profile $\sigma = (\sigma^i)_{i\in N}$ (in a continuous-time stochastic game with player set N) is called *admissible* if for every player i and every pure strategy τ^i (recall that τ_t^i is independent of its past actions), the strategy profile (σ^{-i}, τ^i) (where $\sigma^{-i} = (\sigma^j)_{j\neq i}$) defines a unique distribution P_{σ^{-i},τ^i} on plays $h = (z_t, x_t)_{t\geq 0}$ with right continuous state function $[0, \infty) \ni t \mapsto z_t$.

The uniqueness applies only to the the space of plays with right continuous state functions. Therefore we henceforth assume that the state function of a play is right continuous.

An example of an admissible strategy profile in a continuous-time stochastic game is a profile of pure stationary strategies.

An example of a strategy profile that is not admissible is $(\sigma^1, 0)$, where σ^1 is the strategy of player 1 that is defined in the previous section and 0 is the strategy of player 2 that always plays 0. The play with $a_t = (1, 0)$ for every t is the only one that is compatible with $(\sigma^1, 0)$. However, if player 2 deviates to the strategy σ^2 (of the previous section) then there is no play that is compatible with (σ^1, σ^2). Similarly, for any strategy η^1 of player 1, (η^1, σ^2) is not an admissible strategy profile.

An *ε-equilibrium* is an admissible strategy profile for which no single player can benefit by more than ε from a unilateral deviation. An *equilibrium* is a 0-equilibrium.

For example, if $\overrightarrow{w} = (w^i)_{i\in N}$ is a profile of measures on $[0,\infty]$, then a strategy profile σ is an ε-equilibrium of $\Gamma_{\overrightarrow{w}}$ if σ is admissible and for every player i and every strategy τ^i of player i, we have

$$\underline{\gamma}^i_{\overrightarrow{w}}(z,\sigma) \geq \bar{\gamma}^i_{\overrightarrow{w}}(z,\sigma^{-i},\tau^i) - \varepsilon, \tag{1}$$

where

$$\begin{aligned}
\underline{\gamma}^i_{\overrightarrow{w}}(z,\sigma) &:= E^z_{P_\sigma}\int_{[0,\infty]} \underline{g}^i_t\, dw^i(t), \quad \bar{\gamma}^i_{\overrightarrow{w}}(z,\sigma^{-i},\tau^i) := E^z_{P_{\sigma^{-i},\tau^i}}\int_{[0,\infty]} \bar{g}^i_t\, dw^i(t),\\
\underline{g}^i_t = \bar{g}^i_t &:= g^i(z_t,x_t) \text{ for } t<\infty,\\
\underline{g}^i_\infty &:= \liminf_{s\to\infty}\frac{1}{s}\int_0^s g^i_t\, dt, \text{ and } \bar{g}^i_\infty := \limsup_{s\to\infty}\frac{1}{s}\int_0^s g^i_t\, dt.
\end{aligned}$$

Obviously, if $\overrightarrow{w}$ is supported on $[0,\infty)$, as in the case of stationary discounting or finite-horizon games, $\underline{\gamma}^i_{\overrightarrow{w}}(z,\sigma) = \bar{\gamma}^i_{\overrightarrow{w}}(z,\sigma)$, and then we write $\gamma^i_{\overrightarrow{w}}(z,\sigma)$ for short.

In the remainder of this section we introduce the continuous-time equilibrium conditions, which are analogous to those in the discrete-time model, for three special cases: the time-independent discounting game, the finite horizon game, and the limiting-average game.

We show that each one of these equilibrium conditions corresponds to inequality (1) for a suitable profile of measures $\overrightarrow{w}$.

Fix a profile $\overrightarrow{\rho} = (\rho_i)_{i\in N}$ of discounting rates. The strategy profile σ is an ε-equilibrium, $\varepsilon \geq 0$, of the normalized $\overrightarrow{\rho}$-discounted game if σ is an admissible strategy profile and for every state z, every player i, and every strategy τ^i of player i, we have

$$E^z_\sigma \int_0^\infty g^i_t \rho_i e^{-\rho_i t}\, dt \geq E^z_{\sigma^{-i},\tau^i}\int_0^\infty g^i_t \rho_i e^{-\rho_i t}\, dt - \varepsilon.$$

This inequality corresponds to inequality (1) with the profile $\overrightarrow{w} = (w^i)_{i\in N}$ of measures where $w^i([t,\infty)) = e^{-\rho_i t}$ or, equivalently, where w^i is the probability measure on $[0,\infty)$ with density $\rho_i e^{-\rho_i t}$. Indeed, the left-hand side of the above inequality corresponds to $\gamma^i_{\overrightarrow{w}}(z,\sigma)$ and the right-hand side of the above inequality corresponds to $\gamma^i_{\overrightarrow{w}}(z,\sigma^{-i},\tau^i)$.

Fix a finite horizon $s > 0$. The strategy profile σ is an ε-equilibrium, $\varepsilon \geq 0$, of the normalized s-horizon game if σ is an admissible strategy profile

and for every state z, every player i, and every strategy τ^i of player i, we have

$$\gamma_s^i(z,\sigma) := E_\sigma^z \frac{1}{s}\int_0^s g_t^i\,dt \geq E_{\sigma^{-i},\tau^i}^z \frac{1}{s}\int_0^s g_t^i\,dt - \varepsilon.$$

This inequality corresponds to inequality (1) with the profile $\overrightarrow{w} = (w^i)_{i\in N}$ of measures where w^i is the uniform probability distribution on $[0,s]$, namely, where w^i is the probability measure on $[0,s]$ with density $\frac{1}{s}$. Indeed, for this profile of measures the left-hand side of the above inequality corresponds to $\gamma_{\overrightarrow{w}}^i(z,\sigma)$ and the right-hand side of the above inequality corresponds to $\gamma_{\overrightarrow{w}}^i(z,\sigma^{-i},\tau^i)$.

The strategy profile σ is an *ε-equilibrium, $\varepsilon \geq 0$, of the limiting-average game* if σ is an admissible strategy profile and for every state z, every player i, and every strategy τ^i of player i, we have

$$E_\sigma^z \liminf_{s\to\infty} \frac{1}{s}\int_0^s g_t^i\,dt \geq E_{\sigma^{-i},\tau^i}^z \limsup_{s\to\infty} \frac{1}{s}\int_0^s g_t^i\,dt - \varepsilon.$$

This inequality corresponds to inequality (1) with the profile $\overrightarrow{w} = (w^i)_{i\in N}$ of probability measures on $[0,\infty]$ where $w^i(\infty) = 1$. Indeed, for this profile of probability measures the left-hand side of the above inequality corresponds to $\underline{\gamma}_{\overrightarrow{w}}^i(z,\sigma)$ and the right-hand side of the above inequality corresponds to $\bar{\gamma}_{\overrightarrow{w}}^i(z,\sigma^{-i},\tau^i)$.

1.8 Uniform equilibrium

In this section we define a uniform equilibrium payoff of a continuous-time stochastic game. The definition is the natural concept that is analogous to the much-studied uniform equilibrium payoffs of discrete-time stochastic games.

An *ε-uniform equilibrium payoff* of the non-zero-sum continuous-time stochastic game is a vector $u = (u^i(z))_{i\in N, z\in S} \in \mathbb{R}^{S\times N}$ such that there is an admissible strategy profile σ_ε and a sufficiently long finite horizon s_ε such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy σ^i of player i, we have

$$\gamma_s^i(z,\sigma_\varepsilon^{-i},\sigma^i) - \varepsilon \leq u^i(z) \leq \gamma_s^i(z,\sigma_\varepsilon) + \varepsilon.$$

A *uniform equilibrium payoff* of the non-zero-sum continuous-time stochastic game is a vector $u = (u^i(z))_{i\in N, z\in S} \in \mathbb{R}^{S\times N}$ that is an ε-uniform equilibrium payoff for every $\varepsilon > 0$.

An *ε-uniform equilibrium strategy profile* of the non-zero-sum continuous-time stochastic game is an admissible strategy profile σ_ε such that there is a sufficiently long finite horizon s_ε such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy σ^i of player i, we have

$$\gamma_s^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq u^i(z) \leq \gamma_s^i(z, \sigma_\varepsilon) + \varepsilon,$$

where u is a uniform equilibrium payoff.

It is important to note that a strategy profile σ_ε that is an ε-equilibrium strategy in all sufficiently long finite-horizon games – for which there is a sufficiently long finite-horizon s_ε such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy σ^i of player i, we have $\gamma_s^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq \gamma_s^i(z, \sigma_\varepsilon)$ – need not be a 2ε-uniform equilibrium strategy profile. The reason is that the payoff $\gamma_s^i(z, \sigma_\varepsilon)$ need not converge as the horizon s goes to ∞.

The uniform equilibrium and the limiting-average equilibrium concepts are both tailored for the long-horizon undiscounted games. However, they are two different concepts.

An ε-uniform equilibrium strategy need not be a 2ε-limiting-average equilibrium strategy, and, vice versa, an ε-limiting-average equilibrium strategy need not be a 2ε-uniform equilibrium strategy.

A third equilibrium concept' which extends simultaneously both of the above, is the uniform-limiting-average, or uniform-l-average for short.

An *ε-uniform-l-average equilibrium payoff* of the non-zero-sum continuous-time stochastic game is a vector $u = (u^i(z))_{i\in N, z\in S} \in \mathbb{R}^{S\times N}$ such that there is an admissible strategy profile σ_ε and a sufficiently long finite-horizon s_ε such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy τ^i of player i, we have

$$\gamma_s^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq u^i(z) \leq \gamma_s^i(z, \sigma_\varepsilon) + \varepsilon \quad \text{and}$$

$$\bar{\gamma}^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq u^i(z) \leq \underline{\gamma}^i(z, \sigma_\varepsilon) + \varepsilon.$$

An *ε-uniform-l-average equilibrium strategy profile* of the non-zero-sum continuous-time stochastic game is an admissible strategy profile σ_ε for which

there is a sufficiently long finite-horizon s_ε such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy τ^i of player i, we have

$$\gamma_s^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq \underline{\gamma}^i(z, \sigma_\varepsilon) \leq \bar{\gamma}^i(z, \sigma_\varepsilon) \leq \gamma_s^i(z, \sigma_\varepsilon) + \varepsilon \quad \text{and}$$

$$\bar{\gamma}^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq \underline{\gamma}^i(z, \sigma_\varepsilon).$$

A *uniform-l-average equilibrium payoff* of the non-zero-sum continuous-time stochastic game is a vector $u = (u^i(z))_{i \in N, z \in S} \in \mathbb{R}^{S \times N}$ that is an ε-uniform-l-average equilibrium payoff for every $\varepsilon > 0$.

Obviously, a uniform-l-average equilibrium payoff is both a uniform equilibrium payoff and a limiting-average equilibrium payoff.

1.9 Robust equilibria

Our main objective is the study of approximate[3] equilibria (of discounted continuous-time stochastic games) that are insensitive to small imprecision in the specification of players' evaluations of streams of payoffs.

The interest in such a study stems from the fact that in most real-life interactions, the exact duration and/or the exact discounting rate is unknown (and definitely not commonly known).

Moreover, equilibrium analysis of game models whose duration is known, but not commonly known, to all players leads to results that are completely different from those of the model with a fixed finite duration; see [23].

For each continuous-time stochastic game Γ, we wish to identify a large family of sequences of profiles $\overrightarrow{w}_k$ of measures for which the following condition holds.

For every $\varepsilon > 0$ there is a pair consisting of a strategy profile σ and a vector payoff v such that for k sufficiently large σ is an ε-equilibrium of $\Gamma_{\overrightarrow{w}_k}$ with a payoff within ε of v. If the condition holds for a sequence of profiles $\overrightarrow{w}_k$ we say that Γ has $(\overrightarrow{w}_k)$-*robust-equilibria*.

The sequence of profiles $\overrightarrow{w}_k$ of measures might be interpreted as an attempt to construct "better and better" approximations to a desired profile of measures, or to a small imprecision in the profile of measures that define the time-separable payoffs of the players.

[3]It is impossible to find a strategy profile (or a vector payoff) that is an exact equilibrium (or an exact equilibrium payoff) in all sufficiently small changes in the payoff valuations.

The "better and better" approximations are subject to the usual caveats of converging sequences. The convergence can be with respect to a metric or a topology. Note that the smaller the metric, respectively, the weaker the topology, the larger the set of converging sequences of profiles of measures.

Before turning to our definition of converging sequences and the corresponding concept of robust equilibrium, we introduce a simple example of robustness.

Let $\overrightarrow{w}_k$ be a sequence of profiles of measures on $[0,\infty)$ such that $\overrightarrow{w}_k^i$ is, for every i, a cauchy sequence in the norm distance.[4] Note that the sequence of payoffs $\gamma^i_{\overrightarrow{w}_k}(z,\sigma)$ converges uniformly in σ and that it is easy to prove that for any profile of measures $\overrightarrow{w}$ on $[0,\infty)$ and $\varepsilon > 0$, a continuous-time stochastic game Γ (with finitely many states and actions) has an ε-equilibrium. Therefore, for sufficiently large k, an ε-equilibrium σ of $\Gamma_{\overrightarrow{w}_k}$ is, for every $l \geq k$, a 2ε-equilibrium of $\Gamma_{\overrightarrow{w}_l}$ with a payoff within ε of $\gamma_{\overrightarrow{w}_k}(z,\sigma)$. Therefore Γ has $(\overrightarrow{w}_k)$-robust equilibria.

We continue with the introduction of a few examples of sequences of profiles of measures. Let us state upfront that our results prove that a continuous-time stochastic game with finitely many states and action has $(\overrightarrow{w}_k)$-robust equilibria for each one of the sequences $(\overrightarrow{w}_k)$ that appear in examples 1–5 below.

Example 1. Fix a profile $\overrightarrow{w}$ of probability measures that are supported on $[0,\infty)$. Consider the family of all sequences of profiles of probability measures $\overrightarrow{w}_k$ such that $\overrightarrow{w}_k([0,t)) \to_{k\to\infty} \overrightarrow{w}([0,t))$ for every point t with $\overrightarrow{w}(t) = 0$.

Example 2. The sequence of profiles $\overrightarrow{w}_k$ of measures, where w^i_k is the uniform probability measure on $[0,k]$. Note that for this sequence of profiles of measures, the statement that Γ has a uniform equilibrium payoff is equivalent to the statement that Γ has $(\overrightarrow{w}_k)$-robust equilibria.

Example 3. The family of sequences of profiles $\overrightarrow{w}_{\overrightarrow{\rho}_k}$ of measures, $\rho^i_k \downarrow 0$ $\forall i$, where $w^i_{\rho^i_k}([t,\infty)) = e^{-\rho^i_k t}$.

Example 4. The family of sequences of profiles $\overrightarrow{w}_k$ of measures such that $\forall i$, $w^i_k([a+c,b+c))$ is nonincreasing in c ($\forall 0 \leq a < b < \infty$) and $w^i_k([0,1)) \to_{k\to\infty} 0$.

[4]The norm distance between two probability measures P and Q on a measurable space $(X,\mathcal{X})$ is defined by $\|P-Q\| := 2\sup_{B\in\mathcal{X}} |P(B)-Q(B)|$. If X is finite (or countable) and $\mathcal{X}$ consists of all subsets of X, then $\|P-Q\| = \sum_{x\in X} |P(x)-Q(x)|$.

Note that each one of the sequences of the profiles of measures in Examples 2 and 3 is in the family of sequences described in Example 4.

Example 5. The family of sequences of profiles $\overrightarrow{w}_k$ of measures that are supported on $[0,\infty)$ and such that $\forall i$, $w^i_k([a+c,b+c))$ is nonincreasing in c ($\forall 0 \leq a < b < \infty$) and $w^i_k([t,\infty)) \to_{k\to\infty} \theta^i e^{-\rho^i t}$, where $\overrightarrow{\rho}$ is a fixed profile of nonnegative numbers $\rho^i \geq 0$ and $\overrightarrow{\theta}$ is a fixed profile with $0 \leq \theta^i \leq 1$.

Note that each one of the sequences of the profiles of measures in Example 4 is in the family of sequences described in Example 5.

Our robustness results will require that the sequence $\overrightarrow{w}_k$ w^*-converge; i.e., for every continuous function $[0,\infty] \ni t \mapsto g(t) \in \mathbb{R}$ the integrals $\int_0^\infty g(t)\, dw^i_k(t)$ converge as $k \to \infty$.

In a w^*-converging sequence of measures $\overrightarrow{w}_k$ on $[0,\infty)$, part of the mass of w^i_k may be pushed to infinity. Therefore, there need not be a profile $\overrightarrow{w}$ of measures on $[0,\infty)$ that represent the limit.

The limit is represented by a profile $\overrightarrow{w}$ of measures on $[0,\infty]$ such that for every player i and every continuous function $g : [0,\infty] \to \mathbb{R}$, we have $\int_0^\infty g(t)\, dw^i_k(t) \to_{k\to\infty} \int_{[0,\infty]} g(t)\, dw^i(t)$.

We equip the space of measures W on $[0,\infty]$ with the minimal topology (called the w^*-topology) such that for every continuous function $[0,\infty] \ni t \mapsto f_t$, the function $W \ni w \mapsto \int_{[0,\infty]} f_t\, dw(t)$ is continuous. $\overrightarrow{W} := W^N$ is equipped with the product topology.

Given $\overrightarrow{w} \in \overrightarrow{W}$, we say that Γ has *$\overrightarrow{w}$-robust equilibria* if there is a vector payoff v (called a *$\overrightarrow{w}$-time-separable-robust equilibrium payoff*) such that for every $\varepsilon > 0$ there is a strategy profile σ (called a *$\overrightarrow{w}$-robust ε-equilibrium strategy*) and a neighborhood U of $\overrightarrow{w}$, such that for every $\overrightarrow{u} \in U$, σ is an ε-equilibrium of $\Gamma_{\overrightarrow{u}}$ with a payoff within ε of v.

A *discounting measure* is a measure $w \in W$ with $w([a+c,b+c))$ nonincreasing in $c \geq 0$ for every $0 \leq a < b < \infty$. A discounting measure w is absolutely continuous[5] on $(0,\infty)$ with $\frac{dw}{dt}(t)$ nonincreasing. It may have atoms at 0 and ∞. Small values of $w([0,1])/w([0,\infty])$ correspond to valua-

[5]Recall that a finite measure μ on Borel subsets of the real line is *absolutely continuous* if for every positive number $\varepsilon > 0$ there is a positive number $\delta > 0$ such that $\mu(A) < \varepsilon$ for all Borel sets A of Lebesgue measure less than δ. If μ is absolutely continuous than there exists a Lebesgue integrable function g, denoted by $\frac{d\mu}{dt}$, on the real line such that $\mu(A) = \int_A g\, dt$ for all Borel subsets A of the real line.

tions of patient players. The space of all discounting measures is denoted by W_d.

Given $\overrightarrow{w} \in \overrightarrow{W}_d := (W_d)^N$, we say that Γ has $\overrightarrow{w}$-*discounting-robust equilibria* if there is a vector payoff v (called a $\overrightarrow{w}$-*discounting-robust equilibrium payoff*) such that for every $\varepsilon > 0$ there is a strategy profile σ called a $\overrightarrow{w}$-*discounting-robust* ε-*equilibrium* (*strategy*) and a neighborhood U of $\overrightarrow{w}$, such that for every $\overrightarrow{u} \in U \cap \overrightarrow{W}_d$, σ is an ε-equilibrium of $\Gamma_{\overrightarrow{u}}$ with a payoff within ε of v.

The continuous-time stochastic game Γ has *discounting-robust* equilibria if for every $\overrightarrow{w} \in \overrightarrow{W}_d$, Γ has $\overrightarrow{w}$-discounting-robust equilibria.

1.10 The results

The main result is that every continuous-time stochastic game with finitely many states and actions has discounting-robust equilibria.

The set $\overrightarrow{W}_d^1$ of all profiles of probability discounting measures is infinite. However, as it is a compact subset of $\overrightarrow{W}$, the existence of discounted-robust equilibria implies (and is equivalent to) the following finiteness result.

For every $\varepsilon > 0$, there is a finite open cover $(U_\alpha)_{\alpha \in \mathcal{A}}$ of $\overrightarrow{W}_d^1$, a finite list of strategy profiles $(\sigma_\alpha)_{\alpha \in \mathcal{A}}$, and a finite list of vector payoffs $(v_\alpha)_{\alpha \in \mathcal{A}}$, such that for every profile of discounting measures $\overrightarrow{w}$ in U_α, σ_α is an ε-equilibrium of $\Gamma_{\overrightarrow{w}}$ with a payoff within ε of v_α.

The discounting measure can vary among players. Therefore the main result applies also to cases where some of the players are extremely patient (e.g., a government) and other players' patience can vary (e.g., as a function of each player's uncertain life horizon).

The discounting measure of player i, w^i, may have a positive mass at infinity, representing the weight of the discounting measure on the distant future.

If both $w^i([0,\infty))$ and $w^i(\{\infty\})$ are positive, the discounting measure w^i represents a valuation that is an average of a classical (not necessarily stationary) discounting valuation and a valuation of an extremely patient player.

The existence of $\overrightarrow{w}$-discounting-robust equilibria, where each discounting measure w^i is supported on $\{\infty\}$, implies the existence of a uniform-l-average equilibrium payoff.

The existence of a uniform (or a limiting-average) equilibrium payoff in any continuous-time stochastic game with finitely many[6] states and actions, is in sharp contrast to our knowledge about uniform equilibrium in discrete-time stochastic games, where it is still unknown whether all discrete-time stochastic games with finitely many states and actions have a uniform (or a limiting-average) equilibrium payoff.

A continuous-time stochastic game with finitely many states and actions need not have a $\overrightarrow{w}$-robust equilibrium payoff for every $\overrightarrow{w} \in \overrightarrow{W}$. However, if the profile of measures $\overrightarrow{w} \in \overrightarrow{W}$ is supported on $[0, \infty)$, then it has a $\overrightarrow{w}$-robust equilibrium.

Earlier studies of continuous-time stochastic games include [44, 4, 5]. An important difference between the continuous-time game model of the present paper and those of the other papers is that in the present paper we consider all strategies in the continuous-time game, while each one of the other papers confines the strategy space to a subclass of all strategies.

On the other hand, the present paper focuses on the model with finitely many states and actions, while the earlier papers include also more general action and state spaces.

[44] studies two-person zero-sum finite-horizon stochastic games with a terminal payoff, where players are confined to Markov strategies, and proves the existence of the value and optimal strategies.

[4] studies a subclass of two-person zero-sum stochastic games with the average payoff criterion, where players are confined to continuous Markov strategies (see Section 2.1 for the definition), and proves the existence of the value and optimal strategies.

[5] study two-person non-zero-sum discounted stochastic games, where players are confined to continuous[7] Markov strategies, and proves the existence of the value and optimal strategies.

[6]The assumption of finitely many states is essential even in the case of a single player; see, e.g., [25, Section 1.4]. The assumption of finitely many actions is essential even in the two-person zero-sum case with finitely many states; see [40].

[7]The confinement there is more severe, as it requires the strategy of a player to be a Markov strategy such that for any Markov strategy of the other player the transition rates are continuous. See [5, Definition 3.1.]. However, unless the game is degenerate, there are no such strategies. Nevertheless, the proofs there hold also for continuous Markov strategies.

For completeness, we derive the existence of stationary equilibrium in the continuous-time non-zero-sum discounted stochastic game (without the restriction to continuous Markov strategies and more generally without the restriction to (memoryless) Markov strategies).

2 The model

A *continuous-time stochastic game* (with finitely many states and actions) is defined by: a finite set of players N; a finite set of states S; for each $z \in S$ and each player $i \in N$ a finite set of actions $A^i(z)$; a (vector-valued) payoff function $g : \mathcal{A} \to \mathbb{R}^N$, where $\mathcal{A} = \{(z, a) : a \in A(z)\}$ and $A(z) = \times_{i \in N} A^i(z)$; and for each $z' \neq z \in S$ and $a \in A(z)$ a real-valued transition rate $\mu(z', z, a) \geq 0$. The i-th component of g is denoted by g^i.

For notational convenience we set $A^i = \cup_{z \in S} A^i(z)$ and $A = \times_{i \in N} A^i$. The i-th coordinate, $i \in N$, of $a \in A(z)$ is denoted by a^i.

The set $A^i(z)$ represents the set of feasible actions of player i when the state is z. An element $a \in A(z)$, $a = (a^i)_{i \in N}$ with $a^i \in A^i(z)$, is called an *action profile*.

The interpretation of the payoff function and of the transition rates is that when the state is $z \in S$ and players play the action profile $a \in A(z)$ during the infinitesimal time dt, then the payoff to player i is $g^i(z, a)dt$ and the state moves to state $z' \neq z$ with probability $\mu(z', z, a)dt$ and stays in state z with probability $1 + \mu(z, z, a)dt$, where $\mu(z, z, a) := -\sum_{z' \neq z} \mu(z', z, a)$.

A *pure play* of the continuous-time stochastic game is a measurable function $h : [0, \infty) \to S \times A$, $t \mapsto h(t) = (z_t, a_t)$, with $a_t \in A(z_t)$, and $t \mapsto z_t$ right continuous.

A pure play h is identified with a pair of functions $h^S : [0, \infty) \to S$, which is the restriction of h to the first coordinate, and $h^A : [0, \infty) \to A$, which is the restriction of h to the second coordinate.

Given a pure play h we define h_t as the restriction of the first coordinate of h to the time interval $[0, t]$ and the restriction of the second coordinate to $[0, t)$. h_t is identified with the pair $(h^S_{[0,t]}, h^A_{[0,t)})$, where $h^S_{[0,t]}$ is the restriction of h^S to the time interval $[0, t]$ and $h^A_{[0,t)}$ is the restriction of h^A to the time interval $[0, t)$

The unnormalized, respectively, normalized, ρ-discounted payoff ($\rho > 0$) of a pure play h is $\int_0^\infty e^{-\rho t} g(z_t, a_t)\, dt$, respectively, $\rho \int_0^\infty e^{-\rho t} g(z_t, a_t)\, dt$. The

s-horizon normalized payoff of a pure play[8] h is $\frac{1}{s}\int_0^s g(z_t, a_t)\,dt$.

A *play* of the continuous-time stochastic game is a (measurable) function $h : [0,\infty) \to S \times \Delta(A)$, $t \mapsto h(t) = (z_t, x_t)$, with $x_t \in \Delta(A(z_t))$, where $\Delta(*)$ denotes all probabilities on the set $*$.

Given a play h we define h_t, as we did for the pure play, i.e., as the restriction of the first coordinate of h to the time interval $[0,t]$ and the restriction of the second coordinate to $[0,t)$.

The ρ-discounted payoff of a play h is $\int_0^\infty e^{-\rho t} g(z_t, x_t)dt$, where $g(z,x) = \sum_{a\in A(z)} x(a)g(z,a)$ is the linear extension of g. The s-horizon normalized payoff of a play h is $\frac{1}{s}\int_0^s g(z_t, x_t)\,dt$.

2.1 Strategies in continuous-time games

In this section we define and discuss several classes of strategies: stationary strategies, Markov strategies, discretimized strategies, and general strategies.

Each one of these classes of strategies is a proper subclass of the next one in the list. In each equilibrium existence result we desire that the equilibrium strategies be as simple as possible.

However, one has to recall that the equilibrium strategy profile should be such that there is no unilateral beneficial deviation, and a player can deviate to any general strategy. Therefore, an equilibrium strategy profile need be such that, in particular, any unilateral deviation to a general strategy yield a strategy profile that defines a probability distribution on plays.

2.1.1 Markov strategies

Let $X^i(z)$, $X(z)$, X^i, and X denote all probability distributions over $A^i(z)$, $A(z)$, A^i, and A, respectively.

A *Markov strategy* of player i is defined by a measurable function $\sigma^i : S \times [0,\infty) \to X^i$ with $\sigma^i(z,t) \in X^i(z)$. A profile σ of Markov strategies $(\sigma^i)_{i\in N}$ defines a Markov strategy profile $\sigma : S \times [0,\infty) \to X$ with $\sigma(z,t) \in X(z)$ by $\sigma(z,t)[a] = \prod_{i\in N} \sigma^i(z,t)[a^i]$.

[8]Much of the theory developed remains intact even if the integral $\int_a^b f(t)\,dt$ of a real-valued bounded function f refers to a fixed monotonic linear functional (on the space of real-valued bounded functions), with $\int_a^b C\,dt = C(b-a)$ and $\int_a^c f\,dt = \int_a^b f\,dt + \int_b^c f\,dt$ for all $0 \le a, b, c$. This defines the discounted and s-horizon payoffs over all plays, not necessarily measurable ones.

A *Markov correlated strategy* is defined by a measurable function $\sigma : S \times [0, \infty) \to X$ with $\sigma(z, t) \in X(z)$. Note that a profile of Markov strategies is a special case of a Markov correlated strategy.

A Markov strategy σ^i of player i is *continuous* if $\sigma^i(z, t)$ is continuous in t. Similarly, a Markov correlated strategy σ is *continuous* if $\sigma(z, t)$ is continuous in t.

A *stationary* strategy of player i is a Markov strategy σ^i where $\sigma^i(z, t)$ is independent of t.

A profile σ of Markov strategies, or more generally, a Markov correlated strategy σ, and an initial state z_0 define on the space of right-continuous functions $t \mapsto z_t \in S$ a unique probability distribution P_σ that for every $z \neq z_t$ satisfies the equality

$$P_\sigma(z_{t+\delta} = z \mid h_t) = \int_t^{t+\delta} \mu(z, z_t, \sigma(z_t, s))\, ds + o(\delta), \tag{2}$$

where $\mu(z', z, x) := \sum_{a \in A} \mu(z', z, a)x(a)$.

This unique distribution is supported on the space H^S of all functions $t \mapsto z_t$ that have left limits and are right continuous at every $t \geq 0$. (The fact that with P_σ probability 1, the function $t \mapsto z_t$ has left and right limits at any t follows from condition (2) and the right continuity follows from our assumption.)

In order to define the probability P_σ on the space of right continuous functions $t \mapsto z_t$, it suffices to define the distribution of $(s_k, z_{s_k})_{k \geq 0}$ where s_k is the time of the k-th state change.

Conditional on $(s_\ell, z_{s_\ell})_{\ell \leq k}$ the distribution of $s_{k+1} - s_k$ is given by

$$P_\sigma(s_{k+1} - s_k \geq a \mid (s_\ell, z_{s_\ell})_{\ell \leq k}) = e^{\int_{s_k}^{s_k + a} \mu(z_{s_k}, z_{s_k}, \sigma(z_{s_k}, s))\, ds} \text{ for } a \geq 0,$$

and (using the Lebesgue density theorem) conditional on $(s_\ell)_{\ell \leq k+1}$ and $(z_{s_\ell})_{\ell \leq k}$, the distribution of $z_{s_{k+1}}$ is given by

$$P_\sigma(z_{s_{k+1}} = z \mid (s_\ell)_{\ell \leq k+1},\, (z_{s_\ell})_{\ell \leq k}) = \frac{-\mu(z, z_{s_k}, \sigma(z_{s_k}, s_{k+1}))}{\mu(z_{s_k}, z_{s_k}, \sigma(z_{s_k}, s_{k+1}))} \text{ a.e.}$$

The space of mixed strategies that are mixtures of Markov strategies depends on the measurable structure on the space of Markov strategies. Obviously, there are many measurable structures on the space of Markov strategies. One of them is derived from the minimal topology for which, for every

real-valued continuous function f on $S \times X^i$ and all $0 \leq a < b < \infty$, the function $\sigma^i \mapsto \int_a^b f(z, \sigma^i(z,t))\, dt$ is continuous.

2.1.2 Discretimized strategies

Fix a strictly increasing sequence of times $\mathcal{T} = (t_k)_{k \geq 0}$ with $t_0 = 0$ and $\mathbb{R} \ni t_k \to \infty$ as $k \to \infty$.

A $\mathcal{T}$-*discretimized strategy* of player i specifies for each k and $h_{t_k} \in H_{t_k}$ a function $\sigma^i_{h_{t_k}} : S \times [t_k, t_{k+1}) \to X^i$ with $\sigma^i_{h_{t_k}}(z,t) \in X^i(z)$.

The interpretation of this description of σ^i is that at time $t_k \leq t < t_{k+1}$ the strategy σ^i selects an action $\sigma^i_{h_{t_k}}(z_t, t) \in X^i(z)$ as a function of the state z_t at time t and the history of play $h_{t_k} \in H_{t_k}$ up to time t_k.

A profile $\sigma = (\sigma^i)_{i \in N}$ of $\mathcal{T}$-discretimized strategies defines for each k and $h_{t_k} \in H_{t_k}$ a function $\sigma_{h_{t_k}} : S \times [t_k, t_{k+1}) \to X$ with $\sigma_{h_{t_k}}(z,t) := \otimes_{i \in N} \sigma^i_{h_{t_k}}(z,t) \in \times_{i \in N} X^i(z) \subset X(z)$ (where $\otimes$ stands for the product of measures).

A $\mathcal{T}$-*discretimized correlated strategy* specifies for each k and $h_{t_k} \in H_{t_k}$ a function $\sigma_{h_{t_k}} : S \times [t_k, t_{k+1}) \to X$ with $\sigma_{h_{t_k}}(z,t) \in X(z)$. Note that a profile of $\mathcal{T}$-discretimized strategies is a special case of a $\mathcal{T}$-discretimized correlated strategy.

A $\mathcal{T}$-discretimized correlated strategy (with suitable measurability assumptions) defines (for each initial state z_0) a (unique) probability distribution P_σ on plays (with right continuous state function $t \mapsto z_t$) such that for $t_k \leq t < t_{k+1}$ and $z \in S$,

$$x_t = \sigma_{h_{t_k}}(z_t, t) \quad \text{and} \tag{3}$$

$$P_\sigma(z_{t+\delta} = z \mid h_t) = 1_{z=z_t} + \int_t^{t+\delta} \mu(z, z_t, \sigma_{h_{t_k}}(z_t, s))\, ds + o(\delta), \tag{4}$$

where $1_{z=z_t} = 1$ if $z = z_t$, and $1_{z=z_t} = 0$ if $z \neq z_t$.

In all our results, the profile of equilibrium strategies will be a profile of $\mathcal{T}$-discretimized strategies (and recall that $\mathcal{T}$ is a sequence of real numbers).

However, a deviation of a player (even in the single-player case) need not be a $\mathcal{T}^*$-discretimized strategy with $\mathcal{T}^*$ being a sequence of real numbers.

For example, the strategy that for $t \geq s_1$ plays one action $a_t = a$, and for $t < s_1$ plays another action $a_t = b \neq a$, is not a $\mathcal{T}$-discretimized strategy w.r.t. a strictly increasing sequence $\mathcal{T} = (t_k)_{k \geq 0}$ of real numbers.

Let us show, informally, that a profile of strategies that is obtained by a unilateral deviation from a profile of $\mathcal{T}$-discretimized strategies defines a unique probability on plays.

The derivation is not formal as we have not yet introduced the definition of general strategies. However, aside from the needed suitable measurability assumptions, the general form of the behavior of a strategy of player i when confronted with $\mathcal{T}$-discretimized strategies of the other players should be clear.

Following our definition (in the next section) of general strategies, it will be clear that the assumed behavior of a general unilateral deviation holds for any general strategy, and therefore, with the suitable measurability assumptions, the informal arguments below turn into a formal proof.

Let σ be a profile of $\mathcal{T}$-discretimized strategies, where $\mathcal{T} = (t_k)_{k\geq 0}$ is an increasing sequence of real numbers with $t_0 = 0$ and $t_k \to_{k\to\infty} \infty$.

Let $s_0 = 0$ and let s_l be the time of the l-th state change. If there are fewer than l state changes than $s_l = \infty$. Note that s_l is a function of the play.

Define $t_k^l = s_l \vee (s_{l+1} \wedge t_k)$, where $a \vee b$ stands for $\max(a, b)$ and $a \wedge b$ stands for $\min(a, b)$. Note that $s_l \leq t_k^l \leq s_{l+1}$ and $t_k^l \to_{k\to\infty} s_{l+1}$.

Assume that player i is deviating to a strategy τ^i. By induction on l assume that the distribution of h_{s_l} is uniquely defined by the strategy profile (σ^{-i}, τ^i). We will show that this assumption implies that the distribution of $h_{s_{l+1}}$ is uniquely defined.

The first step shows that for $s_l \leq t < s_{l+1}$, the action profile x_t is a well-defined function of h_{s_l} and the strategy profile (σ^{-i}, τ^i).

For any $j \neq i$, the strategy σ^j is a $\mathcal{T} = (t_k)_{k\geq 0}$-discretimized strategy. Therefore, for any $t_k^l \leq t < t_{k+1}^l$, x_t^j is well defined as a function of $h_{t_k^l}$ and σ^j.

Therefore, for any $t_k^l \leq t < t_{k+1}^l$, $x_t^{-i} = (x_t^j)_{j\neq i}$ is well defined as a function of $h_{t_k^l}$ and $\sigma^{-i} = (\sigma^j)_{j\neq i}$.

As x_t^{-i} is well defined for all $t_k^l \leq t < t_{k+1}^l$ as a function of $h_{t_k^l}$ and σ^{-i}, player i does not learn any new information from the actions x_t^{-i}, $t_k^l \leq t < t_{k+1}^l$. Therefore, (given σ^{-i}) the action of player i at time $t_k^l \leq t < t_{k+1}^l$, x_t^i, which is defined by τ^i, is a function of $h_{t_k^l}$.

Therefore, for any $t_k^l \leq t < t_{k+1}^l$, the action profile x_t is for all $t_k^l \leq t < t_{k+1}^l$ a well-defined function of $h_{t_k^l}$ and the strategy profile (σ^{-i}, τ^i).

By induction on k, the action profile x_t is defined for every $s_l \leq t < s_{l+1}$ as a function of h_{s_l} and the strategy profile (σ^{-i}, τ^i).

For $s_l \leq t$, we denote by $x_{l,t}$ the action profile defined by the strategy profile (σ^{-i}, τ^i) at time t conditional on $s_{l+1} > t$. Note that the action profile $x_{l,t}$ is a well-defined function of h_{s_l} and the strategy profile (σ^{-i}, τ^i).

This enables us to define the action profile x_t at time $s_l \leq t < s_{l+1}$ and the conditional distribution of $(s_{l+1}, z_{s_{l+1}})$ as a function of h_{s_l} and the strategy profile $\hat{\sigma} := (\sigma^{-i}, \tau^i)$, as follows.

For $s_l \leq t < s_{l+1}$,

$$x_t = x_{l,t}, \tag{5}$$

$$P_{\hat{\sigma}}(s_{l+1} \geq s_l + \theta \mid h_{s_l}) = e^{\int_{s_l}^{s_l+\theta} \mu(z_{s_l}, z_{s_l}, x_{l,t})\, dt}, \tag{6}$$

and (using the Lebesgue density theorem) for $z' \neq z_{s_l}$,

$$P_{\hat{\sigma}}(z_{s_{l+1}} = z' \mid h_{s_l}, s_{l+1}) = \frac{-\mu(z', z_{s_l}, x_{l,s_{l+1}})}{\mu(z', z', x_{l,s_{l+1}})} \text{ a.e.} \tag{7}$$

By induction on l we deduce, using Tulcea's extension theorem (see, e.g., [39, Theorem 1.1.9]), that the strategy profile (σ^{-i}, τ^i) defines a unique probability distribution on plays.

2.1.3 General strategies

The space of all plays is denoted by H. The space of all (measurable) functions $h : [0, \infty) \to h(t) = (z_t, x_t^{-i})$, with $x_t^{-i} \in \Delta(\times_{N \ni j \neq i} A^j)$ and $z_t \in S$, is denoted by H^{-i}.

A *strategy* σ^i of player i is a (measurable) function $\sigma^i : H^{-i} \times [0, \infty) \to X^i$ with $\sigma^i(h, t) \in X^i(z_t)$ and such that for all triples $(h', h, t) \in H^{-i} \times H^{-i} \times [0, \infty)$ with $h'_t = h_t$ we have $\sigma^i(h, t) = \sigma^i(h', t)$. This last condition asserts that a player cannot decide his current mixed action based on future events. Note that a discretimized strategy is a special case of a (general) strategy.

A play $h \in H$ is *compatible* with the strategy profile $\sigma = (\sigma^i)_{i \in N}$ if $x_t^i = \sigma^i(h^{-i}, t)$ for every $t \geq 0$.

A *correlated strategy* σ is a (measurable) function $\sigma : H^S \times [0, \infty) \to X$ with $\sigma(h, t) \in X(z_t)$ and such that for all triples $((z'_s)_{s \geq 0}, (z_s)_{s \geq 0}, t) \in H^S \times H^S \times [0, \infty)$ with $z'_s = z_s \ \forall s \leq t$ we have $\sigma((z_s)_{s \geq 0}, t) = \sigma((z'_s)_{s \geq 0}, t)$.

A play $h = ((z_s)_{s\geq 0}, (x_s)_{s\geq 0}) \in H$ is *compatible* with the correlated strategy σ if $x_t = \sigma((z_s)_{s\geq 0}, t)$ for every $t \geq 0$.

A profile of strategies need not define unambiguously a probability distribution over plays; see, e.g., Proposition 1.

An *admissible* profile of strategies $\sigma = (\sigma^i)_{i\in N}$ is a strategy profile σ such that for every player i and every strategy τ^i of player i the strategy profile (σ^{-i}, τ^i) defines a (unique and thus unambiguously) probability distribution P_{σ^{-i},τ^i} on plays (with right continuous state function $t \mapsto z_t$).

Throughout, we have used measurability assumptions in our definitions. The meaning of a play and of a strategy depends on the σ-algebras that define measurability. However, we skipped the spelling out of the σ-algebras for the following reason.

The space of measurable functions from a measurable space $(X, \mathcal{X})$ to a measurable space $(Y, \mathcal{Y})$ is monotonic increasing in $\mathcal{X}$ (namely, for each fixed space X and measurable space $(Y, \mathcal{Y})$, the larger the σ-algebra $\mathcal{X}$, the larger is the space of measurable functions from the measurable space $(X, \mathcal{X})$ to the measurable space $(Y, \mathcal{Y})$) and monotonic decreasing in $\mathcal{Y}$.

In constructing a strategy profile in an existence result, it is desirable that this strategy be in a space of strategies whose meaning, implementation, and interpretation are as simple as possible. To this end, we wish to take a minimal reasonable σ-algebra on the space of plays in the domain of a strategy, and a maximal reasonable σ-algebra in the image of a strategy.

On the other hand, when we wish to demonstrate that there are no unilateral beneficial deviations, it is desirable to demonstrate it for the most general deviation. To this end, we wish to take a maximal reasonable σ-algebra on the space of plays in the domain of a strategy, and a minimal reasonable σ-algebra on the image of a strategy.

2.2 Discussion of strategies

The realization of an increasing sequence of real numbers is a well-ordered subset of the reals. This enables us to define unambiguously the distribution on plays that is defined by a profile of strategies that is obtained by a unilateral deviation from a profile of discretimized strategies.

In this section we illustrate the difficulties that arise when the set of decision times is not well ordered.

2.2.1 Prelude: the set-theoretic setup

Consider a totally ordered set $\mathcal{T}$. For every element $t \in \mathcal{T}$ let $\mathcal{P}_t := \{t' \in T : t' < t\}$. Fix a set A with at least two elements. If we interpret the set $\mathcal{T}$ as the set of action times, and A as the set of single-stage action profiles, then a *play* is an element $h \in A^{\mathcal{T}}$, and the value of h at time t, $h(t)$, is interpreted as the action profile at time t. A history of play up to (and not including) time t is an element $A^{\mathcal{P}_t}$, and given a play h we denote by h_t its restriction to $A^{\mathcal{P}_t}$. A *local pure strategy profile* σ is a list of functions $\sigma_t : A^{\mathcal{P}_t} \to A$, $t \in \mathcal{T}$. The *integral* of a local pure strategy profile σ is the set of all plays h such that $h(t) = \sigma_t(h_t)$ for every t. A local pure strategy profile is *integrable* if its integral is nonempty; i.e., there is at least one play $h \in A^{\mathcal{T}}$ such that for every $t \in \mathcal{T}$ we have $h(t) = \sigma_t(h_t)$.

The following simple proposition demonstrates that when the set of times is not well ordered then there are local strategies that are not integrable, and there are local strategies whose integral contains more than one element. On the other hand, if the set of times is well ordered, the integral contains a single element.

Proposition 1. *The following conditions are equivalent: 1) every local pure strategy profile σ is integrable, 2) the set of times $\mathcal{T}$ is well ordered, and 3) the integral of every local pure strategy contains at most one element.*

Proof. Assume that $\mathcal{T}$ is well ordered. We define for every $t \in \mathcal{T}$ the action profile $h(t)$ by transfinite induction: (1) for the first element t^* of $\mathcal{T}$, $h_{t^*} = \emptyset$ and therefore we set $h(t^*) = \sigma_{t^*}(\emptyset)$; (2) if $h(t')$ has been defined for every $t' < t$, then h_t has been defined and we set $h(t) = \sigma_t(h_t)$. As $\mathcal{T}$ is well ordered the action profile $h(t)$ is defined uniquely for every $t \in \mathcal{T}$. Therefore, the play h is the unique element in the integral of σ.

If $\mathcal{T}$ is not well ordered, there is an infinite decreasing sequence $t_1 > t_2 > \ldots$ of times in $\mathcal{T}$. Let a, b be two distinct elements of A. Define the local pure strategy profile σ by $\sigma_t(*) = a$ if $t \notin \{t_j : j \geq 1\}$, $\sigma_{t_i}(h_{t_i}) = b$ if $\limsup_{k\to\infty} |\{i < j \leq i + k : h(t_j) = a\}|/k \geq 1/2$, and $\sigma_{t_i}(h_{t_i}) = a$ if $\limsup_{k\to\infty} |\{i < j \leq i+k : h(t_j) = a\}|/k < 1/2$. Note that for every play h, $\limsup_{k\to\infty} |\{i < j \leq i + k : h(t_j) = a\}|/k$ is independent of i. Therefore, if h is in the integral of σ, $h(t_i)$ is a constant and $\sigma_{t_i}(h_{t_i}) \neq h(t_i)$, contradicting the local definition of σ.

Define the local pure strategy τ by $\tau_{t_i}(h_{t_i}) = b$ if $|\{j : i < j \text{ and } h(t_j) = a\}| < \infty$, and $\tau_t(*) = a$ otherwise. The play that plays a everywhere, and

the play that plays b at time $t = t_j$, $j \geq 1$, and a elsewhere, i.e., at times $t \notin \{t_j : j \geq 1\}$, are in the integral of τ. □

A *delayed local pure strategy profile* is a local strategy σ such that there is a well-ordered[9] subset $\mathcal{T}^*$ of $\mathcal{T}$ such that for every $t \in \mathcal{T}$ that is not a maximal element of $\mathcal{T}$ there is $t^* \in \mathcal{T}^*$ such that $t^* \leq t < \bar{t}^*$, where $\bar{t}^*$ is the least element in $\mathcal{T}^*$ that is $> t^*$, and for $t^* \leq t < \bar{t}^*$ we have

$$\sigma_t(h_t) = \sigma_t(h'_t) \text{ whenever } h_{t^*} = h'_{t^*}.$$

Proposition 2. *A delayed local pure strategy profile is integrable and its integral contains a unique element.*

Proof. The restriction of the function $h(t)$ to the interval $t^* \leq t < \bar{t}^*$ is defined by transfinite induction on $t^* \in \mathcal{T}^*$. □

Proposition 2 is standard and classical, and is used in the theory of differential games, when the set of decision times $\mathcal{T}$ is either a finite interval $[0, T]$ or the nonnegative real numbers $[0, \infty)$.

2.3 Strategies that observe and select mixed actions

In discrete-time games with perfect monitoring we focus on strategies that observe the pure past actions and select a pure action. Below we discuss our choice of the model, where players observe past mixed actions and strategies select mixed actions as a function of past observed variables.

These assumptions are conceptually innocuous in the case where each player is a continuum of agents and the observable variables are the statistics of actions of the continuum of agents. In the continuous-time model, we find these assumptions natural even in the case where each player represents a single decision maker. This is motivated in part by the limitations on the players' perception.

The perception of players is not without limitation. A person watching a sequence of blue and yellow pictures will be unable to tell the exact times when the blue ones were presented if the switching rate crosses some threshold. In fact, he will observe a greenish picture, and the greenness will change as a function of the fraction of time in which the blue pictures were

[9] A well-ordered subset of an ordered set is a subset such that every nonempty subset of it has a minimal element.

presented. It is therefore desirable to model the observation of the past by the observation of time averages of pure actions, and time averages of pure actions correspond to mixed actions. Given this limitation on perception, it is also natural to model the interaction by assuming that the players select mixed actions.

These assumptions – of observing and selecting mixed actions – are also technically convenient. They result in a tractable analytic model. In addition, the analysis of this analytic model leads to analogous results in the asymptotic theory of discrete-time stochastic games with short-stage durations and in the (discretimized approach to) continuous-time stochastic games where players are subject to a short delay in observing other players' actions; see [27, 28].

3 Non-zero-sum continuous-time stochastic games

3.1 Useful inequalities

Fix a continuous-time stochastic game $\Gamma = \langle N, S, A, g, \mu \rangle$. In this section we introduce several inequalities that are often used in the paper.

Recall that H^S is the space of all functions $t \mapsto z_t \in S$ that are right continuous and have left limits everywhere. Let $\mathcal{H}_t^S$, respectively, $\mathcal{H}^S$, be the σ-algebra of subsets of H^S that is generated by all the S-valued random variables z_s, $s \leq t$, respectively, $s < \infty$.

The concept of a $\mathcal{T}$-discretimized strategy with respect to an increasing sequence of real numbers, which was defined earlier, extends naturally to the case where $\mathcal{T}$ is an increasing sequence of $(\mathcal{H}_t^S)_{t\geq 0}$-stopping times.

Fix a strictly increasing sequence of $(\mathcal{H}_t^S)_{t\geq 0}$-stopping times $\mathcal{T} = (t_k)_{k\geq 0}$ with $t_0 = 0$ and $t_k \to \infty$ as $k \to \infty$.

A special case of interest is where $\mathcal{T} = (s_k)_{k\geq 0}$, where $s_0 = 0$ and $s_k \leq \infty$ is the time of the k-th state change.

In Section 2.1.2 we showed that if σ is a profile of $\mathcal{T}$-discretimized strategies, where $\mathcal{T}$ is an increasing sequence of real numbers, and τ^i is a strategy of player i, then, for $s_l \leq t < s_{l+1}$, the action-profile x_t is a well defined function of h_{s_l} and the strategy (σ^{-i}, τ^i). Therefore, there is an $(s_k)_{k\geq 0}$-discretimized correlated strategy $\hat{\sigma}$ such that $\hat{\sigma}$ and (σ^{-i}, τ^i) define the same distribution on plays.

The same property holds also in the case where each strategy σ^j is a discretimized strategy with respect to a strictly increasing sequence (or even a well-ordered countable set) of $(\mathcal{H}_t^S)_{t\geq 0}$-stopping times.

The lemmas in this section assume that σ is a $(s_k)_{k\geq 0}$-discretimized correlated strategy. Therefore, the conclusions of the lemmas hold for any profile that is obtained by a unilateral deviation from a profile of discretimized strategies.

For such a correlated strategy, the mixed action-profile at time $t < s_1$ is well defined as a function of z_0 (and σ) and is denoted by $x_{0,t}$. Note that $x_{0,t}$ is the mixed action-profile that is selected by the correlated strategy σ at time t conditional on no state change in the interval $[0, t]$. Note that $x_{0,t}$ is defined for every $t \geq 0$ as a function of z_0 (and σ).

Using the notations that are used in the definition of a general correlated strategy, $x_{0,t} = \sigma(h, t)$, where h_t is the unique play that is compatible with σ and $z_s = z_0$ for every $s \leq t$.

Similarly, the mixed action-profile at time $s_1 \leq t < s_2$ is well defined as a function of (s_1, z_{s_1}) (and σ) and is denoted by $x_{1,t}$. In other words, $x_{1,t}$ is the mixed action-profile that is selected by the correlated strategy σ at time t conditional on $t \geq s_1$ and no state change in the interval $[s_1, t]$. Note that $x_{1,t}$ is defined for every $t \geq s_1$ as a function of (s_1, z_{s_1}) and σ.

Using the notations that are used in the definition of a general correlated strategy, for every (s_1, z_{s_1}) and $t > s_1$, $x_{1,t} = \sigma(h, t)$, where h_t is unique play up to time t that is (1) compatible with σ, (2) $z_s = z_0$ for every $s < s_1$, and (3) $z_s = z_{s_1}$ for every $s_1 < s \leq t$.

Let $\|\mu\| := \max_{(z,a)\in\mathcal{A}} |\mu(z, z, a)|$. Obviously, $\|\mu\|$ is finite as it is the maximum over finitely manu real numbers.

The following Lemma shows that the probability of at least one state change in a time interval is bounded by a constant ($\|\mu\|$) times the length of the time interval, and that the probability of at least two state changes in a time interval is bounded by a constant ($\|\mu\|^2/2$) times the square of the length of the time interval.

The second bound implies that for any (finite or) countable collection of time intervals the probability of at least two state changes in one of these time intervals is bounded by a constant ($\|\mu\|^2/2$) times the sum of the squares of the lengths of the time intervals.

Lemma 1. *For every $\theta > 0$,*

$$P_\sigma(s_1 \geq \theta) \geq e^{-\|\mu\|\theta} \geq 1 - \theta\|\mu\| \tag{8}$$

and

$$P_\sigma(s_2 \leq \theta) \leq \theta^2\|\mu\|^2/2. \tag{9}$$

Proof. As $P^z_\sigma(s_1 \geq \theta) = e^{\int_0^\theta \mu(z,z,x_{0,t})\,dt}$ and $0 \geq \int_0^\theta \mu(z,z,x_{0,t})\,dt \geq -\|\mu\|\theta$, $P^z_\sigma(s_1 \geq \theta) \geq e^{-\|\mu\|\theta} \geq 1 - \theta\|\mu\|$, proving inequality (8).

The conditional probability, given (s_1, z_{s_1}), that $s_2 \leq \theta - s_1$ is equal, on $s_1 \leq \theta$, to $1 - e^{\int_{s_1}^\theta \mu(z_{s_1},z_{s_1},x_{1,t})\,dt} \leq \|\mu\|(\theta - s_1)$.

The density of s_1 at $\eta \in \mathbb{R}_+$ is given by $-\mu(z,z,x_{1,\eta})e^{\int_0^\eta \mu(z,z,x_{0,t})\,dt} \leq \|\mu\|$.

Therefore, $P_\sigma(s_2 \leq \theta) \leq \int_0^\theta \|\mu\|(\theta - s)\|\mu\|\,ds = \|\mu\|^2\theta^2/2$, proving inequality (9). □

Several proofs in the paper use backward induction arguments. In these arguments we will need to consider the expectation of the sum of the accumulation of payoffs in a (short) time interval and a function of the state at the end of the interval.

Without loss of generality we can assume that the time interval starts at time 0 and ends at time δ. Therefore, we are led to estimate the expectation of $\int_{[0,\delta)} g^i_t\,dw(t) + u(z_\delta)$, where w is a nonnegative measure on $[0,\delta)$ and u is a real-valued function that is defined on the state space. The expectation depends (obviously) on the strategy profile, or, more generally, the correlated strategy σ.

The next lemma uses the inequalities of Lemma 1 to approximate each one of the terms $E_\sigma \int_{[0,\delta)} g^i_t\,dw(t)$ and $E_\sigma u(z_\delta)$, and therefore also their sum, by functions that depend on the choice of mixed action-profile of the correlated strategy σ conditional on no state change in the interval $[0,\delta]$. Recall that this choice is denoted by $x_{0,t}$.

Recall that $\|g\| = \max_{i,z,a} |g^i(z,a)|$.

Lemma 2. *Let $\delta > 0$, w be a finite nonnegative measure on $[0,\delta)$, and $u : S \to \mathbb{R}$. Let σ be a correlated strategy. Then, for every $z \in S$,*

$$|E^z_\sigma\left(\int_{[0,\delta)} g^i_t\,dw(t)\right) - \int_{[0,\delta)} g^i(z,x_{0,t})\,dw(t)| \leq 2\|\mu\|\delta\|g\|w([0,\delta)),$$

$$|E^z_\sigma u(z_\delta) - u(z) - \sum_{z' \in S} u^i(z') \int_{[0,\delta)} \mu(z',z,x_{0,t})\,dt| \leq 2\|\mu\|^2\delta^2\|u\|,$$

and, therefore, by setting $\varepsilon = \varepsilon(\delta, \|\mu\|, w) = 2\|\mu\|\delta\|g\|w([0,\delta)) + 2\|\mu\|^2\delta^2\|u\|$,

$$\begin{aligned} & \varepsilon + E^z_\sigma\left(\int_{[0,\delta)} g^i_t\, dw(t) + u(z_\delta)\right) \\ \geq\ & \int_{[0,\delta)} g^i(z, x_{0,t})\, dw(t) + u(z) + \sum_{z'\in S} u(z') \int_0^\delta \mu(z', z, x_{0,t})\, dt \\ \geq\ & E^z_\sigma\left(\int_{[0,\delta)} g^i_t\, dw(t) + u(z_\delta)\right) - \varepsilon. \end{aligned}$$

Proof. As $x_t = x_{0,t}$ on $t < t_* := \delta \wedge s_1$, $|g^i(z_t, x_t) - g^i(z, x_{0,t})| \leq 2\|g\|\,\mathbb{I}(t \geq s_1)$, where $\mathbb{I}(t \geq s_1)$ is the indicator of the event $\{t \geq s_1\}$. Therefore,

$$\begin{aligned} & |E^z_\sigma \int_{[0,\delta)} g^i_t\, dw(t) - \int_{[0,\delta)} g^i(z, x_{0,t})\, dw(t)| \\ \leq\ & E^z_\sigma \int_{[0,\delta)} |g^i_t - g^i(z, x_{0,t})|\, dw(t) \leq E^z_\sigma \int_{[0,\delta)} 2\|g\|\,\mathbb{I}(t \geq s_1)\, dw(t) \\ \leq\ & \int_{[0,\delta)} 2\|g\|\, P^z_\sigma(s_1 < \delta)\, dw(t) \leq 2\|g\|\delta\|\mu\| w([0,\delta)). \end{aligned}$$

The event $\{z_{t_*} \neq z_\delta\}$ is a subset of the event $\{s_2 \leq \delta\}$. Therefore, by inequality (9), $P^z_\sigma(z_{t_*} \neq z_\delta) \leq \delta^2\|\mu\|^2/2$. For every $z' \in S$,

$$P(z_{t_*} = z') = 1_{z'=z} + \int_0^\delta \mu(z', z, x_{0,t}) e^{\int_0^t \mu(z,z,x_{0,s})ds}\, dt.$$

Therefore, as $|e^{\int_0^t \mu(z,z,x_{0,s})ds} - 1| \leq t\|\mu\|$ and thus $\sum_{z'\in S} \int_0^\delta |\mu(z', z, x_{0,t})| t\|\mu\|\, dt \leq \delta^2\|\mu\|^2$,

$$|\sum_{z'\in S} P(z_{t_*} = z')u(z') - u(z) - \sum_{z'\in S} u(z') \int_0^\delta \mu(z', z, x_{0,t})\, dt| \leq \|u\|\delta^2\|\mu\|^2,$$

which together with the inequality $|E^z_\sigma u(z_\delta) - E^z_\sigma u(z_{t_*})| \leq 2\|u\| P^z_\sigma(z_\delta \neq z_{t_*}) \leq \|\mu\|^2\delta^2\|u\|$ completes the proof of the second part of the lemma.

The last part of the lemma is obtained by summing the inequalities of the first two parts. □

3.2 The stationary discounting games

Fix a profile $\overrightarrow{\rho} = (\rho_i)_{i\in N}$ of discounting rates. We say that the strategy profile σ is an ε-*equilibrium*, $\varepsilon \geq 0$, of the $\overrightarrow{\rho}$-discounted continuous-time stochastic game $\Gamma_{\overrightarrow{\rho}}$, if σ is an admissible strategy profile such that for every player i and every strategy τ^i of player i we have

$$E_\sigma \int_0^\infty e^{-\rho_i t} g^i(z_t, x_t)\, dt \geq E_{\sigma^{-i},\tau^i} \int_0^\infty e^{-\rho_i t} g^i(z_t, x_t)\, dt - \varepsilon$$

where E_σ is the expectation with respect to the probability distribution P_σ defined by σ on plays, and σ^{-i}, τ^i is the strategy profile whose i-th component is τ^i and for $j \neq i$ the j-th component is σ^j.

The strategy profile σ is an *equilibrium* if it is a 0-equilibrium.

Theorem 1. *Every $\overrightarrow{\rho}$-discounted continuous-time stochastic game $\Gamma_{\overrightarrow{\rho}}$ (with finitely many states and actions) has a profile of stationary strategies that is an equilibrium of $\Gamma_{\overrightarrow{\rho}}$.*

Proof. Define $\|g^i\| := \max_{z\in S} \max_{a\in A(z)} |g^i(z,a)|$, $J^i(z) = [-\|g^i\|/\rho_i, \|g^i\|/\rho_i]$, and $J = \times_{(i,z)\in N\times S} J^i(z)$. The (i,z)-th coordinate of $v \in \mathbb{R}^{N\times S} \supset J$ is denoted by $v^i(z)$. Recall that $X^i(z) = \Delta(A^i(z))$. Let $Y^i = \times_{z\in S} X^i(z)$ and $Y = \times_{i\in N} Y^i$.

For every $i \in N$, $z \in S$, $a \in A(z)$, $v \in \mathbb{R}^{N\times S}$, and $x \in Y$, $G^i_z[v](x)$, and $f_{z,i}$ are the real-valued functions defined on $Y \times J$ as follows.

$$G^i_z[v](a) = \frac{1}{\|\mu\| + \rho_i} \left(g^i(z,a) + \sum_{z'\in S} \mu(z', z, a) v^i(z') + \|\mu\| v^i(z) \right),$$

where $\|\mu\| = \max_{z,a} |\mu(z,z,a)|$.

$$\begin{aligned} G^i_z[v](x) &= \sum_{a\in A(z)} G^i_z[v](a) \prod_{j\in N} x^j(z)(a^j), \quad \text{and} \\ f_{z,i}(x,v) &= \max_{y^i \in X^i(z)} G^i_z[v](x^{-i}, y^i), \end{aligned}$$

where $(x^{-i}(z), y^i)$ is the profile of mixed actions whose j-th coordinate for $j \neq i$ is $x^j(z)$ and whose i-th coordinate is y^i. Let $F_{z,i}$ be the correspondence from $Y \times J$ to $X^i(z)$ given by

$$F_{z,i}(x,v) = \arg \max_{y^i \in X^i(z)} G^i_z[v](x^{-i}, y^i).$$

The cartesian product $Y \times J$ is a product of nonempty convex compact sets and therefore it is nonempty, convex, and compact.

The function $(x, v) \mapsto G^i_z[v](x)$ is a polynomial in the coordinates of (x, v) and therefore continuous, and therefore $f_{z,i}(x, v)$ is also a continuous function of (x, v).

If $\|v\| \leq \|g^i\|/\rho_i$ then $|G^i_z[v](a)| \leq \frac{1}{\|\mu\|+\rho_i}(\|g^i\| + \|\mu\|\|g^i\|/\rho_i) \leq \|g^i\|/\rho_i$. Therefore $|G^i_z[v](x)| \leq \|g^i\|/\rho_i$ and therefore also $|f_{z,i}(x, v)| \leq \|g^i\|/\rho_i$.

We deduce that $f_{z,i}$ is a continuous function from $X \times J$ to $J^i(z)$.

The correspondence $F_{z,i}$ from $Y \times J$ to $X^i(z)$ is nonempty, convex-valued, and upper semicontinuous . Therefore the correspondence F defined on $Y \times J$ by $F(x, v) = \times_{z,i}(F^z_i(x, v) \times \{f_{z,i}(x, v)\})$ is a nonempty, convex-valued, upper semicontinuous correspondence from the nonempty convex compact set $Y \times J$ to itself, and therefore has a fixed point.

Let (x, V) be a fixed point of F. We claim that the profile of stationary strategies σ^i with $\sigma^i(h_t) = x^i(z_t)$ is an equilibrium of the $\overrightarrow{\rho}$-discounted game with equilibrium payoff V.

Fix a player $i \in N$, a strategy τ^i of player i, and an initial state $z = z_0$. First we prove that

$$E^z_{\sigma^{-i},\tau^i} \int_0^\infty e^{-\rho_i t} g(z_t, x_t)\, dt \leq V^i(z). \tag{10}$$

For every correlated strategy, or a strategy profile that defines a unique distribution on plays, τ, state $z \in S$, and $s \geq 0$, define $f^z_\tau(s) = E^z_\tau(\int_0^s e^{-\rho_i t} g^i(z_t, x_t)\, dt + e^{-\rho_i s} V^i(z_s))$.

Set $f(s) = f^z(s) = f^z_{\sigma^{-i},\tau^i}(s)$. Note that $f(0) = V^i(z)$ and that $f(s) \to_{s\to\infty} E^z_{\sigma^{-i},\tau^i} \int_0^\infty e^{-\rho_i t} g^i(z_t, x_t)\, dt$. Therefore, in order to prove (10), it suffices to prove that f is Lipschitz and that the upper-right derivative of f is nonpositive a.e.

Let $y_t := x_{0,t}(x, \sigma^{-i}, \tau^i)$, namely, $y_t = (\sigma^{-i}, \tau^i)(h, t)$, where h is a play that is (1) compatible with σ and (2) $z_s = z$ for every $0 \leq s < \delta$. By Lemma 2,

$$
\begin{aligned}
f(\delta) - f(0) &= E^z_{\sigma^{-i},\tau^i}\left(\int_0^\delta e^{-\rho_i t} g(z_t, x_t)\, dt + e^{-\rho_i \delta} V^i(z_s)\right) \\
&\leq \int_0^\delta e^{-\rho_i t} g(z_0, y_t)\, dt \\
&\quad + e^{-\rho_i \delta}\left(V^i(z_0) + \sum_{z' \in S} V^i(z') \int_0^\delta \mu(z', z_0, y_t)\, dt\right) + O(\delta^2) \\
&\leq \int_0^\delta \left(g(z_0, y_t) \sum_{z' \in S} V^i(z') \mu(z', z_0, y_t) - \rho_i V^i(z_0)\right) dt + O(\delta^2).
\end{aligned}
$$

Therefore, the function $s \mapsto f(s)$ is Lipschitz ($|f(s+\delta) - f(s)| \leq \delta e^{-\rho_i s}(2\|\mu\|\|V^i\|_\infty + \|g\| + O(\delta))$). Therefore, in order to prove (10), it suffices to prove that the upper-right derivative of f is nonpositive a.e.

The choice of σ implies that $g(z_0, y_t) + \sum_{z' \in S} \mu(z', z_0, y_t) V^i(z') - \rho_i V^i(z_0) \leq 0$ and therefore the upper-right derivative of f at 0, $\limsup_{\delta \to 0+} \frac{f(\delta) - f(0)}{\delta}$, is nonpositive.

Similarly, for a.e. $s > 0$, the upper-right derivative of f at s is nonpositive. We conclude that f is nonincreasing.

Now setting $\tau^i = \sigma^i$, we can change $\leq$ in the above-derived inequalities to $\geq$ (and replace the terms $+O(\delta^2)$ by $-O(\delta^2)$) in order to deduce that $f^z_\sigma(x)$ is nondecreasing and thus

$$
E^z_\sigma \int_0^\infty e^{-\rho_i t} g(z_t, x_t)\, dt \geq V^i(z),
$$

which together with inequality (10) completes the proof of the theorem. □

Covariance properties. In the case where $\rho_i = \rho$ for every i, an alternative interpretation of the ρ-discounted game is as a game with an uncertain duration d in which each player maximizes his expected total payoff. Under this interpretation, ρ is a parameter of the distribution of the random duration d, $P(d \geq t) = e^{-\rho t}$. Multiplying ρ by α means that the game terminates faster at a rate of α. This can be compensated by multiplying all transitions and payoffs by α as well. Therefore, any stationary ρ-discounted equilibrium in the game $\langle N, S, A, \mu, g\rangle$ is also a stationary $\alpha\rho$-discounted equilibrium in the game $\langle N, S, A, \alpha\mu, \alpha g\rangle$. In addition, if $\bar{p}$ is the transition matrix such that

$\bar{p}(z' \mid z, a) = \frac{1}{1-\rho}\mu(z', z, a)$ for all $z' \neq z$ and $\bar{p}(z \mid z, a) = 1 + \frac{1}{1-\rho}\mu(z, z, a)$, then equations defining a fixed point (x, V) of the correspondence F (used in the proof of the theorem) are equivalent[10] to the equations defining a stationary equilibrium of the discrete-time stochastic game with a stage payoff function g and transition probabilities $\bar{p}$.

Formally, and more generally, fix $\alpha > 0$. Note that the auxiliary function $G^i_z[v]$ of the $\overrightarrow{\rho}$-discounted game $\Gamma = \langle N, S, A, \mu, g\rangle$ and the $\alpha\overrightarrow{\rho}$-discounted game $\Gamma = \langle N, S, A, \alpha\mu, \alpha g\rangle$ are the same. Therefore, a point $(x, V) \in \times_{z\in S, i\in N}(X^i(z) \times [-\|g^i\|/\rho_i, \|g^i\|/\rho_i])$ is a stationary equilibrium (strategies and payoffs) of the continuous-time $\overrightarrow{\rho}$-discounted game $\Gamma = \langle N, S, A, \mu, g\rangle$ if and only if (x, V) is a stationary equilibrium of the continuous-time $\alpha\overrightarrow{\rho}$-discounted game $\Gamma = \langle N, S, A, \alpha\mu, \alpha g\rangle$.

In addition, if $0 < \rho < 1$ and $\|\mu\| \leq 1 - \rho$, then a point $(x, V) \in \times_{z\in S, i\in N}(X^i(z) \times [-\|g^i\|/\rho_i, \|g^i\|/\rho_i])$ is a stationary equilibrium (strategies and payoffs) of the continuous-time $\overrightarrow{\rho}$-discounted game $\Gamma = \langle N, S, A, \mu, g\rangle$ if and only if it is a stationary equilibrium of the discrete-time ρ-discounted (where the discount factor is $1 - \rho$) game $\bar{\Gamma} = \langle N, S, A, \bar{p}, g\rangle$, where $\bar{p}$ is the transition probability that is given by $\bar{p}(z' \mid z, a) = \frac{1}{1-\rho}\mu(z', z, a)$ for all $z' \neq z$.

The discounting minmax. The next theorem is used in the proof of Theorem 3, which is used (in the punishing phases) in the proof of the main result. It asserts that (1) the minmax $\bar{V}^i_\rho$ of player i in the unnormalized ρ-discounted game exists, (2) the function $\rho \mapsto \bar{V}^i_\rho$ is semialgebraic,[11] and (3) there are minimaxing strategies that are stationary.

Theorem 2. *Fix a ρ-discounted continuous-time stochastic game (with finitely many states and actions) Γ_ρ and a player $i \in N$.*
a) The set of equations in the real variables $v(z)$, $z \in S$,

$$\rho v(z) = \min_{x^{-i}} \max_{x^i} \left(g^i(z, x^{-i} \otimes x^i) + \sum_{z'\in S} \mu(z', z, x^{-i} \otimes x^i) v(z') \right), \qquad (11)$$

where the minimum is over all $x^{-i} = (x^j)_{j\neq i} \in \times_{j\neq i}\Delta(A^j(z))$ and the maximum is over all $x^i \in \Delta(A^i(z))$, has a unique solution $\bar{V}^i_\rho$.

[10] This equivalence can be used to provide an alternative proof of the existence of a fixed point of F.

[11] See, e.g., [24] for the definition of semialgebraic functions, and their applications to stochastic games.

b) The function $\rho \mapsto \bar{V}^i_\rho$ is semialgebraic and $\|\bar{V}^i_\rho\| \leq \|g\|/\rho$.
c) There are stationary strategies σ^j, $j \neq i$, such that for every strategy σ^i of player i and every initial state z we have

$$E^z_{\sigma^{-i},\sigma^i}\left(\int_0^s e^{-\rho t} g^i(z_t, x_t)\, dt + e^{-\rho s}\bar{V}^i_\rho(z_s)\right) \leq \bar{V}^i_\rho(z) \;\; \forall s \geq 0, \;\; \textit{and}$$

$$E^z_{\sigma^{-i},\sigma^i}\int_0^\infty e^{-\rho t} g^i(z_t, x_t)\, dt \leq \bar{V}^i_\rho(z). \tag{12}$$

Proof. Let Q be the map from $\mathbb{R}^S$ to itself that is given by $Qv(z) = \min_{x^{-i}} \max_{x^i} G^i_z[v](x^{-i}, x^i)$, where

$$G^i_z[v](x^{-i}, x^i) = \frac{1}{\|\mu\| + \rho}\left(g^i(z, x^{-i} \otimes x^i) + \sum_{z' \in S} \mu(z', z, x^{-i} \otimes x^i)v(z') + \|\mu\| v(z)\right).$$

Q is a strict contraction and therefore has a unique fixed point V.

A vector $V = (\bar{V}^i_\rho(z))_{z \in S} \in \mathbb{R}^S$ is a solution of $Qv = v$ if and only if it is a solution of equality (11). Therefore, equality (11) has a unique solution. This completes the proof of part a).

Observe that any solution $(\bar{V}^i_\rho(z))_{z \in S}$ of equality (11) obeys $|\bar{V}^i_\rho(z)| \leq \|g\|/\rho$, and the semialgebraicity of equality (11) guarantees that the function that maps ρ to the unique solution $(\bar{V}^i_\rho(z))_{z \in S}$ is semialgebraic. This completes the proof of part b).

Let $x^{-i}(h_t) = (x^j(h_t))_{j \neq i}$ be the minimizer of the right-hand side of (11) for $z = z_t$.

For every $j \neq i$ we define the strategy σ^j by $\sigma^j_t(h) = x^j(h_t)$. As in the proof of Theorem 1, it follows that for every strategy σ^i we have

$$E^z_{\sigma^i,\sigma^{-i}}\int_0^\infty e^{-\rho t} g^i(z_t, x_t)\, dt \leq V(z),$$

which completes the proof of the theorem. □

Theorem 2 implies that the function $\rho \mapsto v^i_\rho(z) := \rho\bar{V}^i_\rho$ is a bounded semialgebraic function, and therefore it converges to a limit $v^i(z)$ as $\rho \to 0+$.

3.3 The undiscounted minmax

Fix a continuous-time stochastic game $\Gamma = \langle N, S, A, \mu, g \rangle$ with finitely many states and actions.

A correlated strategy σ defines for every state $z \in S$ a unique probability distribution P_σ^z on the plays of Γ. The expectation with respect to the probability distribution P_σ^z is denoted by E_σ^z.

For every player $i \in N$, we set

$$\gamma_s^i(z,\sigma) := E_\sigma^z \frac{1}{s} \int_0^s g_t^i \, dt, \text{ where } g_t^i := g^i(z_t, x_t),$$

$$\bar{\gamma}^i(z,\sigma) := E_\sigma^z \limsup_{s\to\infty} \frac{1}{s} \int_0^s g_t^i \, dt, \quad \text{and} \quad \underline{\gamma}^i(z,\sigma) := E_\sigma^z \liminf_{s\to\infty} \frac{1}{s} \int_0^s g_t^i \, dt.$$

Recall that the profile of strategies that is obtained by a unilateral deviation from a profile of admissible strategies (e.g., from a profile of discretimized strategies) is equivalent, in terms of the probability distribution that it defines on the plays, to a correlated strategy.

In the previous section, we showed that there are stationary minimaxing strategies in the ρ-discounted games.

There are continuous-time stochastic games (as in the discrete-time case) for which Markov strategies can serve neither as approximate minimaxing strategies in the limiting-average game, nor as approximate minimaxing strategies in all sufficiently long finite-horizon games.

The next result shows that such minimaxing strategies exist in the class of discretimized strategies. Moreover, the discretimization can be done with the sequence of nonnegative integers. An $\mathbb{N}$-discretimized strategy is a $(t_k)_{k\geq 0}$-discretimized strategy where $t_k = k$.

Theorem 3. *For every $\varepsilon > 0$ and every player i there are $\mathbb{N}$-discretimized strategies σ^j, $j \neq i$, and a time s_ε such that for every strategy σ^i of player i and every $s > s_\varepsilon$ we have*

$$\gamma_s^i(z, \sigma^{-i}, \sigma^i) \leq v^i(z) + \varepsilon \quad \textit{and} \quad \bar{\gamma}^i(z, \sigma^{-i}, \sigma^i) \leq v^i(z) + \varepsilon.$$

Proof. For every nonnegative integer $k \geq 0$, let $r_k := \int_{t=k}^{k+1} g^i(z_t, x_t) \, dt$.

Fix a player $i \in N$. For every sequence ρ_k, $k \geq 0$, where $\rho_{k+1} > 0$ is a function of h_{k+1} (e.g., a function of ρ_k, z_{k+1}, and r_k), there is by Theorem 2

an $N \setminus \{i\}$ profile of strategies $\sigma^{-i} = (\sigma_j)_{j \neq i}$, so that for every strategy σ^i of player i we have

$$E_\sigma^{h_k} \left(\int_k^{k+1} \rho_k e^{-\rho_k (t-k)} g_t^i \, dt + e^{-\rho_k} v_{\rho_k}^i (z_{k+1}) \right) \leq v_{\rho_k}^i (z_k),$$

where σ is the strategy profile (σ^{-i}, σ^i), and $E_\sigma^{h_k} f$ is the conditional expectation, given h_k, of the random variable f w.r.t. the probability on plays defined by the strategy profile σ.

As $|g_t^i| = |g^i(z_t, x_t)| \leq \|g\|$ and $\|v_\rho^i\| := \max_{z \in S} |v_\rho^i(z)| \leq \|g\|$, the inequality $\int_0^1 |1 - e^{-\rho t}| \rho \, dt + |1 - \rho - e^{-\rho}| = 2|1 - \rho - e^{-\rho}| \leq O(\rho^2)$ implies that

$$E_\sigma^{h_k} \left(\rho_k r_k + (1 - \rho_k) v_{\rho_k}^i (z_{k+1}) \right) \leq E_\sigma^{h_k} \left(\int_k^{k+1} \rho_k e^{-\rho_k (t-k)} g_t^i \, dt + e^{-\rho_k} v_{\rho_k}^i (z_{k+1}) \right) + O(\rho_k^2).$$

Therefore, for sufficiently small $\theta > 0$, the inequality $\rho_k \leq \theta$ implies that

$$E_\sigma^{h_k} \left(\rho_k r_k + (1 - \rho_k) v_{\rho_k}^i (z_{k+1}) \right) \leq v_{\rho_k}^i (z_k) + \varepsilon(\theta) \rho_k,$$

where $\varepsilon(\theta) \to 0$ as $\theta \to 0+$.

By the proof in [17] (or using the statement of [25, Lemma 1]) and the inequality $\frac{1}{s} \int_0^s g^i(z_t, x_t) \, dt \leq \frac{1}{[s]} \int_0^{[s]} g^i(z_t, x_t) \, dt + O(\varepsilon)$, we deduce that there is a sequence of history-dependent (hence, random) discount rates ρ_k (where ρ_0 is a constant and for a positive integer k the discount rate ρ_k is a function of z_k, ρ_{k-1}, and r_{k-1}) such that the corresponding strategy profile σ^{-i} satisfies, for all s sufficiently large and all strategies σ^i of player i,

$$\gamma_s^i(z, \sigma^{-i}, \sigma^i) \leq v^i(z) + O(\varepsilon) \quad \text{and} \quad \bar{\gamma}^i(z, \sigma^{-i}, \sigma^i) \leq v^i(z) + O(\varepsilon).$$

□

The choice of $\mathbb{N}$ in the construction of the $\mathbb{N}$-discretimized minimaxing strategies is for convenience. In fact, for any increasing sequence (t_k) of reals with $t_0 = 0$ and, e.g., bounded difference $t_{k+1} - t_k$, there are undiscounted minimaxing strategies that are (t_k)-discretimized.

3.4 Uniform l-average equilibria

One of the central open problems in stochastic games is the existence of a uniform (or a limiting-average) equilibrium payoff in (discrete-time) stochastic games with finitely many states and actions.

The present paper proves the existence of uniform and limiting-average equilibrium payoffs in all continuous-time stochastic games with finitely many states and actions.

Theorem 4. *Let $\Gamma = \langle N, S, A, g, \mu \rangle$ be a continuous-time stochastic game with finitely many states and actions. There exists a vector payoff $u \in \mathbb{R}^{S \times N}$ such that for every $\varepsilon > 0$ there are (discretimized) strategies σ^i, $i \in N$, and a positive number $s_0 > 0$, such that for every $s > s_0$, every player i, every state $z \in S$, and every strategy τ^i of player i, we have*

$$-\varepsilon + \gamma_s^i(z, \sigma^{-i}, \tau^i) \leq u^i(z) \leq \gamma_s^i(z, \sigma) + \varepsilon, \quad \text{and} \tag{13}$$

$$-\varepsilon + \bar{\gamma}^i(z, \sigma^{-i}, \tau^i) \leq u^i(z) \leq \underline{\gamma}^i(z, \sigma) + \varepsilon. \tag{14}$$

Such a vector payoff u is called a *uniform-l-average equilibrium payoff*, and the corresponding strategy σ is called a *uniform-l-average ε-equilibrium strategy*.

A vector payoff u for which $\forall \varepsilon > 0$ $\exists \sigma, s_0$ s.t. $\forall s > s_0, i, \tau^i$ the system of inequalities (13) holds is called a *uniform equilibrium payoff*, and a vector payoff u for which $\forall \varepsilon > 0$ $\exists \sigma$ s.t. $\forall i, \tau^i$ the system of inequalities (14) holds is called a limiting-average equilibrium payoff.

The corresponding strategies are called a *uniform ε-equilibrium strategy* and a *limiting-average ε-equilibrium strategy*.

The proof of Theorem 4 is given in Section 4. The proof follows a similar outline to that used in Solan and Vieille [36] for the proof of the existence of uniform extensive-form correlated equilibria in discrete-time stochastic games.

3.5 Robust equilibria

Given a correlated strategy profile σ (which, as said earlier, defines for every state $z \in S$ a unique probability distribution P_σ^z on the plays of the continuous-time stochastic game $\Gamma = \langle N, S, A, \mu, g \rangle$), a profile of discounting measures $\overrightarrow{w} \in W^N$, and a player $i \in N$, we set

$$\underline{\gamma}_{\overrightarrow{w}}^i(z, \sigma) := E_\sigma^z \int_{[0,\infty)} g_t^i \, dw^i(t) + w^i(\infty) \underline{\gamma}^i(z, \sigma) = E_\sigma^z \int_{[0,\infty]} \underline{g}_t^i \, dw^i(t),$$

where $\underline{g}^i = g^i_t$ if $t < \infty$ and $\underline{g}^i_\infty = \liminf_{s\to\infty} \frac{1}{s}\int_0^s g^i_t\,dt$, and

$$\bar{\gamma}^i_{\overrightarrow{w}}(z,\sigma) := E^z_\sigma \int_{[0,\infty)} g^i_t\,dw^i(t) + w^i(\infty)\bar{\gamma}^i(z,\sigma) = E^z_\sigma \int_{[0,\infty]} \bar{g}^i_t\,dw^i(t),$$

where $\bar{g}^i_t = g^i_t$ if $t < \infty$ and $\bar{g}^i_\infty = \limsup_{s\to\infty} \frac{1}{s}\int_0^s g^i_t\,dt$.

Theorem 5. *Let $\Gamma = \langle N, S, A, g, \mu\rangle$ be a continuous-time stochastic game with finitely many states and actions. For every $\varepsilon > 0$, there is a finite set $\mathcal{A}$ and an open cover $(U_\alpha)_{\alpha\in\mathcal{A}}$ of $\overrightarrow{W}^1_d$, such that for every $\alpha \in \mathcal{A}$ there is a vector payoff $u_\alpha \in \mathbb{R}^{S\times N}$ and a profile of (discretimized) strategies σ_α, such that for every $\overrightarrow{u} \in U_\alpha$, σ_α is an ε-equilibrium of $\Gamma_{\overrightarrow{u}}$ with a payoff ε-close to u_α; alternatively, such that for every player i, every state $z \in S$, and every strategy τ^i of player i we have*

$$-\varepsilon + \bar{\gamma}^i_{\overrightarrow{u}}(z,\sigma^{-i}_\alpha,\tau^i) \le u^i(z) \le \underline{\gamma}^i_{\overrightarrow{u}}(z,\sigma_\alpha) + \varepsilon. \tag{15}$$

The concept of robust equilibria depends on the topology defined on (W, hence on) W_d. Note that the coarser the topology is, the stronger the corresponding robustness result is.

A basis for the w^*-neighborhoods of W^N are sets of the form $\{\overrightarrow{w} = (w^i)_{i\in N} \in W^N : |\int_{[0,\infty]} f^i_j(t)\,dw^i(t) - c^i_j| < d^i_j\ \forall j \in J, i \in N\}$, where J is a finite set, $f^i_j \in C([0,\infty])$ with $0 \le f^i_j \le 1$, and $c^i_j, d^i_j \in [0,1]$.

Therefore, the robustness result is equivalent to the following statement.

For every $\varepsilon > 0$, there is $\delta > 0$ and a finite set $\mathcal{A}$, where $\alpha \in \mathcal{A}$ is a list $(f_\alpha, c_\alpha, v_\alpha, \sigma_\alpha)$ with $f_\alpha = (f^i_\alpha)_{i\in N}$ a vector of $[0,1]$-valued continuous functions on $[0,\infty]$, $\{c_\alpha : \alpha \in \mathcal{A}\}$ being a finite δ net of $[0,1]^N$, $v_\alpha \in \mathbb{R}^{S\times N}$, and σ_α an admissible strategy profile, such that for every $\overrightarrow{w} \in \overrightarrow{W}^1_d$ with $|\int_{[0,\infty]} f^i_\alpha(t)\,dw^i(t) - c^i_\alpha| < \delta$, σ_α is an ε-equilibrium of $\Gamma_{\overrightarrow{w}}$ with a payoff ε-close to u_α.

A crucial step in the proof of Theorem 5 is Theorem 4, which is of independent interest.

4 Proof of Theorem 4

Theorem 4 is straightforward if $\|\mu\| = 0$. Indeed, if $\|\mu\| = 0$, there are no state changes in a play; i.e., conditional on the initial state z we have a

continuous-time supergame. Therefore, if σ is a stationary strategy profile with $\sigma(z)$ an equilibrium of the single stage game with payoff function $a \mapsto g(z,a)$ and $u(z) = g(z,\sigma(z))$, then inequalities (13) and (14) hold for every $\varepsilon \geq 0$. Therefore, we assume that $\|\mu\| > 0$.

Without loss of generality we can assume that $0 \leq g^i \leq 1$ and $\|\mu\| = 1$. Indeed, the continuous-time stochastic game $\Gamma = \langle N, S, A, \mu, g\rangle$ has a uniform-l-average equilibrium payoff if and only if $\bar{\Gamma} = \langle N, S, A, \bar{\mu}, \bar{g}\rangle$ does, where $\bar{\mu} = \alpha\mu$, and $\bar{g}^i = \beta_i + \alpha_i g^i$ for some (and therefore for all) $\alpha, \alpha_1, \ldots, \alpha_n > 0$ and $\beta_1, \ldots, \beta_n \in \mathbb{R}$. Therefore, we assume that $0 \leq g^i(z,a) \leq 1$ and $\|\mu\| = 1$.

In order to prove Theorem 4, it suffices to prove it with the ε-independent vector payoff u being replaced by an ε-dependent u_ε. Namely, it suffices to prove that $\forall \varepsilon > 0$, $\exists u_\varepsilon, s_\varepsilon, \sigma_\varepsilon$ s.t. $\forall s_\varepsilon < s < \infty$, (13) and (14) hold when u, s_0, and σ are replaced there by u_ε, s_ε, and σ_ε, respectively. Indeed, let u be a limit point of u_ε; then $\exists \varepsilon_k \to_{k\to\infty} 0+$ s.t. $u_{\varepsilon_k} \to_{k\to\infty} u$. For $\varepsilon > 0$, (13) and (14) hold with $s_0 = s_{\varepsilon_k}$ and $\sigma = \sigma_{\varepsilon_k}$ and k sufficiently large so that $\varepsilon_k < \varepsilon/2$ and $\|u - u_\varepsilon\| < \varepsilon/2$.

4.1 An informal outline of the proof

The proof constructs a profile of discretimized pure-action strategies $\hat{\tau}$ that depends on $\varepsilon > 0$ and satisfies several properties. The discretimization is with respect to a refinement of a sequence of real numbers $\mathcal{T} = (t_j)_{j\geq 0}$, where $t_0 = 0$ and $t_j < t_{j+1} \to_{j\to\infty} \infty$.

The approximate equilibrium strategy profile σ consists of following $\hat{\tau}$ as long as no player deviates from playing according to $\hat{\tau}$. If no player deviates before time t_{j-1} and a single player deviates in the time interval $[t_{j-1}, t_j)$, then all the other players switch at time t_j to punishing the deviating player in the uniform-l-average game.

The constructed strategy profile $\hat{\tau}$ satisfies the following property. For any nonnegative integer j and $s \geq t_j$, the following inequality hold.

$$E^z_{\hat{\tau}}\left(\frac{1}{s-t_j}\int_{t_j}^{s} g^i_t\, dt \mid \mathcal{H}_{t_j}\right) \geq E^z_{\hat{\tau}} v^i(z_{t_j}) - O(\varepsilon) - O\left(\frac{1}{s-t_j}\right). \tag{16}$$

A deviation from $\hat{\tau}$ in the time interval $[t_{j-1}, t_j)$ cannot change the expectation of $v^i(z_{t_j})$ or the payoff of a player in this time interval by more than $O(t_j - t_{j-1})$.

Therefore, if $\sup_j(t_{j+1}-t_j)\leq\varepsilon$, then for any strategy τ^i of player i,

$$\gamma_s^i(z,\sigma^{-i},\tau^i)=E^z_{\sigma^{-i},\tau^i}\,\frac{1}{s}\int_0^s g_t^i\,dt\leq\gamma_s^i(z,\hat{\tau})+O(\varepsilon)+O(\frac{1}{s}). \tag{17}$$

Therefore, if $\sup_j(t_{j+1}-t_j)$ is sufficiently small, σ is an approximate equilibrium of the s-horizon game with a payoff within $O(\varepsilon)$ of $\gamma_s(z,\hat{\tau})$ for all sufficiently large s.

A priori, the limit of $\gamma_s(z,\hat{\tau})$, as $s\to\infty$, need not exist. Moreover, the difference between the limit superior and the limit inferior of $\gamma_s(z,\hat{\tau})$ need not be small. Therefore, condition (17) by itself does not guarantee the existence of a uniform equilibrium payoff.

In order to prove the existence of a uniform equilibrium payoff, the constructed strategy $\hat{\tau}$ will be such that in addition to obeying (16) it satisfies

$$\gamma_s^i(z,\hat{\tau})=E^z_{\hat{\tau}}\,\frac{1}{s}\int_0^s g_t\,dt,\ \text{converges as } s\to\infty. \tag{18}$$

The existence, for every $\varepsilon>0$, of a profile of $\mathcal{T}=(t_j)_{j\geq0}$-discretimized pure-action strategies $\hat{\tau}$ (with $\sup_j(t_{j+1}-t_j)$ sufficiently small) that satisfies (16) and (18) guarantees the existence of a uniform equilibrium payoff.

In order to prove the existence of a uniform-l-average equilibrium payoff, the constructed strategy $\hat{\tau}$ will be such that, in addition to satisfying (16) and (18), it satisfies

$$\frac{1}{s}\int_0^s g_t\,dt\ \text{converges as } s\to\infty\ \text{with } P^z_{\hat{\tau}}-\text{probability 1}, \tag{19}$$

and

$$\lim_{s\to\infty}\gamma_s^i(z,\hat{\tau})=E^z_{\hat{\tau}}\lim_{s\to\infty}\frac{1}{s}\int_0^s g_t\,dt. \tag{20}$$

The pure-action strategy $\hat{\tau}$ is derived by a "purification" of a $\mathcal{T}$-discretimized correlated strategy $\tilde{\tau}$.

If $\tilde{\tau}$ satisfies conditions (16)–(20), then, by Lemma 2, if in each time interval $[t_{j-1},t_j)$, the empirical distribution of the actions of the pure Markov strategy $\hat{\tau}$ is sufficiently close to the average of the mixed action profile of $\tilde{\tau}$, then $\hat{\tau}$ will satisfy these conditions with a possible error of $O(\sum_j(t_j-t_{j-1})^2)$.

Therefore, the existence of a $\mathcal{T}$-discretimized correlated strategy $\tilde{\tau}$ that satisfies conditions (16)–(20) will imply the existence of a uniform-l-average equilibrium.

Let us mention that the existence of a continuous-time strategy that obeys conditions (16)–(20) follows easily from the existence of a stationary correlated strategy in the discrete-time game that obeys the discrete-time analogous conditions (16*)–(20*).

Moreover, if $\tilde{\tau}$ is a discrete-time stationary correlated strategy, then conditions (18*), (19*), and (20*) hold, and condition (16*) holds if and only if condition (17*) holds.

However, the existence of a continuous-time correlated Markov strategy that obeys conditions (16)–(20) does not follow easily from the existence of a correlated Markov strategy (in the discrete-time game) that obeys conditions (16*)–(20*). The existence of such a correlated Markov strategy in the discrete-time game is proved in [36].

Therefore, we cannot use the result of [36] directly. However, we follow similar ideas to those used in [36].

The strategy $\hat{\tau}$ is constructed in a sequence of steps.

First, one defines a strategy profile τ that is discretimized w.r.t. the sequence $\mathcal{T}_\ell = (j/\ell)_{j\geq 0}$, where ℓ is a sufficiently large integer, and a sufficiently large positive integer N_0, such that

$$E_\tau^z v^i(z_T) \geq v^i(z) - \varepsilon_1/2 \quad \text{for every finite } (\mathcal{H}_t)_{t\geq 0}\text{-stopping time } T \tag{21}$$

and

$$\gamma_{N_0}^i(z,\tau) \geq v^i(z) - \varepsilon_1/2, \tag{22}$$

where ε_1 is sufficiently small. See Section 4.4.

The second step defines a correlated strategy $\tilde{\tau}$ that is discretimized w.r.t. the sequence $\mathcal{T}_\ell = (j/\ell)_{j\geq 0}$ and (1) coincides with τ if the initial state is in some subset S_2 of states, where the set S_2 depends on τ, (2) coincides with a profile of stationary strategies y if the initial state is in some subset $\bar{S}$ of states, where y and $\bar{S}$ depend on the stochastic game (but not on τ), and (3) coincides with a discretimized correlated strategy whose correlated mixed action is close to that of τ but avoids certain possible transitions if the initial state is in the complement S_1 of $\bar{S} \cup S_2$. See Section 4.7.

The third step "approximates" the correlated strategy $\tilde{\tau}$ by a pure-action strategy $\bar{\tau}$. The "approximation" is so that in any time interval of the form $[\frac{k}{\ell}, \frac{k+1}{\ell})$, where k a nonnegative integer, the empirical distribution of the pure actions of $\bar{\tau}$ is within $O(1/\ell)$ of the average of the correlated mixed actions

of $\tilde{\tau}$. The "approximation" enables us to use Lemma 2 in order to derive properties of $\bar{\tau}$ from the corresponding properties of $\tilde{\tau}$. See Section 4.8.

Finally, one defines the pure-action strategy $\hat{\tau}$ that is played in blocks of play-dependent random duration that are $\leq N_0$. At the start of each block, $\hat{\tau}$ restarts with the strategy $\bar{\tau}$. The periodicity of $\hat{\tau}$ guarantees that $\hat{\tau}$ obeys properties (18)–(20). See Section 4.8.

The approximate equilibrium strategy σ follows $\hat{\tau}$ until the first time $t = j/\ell$, with j a positive integer, where a deviation by a single player was observed, and thereafter all other players punish the deviating player in the uniform-l-average game. See Section 4.10.

4.2 A few auxiliary functions of Γ

In this subsection we define v_ρ, v, C_z, w, and $d > 0$, as a function of Γ. In short, $v^i_\rho(z)$ is a payoff level where the other players can force the expected payoff of player i in the normalized ρ-discounted game that starts at state z to be no more than this level; $v^i(z)$ is the limit of $v^i_\rho(z)$ as $\rho \to 0$; C_z is the subset of states z' with $v(z') = v(z)$ (namely, $v^i(z') = v^i(z)$ for all i); $w \in \mathbb{R}^S$ is a positive linear combination of the vectors $v^i \in \mathbb{R}^S$, with $0 \leq w(z) \leq 1$ and such that $v(z) \neq v(z')$ implies that $w(z) \neq w(z')$ (and therefore $w(z) = w(z')$ whenever $v(z) = v(z')$); and $d > 0$ is a positive constant such that $w(z) \neq w(z')$ implies that $|w(z) - w(z')| > d$.

Definition of v_ρ and v. Let $V^i_\rho = (V^i_\rho(z))_{z\in S}$ be the (unique) solution $V \in \mathbb{R}^S$ of the system of equations

$$\rho V(z) = \min_{x^{-i}} \max_{x^i} \left(g^i(z, x^{-i} \otimes x^i) + \sum_{z' \in S} \mu(z', z, x^{-i} \otimes x^i) V(z') \right), \quad z \in S,$$

where the minimum ranges over all $x^{-i} = (x^j)_{j\neq i} \in X^{-i}(z) := \times_{j\neq i}\Delta(A^j(z))$, and the maximum ranges over all $x^i \in X^i(z)$. Set $v^i_\rho = \rho V^i_\rho$. The function $\rho \mapsto v^i_\rho(z)$ is a bounded semialgebraic function, and therefore it converges to a limit $v^i(z)$ as $\rho \to 0+$. The assumption $0 \leq g^i \leq 1$ implies that $0 \leq v^i_\rho(z) \leq 1$, and therefore $0 \leq v^i(z) \leq 1$, for every $\rho > 0$, $i \in N$, and $z \in S$.

Definition of C_z. For every state $z \in S$ let $C_z = \{z' \in S \text{ s.t. } v(z') = v(z)\}$.

Definition of w and d. For $i \in N$, with $\max_{z\in S} v^i(z) > \min_{z\in S} v^i(z)$, we set $d_i = \min\{v^i(z) - v^i(z') : z, z' \in S, \text{ and } v^i(z) > v^i(z')\}$. For $i \in N$ with $v^i(z) = v^i(z')$ for every $z, z' \in S$, we set $d_i = 0$. By renaming the players, we can assume without loss of generality that $N = \{1, 2, \ldots, n\}$ and $d_i \geq d_{i+1}$ for $1 \leq i < n$. We will see that if $d_1 = 0$ (and then $v(z) = v(z')$ for every $z, z' \in S$) the theorem follows easily. So assume that $d_1 > 0$ and let i_0 be the maximal positive integer that is less than or equal to n and with $d_{i_0} > 0$.

For $z \in S$, we denote by $v(z)$ the vector $(v^i(z))_{i\in N} \in \mathbb{R}^N$. We define a $[0, 1]$-valued function w on S and a positive number $d > 0$ such that for every two states $z, z' \in S$ and player $i_1 \in N$, we have

$$v(z) \neq v(z') \implies |w(z) - w(z')| \geq d, \quad \text{and} \tag{23}$$

$$\left(\ v^i(z) = v^i(z') \quad \forall i < i_1 \text{ and } v^{i_1}(z) > v^{i_1}(z') \ \right) \implies w(z) > w(z'). \tag{24}$$

For example, the function $w(z) = \sum_{i=1}^{i_0} 2^{-i}(\prod_{j<i} d_j)v^i(z)$ is $[0, 1]$-valued and satisfies (24), and if $d = \min\{w(z) - w(z') : z, z' \in S, \text{ and } w(z) > w(z')\}$, then $d \geq 2^{-n} \prod_{j\leq i_0} d_j > 0$, and the function w and the positive constant d satisfy (23) and (24).

4.3 The auxiliary mixed actions and their basic properties

In this subsection we first select profiles of mixed actions $x(z, \overrightarrow{\rho}) \in X(z)$, $z \in S$, that depend on the vector $\overrightarrow{\rho} = (\rho^i)_{i\in N}$ of individual discount rates, and $X^*(z) \subset X(z)$ consists of all limit points of $x(z, \overrightarrow{\rho})$ as $\overrightarrow{\rho} \to 0$.

The subset of states $\bar{S}$ is defined as all states z for which there is a stationary strategy y, $z \mapsto y(z) \in X^*(z)$, such that $P_y^z(z_1 \notin C_z) > 0$.

We select a stationary strategy $y : z \mapsto y(z) \in X^*(z)$ and a positive number $0 < \delta < 1$ that satisfy

$$P_y^z(z_1 \notin C_z) > 4\delta \quad \forall z \in \bar{S}. \tag{25}$$

Definition of $x(z, \overrightarrow{\rho})$. For every $\overrightarrow{\rho} = (\rho^i)_{i\in N} \in (0, \infty)^N$, let $x(z, \overrightarrow{\rho})$, $z \in S$, be an equilibrium of the single-stage game with the payoff function to player i

$$a \mapsto \rho^i g^i(z, a) + \sum_{z'\in S} \mu(z', z, a) v^i_{\rho^i}(z').$$

Then, using the definition of $v^i_{\rho^i}$ and the fact that $x(z, \overrightarrow{\rho})$ is a profile of mixed actions and therefore $x^{-i}(z, \overrightarrow{\rho}) \in X^{-i}(z)$, we have,

$$\begin{aligned}
\rho^i v^i_{\rho^i}(z) &= \min_{x^{-i}} \max_{x^i} \left(\rho^i g^i(z, x^{-i} \otimes x^i) + \sum_{z' \in S} \mu(z', z, x^{-i} \otimes x^i) v^i_{\rho^i}(z') \right) \\
&\leq \max_{x^i} \left(\rho^i g^i(z, x^{-i}(z, \overrightarrow{\rho}) \otimes x^i) + \sum_{z' \in S} \mu(z', z, x^{-i}(z, \overrightarrow{\rho}) \otimes x^i) v^i_{\rho^i}(z') \right) \\
&= \rho^i g^i(z, x(z, \overrightarrow{\rho})) + \sum_{z' \in S} \mu(z', z, x(z, \overrightarrow{\rho})) v^i_{\rho^i}(z'), \qquad (26)
\end{aligned}$$

where the minimum is over all $x^{-i} \in X^{-i}(z) := \times_{i \neq j \in N} \Delta(A^j(z))$ and the maximum is over all $x^i \in X^i(z) := \Delta(A^i(z))$.

Definition of $X^*(z)$. For every $\theta > 0$ let $X(z, \theta) \subset X(z)$ be the closure of the set $\{(x(z, \overrightarrow{\rho})) : \max_{j \in N} \rho_j < \theta\}$. Let $X^*(z) \subset X(z)$ be the intersection of the sets $X(z, \theta)$, namely $X^*(z) := \cap_{\theta > 0} X(z, \theta)$.

As the sets $X(z, \theta)$ and $X(z)$ are nonempty and compact, and $X(z, \theta) \subseteq X(z, \theta')$ whenever $0 < \theta < \theta'$, the set $X^*(z)$ is nonempty and compact.

Note that if $y \in X^*(z)$ then there is a sequence $(\overrightarrow{\rho}(k))$ s.t. $\rho^i(k) \to_{k \to \infty} 0$ and $x(z, \overrightarrow{\rho}(k)) \to_{k \to \infty} y$. Therefore, by passing to the limit in inequality (26), for every $z \in S$, $y \in X^*(z)$, and $i \in N$, we have

$$\sum_{z' \in S} \mu(z', z, y) v^i(z') \geq 0. \qquad (27)$$

Definition of $\bar{S}$, y, and δ. The set of states $\bar{S}$, the stationary strategy profile $y : z \mapsto y(z) \in X^*(z)$, and $0 < \delta < 1$ are such that

$$P^z_y(z_1 \notin C_z) > 4\delta \ \ \forall z \in \bar{S},$$

$$\mu((S \setminus C_z) \cup \bar{S}, z, x) = 0 \ \ \forall z \notin \bar{S}, x \in X^*(z).$$

An explicit construction of $\bar{S}$, y, and δ is as follows.

The set $\bar{S}$ is the set of all states z for which there is a sequence $z = z_0, y_1, z_1, \ldots, y_k, z_k$ such that $v(z_k) \neq v(z)$, $y_j \in X^*(z_{j-1})$, and $\mu(z_j, z_{j-1}, y_j) > 0$, and $y : z \mapsto y(z) \in X^*(z)$ is a stationary strategy such that for every state $z \in \bar{S}$ there is a sequence $z = z_0, z_1, \ldots, z_k$ such that $v(z_k) \neq v(z)$ and $\mu(z_j, z_{j-1}, y(z_{j-1})) > 0$.

An alternative longer definition, but somehow more explicit, can be obtained by defining the sequence of disjoint subsets of states $\bar{S}_k$ with $z \in \bar{S}_k$ iff k is the minimal positive integer for which there is a sequence $z = z_0, y_1, z_1, \ldots, y_k, z_k$ such that $v(z_k) \neq v(z)$, $y_j \in X^*(z_{j-1})$, and $\mu(z_j, z_{j-1}, y_j) > 0$. In that case,

$$\bar{S}_1 := \{z : \exists y \in X^*(z) \text{ s.t. } \mu(S \setminus C_z, z, y) > 0\},$$

and for every $z \in \bar{S}_1$ we select an element $y(z) \in X^*(z)$ with $\mu(S \setminus C_z, z, y) > 0$. Inductively, for $k \geq 1$,

$$\bar{S}_{k+1} \;:=\; \{z \notin \cup_{j \leq k} \bar{S}_j : \exists y \in X^*(z) \text{ s.t. } \mu(\bar{S}_k, z, y) > 0\},$$

and for every $z \in \bar{S}_{k+1}$ we select an element $y(z) \in X^*(z)$ with $\mu(\bar{S}_k, z, y(z)) > 0$. We set

$$\bar{S} = \cup_{k=1}^{|S|} \bar{S}_k.$$

Next, we show that $\bar{S}$ is a proper subset of S. Let $z \in \arg\max_{z' \in S} w(z')$, i.e., $w(z) \geq w(z')\ \forall z' \in S$. By inequality (27) and implication (24), we deduce that (1) $w(z) < \max_{z' \in S} w(z')$ for every $z \in \bar{S}_1$, and (2) $w(z) < \max_{z' \in \bar{S}_k} w(z')$ for every $z \in \bar{S}_{k+1}$. Therefore, any state $z \in \arg\max_{z' \in S} w(z')$ (i.e., $w(z) \geq w(z')\ \forall z' \in S$) is not in $\bar{S}$.

For every $z \in S \setminus \bar{S}$ we select an (arbitrary) element $y(z) \in X^*(z)$, and define the stationary strategy y by $y : z \mapsto y(z)$.

As the probability P_y^z that the state z_1 at time 1 is not in C_z is > 0, i.e., $P_y^z(z_1 \notin C_z) > 0$, for every $z \in \bar{S}$ we can select $\delta > 0$ such that $P^z(z_1 \notin C_z) > 4\delta$ for all $z \in \bar{S}$.

Note that as $y(z) \in X^*(z)$ for every $z \in S$, inequality (27) implies that

$$E_y^z\, v(z_1) \geq v(z)\ \ \forall z \in S.$$

Definition of $B(z)$. For every $z \in S$ let

$$B(z) \;:=\; \{a \in A(z) \text{ s.t. } \mu((S \setminus C_z) \cup \bar{S}, z, a) > 0\}.$$

Fix $z \notin \bar{S}$. By the definition of $\bar{S}$, for every $x \in X^*(z)$, we have $\mu((S \setminus C_z) \cup \bar{S}, z, x) = 0$. Therefore, for any action profile a in the support of $x \in X^*(z)$, we have $\mu((S \setminus C_z) \cup \bar{S}, z, x) = 0$. As $X^*(z)$ is nonempty, we deduce that for any $z \notin \bar{S}$, $B(z)$ is a proper subset of $A(z)$.

Therefore, for $z \notin \bar{S}$, for every $x \in X(z)$ there is a point $x^* \in X(z)$ with $x^*(B(z)) = 0$ and $x^*(a) \geq x(a)$ for all $a \in A(z) \setminus B(z)$. The norm distance between x and x^*, $\|x - x^*\| := \sum_{a \in A(z)} |x^*(a) - x(a)|$ $(= x^*(A(z) \setminus B(z)) - x(A(z) \setminus B(z)) + \sum_{a \in B(z)} |x^*(a) - x(a)|)$, is equal to $2x(B(z))$.

Definition of ξ. Let $\xi > 0$ be a positive constant such that $\mu(z', z, a) > 0$ implies that $\mu(z', z, a) > \xi$.

4.4 The strategy τ and the time N_0

Lemma 3. *For every $1 > \varepsilon_1 > 0$, $\alpha_0 > 0$, and a positive integer $\ell > 8/\varepsilon_1$, there are functions $(z_{k/\ell})_{k=0}^{k=j} \mapsto \overrightarrow{\rho}_j \in (0, \alpha_0)^N$ (where k and j are nonnegative integers), and a sufficiently large positive integer N_0, such that if τ is the profile of strategies with $\tau(h, t) = x(z_t, \overrightarrow{\rho}_{[t\ell]})$, where $[t\ell]$ is the largest integer that is $\leq t\ell$, then for all $z \in S$ and $i \in N$,*

$$E_\tau^z v^i(z_T) \geq v^i(z) - \varepsilon_1/2 \quad \textit{for every } (\mathcal{H}_t)_{t \geq 0}\textit{-stopping time } T \tag{28}$$

and

$$\gamma_{N_0}^i(z, \tau) \geq v^i(z) - \varepsilon_1/2. \tag{29}$$

Proof. For every fixed vector $\overrightarrow{\rho} = (\rho^i)_{i \in N} \in (0, 1)^N$, the stationary strategy $\alpha(\overrightarrow{\rho}) : z \mapsto x(z, \overrightarrow{\rho})$ satisfies

$$E_\alpha^z\Big(\int_0^t \rho^i e^{-\rho^i t} g^i(z_t, x_t)\, dt + e^{-\rho^i t} v_{\rho^i}^i(z_t)\Big) \geq v_{\rho^i}^i(z) \ \ \forall t \geq 0. \tag{30}$$

Let $\varepsilon_2 \leq \varepsilon_1/2$; e.g., $\varepsilon_2 = \varepsilon_1/2$. Following the proof in [17], we select an integrable function $\psi : (0, 1] \to \mathbb{R}_+$ such that $\|dv_\rho(x)\| \leq \psi(x)$ $\forall x \in (0, 1]$ (equivalently, $|\int_\rho^{\bar{\rho}} \psi(x)\, dx| \geq \|v_\rho^i - v_{(\bar{\rho}}^i\|$ $\forall 0 < \rho, \bar{\rho} \leq 1$ and $\forall i \in N$), and define the function $s : (0, 1] \to [1, \infty)$ by

$$s(y) = \frac{12}{\varepsilon_2} \int_y^1 \frac{\psi(x)}{x} dx + y^{-1/2}.$$

The function s is strictly increasing and onto and its inverse function is denoted by λ.

Letting $\tilde{z}_j = z_{\frac{j}{\ell}}$ and $\tilde{\mathcal{H}}_j = \mathcal{H}_{\frac{j}{\ell}}$, we set

- $s_0^i \geq M$ with M sufficiently large,

- $\rho_j^i = \lambda(s_j^i)$,
- $u_j^i = v^i_{\rho_j^i}$ and $y_j = \ell E^z_{\alpha(\overrightarrow{\rho_j})}(\int_{\frac{j}{\ell}}^{\frac{j+1}{\ell}} g^i(z_t, x_t)\, dt \mid \tilde{\mathcal{H}}_j)$, and
- $s_{j+1}^i = \max\{M, s_j^i + y_j^i - u_j^i(\tilde{z}_{j+1}) + 4\varepsilon_2\}$.

The inequalities that follow depend on a fixed player i and for ease of notation we suppress the superscript i.

Inequality (30) implies, as in the proof of Theorem 3, that for every $\theta > 0$ there is α_0 sufficiently small such that for $\rho_j^i < \alpha_0$ and $\lambda_j = \rho_j^i/\ell$, we have

$$E^z_{\alpha(\overrightarrow{\rho}_j)}(\lambda_j y_j + (1-\lambda_j)u_j(\tilde{z}_{j+1}) \mid \tilde{\mathcal{H}}_j) \geq u_j(\tilde{z}_j) - \theta\frac{\rho_j^i}{\ell}.$$

Equivalently,

$$E^z_{\alpha}(u_j(\tilde{z}_{j+1}) - u_j(\tilde{z}_j) + \lambda_j(y_j - u_j(\tilde{z}_{j+1})) \mid \tilde{\mathcal{H}}_j) \geq -\theta\lambda_j. \tag{31}$$

Note the similarity between this last inequality and inequality (2.1) in [17]. One minor difference, however, is the added term of $-\theta\lambda_j$ on the right hand side of the inequality. Another one is that y_j depends $\overrightarrow{\rho}_j$ and thus measurable with respect to $\tilde{\mathcal{H}}_j$, while in inequality (2.1) in [17] the corresponding term is not a function of the past.

We first observe the following:

$$|s_{j+1} - s_j| \leq 6 \tag{32}$$

$$y_j - u_j(\tilde{z}_{j+1}) + 4\varepsilon_2 \leq s_{j+1} - s_j \leq y_j - u_j(\tilde{z}_{j+1}) + 4\varepsilon_2 + 21_{s_{j+1}=M}. \tag{33}$$

Since for every $1 > \eta > 0$, $s((1-\eta)y) - s(y) \to_{y\to 0+} \infty$, (32) implies that there is M_0 such that for every $M \geq M_0$

$$|\rho_{j+1}^i - \rho_j^i| \leq \varepsilon_2\rho_j/6. \tag{34}$$

Thus for $M \geq M_0$,

$$\begin{aligned} 6 &\geq |s_{j+1} - s_j| \geq \frac{12}{\varepsilon_2}\,|\int_{\rho_j^i}^{\rho_{j+1}^i} \frac{\psi(x)}{x}\, dx| \geq \frac{6}{\varepsilon_2\rho_j^i}|\int_{\rho_j^i}^{\rho_{j+1}^i} \psi(x)\, dx| \\ &\geq \frac{6}{\varepsilon_2\rho_j^i}\|v_{\rho_{j+1}^i} - v_{\rho_j^i}\|. \end{aligned}$$

Hence,

$$\|v^i_{\rho^i_{j+1}} - v^i_{\rho^i_j}\| \le \varepsilon_2 \rho^i_j. \tag{35}$$

By the mean value theorem and inequality (33),

$$\begin{aligned} \int_{s_j}^{s_{j+1}} \lambda(y)dy &= \int_{s_j}^{\infty} \lambda(y)dy - \int_{s_{j+1}}^{\infty} \lambda(y)dy \ge \rho^i_j(s_{j+1} - s_j) - \varepsilon_2\rho^i_j \\ &\ge \rho^i_j(y_j - u_j(\tilde{z}_{j+1})) + 3\varepsilon_2\rho^i_j. \end{aligned} \tag{36}$$

Replacing in (31), $u_j(\tilde{z}_{j+1})$ by the larger number $u_{j+1}(\tilde{z}_{j+1}) + \varepsilon_2\rho^i_j$ (see (35)), $\rho^i_j(y_j - u_j(\tilde{z}_{j+1}))$ by the larger number $\int_{s_j}^{\infty} \lambda(y)dy - \int_{s_{j+1}}^{\infty} \lambda(y)dy - 3\varepsilon_2\rho^i_j$ (see (36)), and letting $Y_j = u_j(\tilde{z}_j) - \int_{s_j}^{\infty} \lambda(y)dy$, we have (for $\theta < \varepsilon_2$)

$$E_\tau(Y_{j+1} - Y_j \mid \tilde{\mathcal{H}}_j) \ge \varepsilon_2\rho^i_j. \tag{37}$$

Y_j is thus a bounded ($|Y_j| \le 1$) submartingale (by (37)), and therefore for any $\tilde{\mathcal{H}}_j$-stopping-time T, $E^z_\tau Y_T \ge Y_0$.

By summing the right-hand side inequalities of (33) for $0 \le j < \ell n$, we have

$$\sum_{j<\ell n} y_j \ge \sum_{j<\ell n} u_j(\tilde{z}_{j+1}) + s_{\ell n} - s_0 - 2\sum_{j<\ell n} 1_{s_{j+1}=M} - 4\ell n\varepsilon_2. \tag{38}$$

For sufficiently large M, we have $E^z_\tau u_j(\tilde{z}_{j+1}) \ge v^i(z) - \varepsilon_2$ and

$$E^z_\tau \sum_{j<\ell n} 1_{s_{j+1}=M} \le E^z_\tau \sum_{j<\ell n} \frac{1}{\varepsilon_2\lambda(M_0)}\varepsilon_2\rho^i_j \le \frac{1}{\varepsilon_2\lambda(M_0)} E^z_\tau(Y_{\ell n} - Y_0).$$

Therefore, for all sufficiently large n,

$$E^z_\tau \int_0^n g^i(z_t, x_t)\,dt = \frac{1}{\ell n} E^z_\tau \sum_{j<\ell n} y_j \ge v^i(z) - 6\varepsilon_2.$$

This completes the proof of inequality (29).

For every real number x, let $\lceil x \rceil$ denote the smallest integer that is $\ge x$.

Note that if T is an $\mathcal{H}_t$-stopping time, then $\lceil \ell T \rceil$ is an $\tilde{\mathcal{H}}_j$-stopping time. As Y_j is a submartingale, we have $E^z_\tau Y_{\lceil \ell T \rceil} \ge Y_0$. Therefore, as $|Y_j - v^i(\tilde{z}_j)| < \varepsilon_2/16$ for M sufficiently large, we deduce that

$$E^z_\tau v^i(\tilde{z}_{\lceil \ell T \rceil}) > v^i(z) - \varepsilon_2/8.$$

By Lemma 1 and the assumption $\|\mu\| = 1$, we have $E\,|v^i(z_T) - v^i(z_{\frac{\lceil \ell T \rceil}{\ell}})| \leq 1/\ell \leq \varepsilon_2/8$.

Therefore,

$$E^z_\tau v^i(z_T) > v^i(z) - \varepsilon_2/4 \geq v^i(z) - \varepsilon_1/2.$$

This completes the proof of inequality (28). □

4.5 An auxiliary lemma

Lemma 4. *Let $C \subset S$ be a subset of states. Let $T = T_C = \min_{s \geq t}\{s : z_s \in C\}$. Then for every $(s_k)_{k\geq 0}$-discretimized correlated strategy σ and $n > t$,*

$$P_\sigma(t < T \leq n \mid \mathcal{H}_t) = E_\sigma\left(\int_t^{T\wedge n} \mu(C, z_s, x_s)\, ds \mid \mathcal{H}_t\right) \tag{39}$$

where $T \wedge n := \min(T, n)$.

Proof. Equality (39) holds trivially on $z_t \in C$. Assume that $z_t \notin C$. It suffices to prove the lemma for $t = 0$.

We provide two proofs. The first one approximates the probability that $T \leq n$ by summing the approximations of the probabilities that $s < T \leq s+\delta$, where $\delta > 0$ is small, and s ranges over a grid of n/δ points. The error term of each approximation is $O(\delta^2)$. The sum of the approximations is the desired formula and the cumulative error term is $O(\delta)$. The second one derives the formula without an approximation.

The mixed action profile x_s that the $(s_k)_{k\geq 0}$-discretimized correlated strategy σ selects at time s is a function of the random list $(s_k \wedge s, z_{s_k \wedge s})$, $k \geq 0$.

For $s_j \leq s < s_{j+1}$ we denote this action profile by $x_{j,s}$. Note that $x_{j,s}$ is measurable w.r.t. $\mathcal{H}^S_{s_j}$ and is defined for every $s_j \leq s$. It coincides with x_s on $s_j \leq s < s_{j+1}$, but may differ from x_s on $s_{j+1} \leq s$.

Given $s_j \leq s < s_{j+1}$ and $T > s$, the conditional P^z_σ probability, given $\mathcal{H}^S_s$, that $s < T = s_{j+1} \leq s + \delta$ equals

$$\int_s^{s+\delta} \mu(C, z_{s_j}, x_{j,t}) e^{\int_s^t \mu(z_{s_j}, z_{s_j}, x_{j,u})\, du}\, dt.$$

If there is no state change in the interval $[s, s + \delta]$, then $x_{j,t} = x_t$ for every $s \leq t \leq s + \delta$. Therefore, given $s_j \leq s < s_{j+1}$, the P^z_σ probability that $x_t = x_{j,t}$ for every $s \leq t \leq s + \delta$ is at least $1 - \delta\|\mu\|$.

Therefore, the above conditional probability, given $\mathcal{H}_s^S$, is within $O(\delta^2)$, the conditional probability, given $\mathcal{H}_s^S$, of $\int_s^{s+\delta} \mu(C, z_t, x_t)\, dt$.

Therefore, the P_σ^z probability that $T \leq n$ is, within $O(\delta)$, the expectation w.r.t. P_σ^z of $\int_0^{T\wedge n} \mu(C, z_t, x_t)\, dt$. As this holds for any $\delta > 0$ the result follows. □

Now we present the alternative proof.

Note that on $s_j < n$ the conditional expectation of the indicator of the event $s_j < T = s_{j+1} \leq n$, given the history up to time s_j, equals $\int_{s_j}^n \mu(C, z_{s_j}, x_{j,t}) e^{\int_{s_j}^t \mu(z_{s_j}, z_{s_j}, x_{j,u})\, du}\, dt$. In symbols,

$$E_\sigma(\mathbb{I}(s_j < T = s_{j+1} \leq n) \mid \mathcal{H}_{s_j}) = \int_{s_j}^n \mu(C, z_{s_j}, x_{j,t}) e^{\int_{s_j}^t \mu(z_{s_j}, z_{s_j}, x_{j,u})\, du}\, dt,$$

where $\mathbb{I}(*)$ denotes the indicator of the event $*$.

The term $e^{\int_{s_j}^t \mu(z_{s_j}, z_{s_j}, x_{j,u})\, du}$ equals the conditional probability of the event $t < s_{j+1}$, given the history up to time s_j. Therefore,

$$\int_{s_j}^n \mu(C, x_{s_j}, x_{j,t}) e^{\int_{s_j}^t \mu(z_{s_j}, z_{s_j}, x_{j,u})\, du}\, dt = \int_{s_j}^n \mu(C, z_{s_j}, x_{j,t}) E_\sigma(\mathbb{I}(t < s_{j+1}) \mid \mathcal{H}_{s_j})\, dt.$$

On $s_j \leq t < s_{j+1}$, $z_{s_j} = z_t$ and $x_{j,t} = x_t$. Therefore,

$$\int_{s_j}^n \mu(C, z_{s_j}, x_{j,t}) E_\sigma(\mathbb{I}(t < s_{j+1}) \mid \mathcal{H}_{s_j})\, dt = E_\sigma(\int_{s_j}^{s_{j+1}} \mu(C, z_t, x_t)\, dt \mid \mathcal{H}_{s_j}).$$

We conclude that

$$E_\sigma(\mathbb{I}(s_j < T = s_{j+1} \leq n) \mid \mathcal{H}_{s_j}) = E_\sigma(\int_{s_j}^{s_{j+1}} \mu(C, z_t, x_t)\, dt \mid \mathcal{H}_{s_j}).$$

Taking the expectation on both sides and summing over all nonnegative integers j, we have

$$\begin{aligned} P_\sigma^z(T \leq n) &= E_\sigma^z(\sum_{j=0}^{\infty} \mathbb{I}(s_j < T = s_{j+1} \leq n)) \\ &= E_\sigma^z(\int_0^{T\wedge n} \mu(C, z_t, x_t)\, dt). \end{aligned}$$

□

4.6 The partitioning of the state space: The sets S_1 and S_2

The subset S_1 of $S \setminus \bar{S}$ is defined below so as to guarantee the existence of a correlated strategy $\tilde{\tau}$ such that for $z \in S_1$ (1) the distance between the P_τ^z-distribution and the $P_{\tilde{\tau}}^z$-distribution of the state process up to time N_0 is small, (2) the difference between the expectation w.r.t. P_τ^z and the expectation w.r.t. $P_{\tilde{\tau}}^z$ of the accumulated payoff up to time N_0 is small, and (2) the state process remains in $C_z \setminus \bar{S}$ with $P_{\tilde{\tau}}^z$-probability 1. See Section 4.7.

Fix $\varepsilon_2 > 0$ ($\varepsilon_2 = O(\varepsilon)$; e.g., $\varepsilon_2 = \varepsilon\xi/2$ or $\varepsilon_2 = \varepsilon$). The subset S_1 of $S \setminus \bar{S}$ depends on the constant $\varepsilon_2 > 0$ and the strategy τ as follows. Define the $(\mathcal{H}_t^S)_{t\geq 0}$-stopping time T by $T = \min\{t : z_t \in (S \setminus C_z) \cup \bar{S}\}$. Set

$$S_1 = \{z \in S : P_\tau^z(T \leq N_0) \leq 2\varepsilon_2\}$$

and

$$S_2 = S \setminus (\bar{S} \cup S_1) = \{z \in S \setminus \bar{S} : P_\tau^z(T \leq N_0) > 2\varepsilon_2\}.$$

Note that $T = 0$ on $z_0 \in \bar{S}$. Therefore, by the definition of S_1, we have $S_1 \cap \bar{S} = \emptyset$.

Lemma 5. *For every $z \in S_1$,*

$$E_\tau^z \left(\int_0^{T\wedge N_0} x(z_t, \overrightarrow{\rho}_t)[B(z_t)]\, dt \right) \leq 2\varepsilon_2/\xi. \tag{40}$$

Proof. It follows from the selection of $\xi > 0$ that

$$\mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \geq \xi x(z_t, \overrightarrow{\rho}_{[t]})[B(z_t)] \quad \text{on } t < T. \tag{41}$$

Therefore (using Lemma 4), if $z \in S_1$, then

$$\begin{aligned} 2\varepsilon_2 &\geq P_\tau^z(T \leq N_0) = E_\tau^z \left(\int_0^{T\wedge N_0} \mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]}))\right) dt \\ &\geq E_\tau^z \left(\int_0^{T\wedge N_0} \xi x(z_t, \overrightarrow{\rho}_{[t]})[B(z_t)]\, dt \right). \end{aligned} \tag{42}$$

□

4.7 Definition of the correlated strategy $\tilde{\tau}$

We start by defining an auxiliary correlated strategy $\tilde{\tau}$.

$$\tilde{\tau}(h,t) = \begin{cases} y(z_t) & \text{if } z_0 \in \bar{S} \\ \tau(h,t) & \text{if } z_0 \in S_2 \\ x^*(z_t, \overrightarrow{\rho}_{[t]}) & \text{if } z_0 \in S_1, \end{cases}$$

where $x^*(z_t, \overrightarrow{\rho}_{[t]})$ is a point in $\{x \in X(z_t) : x(B(z_t)) = 0\}$ with $\|x^*(z_t, \overrightarrow{\rho}_{[t]}) - x(z_t, \overrightarrow{\rho}_{[t]})\| \leq 2x(z_t, \overrightarrow{\rho}_{[t]})[B(z)]$ if $x(z_t, \overrightarrow{\rho}_{[t]})[B(z)] > 0$, and $x^*(z_t, \overrightarrow{\rho}_{[t]}) = x(z_t, \overrightarrow{\rho}_{[t]})$ otherwise. Recall that for $z \notin \bar{S}$, $B(z)$ is a proper subset of $A(z)$, and therefore $\{x \in X(z) : x(B(z)) = 0\}$ is nonempty.

Note $\tau(h,t) = x(z_t, \overrightarrow{\rho}_{[t]})$ is a profile of mixed actions. However, $x^*(z_t, \overrightarrow{\rho}_{[t]})$ need not be a profile of mixed actions. Therefore, the correlated strategy $\tilde{\tau}$ need not be a profile of strategies.

Recall that by the definition of $\bar{S}$ it follows that for every $z \in S \setminus \bar{S}$ and every $x \in X^*(z)$, we have $\mu(\bar{S}, z, x) = 0$ and $\sum_{z' \in S} \mu(z', z, x) v(z') = 0$.

Lemma 6. *For every $z \in S_1$, $P^z_{\tilde{\tau}}(z_t \in C_z \setminus \bar{S} \ \ \forall 0 \leq t \leq N_0) = 1$ and therefore $P^z_{\tilde{\tau}}(z_{N_0} \in C_z \setminus \bar{S}) = 1$.*

Proof. Follows from the definition of $\tilde{\tau}$ on $z_0 \in S_1$. □

Lemma 7. *For every $z \in S_1$, the norm distance[12] between the P^z_{τ}-distribution and the $P^z_{\tilde{\tau}}$-distribution of the state process $(z_t)_{0 \leq t \leq T \wedge N_0}$ is $\leq 4\varepsilon_2/\xi$.*

Proof. By the definition of $\tilde{\tau}$, $\|\tilde{\tau}(h,t) - \tau(h,t)\| \leq 2x(z_t, \overrightarrow{\rho}_{[t]})[B(z)]$ on $z_0 = z \in S_1$. Therefore, the result follows from Lemmas 12 (in Section 6) and 5. □

Lemma 8. *For every $z \in S_1$,*

$$\begin{aligned} E^z_{\tilde{\tau}} \int_0^{N_0} g^i_t \, dt &\geq E^z_{\tau} \int_0^{N_0} g^i_t \, dt - 2(2N_0 + 2)\varepsilon_2/\xi \\ &\geq N_0 \left(v^i(z) - \varepsilon_1/2 - O(\varepsilon_2) \right). \end{aligned}$$

[12]Recall that the norm distance between two probability measures P and Q on a measurable space $(X, \mathcal{X})$ is defined by $\|P - Q\| := 2 \sup_{B \in \mathcal{X}} |P(B) - Q(B)|$.

Proof. In the following sequence of equations, equality (43) follows from the fact that on the plays that are compatible with the strategy $\tilde{\tau}$ we have $g_t^i = g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]}))$.

Inequality (44) follows from:
(i) the inequality $0 \leq g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \leq 1$,
(ii) $\overrightarrow{\rho}_{[t]}$ and (therefore also) $g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]}))$ depend only on the state process in time $s \in [0, N_0]$, and
(iii) half the norm distance between the distributions defined by τ and $\tilde{\tau}$ on this state process is (by Lemma 7) less than or equal to $2\varepsilon_2/\xi$.

Inequality (45) follows from the inequality

$$g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]})) \geq g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]})) - \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|.$$

Inequality (46) follows from the equality

$$\begin{aligned} E_\tau^z \int_0^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|\, dt \;=\; & E_\tau^z \int_0^{T\wedge N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|\, dt \\ & + E_\tau^z \int_{T\wedge N_0}^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|\, dt \end{aligned}$$

and the inequality

$$E_\tau^z \int_{T\wedge N_0}^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|\, dt \leq P_\tau^z(T < N_0) 2N_0.$$

Inequality (47) follows from:
(i) the inequality $P_\tau^z(T < N_0) \leq 2\varepsilon_2$ for $z \in S_1$,
(ii) the inequality $\|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \leq 2x(z_t, \overrightarrow{\rho}_{[t]})[B(z_t)]$, and
(iii) the inequality $E_\tau^z \int_0^{T\wedge N_0} 2x(z_t, \overrightarrow{\rho}_{[t]})[B(z_t)]\, dt \leq 4\varepsilon_2/\xi$.

Finally, inequality (48) follows from inequality (29).

$$
\begin{aligned}
E^z_{\tilde\tau}\int_0^{N_0} g^i_t\,dt &= E^z_{\tilde\tau}\int_0^{N_0} g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]}))\,dt && (43)\\
&\geq E^z_{\tau}\int_0^{N_0} g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]}))\,dt - 2N_0\varepsilon_2/\xi && (44)\\
&\geq E^z_{\tau}\int_0^{N_0} g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]}))\,dt - 2N_0\varepsilon_2/\xi && (45)\\
&\quad -E^z_{\tau}\int_0^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|\,dt\\
&\geq E^z_{\tau}\int_0^{N_0} g^i_t\,dt - 2N_0\varepsilon_2/\xi - P^z_\tau(T<N_0)2N_0 && (46)\\
&\quad -E^z_{\tau}\int_0^{T\wedge N_0} \|x(z_t, \overrightarrow{\rho}_{t}) - x^*(z_t, \overrightarrow{\rho}_{t})\|\,dt\\
&\geq E^z_{\tau}\int_0^{N_0} g^i_t\,dt - 4\varepsilon_2 N_0 - \varepsilon_2/\xi(4+2N_0) && (47)\\
&\geq N_0\left(v^i(z) - \varepsilon_1/2 - O(\varepsilon_2)\right). && (48)
\end{aligned}
$$

□

4.8 The discretimization $\bar\tau$ of $\tilde\tau$ and the strategy $\hat\tau$

Recall that conditional on h_j, the strategy $\tilde\tau$ on the time-interval $[j, j+1)$ coincides with a stationary strategy, denoted by $\tilde\tau[h_j] : z \mapsto \tilde\tau[h_j](z)$.

We discretimize $\tilde\tau$ into a pure-action strategy $\bar\tau$ as follows. Let ℓ be a sufficiently large positive integer, and conditional on $\overrightarrow{z}_j := (z_0, z_1, \ldots, z_j)$, j a nonnegative integer. The average of $\bar\tau$ on time intervals $[j+k/\ell, j+(k+1)/\ell]$, k a positive integer with $0 \leq k < \ell$, is within $O(1/\ell)$ of $\tilde\tau[h_j]$ – the average mixed action of $\tilde\tau$ over that interval. For example, define nonnegative integers $\ell(z,a) \geq \ell\,\tilde\tau[h_j](z)[a] - 1$, where $z \in S$ and $a \in A(z)$, such that $\sum_{a\in A(z)} \ell(z,a) = \ell$. For $B \subset A(z)$ we set $\ell(z,B) := \sum_{a\in B}\ell(z,a)$. Let $\bar\tau$ be a pure-action strategy that conditional on h_j, (1) $\bar\tau[h_j](z,t) = \bar\tau[h_j](z, j+i/\ell^2)$ if $j \leq j+i/\ell^2 \leq t < j+(i+1)/\ell^2 \leq j+1$, (2) for every $0 \leq k < \ell$, $z \in S$, and $a \in A(z)$, $\sum_{0\leq i<\ell}\bar\tau[h_j](z, j+k/\ell+i/\ell^2)[a] = \ell(z,a)$, and (3) $\bar\tau(h_t)[B(z_t)] = 0$ on $z_t \notin \bar S$. Therefore, conditional on $\overrightarrow{z}_j$, for every $z \in S$,

$B \subset A(z)$, and $0 \leq k < \ell$, we have $\int_{j+k/\ell}^{j+(k+1)/\ell} \bar{\tau}[h_j](z,t)[B]\,dt = \frac{1}{\ell^2}|\{0 \leq i < \ell : \bar{\tau}[h_j](z, j+k/\ell+i/\ell^2)[B] = 1\}| \geq \ell(z,B)/\ell^2 \geq \tilde{\tau}[h_j](z)[B]/\ell - 1/\ell^2$. Therefore, for $z \notin \bar{S}$,

$$\begin{aligned} \|\frac{1}{\ell}\tilde{\tau}[h_j](z) - \int_{j+k/\ell}^{j+(k+1)/\ell} \bar{\tau}[h_j](z,t)\,dt\| &= 2 \max_{B \subset A(z)\setminus B(z)} \frac{\tilde{\tau}[h_j](z)[B] - \ell(z,B)/\ell}{\ell} \\ &\leq 2|A(z)|/\ell^2 = O(1/\ell^2). \end{aligned} \tag{49}$$

For two nonnegative integers k, ℓ we denote by z_k^ℓ the sequence of states $z_0, z_{1/\ell}, z_{2/\ell}, \ldots, z_{k/\ell}$. By Lemma 12 in Section 6, conditional on $z_0 \in S_1$,

$$\sum_{z_{\ell N_0}^\ell} |P_{\tilde{\tau}}^z(z_{\ell N_0}^\ell) - P_{\bar{\tau}}^z(z_{\ell N_0}^\ell)| \leq N_0(4+2|A|)/\ell = O(1/\ell). \tag{50}$$

Lemma 9. *For every $z \in S_1$, $P_{\bar{\tau}}^z(z_t \in C_z \setminus \bar{S} \ \ \forall\, 0 \leq t \leq N_0) = 1$, and*

$$\gamma_{N_0}^i(z, \bar{\tau}) \geq v^i(z) - \varepsilon_1 - O(\varepsilon_2) \ \ \forall \ell \text{ sufficiently large.}$$

Proof. The first assertion follows from $\bar{\tau}$ being a pure-action strategy and from the imposition of $\bar{\tau}(h,t) \notin B(z_t)$ on $z_t \notin \bar{S}$.

For every nonnegative integer k (for every strategy profile and every initial state), the conditional probability, given $h_{k/\ell}$, that $z_t = z_{k/\ell}$ for all $k/\ell \leq t < (k+1)/\ell$, is greater than or equal to $1 - 1/\ell$. On $z_t = z_{k/\ell}$ for all $k/\ell \leq t < (k+1)/\ell$, we have (by (49))

$$\int_{k/\ell}^{(k+1)/\ell} g^i(z_t, \bar{\tau}(h,t))\,dt \geq \int_{k/\ell}^{(k+1)/\ell} g^i(z_t, \tilde{\tau}(h,t))\,dt - |A|/\ell^2.$$

Therefore, for $z \in S_1$ and $z_{k/\ell} \in C_z \setminus \bar{S}$,

$$E_{\bar{\tau}}^z(\int_{k/\ell}^{(k+1)/\ell} g_t^i\,dt \mid z_k^\ell) \geq E_{\tilde{\tau}}^z(\int_{k/\ell}^{(k+1)/\ell} g_t^i\,dt \mid z_k^\ell) - (|A|+1)/\ell^2. \tag{51}$$

Therefore, using (51) and (50),

$$
\begin{aligned}
N_0\gamma^i_{N_0}(z,\bar\tau) &= \sum_{k=0}^{\ell N_0-1} E^z_{\bar\tau}E^z_{\bar\tau}\Big(\int_{k/\ell}^{(k+1)/\ell} g^i_t\,dt \mid z^\ell_k\Big)\\
&\geq \sum_{k=0}^{\ell N_0-1} E^z_{\bar\tau}E^z_{\tilde\tau}\Big(\int_{k/\ell}^{(k+1)/\ell} g^i_t\,dt \mid z^\ell_k\Big) - \frac{(|A|+1)N_0}{\ell}\\
&= \sum_{z^\ell_{\ell N_0}} P^z_{\bar\tau}(z^\ell_{\ell N_0}) \sum_{k=0}^{\ell N_0-1} E^z_{\tilde\tau}\Big(\int_{k/\ell}^{(k+1)/\ell} g^i_t\,dt \mid z^\ell_k\Big) - \frac{(|A|+1)N_0}{\ell}\\
&\geq \sum_{z^\ell_{\ell N_0}} P^z_{\tilde\tau}(z^\ell_{\ell N_0}) \sum_{k=0}^{\ell N_0-1} E^z_{\tilde\tau}\Big(\int_{k/\ell}^{(k+1)/\ell} g^i_t\,dt \mid z^\ell_k\Big) - \frac{(|A|+1)N_0}{\ell}\\
&\quad - \sum_{z^\ell_{\ell N_0}} |P^z_{\tilde\tau}(z^\ell_{\ell N_0}) - P^z_{\bar\tau}(z^\ell_{\ell N_0})|N_0\\
&\geq N_0\gamma^i_{N_0}(z,\tilde\tau) - \frac{(|A|+1)N_0}{\ell} - \frac{(2|A|+4)N_0^2}{\ell}.
\end{aligned}
$$

Using Lemma 8 in Section 4.7 we conclude that for all sufficiently large ℓ, we have $\gamma^i_{N_0}(z,\bar\tau) \geq v^i(z) - \varepsilon_1/2 - O(\varepsilon_2)$. □

The strategy $\hat\tau$ plays in blocks of random sizes that are $\leq N_0$, and in each block it follows $\bar\tau$ as if the game were starting at the state of the start of the block.

The duration of a block that starts at a state $z \in S_1$ is N_0.

The duration of a block that starts at a state $z \in \bar S$ is 1.

A block that starts at time t in state $z \in S_2$ continues until the first time $t+j/\ell$, with j being a positive integer and $z_s \notin C_z \backslash \bar S$ for some $t < s \leq t+j/\ell$, if this time is $\leq t + N_0$. Otherwise it continuous until time $t + N_0$.

Let $\bar t_k$ be the time at the end of the k-th block ($\bar t_0 = 0$), and $\bar z_k$ be the state at the end of the block, i.e., $\bar z_k = z_{\bar t_k}$.

The strategy $\hat\tau$ is given by $\hat\tau(h,t) = \bar\tau(\bar t_k * h, t - \bar t_k)$, whenever $\bar t_k \leq t < \bar t_{k+1}$, and where $\bar t_k * h$ is the left translation of the play h by $\bar t_k$; i.e., if $h = (z_t, x_t)_{t\geq 0}$, then $\bar t_k * h = (z_{\bar t_k+s}, x_{\bar t_k+s})_{s\geq 0}$.

Let us formally define the stopping times $\bar t_k$ inductively. $\bar t_{k+1}$ is a function of $\bar t_k$ and $z_{\bar t_k}$, and of the process that follows time $\bar t_k$. If $z_{\bar t_k} \in S_1$ then $\bar t_{k+1} = \bar t_k + N_0$. If $z_{\bar t_k} \in S_2$ then $\bar t_{k+1} = N_0 \wedge \inf\{j/\ell : \mathbb{N} \ni j, \exists j/\ell \geq s > \bar t_k \text{ s.t. } z_s \in (S \setminus C_z) \cup \bar S\}$. If $z_{\bar t_k} \in \bar S$ then $\bar t_{k+1} = \bar t_k + 1$.

Lemma 10. *The discrete-time process $\bar{z}_0, \bar{z}_1, \ldots$, with the probability $P_{\hat{\tau}}$ is a homogeneous Markov chain process with a transition matrix F that obeys the following properties.*

1) $F_{z,C_z} := \sum_{z' \in C_z} F_{z,z'} = 1$ *for* $z \in S_1$,

2) $F_{z,S \setminus C_z} > 4\delta - O(1/\ell) \geq 3\delta$ *for* $z \in \bar{S}$ *and* ℓ *sufficiently large,*

3) $F_{z,S \setminus C_z} + F_{z,\bar{S}} \geq 2\varepsilon_2 - O(1/\ell) \geq \varepsilon_2$ *for* $z \in S_2$ *and* ℓ *sufficiently large, and*

4) $\sum_{z' \in S} F_{z,z'} v^i(z') \geq v^i(z) - \varepsilon_1 \mathbb{I}(z \in S_2 \cup \bar{S})$ *for every* $z \in S$, $i \in N$, *and* ℓ *sufficiently large.*

Proof. By the definition of $\hat{\tau}$, the $P_{\hat{\tau}}$ conditional distribution of $\bar{z}_{k+1}$, given $\bar{z}_0, \ldots, \bar{z}_k$, depends only on $\bar{z}_k$. Therefore, the stochastic $\bar{z}_0, \ldots, \bar{z}_k, \ldots$, defined by $\hat{\tau}$, is a homogeneous Markov chain.

Equality 1) follows from the definition of $\bar{\tau}$ on $z_0 \in S_1$. Recall that the norm distance between the $P^z_{\bar{\tau}}$ and the $P^z_{\hat{\tau}}$ distribution of $z_0, z_1, \ldots, z_{N_0}$ is $\leq O(1/\ell)$. For $z \in \bar{S}$, $F_{z,S \setminus C_z} = P^z_{\hat{\tau}}(z_1 \in S \setminus C_z) = P^z_{\bar{\tau}}(z_1 \in S \setminus C_z) \geq P^z_{\hat{\tau}}(z_1 \in S \setminus C_z) - O(1/\ell) > 4\delta - O(1/\ell) \geq 3\delta$ for sufficiently large ℓ. Therefore, 2) holds. For $z \in S_2$, $F_{z,S \setminus C_z} + F_{z,\bar{S}} \geq P^z_{\bar{\tau}}(\exists j \leq \ell N_0 \text{ s.t. } z_{j/\ell} \in (S \setminus C_z) \cup \bar{S}) \geq P^z_{\bar{\tau}}(\exists j \leq \ell N_0 \text{ s.t. } z_{j/\ell} \in (S \setminus C_z) \cup \bar{S}) - O(1/\ell) \geq P^z_{\bar{\tau}}(T \leq N_0) - O(1/\ell)$ (where the last inequality uses the fact that for every stopping time T and every strategy σ, $P_\sigma(z_s = z_T \forall T \leq s \leq T + 1/\ell) \geq 1 - O(1/\ell)$. Therefore, as $P^z_{\bar{\tau}}(T \leq N_0) \geq 2\varepsilon_2$, part 3) follows.

By 1), for $z \in S_1$ we have $\sum_{z' \in S} F_{z,z'} v^i(z') = v^i(z)$; for $z \in \bar{S}$ we have $\sum_{z' \in S} F_{z,z'} v^i(z') = \sum_{z' \in S} P^z_{\bar{\tau}}(z_1 = z') v^i(z') \geq \sum_{z' \in S} P^z_{\hat{\tau}}(z_1 = z') v^i(z') - O(1/\ell) \geq v^i(z) - O(1/\ell)$, and for $z \in S_2$, using (28), we have $\sum_{z' \in S} F_{z,z'} w(z') \geq \sum_{z' \in S} F_{z,z'} w(z) - \varepsilon_1/2 - O(1/\ell)$. Therefore, 4) follows. □

4.9 A probabilistic lemma

The following lemma appears implicitly in [36].

Lemma 11. *Assume that* $0 < \varepsilon := \varepsilon_2 \leq 1$ *and* $\varepsilon_1 < \delta d^2 \varepsilon / 4$. *Then*

(a) All ergodic classes of the Markov chain $(\bar{z}_j)$ are subsets of S_1,

(b) $E(|\{0 \leq j < \infty : \bar{z}_j \notin S_1\}|) < \infty$, *and*

(c) $w_j := w(\bar{z}_j) \to_{j\to\infty} w_\infty$, *and therefore* $v^i_j := v^i(\bar{z}_j) \to_{j\to\infty} v^i_\infty$ $\forall i \in N$,

and, if $\varepsilon_1 < \frac{\varepsilon^2 d^2 \delta}{16}$, *then*

(d) $E|\{0 \le j < \ell : z_j \in \bar{S} \cup S_2\}| \le \frac{16}{\delta d^2 \varepsilon}$, *and*

(e) $\forall i \in N$, $E(v^i_j \mid z_0) \ge v^i_0 - \varepsilon$ *and* $E(v^i_\infty \mid z_0) \ge v^i_0 - \varepsilon$.

Proof. Let $\mathcal{F}_j$, $j \ge 0$, be the algebra generated by $\bar{z}_0, \ldots, \bar{z}_j$. Consider the sequences w_j and Y_j of real-valued random variables $w_j = w(\bar{z}_j)$ and $Y_j = w_j^2 + \delta_j$, where $\delta_j = \delta(\bar{z}_j)$ and $\varepsilon \ge \delta(z) \ge 0$ is defined as follows. For $z \in \bar{S}$, we set $\delta(z) = \delta d^2 := \delta(0)$, for $z \in S_2$ we set $\delta(z) = \delta(2) = \delta\varepsilon d^2/4 - \varepsilon_1$, and for $z \in S_1$ we set $\delta(z) = 0$.

We claim that

$$E(Y_{j+1} \mid \mathcal{F}_j) \ge Y_j + \delta(\bar{z}_j) \ge Y_j + \delta(2)\mathbb{I}(\bar{z}_j \in S_2 \cup \bar{S}). \tag{52}$$

As the process $(\bar{z}_j)_{j\ge 0}$ is a stationary Markov chain, it suffices to prove inequality (52) for $j = 0$.

On $\bar{z}_0 \in S_1$ we have $Y_0 = w_0^2$, $w_1 = w_0$ a.e., and $Y_1 \ge w_1^2 = Y_0 + \delta(\bar{z}_0)$ a.e. Therefore, inequality (52) holds on $\bar{z}_0 \in S_1$.

On $\bar{z}_0 \in \bar{S}$, we have $E(w_1 \mid z_0) = w_0$ and the probability that $|w_1 - w_0| \ge d$ is $\ge 3\delta$. Therefore, $E(w_1^2 \mid z_0) - w_0^2 \ge 3\delta d^2$, implying that $E(Y_1 \mid z_0) \ge w_0^2 + 3\delta d^2 = Y_0 + 2\delta(0) \ge Y_0 + \delta(0)$.

We partition S_2 into two subsets. $S_2(\neq) := \{z \in S_2 : P(w(\bar{z}_1) \ne w(\bar{z}_0) \mid \bar{z}_0 = z) \ge \varepsilon/2\} = \{z \in S_2 : F_{z,S\setminus C_z} \ge \varepsilon/2\}$ and $S_2(=) := S_2 \setminus S_2(\neq)$. The inequality $F_{z,S\setminus C_z} + F_{z,\bar{S}} \ge \varepsilon$ for $z \in S_2$ implies that for $z \in S_2 \setminus S_2(\neq)$ we have $P(\bar{z}_1 \in \bar{S} \mid \bar{z}_0 = z) \ge \varepsilon/2$.

On $\bar{z}_0 \in S_2(\neq)$, the (conditional) probability that $|w_1 - w_0| \ge d$ is $\ge \varepsilon/2$, and $E(w_1 \mid \bar{z}_0) \ge w_0 - \varepsilon_1$. Therefore, on $\bar{z}_0 \in S_2(\neq)$,

$$\begin{aligned} E(w_1^2 \mid \bar{z}_0) &= E((w_1 - w_0)^2 + 2w_0 w_1 - w_0^2 \mid \bar{z}_0) \ge d^2\varepsilon/2 + 2w_0(w_0 - \varepsilon_1) - w_0^2 \\ &\ge w_0^2 - 2\varepsilon_1 + d^2\varepsilon/2 \ge w_0^2 + 2\delta(2) = Y_0 + \delta(2), \end{aligned}$$

where the last inequality follows from $2\delta(2) = \delta\varepsilon d^2/2 - 2\varepsilon_1 \le d^2\varepsilon/2 - 2\varepsilon_1$.

Everywhere, and thus also on $z_0 \in S_2 \setminus S_2(\neq)$, $E(w_1 \mid z_0) \ge w_0 - \varepsilon_1$. The conditional probability that $\bar{z}_1 \in \bar{S}$, given $\bar{z}_0 \in S_2 \setminus S_2(\neq)$, is $\ge \varepsilon/2$. Therefore, on $\bar{z}_0 \in S_2 \setminus S_2(\neq)$, using the convexity of $x \mapsto x^2$,

$$\begin{aligned} E(Y_1 \mid \bar{z}_0) &\ge (E(w_1 \mid \bar{z}_0))^2 + \delta(0)\varepsilon/2 \ge (w_0 - \varepsilon_1)^2 - \varepsilon_1^2 + \delta d^2\varepsilon/2 \\ &= w_0^2 - 2\varepsilon_1 + \delta d^2\varepsilon/2 = w_0^2 + 2\delta(2) = Y_0 + \delta(2). \end{aligned}$$

This completes the proof of inequality (52).

Inequality (52) shows that Y_j is a bounded ($0 \leq Y_j \leq 2$) submartingale, and therefore it converges a.e. to Y_∞. By taking the expectation in inequality (52), and summing over all integers $0 \leq j < \ell$, we have $2 \geq EY_\ell \geq Y_0 + \delta(2)E|\{0 \leq j < \ell : z_j \in \bar{S} \cup S_2\}|$, implying that $E|\{0 \leq j < \ell : z_j \in \bar{S} \cup S_2\}|$ is bounded by $2/\delta(2)$. This completes the proof of (b). Claim (b) implies that with probability 1, $|\{0 \leq j < \infty : \bar{z}_j \notin S_1\}| < \infty$, which together with part 1) of Lemma 10 proves part (c).

By selecting $\varepsilon_1 < \frac{\varepsilon^2 d^2 \delta}{16}$ we have $\delta(2) \geq \varepsilon d^2 \delta/8$. Therefore, $2/\delta(2) \leq \frac{16}{\varepsilon d^2 \delta}$, proving part (d). For every $i \in N$ and $j \geq 0$, by Lemma 10 (part 4), we have

$$E(v^i_{j+1} \mid \mathcal{F}_j) \geq w_j - \varepsilon_1 \mathbb{I}(z_j \in S_2 \cup \bar{S}).$$

Summing these inequalities over all $0 \leq j < \ell$, we have

$$E(v^i_\ell \mid z_0) \geq v^i(z_0) - \varepsilon_1 E \sum_{0 \leq j < \ell} \mathbb{I}(z_j \in S_2 \cup \bar{S}) \geq v^i(z_0) - \varepsilon, \quad \text{and therefore}$$

$$E(v^i_\infty \mid z_0) \geq v^i(z_0) - \varepsilon.$$

This completes the proof of part (e).

□

4.10 The uniform-l-average equilibrium strategy σ

For every player $i \in N$ let σ_ε^{-i} be an $N \setminus \{i\}$ strategy profile and s_ε a positive number such that for every $s \geq s_\varepsilon$ and every strategy $\tilde{\sigma}^i$ we have $\gamma^i_s(z, \sigma_\varepsilon^{-i}, \tilde{\sigma}^i) \leq v^i(z) + \varepsilon$. Therefore, for every $s \geq 0$ we have

$$\gamma^i_s(z, \sigma_\varepsilon^{-i}, \tilde{\sigma}^i) \leq v^i(z) + \varepsilon + \frac{s_\varepsilon}{s}.$$

The strategy σ follows $\hat{\tau}$ until the first time $t = j/\ell$, with j a positive integer, where a deviation by a single player, say player i, in the time interval $[(j-1)/\ell, j/\ell)$ is observed. Thereafter, all players $i' \neq i$ start playing the $N \setminus \{i\}$ strategy profile σ_ε^{-i}.

Let $s > 0$ and $0 \leq T \leq s$ be a $(\mathcal{H}^S_t)_{t \geq 0}$-stopping time. Recall that $\bar{t}_k$ denotes the time of the end of the k-th block and that $\bar{z}_k = z_{\bar{t}_k}$. Let k_T be the smallest integer k s.t. $\bar{t}_k \geq T$.

Note that $k_T \leq T + N_0$ and if $k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2$ then $\bar{t}_k + N_0 \leq s$. Note that for every $T \leq t \leq s$ we have

$$\begin{aligned} g_t^i &\geq \sum_{k:T \leq \bar{t}_k \leq \bar{t}_{k+1} \leq s} 1_{\{\bar{z}_k \in S_1\}} 1_{\{\bar{t}_k \leq t < \bar{t}_k + N_0\}} g_t^i \\ &\geq \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} 1_{\{\bar{z}_k \in S_1\}} 1_{\{\bar{t}_k \leq t < \bar{t}_k + N_0\}} g_t^i \\ &\geq \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} 1_{\{\bar{t}_k \leq t < \bar{t}_k + N_0\}} (g_t^i - 1_{\{\bar{z}_k \notin S_1\}}). \end{aligned}$$

Therefore, using the additivity and monotonicity of the expectation, together with the fact that the expectation equals the expectation of the conditional expectation (that is used in the equality below), we deduce that

$$\begin{aligned} E_\sigma^z \int_T^s g_t^i \, dt &\geq E_\sigma^z \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} \int_{\bar{t}_k}^{\bar{t}_k + N_0} (g_t^i - 1_{\{\bar{z}_k \notin S_1\}}) \, dt \\ &= E_\sigma^z \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} N_0 (\gamma_{N_0}^i(\bar{z}_k, \bar{\tau}) - 1_{\{\bar{z}_k \notin S_1\}}) \\ &\geq E_\sigma^z \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} N_0 (v^i(\bar{z}_k) - 1_{\{\bar{z}_k \notin S_1\}} - O(\varepsilon)) \\ &\geq E_\sigma^z \sum_{k_T \leq k \leq k_T + \frac{s-T}{N_0} - 2} N_0 (v^i(\bar{z}_{k_T}) - 1_{\{\bar{z}_k \notin S_1\}} - O(\varepsilon)) \\ &\geq E_\sigma^z (s-T)(v^i(\bar{z}_{k_T}) - O(\varepsilon)) - (e+1)N_0, \end{aligned} \tag{53}$$

where e is the maximum over $z \in S$ of the expectation with respect to $P_{\hat{\tau}}^z$ of $|\{k < \infty : \bar{z}_k \notin S_1\}|$. In particular, by setting $T = 0$, we have

$$E_\sigma^z \frac{1}{s} \int_0^s g_t^i \, dt \geq v^i(z) - O(\varepsilon) \text{ for all } s \text{ sufficiently large.} \tag{54}$$

Let $\tilde{\sigma}^i$ be a strategy of player i and let T_ℓ be the minimum of s and the (random) smallest time j/ℓ such that $\tilde{\sigma}^i$ deviates from the play under σ at some time in the interval $[(j-1)/\ell, j/\ell)$. Let $\tilde{\sigma}$ denote for short the strategy

profile $(\sigma^{-i}, \tilde{\sigma}^i)$. Note that the P_σ-distribution and the $P_{\tilde{\sigma}}$-distribution of T_ℓ (and thus also of $s - T_\ell$) coincide. By the additivity of the expectation,

$$\gamma_s^i(z, \sigma^{-i}, \tilde{\sigma}^i) = E_{\tilde{\sigma}}^z \frac{1}{s} \int_0^s g_t^i \, dt = E_{\tilde{\sigma}}^z \frac{1}{s} \int_0^{T_\ell} g_t^i \, dt + E_{\tilde{\sigma}}^z \frac{1}{s} \int_{T_\ell}^s g_t^i \, dt. \tag{55}$$

The P_σ-distribution and the $P_{\tilde{\sigma}}$-distribution of $1_{\{t<T_\ell-1/\ell\}} g_t^i$ (where $1_{\{t<T_\ell-1/\ell\}}(t) = 1$ if $t < T_\ell - 1/\ell$ and $1_{\{t<T_\ell-1/\ell\}}(t) = 0$ otherwise) coincide, and $0 \le g_t^i \le 1$. Therefore

$$E_{\tilde{\sigma}}^z \frac{1}{s} \int_0^{T_\ell} g_t^i \, dt \le E_\sigma^z \frac{1}{s} \int_0^{T_\ell} g_t^i \, dt + \frac{1}{s\ell}. \tag{56}$$

By the definition of $\tilde{\sigma}$,

$$E_{\tilde{\sigma}}^z \frac{1}{s} \int_{T_\ell}^s g_t^i \, dt \le \frac{1}{s} E_{\tilde{\sigma}}^z (s - T_\ell) v^i(z_{T_\ell}) + \varepsilon + \frac{s_\varepsilon}{s}. \tag{57}$$

Recall that $0 \le v^i(z) \le 1$ and that for every stopping time T and strategy profile τ, $P_\tau(z_t = z_T \ \forall T \le t \le T + 1/\ell) \ge e^{-\|\mu\|/\ell} \ge 1 - 1/\ell$. Therefore,

$$\frac{1}{s} E_{\tilde{\sigma}}^z (s - T_\ell) v^i(z_{T_\ell}) \le \frac{1}{s} E_\sigma^z (s - T_\ell) v^i(z_{T_\ell}) + \frac{1}{\ell}. \tag{58}$$

By (53)

$$\frac{1}{s} E_\sigma^z (s - T_\ell) v^i(z_{T_\ell}) \le E_\sigma^z \frac{1}{s} \int_{T_\ell}^s g_t^i \, dt + \varepsilon + \frac{\hat{s}_\varepsilon}{s}. \tag{59}$$

Let ℓ be sufficiently large so that $\frac{1}{\ell} < \varepsilon$. Summing inequalities (55)–(59), we deduce that

$$\gamma_s^i(z, \sigma^{-i}, \tilde{\sigma}^i) \le \gamma_s^i(z, \sigma) + 4\varepsilon \quad \text{for all } s \text{ sufficiently large.} \tag{60}$$

Similarly, the limit of $\frac{1}{s} \int_0^s g_t^i \, dt$ as $s \to \infty$ exists a.e. with respect to P_σ^z, and $E_\sigma^z \lim_{s\to\infty} \frac{1}{s} \int_0^s g_t^i \, dt = u^i(z) \ge -4\varepsilon + E_{\tilde{\sigma}}^z \limsup_{s\to\infty} \frac{1}{s} \int_0^s g_t^i \, dt$.

This completes the proof of Theorem 4.

□

5 Proof of Theorem 5

Theorem 5 is straightforward if $\|\mu\| = 0$. Therefore, we assume that $\|\mu\| > 0$.

In order to prove Theorem 5, it suffices to prove that for every continuous-time stochastic game Γ, $\overrightarrow{w} \in (W_d^1)^N$, and $\varepsilon > 0$, there is a vector payoff $v \in \mathbb{R}^{S\times N}$, a strategy profile σ, and a neighborhood U of $\overrightarrow{w}$ in W_d^N, such that for every $\overrightarrow{w}^* \in U$, a player i, and a strategy profile τ such that $\tau^j = \sigma^j$ for all $j \neq i$, we have

$$\varepsilon + E_\sigma^z \int_{[0,\infty]} \underline{g}_t^i \, dw^*(t) \geq v^i(z) \geq E_\tau^z \int_{[0,\infty]} \bar{g}_t^i \, dw^*(t) - \varepsilon, \tag{61}$$

where w^* stands for $(w^*)^i$.

The idea of the proof is as follows.

We consider an auxiliary k-stage discrete-time game that depends on a uniform-l-average payoff vector u, a sufficiently large t_*, and a partition $t_0 = 0 < t_1 < \ldots < t_k = t_*$ of $[0, t_*]$.

The set of feasible actions $a_{t_j}^i$ of player i at stage $j = 0, \ldots, k-1$ are $A^i(z_{t_j})$, the conditional probability that $z_{t_{j+1}} = z'$ given $z_{t_j} = z \neq z'$ and a_{t_j} is $(t_{j+1} - t_j)\mu(z', z, a_{t_j})$, and the payoff to player i in this auxiliary game is $w^i(\infty)u^i(z_{t_k}) + \sum_{j=0}^{k-1} w^i([t_j, t_{j+1}))g^i(z_{t_j}, a_{t_j})$.

By backward induction, the auxiliary game has a Markov equilibrium strategy τ and a corresponding equilibrium payoff vector v.

We associate to an equilibrium strategy profile τ of this auxiliary game a strategy profile σ of the continuous-time game. At time $t_j \leq t < t_{j+1}$, $0 \leq j < k$, the strategy σ^i plays the j-th stage mixed action of τ^i and starting at time $t_k = t_*$, the strategy σ follows a uniform-l-average $\frac{\varepsilon}{2}$-equilibrium that corresponds to the uniform-l-average equilibrium payoff vector u.

We prove that, if t_* is sufficiently large and $\max_j(t_{j+1} - t_j)$ is sufficiently small, then the equilibrium payoff v of the auxiliary game, the associated strategy σ, and a suitable neighborhood U of $\overrightarrow{w}$, satisfy (61).

Now we turn to the formal proof.

Let $1 > \varepsilon > 0$, $\overrightarrow{w} \in (W_d^1)^N$, and let Γ be a continuous-time stochastic game. By the covariance properties, we assume w.l.o.g. that $0 \leq g_t^i \leq 1$ and $\|\mu\| = 1/2$.

Let $u \in \mathbb{R}^{S\times N}$ be a uniform l-average equilibrium payoff of Γ, and let σ_ε be a strategy profile and $t_\varepsilon > 1$ such that for every $z \in S$, $t \geq t_\varepsilon$, and a strategy profile τ such that $\tau^j = \sigma_\varepsilon^j$ for every $j \neq i$,

$$\varepsilon + \frac{1}{t} E_{\sigma_\varepsilon}^z \int_0^t g_t^i \, dt \geq u^i(z) \geq \frac{1}{t} E_\tau^z \int_0^t g_t^i \, dt - \varepsilon, \quad \text{and} \tag{62}$$

$$\varepsilon + E^z_{\sigma_\varepsilon} \underline{g}^i_\infty \geq u^i(z) \geq E^z_\tau \bar{g}^i_\infty - \varepsilon. \tag{63}$$

Fix t_* such that $2(t_\varepsilon + \varepsilon) \max_{i \in N} w^i([t_*, \infty)) < \varepsilon/3$.

Fix an increasing sequence $0 = t_0 < t_1 < \ldots < t_k = t_*$ and $(\varepsilon_j)_{j=0}^k$, such that $\sum_{0 \leq j \leq k} \varepsilon_j < \varepsilon/2$ and for $1 \leq j < k$, $(t_{j+1} - t_j) w^i_j + (t_{j+1} - t_j)^2 < \varepsilon_j$, where $w^i_j := w^i([t_j, t_{j+1}))$.

The existence of such sequences, $(t_j)_{j=0}^k$ and $(\varepsilon_j)_{j=0}^k$, follows from the fact that for a fixed t_*, $\sum_{j=0}^{k-1} (t_{j+1} - t_j) w^i_j + (t_{j+1} - t_j)^2$ goes to 0 as $\max_{0 \leq j < k} (t_{j+1} - t_j) \to 0$.

We define the strategy profile σ as follows. For $t \geq t_k$, $\sigma(h, t) = \sigma_\varepsilon(L_{t_k} h, t - t_k)$, where $(L_s h)(t) = h(s + t)$, and for $0 \leq j < k$ and $t_j \leq t < t_{j+1}$, $\sigma(h, t) = \sigma_j(z_{t_j})$, where σ_j is defined below.

The definition of the stationary strategies σ_j $(1 \leq j < k)$ and the auxiliary vectors $u^i_j \in [0, 1]^S$ $(i \in N$ and $0 \leq j \leq k)$ is recursive.

$u^i_k = w^i(\infty) u^i$. For $0 \leq j < k$ and $z \in S$, $\sigma_j(z)$ is an equilibrium of the stage game with a payoff function to player i given by

$$a \mapsto w^i_j g^i(z, a) + u^i_{j+1}(z) + (t_{j+1} - t_i) \sum_{z' \in S} \mu(z', z, a) u^i_{j+1}(z'), \quad \text{and}$$

$$u^i_j(z) = w^i_j g^i(z, \sigma_j(z)) + u^i_{j+1}(z) + (t_{j+1} - t_j) \sum_{z' \in S} \mu(z', z, \sigma_j(z_{t_j})) u^i_{j+1}(z').$$

Now we show that for every $0 \leq j \leq k$ we have $0 \leq u^i_j(z) \leq w^i([t_j, \infty])$.

As $0 \leq g^i \leq 1$, $0 \leq u^i(z) \leq 1$. Therefore $0 \leq u^i_k(z) \leq w^i(\infty) \leq w^i([t_k, \infty])$.

Fix $0 \leq j < k$ and assume that $0 \leq u^i_{j+1} \leq w^i([t_{j+1}, \infty])$. As $t_{j+1} - t_j \leq 1$ (which follows from the assumption $(t_{j+1} - t_j)^2 < \varepsilon_j < 1$ on the increasing sequence t_j) we have $0 \leq 1 + (t_{j+1} - t_j) \mu(z, z, a) \leq 1$. In addition, $\mu(z', z, a) \geq 0$ for $z' \neq z$, and $\sum_{z' \in S} \mu(z', z, a) = 0$. Therefore, the sum $u^i_{j+1}(z) + \sum_{z' \in S} (t_{j+1} - t_j) \mu(z', z, a) u^i_{j+1}(z')$ is a convex combination of the nonnegative numbers $u^i_{j+1}(z')$, $z' \in S$. Each one of these nonnegative numbers is $\leq w^i([t_{j+1}, \infty])$ (by assumption) and therefore also their convex combination is $\leq w^i([t_{j+1}, \infty])$.

Therefore, by the definition of $u^i_j(z)$, we have

$$\begin{aligned} 0 \le u^i_j(z) &= w^i_j g^i(z, \sigma_j(z)) \\ &\quad + u^i_{j+1}(z) + \sum_{z' \in S} (t_{j+1} - t_j) \mu(z', z, \sigma_j(z)) u^i_{j+1}(z') \\ &\le w^i_j + w^i([t_{j+1}, \infty]) = w^i([t_j, \infty]). \end{aligned}$$

Therefore, $0 \le u^i_j \le w^i([t_j, \infty])$.

Fix a player i. We claim that there is a neighborhood U of w^i such that for every $w^* \in U$ and every strategy profile τ such that $\tau^j = \sigma^j$ for every player $j \neq i$, we have

$$2\varepsilon + E^z_\sigma \int_{[0,\infty]} \underline{g}^i_t \, dw^i(t) \ge u^i_0(z) \ge E^z_\tau \int_{[0,\infty]} \bar{g}^i_t \, dw^i(t) - 4\varepsilon. \tag{64}$$

For $0 \le j < k$, let $U_j := \{w^* \in W : |w^*_j - w^i_j| < \varepsilon_j\}$ where $w^*_j := w^*([t_j, t_{j+1}))$.

By Lemma 2 and the inequality $(t_{j+1} - t_j) w^i_j + (t_{j+1} - t_j)^2 < \varepsilon_j$, for any $0 \le j < k$ and $w^* \in U_j$,

$$\begin{aligned} & E^z_\sigma \left(\int_{[t_j, t_{j+1})} g^i_t \, dw^*(t) + u^i_{j+1}(z_{t_{j+1}}) \mid h_{t_j} \right) \\ \ge\ & E^z_\sigma \left(\int_{[t_j, t_{j+1})} g^i_t \, dw^i(t) + u^i_{j+1}(z_{t_{j+1}}) \mid h_{t_j} \right) - |w^*_j - w^i_j| \\ \ge\ & g^i(z_{t_j}, \sigma(z_{t_j})) w^i_j + u^i_{j+1}(z_{t_j}) - \sum_{z' \in S} u^i_{j+1}(z')(t_{j+1} - t_j) \mu(z', z, \sigma_j(z)) - 2\varepsilon_j \\ \ge\ & u^i_j(z_{t_j}) - 2\varepsilon_j \end{aligned} \tag{65}$$

and by denoting by $\bar{y}_t$ the mixed action profile of τ at time $t_j \le t < t_{j+1}$, conditional on h_{t_j} and $z_s = z_{t_j}$ for every $t_j \le s \le t$,

$$\begin{aligned} & u^i_j(z_{t_j}) \\ =\ & w^i_j g^i(z_{t_j}, \sigma(z_{t_j})) + u^i_{j+1}(z_{t_j}) + (t_{j+1} - t_j) \sum_{z' \in S} u^i_{j+1}(z') \mu(z', z_{t_j}, \sigma_j(z_{t_j})) \\ \ge\ & w^*_j g^i(z_{t_j}, \bar{y}_t) + u^i_{j+1}(z_{t_j}) + (t_{j+1} - t_j) \sum_{z' \in S} u^i_{j+1}(z') \mu(z', z_{t_j}, \bar{y}_t) - |w^*_j - w^i_j| \\ \ge\ & E^z_\tau \left(\int_{[t_j, t_{j+1})} g^i_t \, dw^*(t) + u^i_{j+1}(z_{t_{j+1}}) \mid h_{t_j} \right) - 2\varepsilon_j. \end{aligned} \tag{66}$$

Therefore, if $w^* \in \cap_{0\le j<k} U_j$, then, by summing inequalities (65), respectively (66), over $0 \le j < k$, we have

$$\varepsilon + E^z_\sigma\Big(\int_{[0,t_k)} g^i_t\, dw^*(t) + u^i_k(z_{t_k})\Big) \ge u^i_0(z_0) \ge E^z_\tau\Big(\int_{[0,t_k)} g^i_t\, dw^*(t) + u^i_k(z_{t_k})\Big) - \varepsilon. \tag{67}$$

Set $\bar t := t_k + t_\varepsilon$ and $U_k = \{w^* \in W_d : |w^*([t_k,\bar t]) - w^i([t_k,\bar t])| + |w^*([t_k,\infty]) - w^i([t_k,\infty])| < \varepsilon_k$. Let $w^* \in U_k$ and define $f(t) := \frac{dw^*}{dt}(t)$. Note that as w^* is a discounting measure, $f(t)$ exists for all $0 < t < \infty$, but countably many exceptions, and f is monotonic nonincreasing with $f(t) \to 0$ as $t \to \infty$. (In what follows, the integrals of the form $\int G(t)df(t)$ refer to the Riemann–Stieltjes integral. Note that $-\int_{[t_k,\infty)}(t - t_k)\, df(t) = w^*([t_k,\infty))$.) Then, by integration by part and by using inequality (62), we have

$$\begin{aligned}
& E^z_\sigma\left(\int_{[t_k,\infty)} g^i_t\, dw^*(t) \mid h_{t_k}\right) = -E^z_\sigma\left(\int_{[t_k,\infty)} (\int_{t_k}^t g^i_t\, dt)\, df(t) \mid h_{t_k}\right) \ge \\
\ge\ & -E^z_\sigma\left(\int_{[\bar t,\infty)} (\int_{t_k}^t g^i_t\, dt)\, df(t) \mid h_{t_k}\right) \ge -\int_{[\bar t,\infty)} (t - t_k)(u^i(z_{t_k}) - \varepsilon)\, df(t) \\
\ge\ & -\int_{[t_k,\infty)} (t - t_k)(u^i(z_{t_k}) - \varepsilon)\, df(t) - 2w^*([t_k,\bar t))(\bar t - t_k) \\
=\ & (u^i(z_{t_k}) - \varepsilon)w^*([t_k,\infty) - 2w^*([t_k,\bar t))(\bar t - t_k) \ge u^i(z_{t_*})w^*(t_k,\infty)) - \varepsilon,
\end{aligned}$$

and if τ is a strategy profile such that $\tau^j = \sigma^j$ for every $j \ne i$ then

$$E^z_\tau\left(\int_{[t_k,\infty)} g^i_t\, dw^*(t) \mid h_{t_k}\right) = -E^z_\tau\left(\int_{[t_k,\infty)} (\int_{t_k}^t g^i_t\, dt)\, df(t) \mid h_{t_k}\right) \le$$

$$\begin{aligned}
\le\ & -E^z_\tau\left(\int_{[\bar t,\infty)} (\int_{t_k}^t g^i_t\, dt)\, df(t) \mid h_{t_k}\right) + w^*([t_k,\bar t))(\bar t - t_k) \\
\le\ & -\int_{[\bar t,\infty)} (t - t_k)(u^i(z_{t_k}) + \varepsilon)\, df(t) + w^*([t_k,\bar t))(\bar t - t_k) \\
\le\ & -\int_{[t_k,\infty)} (t - t_k)(u^i(z_{t_k}) + \varepsilon)\, df(t) + w^*([t_k,\bar t))(\bar t - t_k) \\
=\ & (u^i(z_{t_k}) + \varepsilon)w^*([t_k,\infty) + w^*([t_k,\bar t))(\bar t - t_k).
\end{aligned}$$

These inequalities imply that

$$2\varepsilon+E_\sigma^z\left(\int_{[t_k,\infty)} g_t^i\,dw^*(t)\mid h_{t_k}\right)\geq w^*([t_k,\infty))u_k^i(z_{t_k})\geq E_\tau^z\left(\int_{[t_k,\infty)} g_t^i\,dw^*(t)\mid h_{t_k}\right)-2\varepsilon. \tag{68}$$

As $w^*\in U_k$ (and thus $|w^*([t_k,\infty])-w^i([t_k,\infty])|<\varepsilon_k$), and

$$\varepsilon+E_\sigma^z\left(\underline{g}_\infty^i\,w^*(\infty)\mid h_{t_k}\right)\geq w^*(\infty)u_k^i(z_{t_k})\geq E_\tau^z\left(\bar{g}_\infty^i\,w^*(\infty)\mid h_{t_k}\right)-\varepsilon \tag{69}$$

(by inequality (63)), we deduce that

$$4\varepsilon+E_\sigma^z\left(\int_{[t_k,\infty]} \underline{g}_t^i\,dw^*(t)\mid h_{t_k}\right)\geq w^i([t_k,\infty])u_k^i(z_{t_k})\geq E_\tau^z\left(\int_{[t_k,\infty]} \bar{g}_t^i\,dw^*(t)\mid h_{t_k}\right)-4\varepsilon. \tag{70}$$

Inequalities (67) and (70) imply that for $w^*\in U:=\cap_{j=0}^k U_j$,

$$6\varepsilon+E_\sigma^z\int_{[0,\infty]}\underline{g}_t^i\,dw^*(t)\geq u_0^i(z)\geq E_\tau^z\int_{[0,\infty]}\underline{g}_t^i\,dw^*(t)-6\varepsilon.$$

□

The (first) part of the proof that ends with inequality (67), proves that for every $\overrightarrow{w}\in(\overrightarrow{W})^N$ that is supported on $[0,\infty)$, Γ has a $\overrightarrow{w}$-time-separable robust equilibrium.

6 The continuous-time process

Recall that s_k stands for the k-th time of a state change.

In this section we study properties of the stochastic process of states $(z_t)_{t\geq 0}$ that is defined by a correlated strategy. Recall that the correlated strategy choice at time t, is, for $s_k\leq t<s_{k+1}$, a function of t and the state process up to time s_k.

A special case of a correlated strategy is a Markov correlated strategy. If σ is a correlated strategy and τ is a Markov strategy such that $\sigma(h,t)=\tau(h,t)$ whenever $s_1(h)>t$, then the P_σ and the P_τ distributions of the play up to time s_1 coincide.

Therefore, for a correlated strategy σ, the conditional P_σ distribution of $((z_t)_{s_k<t\leq s_{k+1}},(x_t)_{s_k\leq t<s_{k+1}})$, given $((z_t)_{0\leq t\leq s_k},(x_t)_{0\leq t<s_k})$, is the P_τ distribution of $((z_t)_{0\leq t\leq s_1},(x_t)_{0\leq t<s_1})$ of a Markov correlated strategy τ.

Therefore, commenting on the P_τ distribution of plays defined by a Markov correlated strategy τ is useful for studying the P_σ distribution on plays defined by a correlated strategy σ.

A Markov correlated strategy σ defines transition rates as a function of time t. These transition rates are represented by $S \times S$ matrices $Q(t)$ where $Q_{z,z'}(t) = \mu(z', z, \sigma(z, t))$. If σ is stationary then the transition matrices $Q(t)$ are independent of t. If σ is a continuous Markov correlated strategy then the map $t \mapsto Q(t)$ is continuous.

The next theorem is a classic[13] result on non-homogeneous continuous-time Markov processes. The result shows that the transition matrices $F(s,t)$ of the Markov stochastic process with state space S that is defined by a Markov strategy are well defined. The starting point is the fundamental assumption that $F^\sigma_{z,z'}(s, s+\delta) = 1_{z'=z} + \int_s^{s+\delta} \mu(z', z, \sigma(z,t))\, dt + o(\delta)$.

In addition, we derive a quantified continuity property of the state stochastic process as a function of the Markov correlated strategy σ.

Let S and T be finite sets. An $S \times T$ matrix $Q = (Q_{z,z'})_{(z,z') \in S \times T}$ is a transition matrix if for all $(z, z') \in S \times T$ we have $Q_{z,z'} \geq 0$ and for all $z \in S$ we have $\sum_{z' \in T} Q_{z,z'} = 1$.

Let M be the space of all $S \times S$ matrices Q, let M_0 be its subspace of all matrices Q with $\sum_{z' \in S} Q_{z,z'} = 0$ for every $z \in S$ and $Q_{z,z'} \geq 0$ for all $z \neq z'$, and let M_1 be the subspace of M of all transition matrices. The identity matrix is denoted by I.

The space M is a (noncommutative) Banach algebra with the norm $\|Q\| = \max_{z \in S} \sum_{z' \in S} |Q_{z,z'}|$, and the spaces M_0 and M_1 are closed subalgebras of M. For an ordered list $F_1, \ldots, F_j \in M$ we denote by $\prod_{i=1}^{j} F_i$ the matrix (ordered) product $F_1 F_2 \ldots F_j$.

Theorem 6. *Let $Q : [0, \infty) \to M$ (respectively, $\to M_0$) be measurable, with $\|Q(t)\| \leq C$, and let $Q^\eta : [0, \infty) \to M$ converge w* to Q. Then there is a unique family $F(s,t) = F^Q(s,t)$, $t \geq s \geq 0$, of matrices in M (respectively, in M_1) such that for all $0 \leq s \leq t_1 \leq t$ and $\delta > 0$ we have*

$$F(s,s) = I \tag{71}$$

$$\|F(t, t+\delta) - I - \int_t^{t+\delta} Q(x)\, dx)\| \leq o(\delta), \tag{72}$$

[13]For a proof in the case where $Q(t)$ is continuous in t, see, e.g., [30]. Variants of this result are stated in, e.g., [21, 16]. For the present formulation and its proof, see [27].

where $o(\delta)$ is a function of δ such that $o(\delta)/\delta \to 0$ as $\delta \to 0+$, and

$$F(s,t) = F(s,t_1)F(t_1,t). \tag{73}$$

This unique family satisfies

$$\frac{\partial F}{\partial t}(s,t) = F(s,t)Q(t) \quad \text{for a.e. } t \tag{74}$$

and for $\delta \le 1/C$ we have

$$\|F(t,t+\delta) - I - \int_t^{t+\delta} Q(x)\,dx)\| \le C\delta^2. \tag{75}$$

The above theorem derives the unique transition matrices $F(s,t)$ that are defined by a Markov strategy σ via the transition rate matrices Q, where $Q_{z,z'} = \mu(z', z, \sigma(z,t))$.

It is easy (and classic) to prove that any state process $(z_t)_{t\ge 0}$ with law P such that $P(z_t = z' \mid z_s = z) = F_{z,z'}(s,t)$ has, with probability 1, left and right limits everywhere, and for each fixed t it is, with probability 1, continuous at t.

Therefore, as our game payoffs are of the form $\int g\,dw(t)$, and w has at most countably many atoms, there is no loss of generality in assuming in our game model that the state process is right continuous.

A correlated strategy σ defines a probability distribution P_σ on the state process h^S. The mixed action choice of σ at time t is a function of the state process (up to time t) and the time t, and therefore is denoted by $\sigma(h^S, t)$.

Lemma 12. *Let σ and τ be correlated strategies, $z = z_0$ an initial state, and T a bounded $(\mathcal{H}_t^S)_t$-stopping time. For any positive integer ℓ, let $T(\ell) := \min\{j/\ell : j/\ell \ge T\}$, and let $T_i(\ell)$, or T_i for short, be defined by $T_i = T(\ell) \wedge i/\ell := \min(T(\ell), i/\ell)$.*

For $i \ge 0$ and $T_{i-1} \le t < T_i$ we denote by $\bar{x}_t^\ell$, respectively $\bar{y}_t^\ell$, the mixed action choice of the correlated strategy σ, respectively τ, at time t conditional on no state change in the time interval $[T_{i-1}, t]$.

Then, the norm distance between the P_σ and P_τ distributions of $z_n^\ell := z_0, z_{1/\ell}, \ldots, z_{T_n}$, respectively $(z_t)_{t\le T}$, is bounded by

$$\begin{aligned} & 4n\|\mu\|/\ell^2 + 2\|\mu\| E_\sigma^z \sum_{i=1}^{n} \| \int_{T_{i-1}}^{T_i} (\bar{x}_t^\ell - \bar{y}_t^\ell)\,dt\| \\ \le\ & 4n\|\mu\|/\ell^2 + 2\|\mu\| E_\sigma^z \int_0^{T_n} \|\bar{x}_t^\ell - \bar{y}_t^\ell\|\,dt, \end{aligned}$$

respectively $2\|\mu\| E^z_\sigma \int_0^T \|\sigma(h^S,t)-\tau(h^S,t)\|\,dt$.

The proof uses Lemma 2 and the following lemma.

Lemma 13. *Let* $\{\emptyset,\Omega\}=\mathcal{F}_0\subset\mathcal{F}_1\ldots\subset\mathcal{F}_n$ *be an increasing sequence of* σ*-algebras of subsets of a nonempty set* Ω*, and let* P *and* Q *be two probability measures on the measurable space* $(\Omega,\mathcal{F}_n)$*. Denote by* F_i *the set of all* $[-1,1]$*-valued* $\mathcal{F}_i$*-measurable functions on* Ω *and by* P_i *and* Q_i *the probability measures* P *and* Q *on* $(\Omega,\mathcal{F}_i)$*. Then,*

$$\|P_i-Q_i\|=\sup\{E_Pf-E_Qf: f\in F_i\},\ \textit{and}$$

$$\|P-Q\|\le\sup\{E_P\sum_{i=1}^n(E_P(f_i\mid\mathcal{F}_{i-1})-E_Q(f_i\mid\mathcal{F}_{i-1})): f_i\in F_i\}. \tag{76}$$

Proof. $\|P_i-Q_i\|:=\sup\{P(C)-Q(C)+Q(\Omega\setminus C)-P(\Omega\setminus C): C\in\mathcal{F}_i\}=\sup\{E_P(1_C-1_{\Omega\setminus C})-E_Q(1_C-1_{\Omega\setminus C}): C\in\mathcal{F}_i\}\le\sup\{E_Pf-E_Qf: f\in F_i\}$.

As $L_\infty(P_i+Q_i)\ni f\mapsto E_Pf-E_Qf$ is linear and $1_C-1_{\Omega\setminus C}$, $C\in\mathcal{F}_i$, are the extreme points of $L_\infty(P_i+Q_i)$, the last inequality is an equality. This proves the first part of the lemma, and the second part for $n=1$.

Assume that (76) holds for $n=k$ and let $f\in F_{k+1}$. As $E_Q(f\mid\mathcal{F}_k)\in F_k$, $E_Pf-E_Qf=E_PE_P(f\mid\mathcal{F}_k)-E_QE_Q(f\mid\mathcal{F}_k)=E_P(E_P(f\mid\mathcal{F}_k)-E_Q(f\mid\mathcal{F}_k))+E_P(E_Q(f\mid\mathcal{F}_k)-E_QE_Q(f\mid\mathcal{F}_k)\le E_P(E_P(f\mid\mathcal{F}_k)-E_Q(f\mid\mathcal{F}_k))+\|P_k-Q_k\|\le\sup\{E_P\sum_{i=1}^{k+1}(E_P(f_i\mid\mathcal{F}_{i-1})-E_Q(f_i\mid\mathcal{F}_{i-1}): f_i\in F_i\}$. Therefore, (76) holds for $n=k+1$, and thus by induction for every n. □

Proof of Lemma 12. Set in Lemma 13 $\mathcal{F}_i=\mathcal{H}^S_{T_i}$, where $T_i:=T\wedge i/\ell=\min(T,i/\ell)$.

Note that $\bar{x}^\ell_t$ and $\bar{y}^\ell_t$, $T_{i-1}\le t<T_i$, are measurable with respect to $\mathcal{F}_{i-1}$.

By Lemma 2, for every $f_i: S\to[-1,1]$, $|E^z_\sigma(f_i(z_{T_i})\mid\mathcal{F}_{T_{i-1}})-E^z_\tau(f_i(z_{T_i})\mid\mathcal{F}_{T_{i-1}})|\le 4\|\mu\|^2/\ell^2+|\sum_{z'\in S}f_i(z')\int_{T_{i-1}}^{T_i}\mu(z',z,\bar{x}_t-\bar{y}_t)\,dt|\le 4\|\mu\|^2/\ell^2+2\|\mu\|\,\|\int_{T_{i-1}}^{T_i}(\bar{x}_t-\bar{y}_t)\,dt\|\le 4\|\mu\|^2/\ell^2+2\|\mu\|\int_{T_{i-1}}^{T_i}\|\bar{x}_t-\bar{y}_t\|\,dt$.

By summing these inequalities over $i=1,\ldots,n$, the first bound in the lemma follows from Lemma 13.

As $\|\bar{x}^\ell_t-\bar{y}^\ell_t\|$ is bounded and with P_σ-probability 1 converges, as $\ell\to\infty$, a.e. to $\|\sigma(h^S,t)-\tau(h^S,t)\|$, the second part of the lemma follows from the bounded convergence theorem. □

7 Related literature

7.1 Continuous-time supergames and bargaining

Simon and Stinchcombe [33] develop a general theory of continuous-time games. The pathologies of continuous-time strategies are resolved there by imposing assumptions (i.e., restrictions) on pure strategies that identify a class of strategies that yield a well-defined outcome. The assumptions are, essentially, that 1) the number of action changes is uniformly bounded on every finite time interval, 2) the strategies are piecewise continuous with respect to time, and 3) there is strong right continuity with respect to histories.

Bergin and MacLeod [1] develop a model of continuous-time supergames of complete information and study the strategic equilibrium behavior of these games. The pathologies of continuous-time strategies are resolved there by considering only strategies that have inertia. A strategy with inertia is essentially a limit, in a proper sense, of strategies that as a function of history and time select a constant action over a short time interval. In addition, the pure strategies choose a pure stage action, in contrast to our main model, which allows a strategy to choose a mixed action as a function of history.

Perry and Reny [29] develop a theory of continuous-time bargaining, where players can make offers whenever they like. The pathologies of continuous-time strategies are resolved there by considering only strategies where upon making an offer players must wait a fixed amount of time before making another offer. The continuous-time conclusions are obtained by studying the results when the waiting times go to zero.

7.2 Continuous-time MDP and stochastic games

Zachrisson [44] introduced the theory of continuous-time 2-person 0-sum stochastic games. He proves the existence of a value and optimal strategies in a finite horizon game where the payoff of a play is the sum of the integral of the payoffs $g_t = g(t, z_t, x_t)$ and a terminal payoff that is a function of the state at the end of the game.

In Zachrisson's model it is assumed that the space of strategies of each player is the space of Markov strategies. The assumption that all strategies are Markov strategy is restrictive but indeed innocuous in the model of two-person zero-sum in the sense that zachrisson's solution – the value and optimal strategies – is also a solution when one considers all strategies.

On the other hand, Zachrisson's model allows for time-dependent compact action spaces, transition rates, and payoffs. Explicitly, the action space of player i is a compact subset $X^i(t,z)$ of a Euclidean space that depends continuously (in the Hausdorff metric) on t, the transition rates $\mu(z',z,t,x_t)$ depend continuously on t and x_t, and the payoff function $g(t,z_t,x_t)$ depend also on the time t. The payoff in Zachrisson's T-horizon game is $\int_0^T g(t,z_t,x_t)\,dt + f(z_T)$. A condition on the existence of optimal strategies in a family of auxiliary one-stage games $x_t \mapsto g(t,z_t,x_t) + \sum_z \mu(z,z_t,t,x_t)v(z)$, $v \in \mathbb{R}^S$ is shown to suffice for the existence of a value and optimal strategies.

Jasso-Fuentes [10] studies the equilibrium conditions for Markov games with Polish state and action spaces when the game is restricted to Markov strategies. However, existence is not addressed there.

The relation between the value of the discounted continuous-time and the discounted discrete-time Markov decision process, called the *uniformization technique*, appears also in [8].

Guo and Hernandez-Lerma [5] study two-person non-zero-sum continuous-time stochastic games with stationary discounted payoff criteria (and Borel action spaces), and give conditions that ensure the existence of Nash equilibrium in stationary strategies.

Yehuda Levy [15] proved that a non-zero-sum stochastic game of fixed duration has a Markovian correlated equilibrium.

References

[1] Bergin, J. and W. B. MacLeod (1993), Continuous time repeated games, *International Economic Review*, 34, 21–37.

[2] Blackwell, D. and T. S. Ferguson (1968), The big match, *Annals of Mathematical Statistics*, 39, 159–163.

[3] Dynkin, E. B. and A. A. Yushkevic (1975), *Controlled Markov Processes and their applications*, Nauka, MOscow.

[4] Guo, X. and O. Hernandez-Lerma (2003), Zero-sum games for continuous-time Markov chains with unbounded transitions and average payoff rates, *J. Appl. Probability*, 40, 327–345.

[5] Guo, X. and O. Hernandez-Lerma (2005), Nonzero-sum games for continuous-time Markov chains with unbounded discounted payoffs, *J. Appl. Probability*, 42, 303–320.

[6] Guo, X., O. Hernandez-Lerma, and T. Prieto-Rumeau (2006), A survey of recent results on continuous-time Markov decision processes, *TOP*, 14, 177–243.

[7] Guo, X. and O. Hernandez-Lerma (2009), *Continuous-time Markov Decision Processes*. Berlin: Springer.

[8] Howard, R. (1960), *Dynamic Programming and Markov Processes*. New York: Wiley.

[9] Isaacs, R. (1999), *Differential Games*, Dover.

[10] Jasso-Fuentes, H. (2005), Noncooperative continuous-time Markov Games, *Morfismos*, 9, 39–54.

[11] Kakumanu, P. (1969), Continuous time Markov decision models with applications to optimization problems, Technical Report 63, Dept. of Operations Research, Cornell Universsity.

[12] Kakumanu, P. (1971), Continuously discounted Markov decision model with countable state and action space, *The Annals of Mathematical Statistics*, 42, 919–926.

[13] Kakumanu, P. (1971), Nondiscounted continuous time Markovian decision process with countable state space, *SIAM J. Control*, 10, 210–220.

[14] Kakumanu, P. (1975), Continuous-time Markovian decision processes with average return criterion, *J. Math. Anal. Appl.*, 52, 177–183.

[15] Levy, Y. (2013), Continuous-time stochastic games of fixed duration, *Dynamic Games and Applications*, 3, 279–312.

[16] Martin-Lof, A. (1967), Optimal control of a continuous-time Markov chain with periodic transition probabilities, *Operations Research*, 15, 872–881.

[17] Mertens, J.-F. and A. Neyman (1981), Stochastic games, *International Journal of Game Theory*, 10, 53–66.

[18] Mertens, J.-F., A. Neyman, and D. Rosenberg (2009), Absorbing games with compact action spaces, *Mathematics of Operations Research*, 34, 257–262.

[19] Mertens, J.-F., S. Sorin, and S. Zamir (1994), *Repeated Games*, C.O.R.E. D.P. 9420, 9421, 9422.

[20] Merton, R. C. (1992), *Continuous-time Finance*. Cambridge, MA: Basil Blackwell.

[21] Miller, B. L. (1968), Finite state continuous time Markov decision processes with a finite planning horizon, *Siam J. Control*, 6, 266–280.

[22] Miller, B. L. (1968), Finite state continuous time Markov decision processes with an infinite planning horizon, *J. Math. Anal. Appl.*, 22, 522–569.

[23] Neyman, A. (1999), Cooperation in repeated games when the number of stages is not commonly known, *Econometrica*, 67, 45–64.

[24] Neyman, A. (2003), Real algebraic tools in stochastic games, in *Stochastic Games and Applications*, A. Neyman and S. Sorin (eds.), NATO ASI Series, Kluwer Academic Publishers, pp. 57–75.

[25] Neyman, A. (2003), Stochastic games: Existence of the minmax, in *Stochastic Games and Applications*, A. Neyman and S. Sorin (eds.), NATO ASI Series, Kluwer Academic Publishers, pp. 173–193.

[26] Neyman, A. (2003), Stochastic games and nonexpansive maps, in *Stochastic Games and Applications*, A. Neyman and S. Sorin (eds.), NATO ASI Series, Kluwer Academic Publishers, pp. 397–415.

[27] Neyman, A. (2012), Continuous-time stochastic games, DP 616, Center for the Study of Rationality, Hebrew University.

[28] Neyman, A. (2013), Stochastic games with short-stage duration, *Dynamic Games and Applications*, 3, 236–278.

[29] Perry, M. and P. Reny (1993), A non-cooperative bargaining model with strategically timed offers, *Journal of Economic Theory*, 59, 50–77.

[30] Rosenblatt, M. (1962), *Random Processes*, Oxford: Oxford University Press.

[31] Rykov, V. V. (1966), Markov decision processes with finite spaces of states and decisions, *Theory Prob. Apple.*, 11, 343–351.

[32] Shapley, L. S. (1953), Stochastic games, *Proceedings of the National Academy of Sciences of the U.S.A.*, 39, 1095–1100.

[33] Simon, L. K. and M. B. Stinchcombe (1989), Extensive form games in continuous time: Pure strategies, *Econometrica*, 57, 1171–1214.

[34] Solan, E. (2001), Characterization of correlated equilibrium in stochastic games, *International Journal of Game Theory*, 30, 259–277.

[35] Solan, E. (2003), Continuity of the value of competitive Markov decision processes, *Journal of Theoretical Probability*, 16, 831–845.

[36] Solan, E. and N. Vieille (2002), Correlated equilibrium in stochastic games, *Games and Economic Behavior*, 38, 362–399.

[37] Solan, E. and R. Vohra (2002), Correlated equilibrium and public signalling in absorbing games, *International Journal of Game Theory*, 31, 91–121.

[38] Sorin, S. (1986), Asymptotic properties of a non-zero-sum stochastic games, *International Journal of Game Theory*, 15, 101–107.

[39] Strook, D. W. and S. R. Varadhan (2006), *Multidimentional Diffusion Processes*. Berlin: Springer.

[40] Vigeral, G. (2012), A zero-sum stochastic game with compact action sets and no asymptotic value, CEREMADE, Universite Paris-Dauphine, preprint.

[41] Yushkevich, A. A. (1973), On a class of policies in general controlled Markov models, *Theory of Probability and its Applications* 18, 775–779.

[42] Yushkevich, A. A. (1977), Controlled Markov models with countable state and continuous time, *Theory of Probability and its Applications* 22, 215-235.

[43] Yushkevic, A. A. and E. A. Fainberg (1979), On homogeneous Markov model with continuous time and finite or countable state space, *Theory of Probability and its Applications*, 24, 156-161.

[44] Zachrisson, L. E. (1964), Markov games, in *Advances in Game Theory*, Princeton University Press, Princeton, N.J.