האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

CONTINUOUS-TIME STOCHASTIC GAMES

By

ABRAHAM NEYMAN

Discussion Paper # 616 Aug 2012

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Continuous-time Stochastic Games

Abraham Neyman*

September 3, 2012

Abstract

Every continuous-time stochastic game with finitely many states and actions has a uniform and limiting-average equilibrium payoff.

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 has been forecasted by many economic experts, many other experts claimed just before every occurrence that "this time would 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 cautioning of a forthcoming financial crisis have had no idea of the time of its occurrence and many who correctly predicted a crisis had warned that it was imminent when in fact it occurred many years later.

An analogous observation is in sports, e.g., in the game of soccer. 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 to score the

*Institute of Mathematics, and 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.

next goal than the other team. However, it is impossible to be sure which team will be the first to score the next goal, and it is implausible to forecast the time of the scoring.

In both cases, the state of the economy and the score in the game can dramatically change in a split of a 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 the change is impacted by the players' actions, is a stochastic game. However, no single deterministic-time dynamic game, e.g., a discrete-time stochastic 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 Section 6.

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) the probability of a discontinuous state change can be positive (depending on players' actions) but infinitesimal in any short time interval.

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., [18]. 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., [7].

In order to study the above-mentioned three properties, with a focus on discontinuous state changes, 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 Equilibria of long-term games

Our main objective is the study of equilibria of the long-term game. A long-term interaction is often modeled as a game with a fixed long duration or with a fixed small discounting rate. However, in most real-life interactions, the exact duration and/or the exact discounting rate is unknown (and definitely not commonly known). Moreover, exact equilibrium analysis of the game models where the duration is known, but not commonly known, to all players leads to completely different results than those of the model with a fixed finite long duration; see [21].

The equilibrium concept that is insensitive to the exact long duration and/or discount rate is the *uniform equilibrium.* In a uniform equilibrium a player's objective is to maximize his payoff simultaneously in all games with a sufficiently long duration and with a sufficiently small discounting rate. Unfortunately, in most long-term game models, it is impossible to find a strategy profile and/or a payoff vector that is an equilibrium strategy profile and/or payoff in all sufficiently long duration games. Therefore, the uniform equilibrium is based on simultaneous approximate payoff maximization. A *uniform equilibrium payoff* is a vector payoff v such that for every $\varepsilon > 0$ there is a strategy profile that, independently of the sufficiently long duration and/or the sufficiently small discounting rate, is an ε-equilibrium with a payoff (per stage[1]) within ε of v.

Our main objective is to study the existence of uniform equilibrium payoffs. We are able to prove the existence of a uniform equilibrium payoff in any continuous-time stochastic games with finitely many[2] states and actions. This 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 equilibrium.

1.3 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

[1]The per stage payoff in the game with finite duration T, respectively, with discount rate ρ, is $\frac{1}{T}\int_0^T g_t\,dt$, respectively, $\rho\int_0^\infty e^{-\rho t} g_t\,dt$, where g_t is the payoff (density) at time t.

[2]The assumption of finitely many states is essential even in the case of a single player; see, e.g., [22, Section 1.4], and the assumption of finitely many actions is essential even in the two-person zero-sum case; see [34].

the space of plays that are compatible with these strategies. In addition, what defines a strategy is questionable. In spite of these pathologies, we are able to describe unambiguously the value and optimal strategies in two-person zero-sum continuous-time stochastic games, and equilibrium payoffs and equilibrium strategies in non-zero-sum continuous-time stochastic games.

Earlier studies overcome the above-mentioned pathologies by restricting the study to either strategies with inertia, or to Markov (memoryless) strategies that select an action as a function of (only) the current time and state, or to strategies that correspond to the game, where the only information available to a decision maker is the current state and past states. See, e.g., [1, 28] for the study of continuous-time supergames, [24] for the study of continuous-time bargaining, [37, 26, 9, 10, 5] for the study of continuous-time Markov decision processes, and [37, 2, 3] 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 10.

The common characteristic of these earlier studies is that each considers only a subset of strategies, so that a profile 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 only within the restricted class of strategies. Therefore, there is neither optimality nor nonexistence of beneficial unilateral deviation claims for general strategies.

While the Markov (memoryless) assumption is (essentially) innocuous in the study of Markov decision processes, it is restrictive in the game-theoretic framework. It turns out that in discounted stochastic games (with finitely many states and actions) a stationary 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 either the (undiscounted) limiting average payoff game, where a player maximizes the limit of the average payoff per unit of time, or the uniform game, where a player maximizes uniformly over all sufficiently long duration games the average payoff per unit of time, or the game with patient players, where a player maximizes uniformly over all sufficiently small discount rates the discounted payoff.

For example, there are two-person zero-sum continuous-time stochastic game with no limiting average or uniform value where one restricts the strategies to Markov strategies (or to oblivious strategies). Our main result is that a continuous-time stochastic game with finitely many states and actions has a uniform and limiting average equilibrium payoff, and thus, in particular, a uniform and limiting average value in the two-person zero-sum case.

1.4 Discrete-time stochastic games

We recall the basic properties of the discrete-time stochastic game model, to point out the similarities and differences in 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, 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 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 a. Players' stage-action choices are made simultaneously and are observed by all players following the stage play; equivalently, they are observed only following the stage play. Pure, mixed, and behavioral strategies are well and unambiguously defined in the discrete-time model.

If at time $t = 1, 2, \ldots$, the state is z_t and the action profile played is a_t, the stage payoff is $g_t := g(z_t, a_t)$. The undiscounted accumulation of the payoffs in stages $t = 1, \ldots, s$ is $\sum_{t=1}^{s} g_t$. In this case the λ-discounted payoff is $\sum_{t=1}^{\infty} \lambda^{t-1} g_t$ and the limit average payoff is $\lim_{s\to\infty} \frac{1}{s} \sum_{t=1}^{s} g_t$ if the limit exists.

1.5 Continuous-time stochastic games: payoffs

In a continuous-time game, the payoff is defined as an accumulation of infinitesimal payoffs. If at time $t \in [s, s + ds)$ the state is z and the action 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 at time t, is measurable, then the (undiscounted) accumulation of payoffs in the time interval $[0, s)$ is $\int_0^s g(z_t, a_t)dt$. In this case the ρ-discounted payoff is $\int_0^\infty e^{-\rho t} g(z_t, a_t)\, dt$ and the limit average payoff is

$\lim_{s\to\infty}\frac{1}{s}\int_0^s g(z_t,a_t)\,dt$ if the limit exists. We study also the finite horizon continuous-time stochastic game with a terminal payoff V that depends on the final state. In this case, if s is the duration of the game, the payoff is given by $\int_0^s g(z_t,a_t)\,dt + V(z_s)$.

1.6 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 a. The nonnegative real number $\mu(z',z,a)$ describes the rate of transition from state z to state z' when action a is played at state z. If the state at time t is $z_t = z$ and the action at all times $t\le s<t+\delta$ 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 time-dependent action a_t, $t\ge 0$, and two distinct states $z'\neq z$, the transitions obey

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

1.7 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 t, the selected action at stage t as a function of the information derived by signals from others' (including nature's) moves. In a stochastic game with observable states and actions, a strategy specifies the action at the discrete-stage t 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 set of times $t \geq 0$ 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 *pure strategy* of player i is a function σ^i from plays – function $h : t \mapsto (z_t, a_t)$, $t \geq 0$ – to function $t \mapsto \sigma^i_t(h)$ (with proper measurability assumptions), with the no-anticipation property and independent of its own very recent actions. Explicitly, if $h = (z_t, a_t)_{t \geq 0}$ and $h' = (z'_t, b_t)_{t \geq 0}$ are two plays with $(z_t, a^{-i}_t) = (z'_t, b^{-i}_t)$ for $t \leq s$, and $a^i_t = b^i_t$ for $t \leq s - \varepsilon$ (for some $\varepsilon > 0$), then $\sigma^i_t(h) = \sigma^i_t(h')$ on $t \leq s$.

The assumption that the strategy choice of player i is independent of player i's (very recent) past actions is conceptually innocuous, as the strategy itself and other players' past actions define player i's past actions. However, it (or an analog of it) 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 $\sigma(h) = h$. 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 just below.

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^i_t is independent of its own very recent past actions), the strategy profile (σ^{-i}, τ^i) (where $\sigma^{-i} = (\sigma^j)_{j \neq i}$) defines a unique distribution on plays. For example, a profile of pure stationary strategies in a continuous-time stochastic game is admissible.

1.8 Continuous-time stochastic games: equilibria

An *equilibrium* is an admissible strategy profile for which no single player can benefit from a unilateral deviation. For example, let $\beta_\rho(z, \sigma) = (\beta^i_\rho(z, \sigma))_{i \in N}$ denote the normalized ρ-discounted payoff associated with the strategy profile σ and the initial state z. Namely, $\beta_\rho(z, \sigma)$ is the expectation of $\int_0^\infty \rho e^{-\rho t} g(z_t, a_t)\, dt$ with respect to the probability P^z_σ defined on plays by the strategy profile

σ and the initial state z. Then $\sigma = (\sigma^i)_{i\in N}$ is an equilibrium of the ρ-discounted game if σ is admissible and for every state z, every player i, and every strategy τ^i of player i, we have $\beta^i_\rho(z,\sigma) \geq \beta^i_\rho(z,\sigma^{-i},\tau^i)$. Similarly, if $\gamma_s(z,\sigma)$ is the expectation of $\frac{1}{s}\int_0^s g(z_t,a_t)\,dt$ with respect to P^z_σ, then σ is an equilibrium of the s-stage game if for every state z, every player i, and every strategy τ^i of player i, we have $\gamma^i_s(z,\sigma) \geq \gamma^i_s(z,\sigma^{-i},\tau^i)$.

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 s_ε sufficiently large 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 $\gamma^i_s(\sigma^{-i}_\varepsilon, \sigma^i) - \varepsilon \leq u^i(z) \leq \gamma^i_s(\sigma_\varepsilon) + \varepsilon$. 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 $\varepsilon > 0$.

The main result is that the non-zero-sum continuous-time stochastic game (with finitely many states and actions[3]) has a uniform equilibrium payoff; in particular, the uniform value exists in the zero-sum case. In fact, we prove an even stronger result, whose statement relies on the additional concepts discussed below.

The limiting-average payoff need not exists a.e with respect P^z_σ, and the expectation (with respect to P^z_σ) of $\liminf_{s\to\infty}\frac{1}{s}\int_0^s g^i(z_t,a_t)\,dt$, denoted $\underline{\gamma}^i(z,\sigma)$, may be strictly less than the expectation (with respect to P^z_σ) of $\limsup_{s\to\infty}\frac{1}{s}\int_0^s g^i(z_t,a_t)\,dt$, denoted $\bar{\gamma}^i(z,\sigma)$. An *ε-limiting-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 such that for every initial state z, every player $i \in N$, and every strategy σ^i of player i we have $\bar{\gamma}^i(z,\sigma^{-i}_\varepsilon,\sigma^i) - \varepsilon \leq u^i(z) \leq \underline{\gamma}^i(z,\sigma_\varepsilon) + \varepsilon$. A *limiting-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-limiting-average equilibrium payoff for every $\varepsilon > 0$.

A *ε-uniform-limiting 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 s_ε sufficiently large such that for every initial state z, every player $i \in N$, every $s \geq s_\varepsilon$, and every strategy σ^i of

[3] These assumptions are essential for the conclusion. The need for finitely many states is well known, and the need for finitely many actions follows from the recent result of Vigeral [34].

player i we have $\gamma_s^i(\sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq u^i(z) \leq \gamma_s^i(\sigma_\varepsilon) + \varepsilon$ and $\bar{\gamma}^i(z, \sigma_\varepsilon^{-i}, \sigma^i) - \varepsilon \leq u^i(z) \leq \underline{\gamma}^i(z, \sigma_\varepsilon) + \varepsilon$. A *limiting-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-limiting-average equilibrium payoff for every $\varepsilon > 0$.

1.9 The results

The main result is that the non-zero-sum continuous-time stochastic game (with finitely many states and actions) has a uniform-limiting average equilibrium payoff; in particular, the undiscounted value exists in the zero-sum case.

The existence of a stationary equilibrium in the discounted game restricted to Markov strategies appears in [37] and [3]. For completeness, we derive these results (without the restriction to Markov strategies) for the history-dependent strategy model.

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 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 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. 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 $[0, t)$.

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-stage normalized payoff of a pure play[4] 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 done for a 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-stage normalized payoff of a play h is $\frac{1}{s} \int_0^s g(z_t, x_t)\, dt$.

The continuous-time stochastic game $\Gamma = \langle N, S, g, A, \mu \rangle$ is called a) a *continuous-time supergame* if $|S| = 1$, b) a *continuous-time Markov decision process* if $|N| = 1$, and c) a *continuous-time two-player zero-sum stochastic game* if $|N| = 2$, and then we set $N = \{1, 2\}$, and $g^2 = -g^1$.

2.1 Strategies in continuous-time games

2.1.1 Stationary strategies

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

A *stationary strategy* of player i is defined by a function $\sigma^i : S \to X^i$ with $\sigma^i(z) \in X^i(z)$. A profile σ of stationary strategies $(\sigma^i)_{i \in N}$ defines a function $\sigma : S \to X$ with $\sigma(z) \in X(z)$ by $\sigma(z)[a] = \prod_{i \in N} \sigma^i(z)[a^i]$. For $x \in X(z)$ we define $\mu(z', z, x) := \sum_{a \in A(z)} x[a] \mu(z', z, a)$ (the linear extension of μ). It is easy to show[5] that every profile σ of stationary strategies and an initial state z_0 define on the space of right-continuous functions $t \mapsto z_t \in S$ ($t \geq 0$) a unique probability distribution P_σ that for every $z \neq z_t$ satisfies the equality

$$P_\sigma(z_{t+\delta} = z \mid h_t) = \delta \mu(z, z_t, \sigma(z_t)) + o(\delta). \qquad (1)$$

[4] 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 $0 \leq a, b, c$. This defines the discounted and s-stage payoffs over all plays, not necessarily measurable ones.

[5] See, e.g., Section 7.

This unique distribution is defined by the distribution of $(s_k, z_{s_k})_{k\geq 0}$ where s_k is the time of the k-th state change, and conditional on $(s_\ell, z_{s_\ell})_{\ell\leq k}$ the distribution of $s_{k+1} - s_k$ is given by

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

and therefore $s_{k+1} = \infty$ on $\mu(z_{s_k}, z_{s_k}, \sigma(z_{s_k})) = 0$, and 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

$$\forall z \neq z_{s_k}, \quad P_\sigma(z_{s_{k+1}} = z \mid (s_\ell)_{\ell\leq k+1},\, (z_{s_\ell})_{\ell\leq k}) = \frac{-\mu(z, z_{s_k}, \sigma(z_{s_k}))}{\mu(z_{s_k}, z_{s_k}, \sigma(z_{s_k}))}$$

on $\mu(z_{s_k}, z_{s_k}, \sigma(z_{s_k})) \neq 0$.

2.1.2 Markov strategies

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 $\sigma(z,t)[a] = \prod_{i\in N} \sigma^i(z,t)[a^i]$. A profile σ of Markov strategies 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

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

This unique distribution is defined by the distribution of $(s_k, z_{s_k})_{k\geq 0}$ where s_k is the time of the k-th state change, and 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 Lebesque 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 by the minimal topology for which, for every 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.

2.1.3 Oblivious strategies

Let H^S denote the set of all functions $h^S : t \mapsto z_t$, $t \geq 0$, that are right continuous and have left limits. A point h^S in H^S is identified with the sequence (s_k, z_{s_k}), where $s_0 = 0$ and s_k is the time of the k-th state change. $\mathcal{H}^S$ is the minimal σ-algebra for which all the functions $H^S \ni h^S \mapsto s_k$ and $H^S \ni h^S \mapsto z_{s_k}$, $k \in \mathbb{N}$, are measurable. Equivalently, it is the Borel σ-algebra with respect to the minimal topology for which all functions $H^S \ni h^S \mapsto s_k$ and $H^S \ni h^S \mapsto z_{s_k}$, $k \in \mathbb{N}$, are continuous. Its sub-σ-algebra generated by all $(s_k, z_{s_k})\mathbb{I}(s_k \leq t)$, namely, by the values of $(z_s)_{s \leq t}$, is denoted $\mathcal{H}^S_t$. The increasing family $\mathcal{H}^S_t$, $t \geq 0$, is a filtration of $\mathcal{H}^S$ and s_k is a stopping time with respect to this filtration. Therefore $\mathcal{H}^S_{s_k}$ is well defined.

An *oblivious strategy* of player i specifies the player's action as a function of the state process only. Therefore, its behavior in the time interval $[s_k, s_{k+1})$ is measurable with respect to $\mathcal{H}^S_{s_k}$. An *oblivious strategy* σ^i is represented by a sequence of Markov strategies $(\sigma^i_k)_{k \in \mathbb{N}}$, where σ^i_k is measurable with respect to $\mathcal{H}^S_{s_k}$. Note that there is some redundancy in this representation, as all that matters is the behavior of σ^i_k in the interval $[s_k, s_{k+1})$ and on states $z = z_{s_k}$. However, this representation simplifies the notations. One can obviously define the oblivious strategies as the proper equivalence classes of the redundant representation above.

A profile σ of oblivious strategies $(\sigma^i)_{i \in N}$ defines an oblivious strategy profile $(\sigma_k)_{k \in \mathbb{N}}$ with $\sigma_k(z, t) \in X(z)$. It is easy to show that a profile σ of oblivious strategies and an initial state z_0 define a unique probability distribution P_σ on $(H^S, \mathcal{H}^S)$ that for every k, $s_k \leq t < s_{k+1}$, and $z \neq z_{s_k}$, satisfies the equality

$$P_\sigma(z_{t+\delta} = z \mid \mathcal{H}^S_{s_k}) = \int_t^{t+\delta} \mu(z, z_t, \sigma_k(z_{s_k}, s))\, ds + o(\delta). \tag{3}$$

This unique distribution is defined by

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

and (using the Lebesque 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

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

2.1.4 Discretimized strategies

Fix a strictly increasing sequence of real numbers $\mathcal{T} = (t_k)_{k\geq 0}$ with $t_0 = 0$ and $t_k \to \infty$ 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 $\sigma^i_{h_{t_k}} : S \times [t_k, t_{k+1}) \to X^i$. 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 $\sigma_{h_{t_k}}(z, s) := \otimes_{i\in N}\sigma^i_{h_{t_k}}(z, s) \in X(z)$. It defines a probability distribution P_σ on plays as follows. For $t_k \leq t < t_{k+1}$ and $z \in S$,

$$x_t = \sigma_{h_{t_k}}(z_t, t)$$

and

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

where $\mathbb{I}$ stands for the indicator function; $\mathbb{I}(z = z_t) = 1$ if $z = z_t$, and $\mathbb{I}(z = z_t) = 0$ if $z \neq z_t$.

A more informative, but longer, name of a $\mathcal{T}$-discretimized strategy is a $\mathcal{T}$-discretimized Markov strategy. The reason is that conditional on h_{t_k}, in between two "updating times," t_k and t_{k+1}, the strategy coincides with a Markov strategy, and thus, at time $t_k \leq t < t_{k+1}$, disregards the additional information of the states z_s, $t_k < s < t$, and of the action profiles a_s, $t_k \geq s < t$.

The concept of discretimized strategies applies also to a strictly increasing sequence of $(\mathcal{H}^S_t)_{t\geq 0}$-stopping times. A sequence of stopping times $\mathcal{T} = (t_k)_{k\geq 0}$ (with $t_0 = 0$ and $t_k \to \infty$ as $k \to \infty$) is called *fine* if for every $k \geq 0$ and $t_k < s < t_{k+1}$ we have $z_s = z_{t_k}$ (equivalently, if the stopping times refine the stopping times $(s_k)_{k\in\mathbb{N}}$). Given a fine sequence $\mathcal{T}$, a profile of $\mathcal{T}$-discretimized strategies σ defines a probability P_σ on plays as above.

Discritimized strategies are termed also no anticipation strategies with delay in the differential games literature.

2.1.5 General strategies

The space of all plays is denoted H. The space of plays without the actions of player i is denoted 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.

A profile of strategies need not define unambiguously a probability distribution over plays. 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 unambiguously the probability distribution P_{σ^{-i},τ^i} on plays. A profile of stationary, or Markov, or oblivious, or discretimized strategies $\sigma = (\sigma^i)_{i \in N}$ is an admissible strategy profile.

2.2 Discussion on strategies

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. Interpreting the set $\mathcal{T}$ as the set of action times, and A as the set of single-stage action profiles, 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 $h(t) = \sigma_t(h_t)$. A local pure strategy profile is *integrable* if its integral is nonempty; i.e., there is a 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. The play h, where $h(t)$ is defined by transfinite induction – $h(t) = \sigma_t(h_t)$ – 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 $|\{j : i < j$ and $h(t_j) = a\}| < \infty$, and $\tau_t(*) = a$ otherwise. The play that plays a everywhere, as well as 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 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 $t^* \in \mathcal{T}^*$. □

2.3 Strategies observing and selecting 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, as a function of past observed variables, mixed actions.

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 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 results in the corresponding classical models, as evident from the results in the asymptotic (see Section 9) and discretimized (see Section 8) approaches.

3 Continuous-time two-person zero-sum stochastic games

In this section, $g(z,x)$ stands for (the payoff to player 1 when the mixed action x is played at state z) $g^1(z,x)$. No confusion should result. Recall that $X^i(z)$ denotes all probability distributions on $A^i(z)$.

3.1 The discounted case

We say that the ρ-discounted two-person zero-sum continuous-time stochastic game Γ has a value $V = V_\rho \in \mathbb{R}^S$ if for every $\varepsilon > 0$ there are admissible strategies, σ of player 1 and τ of player 2, such that for every strategy σ^* and τ^* of player 1 and player 2, respectively, and every state z,

$$\varepsilon + E^z_{\sigma,\tau^*} \int_0^\infty e^{-\rho t} g(z_t, x_t)\, dt \geq V(z) \geq E^z_{\sigma^*,\tau} \int_0^\infty e^{-\rho t} g(z_t, x_t)\, dt - \varepsilon.$$

If the game has a value V, then a strategy σ (respectively, τ) of player 1 (respectively, 2) that obeys this inequality for $\varepsilon = 0$ is called an *optimal strategy.*

Theorem 1 *Every discounted two-person zero-sum continuous-time stochastic game has a value. The value of the ρ-discounted game equals the unique* *solution $V \in \mathbb{R}^S$ of the system of S equations, $z \in S$,*

$$\rho v(z) = \min_{x^2 \in X^2(z)} \max_{x^1 \in X^1(z)} \left(g(z, x^1 \otimes x^2) + \sum_{z' \in S} \mu(z', z, x^1 \otimes x^2) v(z') \right), \quad (4)$$

and each player has an optimal stationary strategy.

Proof. Define $\|g\| := \max_{z \in S, a \in A(z)} |g(z,a)|$, $J(z) = [-\|g\|/\rho, \|g\|/\rho]$, and $J = \times_{z \in S} J(z)$.

For every $z \in S$, $a \in A(z)$, $v \in \mathbb{R}^S$, and $x \in X(z)$, $G^z[v](a)$ is defined by

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

where $\|\mu\| = \max_{z,a} |\mu(z,z,a)|$, and $G^z[v](x)$ is defined by

$$G^z[v](x) = \sum_{a \in A(z)} x(a) G^z[v](a).$$

Alternatively, defining $g(z,x) := \sum_{a \in A(z)} x(a) g(z,a)$ (the linear extension of g^z) and $\mu(z', z, x) := \sum_{a \in A(z)} x(a) \mu(z', z, a)$, we have

$$G^z[v](x) = \frac{1}{\|\mu\| + \rho} \left(g(z,x) + \sum_{z' \in S} \mu(z', z, x) v(z') + \|\mu\| v(z) \right). \quad (5)$$

Define the operator Q from $\mathbb{R}^S$ to $\mathbb{R}^S$ by

$$Qv(z) = \max_{x \in X^1(z)} \min_{x^2 \in X^2(z)} G^z[v](x^1 \otimes x^2)$$

where $x^1 \otimes x^2$ is the product distribution $x \in X(z)$ that is given by $x(a) = x^1(a^1) x^2(a^2)$. By the minmax theorem we have

$$Qv(z) = \min_{x^2 \in X^2(z)} \max_{x \in X^1(z)} G^z[v](x^1 \otimes x^2)$$

and therefore v is a solution of $Qv = v$ if and only if it is a solution of (4). Note that $G^z[v + c1_S](x) = G^z[v](x) + \frac{c\|\mu\|}{\|\mu\| + \rho}$, and therefore

$$Q(v + c1_S)(z) = Qv + \frac{\|\mu\|}{c\|\mu\| + \rho}.$$

In addition, Q is monotonic; i.e., $u \geq v$ implies that $Qu \geq Qv$, and therefore for $v, u \in \mathbb{R}^S$ we have

$$\|Qv - Qu\| \leq \frac{\|\mu\|}{\|\mu\| + \rho} \|v - u\|.$$

Therefore Q is a strict contraction and therefore has a unique fixed point. Let V be a unique fixed point of Q. Let σ be a stationary strategy of player 1 with $\sigma(z)$ maximizing $\min_{x^2 \in X^2(z)} G^z[V](\sigma(z) \otimes x^2)$, namely,

$$\min_{x^2 \in X^2(z)} G^z[V](\sigma(z) \otimes x^2) \geq V(z) = \min_{x^2 \in X^2(z)} \max_{x^1 \in X^1(z)} G^z[V](x^1 \otimes x^2) \qquad (6)$$

for every $x^1 \in X^1(z)$.

We claim that for every strategy τ of player 2 and every state z,

$$V(z) \leq E^z_{\sigma,\tau} \int_0^\infty e^{-\rho t} g(z_t, x_t)\, dt. \qquad (7)$$

Fix a strategy τ and an initial state z. Define

$$f(s) = E^z_{\sigma,\tau} \int_0^s e^{-\rho t} g(z_t, x_t)\, dt + E^z_{\sigma,\tau} e^{-\rho s} V(z_s).$$

Note that $f(s) \to_{s \to \infty} E^z_{\sigma,\tau} \int_0^\infty e^{-\rho t} g(z_t, x_t)\, dt$ and that $f(0) = V(z)$. In addition, f is Lipschitz ($|f(s+\delta) - f(s)| \leq \delta e^{-\rho s}(2\|\mu\| \|V\|_\infty + \|g\|)$. Therefore, in order to prove (7), it suffices to prove that the lower-right derivative of f is nonnegative everywhere.

Let $s_\delta := \min\{\delta, \inf\{t > 0 : z_t \neq z_0\}$. Let $\tilde{\tau}$ be a Markov strategy of player 2 so that $\tilde{\tau}(h_t) = \tau(h_t)$ on $0 \leq t < s_\delta$. Note that for $\alpha > 0$ we have $P_{\sigma,\tau}(s_\delta \geq \alpha \mid z_0) \geq e^{-\|\mu\|\alpha}$. Therefore, setting $y_t = (\sigma(z_0), \tilde{\tau}(t, z_0))$, we have

$$\begin{aligned}
E^z_{\sigma,\tau} \int_0^\delta e^{-\rho t} g(z_t, x_t)\, dt &= E^z_{\sigma,\tau} \int_0^\delta e^{-\rho t} g(z_t, x_t)(\mathbb{I}(t \leq s_\delta) + \mathbb{I}(t > s_\delta))\, dt \\
&\geq \int_0^\delta e^{-\rho t} g(z_0, y_t)\, dt - 2E^z_{\sigma,\tau} \int_0^\delta \mathbb{I}(t > s_\delta)\|g\|\, dt \\
&\geq \int_0^\delta e^{-\rho t} g(z_0, y_t)\, dt - \|g\|\delta^2\|\mu\| \\
&\geq \int_0^\delta g(z_0, y_t)\, dt - \|g\|\delta^2(\|\mu\| + \rho/2).
\end{aligned}$$

The last two inequalities use the inequality $1 - e^{-t\theta} \leq t\theta$ and the equality $\int_0^\delta t\,dt = \delta^2/2$.

Using inequality (39) of Theorem 9 (Section 7) (or by following a similar chain of inequalities as above), we have

$$e^{-\rho\delta} E_{\sigma,\tau}^{z_0} V(z_\delta) \;\; \geq \;\; V(z_0) + \int_0^\delta \left(\sum_{z' \in S} V(z')\mu(z', z_0, y_t) - \rho V(z_0) \right) dt - O(\delta^2).$$

Therefore,

$$f(\delta) - f(0) \;\; \geq \;\; \int_0^\delta \left(g(z_0, y_t) + \sum_{z' \in S} \mu(z', z_0, y_t) - \rho V(z_0) \right) dt - O(\delta^2).$$

The choice of σ implies that $g(z_0, y_t) + \sum_{z' \in S} \mu(z', z_0, y_t) V(z') - \rho V(z_0) \geq 0$ and therefore the lower-right derivative of f at 0, $\liminf_{\delta \to 0+} \frac{f(\delta) - f(0)}{\delta}$, is nonnegative.

Similarly, at every point $s > 0$, the lower-right derivative of f at s is nonnegative. We conclude that f is nondecreasing. □

Corollaries.

The algebraic approach. The graph of the correspondence assigning to each ρ the optimal stationary strategies of each player and the value function V_ρ is semialgebraic. Therefore, there is a map $\rho \mapsto (V_\rho, \sigma^\rho, \tau^\rho)$ that is semialgebraic; in particular, it has a convergent expansion in fractional powers of ρ in a right neighborhood of 0 (and a convergent expansion in fractional powers of ρ in any one-sided neighborhood of a point $\rho_0 > 0$). As V_ρ is the fixed point of the strict contraction map $Q(\rho)$ with the contraction factor $\frac{\|\mu\|}{\|\mu\| + \rho}$, we have $\|V_\rho\|_\infty \leq \frac{\|Q(\rho)(0)\|(\|\mu\| + \rho)}{\rho} \leq \|g\|/\rho$. Therefore $\rho \mapsto v_\rho := \rho V_\rho$ is a bounded semi-algebraic function. In particular, there is a positive integer M, and real coefficients $c_k(z)$, and a $\bar{\rho} > 0$, such that for $0 < \rho \leq \bar{\rho}$ the series $\sum_{k=0}^\infty c_k(z)\rho^{k/M}$ converges and

$$v_\rho(z) = \sum_{k=0}^\infty c_k(z)\rho^{k/M}.$$

If the game is one of perfect information, then each player has for each $\rho > 0$ a pure-action optimal stationary strategy in the ρ-discounted game. Therefore (following the classical argument from discrete-time stochastic games) the value function $\rho \mapsto v_\rho(z)$ is a rational function in ρ in a right neighborhood of 0 (and in any one-sided neighborhood of a point $\rho_0 > 0$). It follows that for a continuous-time two-person zero-sum game there is $\bar{\rho} > 0$ and real coefficients $c_k(z)$, and pure-action stationary strategies σ^i, $i = 1, 2$, such that for $\rho \leq \bar{\rho}$ the series $\sum_{k=0}^{\infty} c_k(z)\rho^k$ converges,

$$v_\rho(z) = \sum_{k=0}^{\infty} c_k(z)\rho^k,$$

and σ^i is optimal in the ρ-discounted game.

Covariance properties. Fix the state and action sets S, and A, and consider the value function $V = V_\rho(g, \mu)$ of the ρ-discounted game as a function of ρ, g, and μ. Then obviously, if $g' \geq g$ and α is a nonnegative real number, $V_\rho(g', \mu) \geq V_\rho(g, \mu)$ and $V_\rho(\alpha g, \mu) = \alpha V_\rho(g, \mu)$. For $\alpha > 0$, a vector V satisfies equation (4) if and only if it satisfies the same equation when ρ is replaced with $\alpha\rho$, g is replaced by αg, and μ is replaced by $\alpha\mu$. Therefore, $V_{\alpha\rho}(\alpha g, \alpha\mu) = V_\rho(g, \mu)$. If $\|\mu\| \leq 1$, we assign to the continuous-time game $\Gamma = \langle N, S, A, \mu, g\rangle$ the discrete-time game $\bar{\Gamma} = \langle N, S, A, p(\mu), g\rangle$ where $p = p(\mu)$ is given by $p(z' \mid z, a) = \mathbb{I}(z' = z) + \mu(z', z, a)$. The value $\bar{V}_\rho(g, \mu)$ of the discrete-time ρ-discounted (with discount factor $1 - \rho$) stochastic game $\bar{\Gamma} = \langle \{1, 2\}, A, p(\mu), g\rangle$ is the unique solution of the equations, $z \in S$,

$$\rho V(z) = \min_{x^2 \in X^2(z)} \max_{x^1 \in X^1(z)} g(z, x^1 \otimes x^2) + (1 - \rho) \sum_{z' \in z} \mu(z', z, x^1 \otimes x^2) V(z').$$

Therefore, for $\rho < 1$, $\bar{V}_\rho(g, \mu) = V_\rho(g, (1 - \rho)\mu)$, and for $0 < \rho < 1 - \|\mu\|$, $V_\rho(g, \mu) = \bar{V}_\rho(g, \mu/(1 - \rho))$.

Summarizing, for $\rho > 0$, $\bar{V}_\rho(g, \mu) = V_\rho(g, (1 - \rho)\mu)$, and for $0 < \rho < 1 - \|\mu\|$,

$$V_\rho(g, \mu) \geq V_\rho(g', \mu) \text{ whenever } g \geq g', \tag{8}$$

$$V_\rho(\alpha g, \beta\mu) = \frac{\alpha}{\beta} V_{\rho/\beta}(g, \mu) \text{ whenever } \alpha \geq 0 \text{ and } \beta > 0, \tag{9}$$

$$V_\rho(g, \mu) = \bar{V}_\rho(g, \frac{1}{1 - \rho}\mu) \text{ whenever } \rho \leq 1 - \|\mu\|; \tag{10}$$

equivalently,

$$\bar{V}_\rho(g, \mu) = V_\rho(g, (1 - \rho)\mu) \text{ whenever } \rho < 1 \text{ and } \|\mu\| \leq 1. \tag{11}$$

The relation $\bar{V}_\rho(g, \mu) = V_\rho(g, (1 - \rho)\mu)$, and for $0 < \rho < 1 - \|\mu\|$, in conjunction with [30, Theorem 6], proves the following inequalities. For

$0 < \rho < 1 - \|\mu\|$, we have

$$\|v_\rho(g', \mu') - v_\rho(g, \mu)\|_\infty \quad \leq \quad 4|S| d(\mu, \mu') \min\{\|g\|, \|g'\|\} + \|g - g'\| \quad (12)$$

where

$$d(\mu, \mu') = \max\left\{\frac{\mu(z', z, a)}{\mu'(z', z, a)}, \frac{\mu'(z', z, a)}{\mu(z', z, a)} \mid a \in A(z), z, z' \in S\right\} - 1,$$

and by convention $x/0 = \infty$ for $x > 0$, and $0/0 = 1$.

3.2 The undiscounted finite-horizon case

Let $T(v)(z)$ be the value of the single-stage game $G^z[v]$ that is given by

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

T is a nonexpansive operator on $\mathbb{R}^S$. Let $(Q_t)_{t \geq 0}$ be the (unique) semigroup of operators with $Q_0 = I$, $Q_{t+s} = Q_t \circ Q_s$, and $\lim_{s \to 0+} \frac{Q_s - I}{s} = T - I$.

In the unnormalized s-stage game, player 1 (respectively, 2) maximizes (respectively, mimimizes) the expectation of the integral $\int_0^s g(z_t, x_t)\, dt$.

The *value* of the s-stage game exists and equals V_s if for every $\varepsilon > 0$ there is an admissible strategy pair σ_ε, τ_ε, such that for every strategy pair σ^* and τ^* and every initial state z we have

$$\varepsilon + E^z_{\sigma_\varepsilon, \tau^*} \int_0^s g(z_t, x_t)\, dt \geq V_s(z) \geq E^z_{\sigma^*, \tau_\varepsilon} \int_0^s g(z_t, x_t)\, dt - \varepsilon. \quad (13)$$

An admissible strategy $\sigma = \sigma_\varepsilon$ (respectively, $\tau = \tau_\varepsilon$) is called ε-*optimal*, $\varepsilon \geq 0$, if for every strategy τ^* (respectively, σ^*) it obeys (13), and it is called an *optimal* strategy if it is 0-optimal.

Proposition 3 *The value of the unnormalized s-stage game exists, equals $Q_s(0)$, and each player has an optimal strategy.*

Proof. It is sufficient (by duality) to prove the existence of a strategy σ of player 1 such that for every strategy τ of player 2 and every initial state z we have

$$E^z_{\sigma, \tau}\left(\int_0^s g(z_t, x_t)\, dt\right) \geq Q_s(0)(z). \quad (14)$$

We define the strategy σ. Let σ be a measurable Markov strategy so that for every $t < s$, $\sigma(t, z_t)$ $(= \sigma(h_t))$ is an optimal strategy in the game $G^{z_t}[Q_{s-t}(0)]$. Fix a strategy τ of player 2 and an initial state z. Define

$$f(t) = E^z_{\sigma,\tau}\left(\int_0^t g(z_t, x_t)\, dt + Q_{s-t}(0)(z_t)\right).$$

It follows that $f(0) = Q_s(0)(z)$.

The function $t \mapsto Q_t(0)$, $t \geq 0$, is Lipschitz (with constant $\|T(0)\|$), and for two distributions P and P' on S we have $|E_P Q_t(0)(z) - E_{P'} Q_t(0)(z)| \leq \|P - P'\|\, \|Q_t(0)\|$. Therefore, f is Lipschitz. By the choice of σ, it follows (along similar lines to the proofs in the discounted case) that the lower derivative of f at every point $0 < t < s$ is nonnegative. Therefore, $f(s) \geq f(0)$. As $f(s) = E^z_{\sigma,\tau}(\int_0^s g(z_t, x_t)\, dt)$ inequality (14) follows. □

The undiscounted continuous-time s-stage game with a terminal payoff vector $\hat{V} \in \mathbb{R}^S$ is the game where player 1 maximizes the expectation of

$$\int_0^s g(z_t, x_t)\, dt + \hat{V}(z_s).$$

The *value* of the s-stage game with terminal payoff vector $\hat{V} \in \mathbb{R}^S$ exists and equals V_s if for every $\varepsilon > 0$ there is an admissible strategy pair σ_ε, τ_ε, such that for every strategy pair σ^* and τ^* and every initial state z we have

$$\varepsilon + E^z_{\sigma_\varepsilon,\tau^*}\left(\int_0^s g(z_t, x_t)\, dt + \hat{V}(z_s)\right) \geq V_s(z) \geq E^z_{\sigma^*,\tau_\varepsilon}\left(\int_0^s g(z_t, x_t)\, dt + \hat{V}(z_s)\right) - \varepsilon. \tag{15}$$

An admissible strategy $\sigma = \sigma_\varepsilon$ (respectively, $\tau = \tau_\varepsilon$) is called ε-optimal ($\varepsilon \geq 0$) in the s-stage game with terminal payoff vector $\hat{V} \in \mathbb{R}^S$, if for every strategy τ^* (respectively, σ^*) it obeys (15), and it is called an *optimal* strategy if it is 0-optimal.

Proposition 4 *The value of the unnormalized continuous-time s-stage game with terminal payoff $\hat{V}$ exists and equals $Q_s(\hat{V})$.*

Proof. We define the strategy σ. Let σ be a measurable Markov strategy so that for every $t < s$, $\sigma(t, z_t)$ $(= \sigma(h_t))$ is an optimal strategy in the game $G^{z_t}[Q_{s-t}(\hat{V})]$. Fix a strategy τ of player 2. Define

$$f(t) = E_{\sigma,\tau}\left(\int_0^t g(z_t, x_t)\, dt + Q_{s-t}(\hat{V})(z_t)\right).$$

It follows that $f(0) = Q_s(\hat{V})(z_0)$. f is Lipschitz and by the choice of σ, it follows (along similar lines to the proofs in the discounted case) that the lower derivative of f at every point $0 < t < s$ is nonnegative. Therefore, $f(s) \geq f(0)$. As $f(s) = E_{\sigma,\tau}\left(\int_0^s g(z_t, x_t)\,dt + \hat{V}(z_s)\right)$ we conclude that

$$E_{\sigma,\tau}(\int_0^s g(z_t, x_t)\,dt) \geq Q_s(\hat{V})(z_0).$$

□

3.3 The undiscounted case

We say that the game has a *value* v if for every $\varepsilon > 0$ there are admissible strategies σ and τ and $s_\varepsilon > 0$ such that for every strategy pair (σ^*, τ^*), every initial state z, and every $s > s_\varepsilon$, we have

$$\varepsilon + E^z_{\sigma,\tau^*}\frac{1}{s}\int_0^s g(z_t, x_t)\,dt \geq v(z) \geq E^z_{\sigma^*,\tau}\frac{1}{s}\int_0^s g(z_t, x_t)\,dt - \varepsilon$$

and

$$\varepsilon + E^z_{\sigma,\tau^*}\liminf_{s\to\infty}\frac{1}{s}\int_0^s g(z_t, x_t)\,dt \geq v(z) \geq E^z_{\sigma^*,\tau}\limsup_{s\to\infty}\frac{1}{s}\int_0^s g(z_t, x_t)\,dt - \varepsilon.$$

Theorem 2 *A continuous-time two-person zero-sum stochastic game with finitely many states and actions has a value.*

Proof. Let $r_k = \int_0^1 g(z_{k+t}, x_{k+t})\,dt$, $v_\rho = \rho V_\rho$, and $v_{\rho,k} = v_\rho(z_k)$. Recall that v_ρ is a semialgebraic function of ρ. By the proof in [15] (or by using the statement of [22, Lemma 1]) it suffices to prove that for every $\varepsilon > 0$ there is $\theta > 0$ so that for every sequence $(\lambda_k)_{k=0}^\infty$ with $0 < \lambda_k < \theta$, where λ_k is measurable with respect to $\mathcal{H}_k$, there is a strategy σ of player 1 such that for every strategy τ of player 2 (and every k) we have

$$E_{\sigma,\tau}(\lambda_k r_k + (1-\lambda_k)v_{\lambda_k,k+1} \mid \mathcal{H}_k) \geq v_{\lambda_k,k} - \varepsilon\lambda_k.$$

If σ is a strategy of player 1 that coincides on the time interval $t \in [k, k+1)$ with a stationary optimal strategy in the continuous-time λ_k-discounted game, then for every strategy τ of player 2 we have

$$E_{\sigma,\tau}\left(\int_0^1 e^{-\lambda_k t}\lambda_k g(z_{t+k}, x_{t+k})\,dt + e^{-\lambda_k}v_{\lambda_k,k+1}) \mid \mathcal{H}_k\right) \geq v_{\lambda_k,k}.$$

As $|g(z_{t+k}, x_{t+k})| \leq \|g\|$ and $|v_{\rho,k+1}| \leq \|g\|$, the inequality $| \int_0^1 |1 - e^{-\rho t}\rho| \, dt + |1 - \rho - e^{-\rho}| \leq O(\rho^2)$ implies that

$$\int_0^1 e^{-\rho t}\rho g(z_{t+k}, x_{t+k}) \, dt + e^{-\rho} v_{\rho,k+1} \leq \rho r_k + (1-\rho) v_{\rho,k+1} + O(\rho^2).$$

Therefore, for sufficiently small θ, the inequality $\lambda_k < \theta$ implies that

$$E_{\sigma,\tau}(\lambda_k r_k + (1-\lambda_k) v_{\lambda_k,k+1} \mid \mathcal{H}_k) \geq v_{\lambda_k,k} - \varepsilon \lambda_k.$$

□

3.4 Nonexpansive maps and continuous-time stochastic games

To every discrete-time two-person zero-sum stochastic game is associated[6] a nonexpansive map, $T : \mathbb{R}^S \to \mathbb{R}^S$, from real-valued functions of the state space, such that the unnormalized value of the n-stage game is $T^n 0$, and the value of the unnormalized λ-discounted game is the unique solution of the equation $T(\lambda V) = V$. In the case of finitely many states and actions this map is semialgebraic, and this semialgebraicity enables us to derive asymptotic properties of the value of the λ-discounted game and of the n-stage game (as the discount factor goes to 1 and the number of stages goes to infinity).

In this section we associate to every continuous-time two-person zero-sum stochastic game a nonexpansive map, $T : \mathbb{R}^S \to \mathbb{R}^S$, define its continuous-time orbit, and relate the value of the s-stage ($s \in [0, \infty)$) game and the ρ-discounted game to the orbit of T.

For every nonexpansive map (not necessarily one defined by a continuous-time stochastic game), $T : \mathbb{R}^S \to \mathbb{R}^S$, there is a unique semigroup of operators on $\mathbb{R}^S$, T_s, $s \geq 0$, with $\lim_{\delta \to 0+} \frac{1}{\delta}(T_\delta - I) = T - I$. (Namely, the flow $T_s v$ satisfies the differential equation $\frac{d}{dt} T_t v = T(T_t v) - T_t v$ with the initial condition $T_0 v = v$.) We prove that if T is semialgebraic, then the limit of $\frac{1}{s} T_s v$, as $s \to \infty$, exists, and is independent of v.

We start with the definition of the nonexpansive operator associated with the continuous-time stochastic game. Define $\|g\| := \max_{z \in S, \, a \in A(z)} |g(z,a)|$, $V(z) = [-\|g\|/\rho, \|g\|/\rho]$.

[6] See, e.g., [23, p. 397].

For every $z \in S$, $v \in V$, $a \in A(z)$, and $x \in X(z)$, $G^z[v](a)$ is defined by

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

where $\|\mu\| = \max_{z,a} |\mu(z,z,a)|$, and $G^z[v](x)$ is defined by

$$G^z[v](x) = \sum_{a \in A(z)} x(a) G^z[v](a).$$

Alternatively, defining $g(z,x) := \sum_{a \in A(z)} x(a)g(z,a)$ (the linear extension of g^z) and $\mu(z',z,x) := \sum_{a \in A(z)} x(a)\mu(z',z,a)$, we have

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

Define the operator Q from $\mathbb{R}^S$ to $\mathbb{R}^S$ by

$$Qv(z) = \max_{x \in X^1(z)} \min_{x^2 \in X^2(z)} G^z[v](x^1 \otimes x^2)$$

where $x^1 \otimes x^2$ is the product distribution $x \in X(z)$ that is given by $x(a) = x^1(a^1)x^2(a^2)$. Note that $G^z[v + c1_S](x) = G^z[v](x) + c$, and therefore

$$Q(v + c1_S)(z) = Qv(z) + c\,.$$

In addition, Q is monotonic; i.e., $u \geq v$ implies that $Qu \geq Qv$, and therefore for $v, u \in \mathbb{R}^S$ we have

$$\|Qv - Qu\| \leq \|u - v\|.$$

Therefore Q is nonexpansive. In addition, note that it is semialgebraic.

Any nonexpansive map $Q : \mathbb{R}^S \to \mathbb{R}^S$ is the generator of the (unique) semigroup of operators Q_s, $s \geq 0$, with $Q_0 = I$ (where I is the identity operator), $Q_{t+s} = Q_t \circ Q_s$, and $\lim_{s \to 0+} \frac{1}{s}(Q_s - I) = Q - I$.

Each element of the semigroup is nonexpansive, i.e., $\|Q_s v - Q_s u\| \leq \|v - u\|$. However, Q_1 need not be semialgebraic even if Q is semialgebraic. Therefore, the results on the iterates of a semialgebraic nonexpansive operator Q on $\mathbb{R}^S$ do not imply directly the results on the continuous iterates (i.e., the semigroup generated by Q) of Q. We will however show that the results hold also for the continuous iterates.

Theorem 3 *1) Let Q be the nonexpansive map associated to a continuous-time stochastic game Γ. Then the value of the unnormalized s-stage game Γ_s equals Q_s0.*

2) For every nonexpansive semialgebraic map $Q : \mathbb{R}^S \to \mathbb{R}^S$ the limit $\lim_{t\to\infty} \frac{1}{t} Q_t V$ exists and is independent of V.

The statement and proof of 1) is part of Proposition 3. Let Q be nonexpansive and semialgebraic. For every $0 < r \le 1$ let $x(r)$ denote the unique fixed point of (the strict contraction) $\frac{1}{1+r}Q$. $r \mapsto rx(r)$ is semialgebraic and bounded. Therefore 2) follows from Theorem 4.

Theorem 4 *Let X be a Banach space and $T : X \to X$ a nonexpansive operator, and for every $0 < r \le 1$ let $x(r)$ denote the unique fixed point of $\frac{1}{1+r}T$. Let $Q : [0,\infty) \times X$ be the unique flow with $Q_{s+t} = Q_s \circ Q_t$ and $\|Q_s x - (1-s)x - sTx\| = o(s)$ as $s \to 0$ (uniformly on bounded sets), where I denotes the identity operator. If the function $r \mapsto rx(r)$ is of bounded variation on the interval $(0,1]$, then the limit $\lim_{s\to\infty} \frac{1}{s} Q_s(0)$ exists and equals $\lim_{r\to 0+} rx(r)$.*

Proof. Let $T_\delta = (1-\delta)Ix + \delta Tx$. Then $T^m_{s/m}(x) \to_{m\to\infty} Q_s(x)$ ($\|T^m_{s/m}(x) - Q^m_{s/m}(x)\| \le K(s,x)o(1)$ as $m \to \infty$), where $T^m_\delta(x)$ stands for the more explicit $(T_\delta)^m(x)$. Note that $T_\delta x(r) = (1+\delta r)x(r)$ and thus $T_\delta x(\frac{1}{n\delta}) = (1+1/n)x(\frac{1}{n\delta})$. Fix $\delta > 0$ (sufficiently small). By the triangle inequality and the nonexpansiveness of T_δ,

$$\begin{aligned}
\|T^{n+1}_\delta(0) - x(\frac{1}{\delta(n+1)})\| &\le \|T^{n+1}_\delta(0) - T_\delta x(\frac{1}{\delta n}))\| + \|T_\delta(x(\frac{1}{\delta n})) - x(\frac{1}{\delta(n+1)})\| \\
&\le \|T^n_\delta(0) - x(\frac{1}{\delta n})\| + \|T_\delta(x(\frac{1}{\delta n})) - x(\frac{1}{\delta(n+1)})\| \\
&\le \|T^n_\delta(0) - x(\frac{1}{\delta n})\| + \delta(n+1)\|x_{\delta,n} - x_{\delta,n+1})\|
\end{aligned}$$

where $x_{\delta,n} = \frac{1}{\delta n}x(\frac{1}{\delta n})$. Summing the above inequalities over $n = k, \ldots, m-1$, we deduce that

$$\begin{aligned}
\|T^m_\delta(0) - x(\frac{1}{m\delta})\| &\le \|T^k_\delta(0) - x(\frac{1}{k\delta})\| + \delta \sum_{n=k}^{m-1} (n+1)\|x_{\delta,n} - x_{\delta,n+1}\| \\
&\le \|T^k_\delta(0)\| + \|x(\frac{1}{k\delta})\| + \delta m \sum_{n=k}^{m-1} \|x_{\delta,n} - x_{\delta,n+1}\|.
\end{aligned}$$

T_δ is nonexpansive and $T_\delta(0) = \delta T(0)$. Therefore $\|T_\delta^k(0)\| \leq k\delta\|T(0)\|$. As T is nonexpansive and $T(x(\frac{1}{k\delta})) = (1+\frac{1}{k\delta})x(\frac{1}{k\delta})$, we have $\|x(\frac{1}{k\delta})\| \leq \delta k\|T(0)\|$. Let $V(\alpha)$, $\alpha > 0$, denote the variation of the function $r \mapsto rx(r)$ on the interval $(0,\alpha)$. As $V(\alpha) < \infty$, $V(\alpha) \to_{\alpha\to 0+} 0$. Fix $\varepsilon > 0$ and let s be sufficiently large so that $\varepsilon^2 > 1/s$. Let m be sufficiently large so that $\|T_{s/m}^m(0) - Q_s(0)\| \leq \varepsilon$ and so that the interval $(\frac{m}{\varepsilon s}, m\varepsilon)$ contains an integer k. Then

$$\begin{aligned}\|Q_s(0) - x(\frac{1}{s})\| &\leq \|T_{s/m}^m(0) - x(\frac{1}{s})\| + \varepsilon \\ &\leq 2\frac{ks}{m}\|T(0)\| + sV(\varepsilon) + \varepsilon \\ &\leq s(2\varepsilon + V(\varepsilon)) + \varepsilon.\end{aligned}$$

As $\frac{1}{s}x(\frac{1}{s})$ converges as $s \to \infty$, we deduce that $\frac{1}{s}Q_s(0)$ converges as $s \to \infty$. □

4 Continuous-time non-zero-sum stochastic games

4.1 The discounted case

We say that the strategy profile σ is an *ε-equilibrium*, $\varepsilon \geq 0$, of the ρ-discounted continuous-time stochastic game Γ_ρ, if σ is an admissible strategy profile such that for every player i and every strategy τ^i of player i we have

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

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 5 *Every ρ-discounted continuous-time stochastic game (with finitely many states and actions) Γ_ρ has a profile of stationary strategies that is an equilibrium of Γ_ρ.*

Proof. Define $\|g^i\| := \max_{z\in S}\sup_{a\in A(z)} |g^i(z,a)|$, $J^i(z) = [-\|g^i\|/\rho, \|g^i\|/\rho]$, 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 $v^i(z)$. 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.

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

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

$$G^i_z[v](x) = \sum_{a \in A(z)} G^i_z[v](a) \prod_{j \in N} x^j(z)(a^j)$$

and

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

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 also $f_{z,i}(x,v)$ is a continuous function of (x,v). If $\|v\| \leq \|g^i\|/\rho$ then $|G^i_z[v](a)| \leq \frac{1}{\|\mu\|+\rho}\left(\|g^i\| - \mu(z,z,a)\|g^i\|/\rho + (\|\mu\| + \mu(z,z,a))\|g^i\|/\rho\right) = \|g^i\|/\rho$, and therefore $|G^i_z[v](x)| \leq \|g^i\|/\rho$ and therefore also $|f_{z,i}(x,v)| \leq \|g^i\|/\rho$. 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 uppersemicontinuous. 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 uppersemicontinuous 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 ρ-discounted game with equilibrium payoff V. First we prove that

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

Define $f^z_\sigma(x) = E^z_\sigma(\int_0^x e^{-\rho t} g^i(z_t, x_t)\, dt + e^{-\rho x} V^i(z_x))$. One proves that f^z_σ is Lipschitz and a.e. differentiable with derivative 0 and $f^z_\sigma(0) = V^i(z)$. Next we prove that for every $i \in N$ and every strategy τ^i of player i we have

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

This last part follows along the same lines as the proof of optimality in the zero-sum case. □

Covariance properties. Fix $\alpha, \beta > 0$. A point $(x, V) \in \times_{z\in S, i\in N}(X^i(z) \times [-\|g^i\|/\rho, \|g^i\|/\rho])$ is a stationary equilibrium (strategies and payoffs) of the continuous-time ρ-discounted game $\Gamma = \langle N, S, A, \mu, g\rangle$ if and only if (x, V) is a stationary equilibrium of the continuous-time $\alpha\rho$-discounted game $\Gamma = \langle N, S, A, \alpha\mu, \alpha g\rangle$, and given $0 < \rho < 1$ and $\|\mu\| \le 1 - \rho$, if and only if it is a stationary equilibrium of the discrete-time ρ-discounted 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 $z' \ne z$.

Theorem 6 *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 $v(z)$, $z \in S$,

$$\rho v(z) = \min_{x^{-i}} \max_{x^i} \left(g^i(z, x^{-i} \otimes x^i) + \sum_{z'\in S} \mu(z', z, x^{-i} \otimes x^i) v(z') \right), \qquad (16)$$

where the minimum is over all $x^{-i} = (x^j)_{j\ne i} \in \times_{j\ne i}\Delta(A^j(z))$ and the maximum is over all $x^i \in \Delta(A^i(z))$, has a unique solution $\underline{V}^i_\rho$.
b) The function $\rho \mapsto \underline{V}^i_\rho$ is semalgebraic and $\|\underline{V}^i_\rho\| \le \|g\|/\rho$.
c) There are stationary strategies σ^j, $j \ne i$, such that for every strategy σ^i of player i and every initial state z we have

$$E^z_{\sigma,\sigma^{-i}} \int_0^\infty e^{-\rho t} g^i(z_t, x_t)\, dt \le \underline{V}^i_\rho(z). \qquad (17)$$

Proof. The proof follows the lines of the proof of Theorem 1. 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 $V \in \mathbb{R}^S$ is a solution of $Qv = v$ if and only if it is a solution of equality (16). Let $x^{-i}(h_t) = (x^i(h_t))_{j \neq i}$ be the minimizer of the right-hand side of (16) 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. □

4.2 The undiscounted minmax

Let $V^i_\rho = (V^i_\rho(z))_{z \in S}$ be the (unique) solution 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 min ranges over all $x^{-i} = (x^j)_{j \neq i} \in \times_{j \neq i} \Delta(A^j(z))$, and the max ranges over all $x^i \in X^i(z) := \Delta(A^i(z))$. Set $v^i_\rho = \rho V^i_\rho$. $v^i_\rho(z)$ is a bounded semialgebraic function, and therefore it converges to a limit $v^i(z)$ as $\rho \to 0+$.

Theorem 7 *For every $\varepsilon > 0$ and every player i there are 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*

$$E^z_{\sigma^{-i},\sigma^i} \frac{1}{s} \int_0^s g^i(z_t, x_t)\, dt \leq v^i(z) + \varepsilon$$

and

$$E^z_{\sigma^{-i},\sigma^i} \limsup_{s \to \infty} \frac{1}{s} \int_0^s g^i(z_t, x_t)\, dt \leq v^i(z) + \varepsilon.$$

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), we define 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

$$E^{h_k}_\sigma \left(\int_k^{k+1} e^{-\rho_k (t-k)} g^i_t\, dt + e^{-\rho_k} v^i_{\rho_k}(z_{k+1}) \right) \leq e^{-\rho_k} v^i_{\rho_k}(z_k),$$

where $g^i_t := g^i(z_t, x_t)$, σ is the strategy profile (σ^{-i}, σ^i), and $E^{h_k}_\sigma f$ 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 $|g^i_t| = |g^i(z_t, x_t)| \leq \|g\|$ and $\|v^i_\rho\| := \max_{z\in S} |v^i_\rho(z)| \leq \|g\|$, the inequality $| \int_0^1 |1 - e^{-\rho t}|\rho\, dt + |1 - \rho - e^{-\rho}| \leq O(\rho^2)$ implies that

$$E^{h_k}_\sigma \left(\rho_k r_k + (1-\rho_k) v^i_{\rho_k}(z_{k+1})\right) \leq E^{h_k}_\sigma \left(\int_k^{k+1} e^{-\rho_k(t-k)} g^i_t\, dt + e^{-\rho_k} v^i_{\rho_k}(z_{k+1}) \right) + O(\rho^2).$$

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

$$E^{h_k}_\sigma \left(\rho_k r_k + (1-\rho_k) v^i_{\rho_k}(z_{k+1})\right) \leq v^i_{\rho_k}(z_k) + \varepsilon(\theta)\rho_k$$

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

By the proof in [15] (or using the statement of [22, Lemma 1]), we deduce that there is a sequence of discount rates ρ_k such that the corresponding strategy profile σ^{-i} satisfies, for all s sufficiently large and all strategies σ^i of player i,

$$E^{z_0}_\sigma \frac{1}{s} \int_0^s g^i_t\, dt \leq v^i(z_0) + O(\varepsilon)$$

and

$$E^{z_0}_\sigma \limsup_{s\to\infty} \frac{1}{s} \int_0^s g^i_t\, dt \leq v^i(z_0) + O(\varepsilon).$$

□

4.3 Uniform equilibrium payoffs

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

Given a correlated strategy profile σ that defines for every state $z \in S$ a probability distribution P^z_σ on the plays of the continuous-time stochastic game $\Gamma = \langle N, S, A, \mu, g\rangle$, we set for every player $i \in N$,

$$\gamma^i_s(z, \sigma) := E^z_\sigma \frac{1}{s} \int_0^s g^i_t\, dt,$$

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

where E^z_σ stands for the expectation with respect to P^z_σ.

Theorem 8 *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*

$$-\varepsilon + \gamma^i_s(z, \sigma^{-i}, \tau^i) \le u^i(z) \le \gamma^i_s(z,\sigma) + \varepsilon, \tag{18}$$

and

$$-\varepsilon + \bar{\gamma}^i(z, \sigma^{-i}, \tau^i) \le u^i(z) \le \underline{\gamma}^i(z,\sigma) + \varepsilon. \tag{19}$$

The proof of Theorem 8 is given in the next section. The proof follows a similar outline to that used in Solan and Vieille [31] for the proof of the existence of uniform extensive-form correlated equilibria in discrete-time stochastic games.

5 Proof of Theorem 8

Theorem 8 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 (18) and (19) hold for every $\varepsilon \ge 0$. Therefore, we assume that $\|\mu\| > 0$.

Without loss of generality we can assume that $0 \le g^i \le 1$ and $\|\mu\| = 1$. Indeed, the continuous-time stochastic game $\Gamma = \langle N, S, A, \mu, g\rangle$ has a uniform 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 $0 \le g^i_z(a) \le 1$ and $\|\mu\| = 1$.

5.1 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 that 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 $x^i \in X^i(z) := \Delta(A^i(z))$. 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 $z, z' \in S$, we have

$$v(z) \neq v(z') \implies |w(z) - w(z')| \geq d, \tag{20}$$

and for every player $i_1 \in N$,

$$\big(\; v^i(z) = v^i(z') \quad \forall i < i_1 \text{ and } v^{i_1}(z) > v^{i_1}(z') \;\big) \implies w(z) > w(z'). \tag{21}$$

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 (21), 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 (20) and (21).

5.2 The auxiliary stage strategies, and their basic properties

In this subsection we first select stage strategies $x(z, \overrightarrow{\rho}) \in X(z)$, $z \in S$, 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 $\rho \to 0$. 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 $P^z_y(z_1 \notin C_z) > 0$. We select a stationary strategy $y : z \mapsto y(z) \in X^*(z)$ and a positive number $\delta > 0$ that satisfy

$$P^z_y(z_1 \notin C_z) > 4\delta \quad \forall z \in \bar{S}. \tag{22}$$

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,

$$\begin{aligned}
\rho_i v^i_{\rho_i}(z) &= \min_{x^{-i}} \max_{x^i} \rho_i g^i(z, x^{-i} \otimes x^i) + \sum_{z'\in S} \mu(z', z, x^{-i} \otimes x^i) v^i_{\rho_i}(z') \\
&\leq \max_{x^i} \rho_i g^i(z, x^{-i}(z, \overrightarrow{\rho}) \otimes x^i) + \sum_{z'\in S} \mu(z', z, x^{-i}(z, \overrightarrow{\rho}) \otimes x^i) v^i_{\rho_i}(z') \\
&= \rho_i g^i(z, x(z, \overrightarrow{\rho})) + \sum_{z'\in S} \mu(z', z, x(z, \overrightarrow{\rho})) v^i_{\rho_i}(z') \qquad (23)
\end{aligned}$$

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 $X^*(z)$**.** Let $X^*(z) \subset X(z) := \Delta(A(z))$ be all accumulation points of $(x(z, \overrightarrow{\rho}))$ as $\max_{j\in N} \bar{\rho}_j \to 0$. The set $X^*(z)$ is nonempty and compact. By passing to the limit in inequality (23), for every $z \in S$, $y \in X^*(z)$, and $i \in N$, we have

$$\sum_{z' \in S} \mu(z', z, y) v^i(z') \geq 0. \tag{24}$$

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

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

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

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 disjoint sequence 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

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

By inequality (24), $\bar{S}$ is a proper subset of S. For every $z \in S \setminus \bar{S}$ we select an (arbitrary) element $y(z) \in X^*(z)$. Finally, y is the stationary strategy $y : z \mapsto y(z)$. As $P_y^z(z_1 \notin C_z) > 0$ 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 (24) implies that

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

Definition of $B_1(z)$. For every $z \in S$ let

$$B_1(z) \;:=\; \{a \in A(z) \text{ s.t. } \mu((S \setminus C_z) \cup \bar{S}, z, a) > 0\}.$$

Note that for $z \notin \bar{S}$, $B_1(z)$ 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 $x^*(B_1(z)) = 0$ and such that the norm distance between x and x^*, $\|x - x^*\|$, is equal to $2x[B_1(z)]$.

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

5.3 The strategy τ and the time N_0

For every fixed vector $\overrightarrow{\rho} = (\rho_i)_{i\in N}$ the stationary strategy $\tau : z \mapsto x(z, \overrightarrow{\rho})$ satisfies

$$E^z_\tau \int_0^1 e^{-\rho_i t}\rho_i g^i(z_t, x_t) + e^{-\rho_i} v^i_{\rho_i}(z_1)\, dt \ge v^i_{\rho_i}(z).$$

Therefore, as in the proof of the existence of a value for the undiscounted two-person zero-sum continuous-time stochastic game, for every $\theta > 0$ there is α_0 sufficiently small such that for $\rho_i < \alpha_0$ we have

$$E^z_\tau \int_0^1 \rho_i g^i(z_t, x_t) + (1-\rho_i) v^i_{\rho_i}(z_1)\, dt \ge v^i_{\rho_i}(z) - \theta\rho_i.$$

By the proof in [15], for all $\varepsilon_1 > 0$ and $\alpha_0 > 0$, there are functions $(z_k)_{k=0}^{k=j} \mapsto \overrightarrow{\rho}_j \in (0,\alpha_0)^N$, k, j are nonnegative integers, such that if τ is the profile of strategies with $\tau(h,t) = x(z_t, \overrightarrow{\rho}_{[t]})(z_t)$, where $[t]$ is the largest integer that is $\le t$, then for all $z \in S$ and $i \in N$,

$$E^z_\tau v^i(z_T) \ge v^i(z) - \varepsilon_1/2 \quad \text{for every } (\mathcal{H}_t)_{t\ge 0}\text{-stopping time } T \qquad (25)$$

and

$$\gamma^i_{N_0}(z,\tau) \ge v^i(z) - \varepsilon_1/2, \qquad (26)$$

where N_0 is a positive integer, $z = z_0 \in S$ is the initial state, and $\gamma^i_{N_0}(z,\tau) = \frac{1}{N_0} E^z_\tau \int_0^{N_0} g^{z_t}(x_t)\, dt$.

Note that τ is a $\mathbb{N}$-discretimized strategy profile.

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

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 a stopping time T by $\min\{t : z_t \in (S \setminus C_z) \cup \bar{S}\}$. Set

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

and

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

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 1 *For every $z \in S_1$,*

$$E_\tau^z\left(\int_0^{T\wedge N_0} x(z_t, \overrightarrow{\rho}_t)[B_1(z_t)]\,dt\right) \le 2\varepsilon_2/\xi. \tag{27}$$

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

$$\mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \ge \xi x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z_t)].$$

Therefore (using Corollary 2 in Section 7), if $z \in S_1$, then

$$\begin{aligned} 2\varepsilon_2 &\ge P_\tau^z(T \le N_0) = E_\tau^z\left(\int_0^{T\wedge N_0} \mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]}))\right) dt \\ &\ge E_\tau^z\left(\int_0^{T\wedge N_0} \xi x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z_t)]\,dt\right). \end{aligned}$$

□

5.5 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 $\{x \in X(z_t) : x(B_1(z_t)) = 0\}$ with $\|x^*(z_t, \overrightarrow{\rho}_{[t]}) - x(z_t, \overrightarrow{\rho}_{[t]})\| \leq 2x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z)]$ if $z_t \notin \bar{S}$, and an arbitrary point in $X(z)$ if $z \in \bar{S}$. (Recall that for $z \notin \bar{S}$, $B_1(z)$ is a proper subset of $A(z)$, and therefore $\{x \in X(z) : x(B_1(z)) = 0\}$ is nonempty.) 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 $\sum_{z' \in S} \mu(z', z, x)v(z') = v(z)$.

Lemma 2 *For every* $z \in S_1$,

$$P^z_{\tilde{\tau}}(z_t \in C_z \setminus \bar{S} \ \ \forall 0 \leq t \leq N_0) = 1 \text{ and therefore } P^z_{\tilde{\tau}}(z_{N_0} \in C_z \setminus \bar{S}) = 1,$$

and

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

Proof. The first equality follows from the definition of $\tilde{\tau}$ on $z_0 \in S_1$.

We now prove the inequality. Recall that for $z \notin \bar{S}$,

$$\|x^*(z_t, \overrightarrow{\rho}_{[t]}) - x(z_t, \overrightarrow{\rho}_{[t]})\| \leq 2x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z_t)],$$

and for $z \notin \bar{S}$, Lemma 9 in Section 7 implies that

$$P^z_{\tau}(T \leq N_0) = \int_0^{T \wedge N_0} \mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \, dt.$$

By the definition of ξ and $B_1(z)$, we have

$$\mu((S \setminus C_z) \cup \bar{S}, z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \geq \xi x(z_t, \overrightarrow{\rho}_t)[B_1(z_t)].$$

Therefore, on $z \in S_1$,

$$\begin{aligned} 2\varepsilon_2 &\geq P^z_{\tau}(T \leq N_0) \\ &\geq \xi \int_0^{T \wedge N_0} x(z_t, \overrightarrow{\rho}_t)[B_1(z_t)] \, dt. \end{aligned} \tag{28}$$

In the following sequence of equations, Equality (29) follows from the fact that on the plays that are compatible with the strategy $\tilde{\tau}$ we have $g^i_t = g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]})$. Inequality (30) follows from: (i) the inequality $0 \leq$

$g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \le 1$, (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) the norm distance between the distributions defined by τ and $\tilde{\tau}$ on this state process is less than or equal to $2\varepsilon_2/\xi$. Inequality (31) follows from the inequality $g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]})) \ge g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]})) - \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\|$. Inequality (32) follows from the equality $E^z_\tau \int_0^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \, dt = E^z_\tau \int_0^{T\wedge N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \, dt + E^z_\tau \int_{T\wedge N_0}^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \, dt$ and the inequality $E^z_\tau \int_{T\wedge N_0}^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \, dt \le P^z_\tau(T < N_0) 2N_0$. Finally, Inequality (33) follows from: (i) the inequality $P^z_\tau(T < N_0) \le 2\varepsilon_2$ for $z \in S_1$, (ii) the inequality $\|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \le 2x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z_t)]$, and (iii) the inequality $E^z_\tau \int_0^{T\wedge N_0} 2x(z_t, \overrightarrow{\rho}_{[t]})[B_1(z_t)] \, dt \le 4\varepsilon_2/\xi$.

$$
\begin{aligned}
E^z_{\tilde{\tau}} \int_0^{N_0} g^i_t \, dt &= E^z_{\tilde{\tau}} \int_0^{N_0} g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]})) \, dt && (29) \\
&\ge E^z_\tau \int_0^{N_0} g^i(z_t, x^*(z_t, \overrightarrow{\rho}_{[t]})) \, dt - 2N_0\varepsilon_2/\xi && (30) \\
&\ge E^z_\tau \int_0^{N_0} g^i(z_t, x(z_t, \overrightarrow{\rho}_{[t]})) \, dt - 2N_0\varepsilon_2/\xi && (31) \\
&\quad - E^z_\tau \int_0^{N_0} \|x(z_t, \overrightarrow{\rho}_{[t]}) - x^*(z_t, \overrightarrow{\rho}_{[t]})\| \, dt \\
&\ge E^z_\tau \int_0^{N_0} g^i_t \, dt - 2N_0\varepsilon_2/\xi - P^z_\tau(T < N_0) 2N_0 && (32) \\
&\quad - E^z_\tau \int_0^{T\wedge N_0} \|x(z_t, \overrightarrow{\rho}_t) - x^*(z_t, \overrightarrow{\rho}_t)\| \, dt \\
&\ge E^z_\tau \int_0^{N_0} g^i_t \, dt - 4\varepsilon_2 N_0 - \varepsilon_2/\xi(4 + 2N_0) && (33) \\
&\ge v^i(z) - \varepsilon_1/2 - O(\varepsilon_2).
\end{aligned}
$$

□

5.6 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 $\tilde{\tau}[h_j] : z \mapsto \tilde{\tau}[h_j](z)$.

We discretimize $\tilde{\tau}$ into a pure-action strategy $\bar{\tau}$; ℓ is 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, let $\bar{\tau}$ be a pure-action strategy that conditional on h_j coincides on the time-interval $[j, j+1)$ with a pure-action Markov strategy, denoted $\bar{\tau}[h_j] : (z,t) \mapsto \bar{\tau}[h_j](z,t)$, that is constant on each one of the time-intervals $[j + k/\ell^2, j + (k+1)/\ell^2)$, and such that for every integer $0 \leq i < \ell$ and every state $z \in S$, the number of integers $0 \leq k < \ell$ with $\bar{\tau}[h_j](z, j + i/\ell + k/\ell^2)[a] = 1$, $a \in A(z)$, is $\geq \ell\, \tilde{\tau}[h_j](z)[a] - 1$. In addition, we require that $\bar{\tau}(h_t) \notin B_1(z_t)$ on $z_t \notin \bar{S}$.

It follows that for every state $z \notin \bar{S}$, and all nonnegative integers j, k with $j < N_0$ and $k < \ell$, we have

$$\|\frac{1}{\ell}\tilde{\tau}[h_j](z) - \int_{j+k/\ell}^{j+(k+1)/\ell} \bar{\tau}[h_j](z,t)\,dt\| \leq |A(z)|/\ell^2.$$

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 Corollary 1 in Section 7, conditional on $z_0 \in S_1$, the norm distance between the $\tilde{\tau}$-distribution $P_{\tilde{\tau}}^z(N_0, \ell)$ and the $\bar{\tau}$-distribution $P_{\bar{\tau}}^z(N_0, \ell)$ on $z_{\ell N_0}^\ell$ is bounded by $2N_0/\ell + \ell N_0 |A|/\ell^2$. I.e.,

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

Lemma 3 *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*

$$\gamma_{N_0}^i(z, \bar{\tau}) \geq v^i(z) - \varepsilon_1/2 - O(\varepsilon_2) - O(1/\ell) \geq v^i(z) - \varepsilon_1 - O(\varepsilon_2)\ \ \forall \ell \text{ sufficiently large.}$$

Proof. The first assertion follows from $\bar{\tau}$ being a pure-action strategy and the imposition $\bar{\tau}(h,t) \notin B_1(z_t)$ 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

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

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

$$E^z_{\bar{\tau}}(\int_{k/\ell}^{(k+1)/\ell} g^i_t \, dt \mid z^\ell_k) \geq E^z_{\tilde{\tau}}(\int_{k/\ell}^{(k+1)/\ell} g^i_t \, dt \mid z^\ell_k) - (|A|+1)/\ell^2.$$

Therefore,

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

Using Lemma 2 in Section 5.5 we conclude that

$$\gamma^i_{N_0}(z, \bar{\tau}) \geq v^i(z) - \varepsilon_1/2 - O(\varepsilon_2) \;\; \forall \ell \text{ sufficiently large.}$$

□

The strategy $\hat{\tau}$ plays in blocks of random $\leq N_0$ sizes, and in each block it follows $\bar{\tau}$ as if the game starts 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 $t + j > t$, with j a positive integer and $z_s \notin C_z \setminus \bar{S}$ for some $t < s \leq t + j$, but no later than $t + N_0$.

Let $\bar{t}_k$ be the time at the end of the k-th block ($\bar{t}_0 = 0$), and $\bar{z}_k$ is 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 $\bar{t}_{k+1} = N_0 \wedge \inf\{\lceil s \rceil : s > \bar{t}_k,\ z_s \in (S \setminus C_z) \cup \bar{S}\}$. If $z_{\bar{t}_k} \in \bar{S}$ then $\bar{t}_{k+1} = \bar{t}_k + 1$.

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

5.7 A probabilistic lemma

The following lemma appears implicitly in [31].

Lemma 5 *Assume that* $0 < \varepsilon \leq 1$ *and* $\varepsilon_1 < \delta d^2 \varepsilon / 4$. *Then*

a) *All ergodic classes of the market 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, if $\varepsilon_1 < \frac{\varepsilon d^2 \delta}{16}$, *then*

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

e) $E(w_j \mid z_0) \geq w_0 - \varepsilon$, *and* $E(w_\infty \mid z_0) \geq w_0 - \varepsilon$.

Proof. Let $\mathcal{F}_j$, $j \geq 0$, 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 $\varepsilon \geq \delta(z) \geq 0$ 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

$$E(Y_{j+1} \mid \mathcal{F}_j) \geq Y_j + \delta(z_j) \geq Y_j + \delta(2)\mathbb{I}(\bar{z}_j \in S_2 \cup \bar{S})). \tag{34}$$

As the process $(\bar{z}_j)_{j \geq 0}$ is a stationary Markov chain, it suffices to prove inequality (34) for $j = 0$.

On $\bar{z}_0 \in S_1$ we have $Y_0 = w_0^2$, $w_1 = w_0$ a.e., and $Y_1 \geq w_1^2 = Y_0 + \delta(z_0)$ a.e. Therefore, inequality (34) 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 $|w_1 - w_0| \geq d$ is $\geq 3\delta$. Therefore, $E(w_1^2 \mid z_0) - w_0^2 \geq 3\delta d^2$, implying that $E(Y_1 \mid z_0) \geq w_0^2 + 3\delta d^2 = Y_0 + 2\delta(0) \geq Y_0 + \delta(0)$.

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

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

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

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

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

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

This completes the proof of Inequality (34).

Inequality (34) 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 (34), 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 parts b and 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 $j \geq 0$ we have

$$E(w_{j+1} \mid \mathcal{F}_j) \geq w_j - \varepsilon_1 \mathbb{I}(z_j \in S_2).$$

Summing these inequalities over all $0 \leq j < \ell$, we deduce that for every ℓ we have

$$E(w_\ell \mid z_0) \geq w(z_0) - \varepsilon_1 E \sum_{0 \leq j < \ell} \mathbb{I}(z_j \in S_2) \geq w(z_0) - \varepsilon,$$

and therefore

$$E(w_\infty \mid z_0) \geq w(z_0) - \varepsilon.$$

This completes the proof of part e.

□

5.8 The uniform 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

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

Therefore, for every $s \geq 0$ we have

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

The strategy σ follows $\bar{\tau}$ until the first time $t = j/\ell$, 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 e be the maximum over $z \in S$ of the expectation with respect to $P^z_{\bar\tau}$ of $|\{k < \infty : \bar z_k \notin S_1\}|$. It follows (using Markov's inequality) that for every $s \geq 0$ we have

$$\frac{N_0}{s} \sum_{k \geq 0}^{[s/N_0]+\ell+1} E^z_\sigma \gamma^i_{N_0}(\bar z_k, \bar\tau) + \frac{e}{s\ell} \geq \gamma^i_s(z,\sigma) \quad \geq \quad \frac{N_0}{s} \sum_{k=0}^{[s/N_0]-1} E^z_\sigma \gamma^i_{N_0}(\bar z_k, \bar\tau) - \frac{e}{s\ell}$$

Therefore,

$$\gamma^i_s(z,\sigma) \to_{s\to\infty} u^i(z) := \lim_{\ell\to\infty} \frac{1}{\ell} \sum_{k=0}^{\ell} E^z_\sigma \gamma^i_{N_0}(\bar z_k, \bar\tau),$$

where the limit exists as $(\bar z_k)_{k\geq 0}$ is a homogeneous Markov process.

Let $\tilde\sigma^i$ be a strategy of player i and let T be the smallest time j/ℓ such that player i has deviated from his 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 $E^z_{\tilde\sigma} v^i(z_T) \leq E^z_\sigma v^i(z_T) + O(1/\ell)$. Let ℓ be sufficiently large so that $E^z_{\tilde\sigma} v^i(z_T) \leq E^z_\sigma v^i(z_T) + \varepsilon$. It follows that

$$\begin{aligned}
\gamma^i_s(z, \sigma^{-i}, \tilde\sigma^i) &= \frac{1}{s} E^z_{\tilde\sigma} \int_0^s g^i_t\, dt \\
&= \frac{1}{s} E^z_{\tilde\sigma} \int_0^{T\wedge s} g^i_t\, dt + \frac{1}{s} E^z_{\tilde\sigma} \int_{T\wedge s}^s g^i_t\, dt \\
&\leq \frac{1}{s} E^z_\sigma \int_0^{T\wedge s} g^i_t\, dt + \frac{1}{\ell s} + \frac{1}{s} E^z_{\tilde\sigma} (s-T)(v^i(z_T) + \varepsilon + \frac{s_\varepsilon}{s}) \\
&\leq \frac{1}{s} E^z_\sigma \int_0^{T\wedge s} g^i_t\, dt + \frac{1}{\ell s} + \frac{1}{s} E^z_\sigma (s-T)(v^i(z_T) + 2\varepsilon + \frac{s_\varepsilon}{s}) \\
&\leq \frac{1}{s} E^z_\sigma \int_0^{T\wedge s} g^i_t\, dt + \frac{1}{\ell s} + E^z_\sigma \int_{T\wedge s}^s g^i_t\, dt + 3\varepsilon + \frac{s_\varepsilon}{s} \\
&\leq \gamma^i_s(z,\sigma) + 4\varepsilon \quad \text{for all } s \geq \frac{2s_\varepsilon + 2/\ell}{\varepsilon}.
\end{aligned}$$

Similarly, the limit of $\frac{1}{s}\int_0^s g^i_t\, dt$ as $s \to \infty$ exists a.e. with respect to P^z_σ, and $E^z_\sigma \lim_{s\to\infty} \frac{1}{s}\int_0^s g^i_t\, dt = u^i(z) \leq -4\varepsilon + E^z_{\tilde\sigma} \limsup_{s\to\infty} \frac{1}{s}\int_0^s g^i_t\, dt$.

This completes the proof of Theorem 8.

□

6 Continuous-time stochastic games as limits of discrete-time stochastic games

We consider a family of discrete-time stochastic games $\Gamma_\delta = \langle N, S, A, p_\delta, g_\delta \rangle$, where the positive parameter $\delta > 0$ represents the stage-duration. The sets of players N, states S, and actions A are independent of δ, and the transition probabilities p_δ and the payoff function g_δ depend on the parameter δ. We study the "convergence" of the family $(\Gamma_\delta)_{\delta>0}$, and the presentation of the "limit" as a continuous-time stochastic game Γ.

First, we define the "convergence" of the family $(\Gamma_\delta)_{\delta>0}$ as the convergence of its data. We say that Γ_δ *converges in data* as $\delta \to 0$ if its data, g_δ and p_δ, converge (in the proper scale) to g and μ respectively; i.e., if for every pair of distinct states $z' \neq z$ and an action profile $a \in A(z)$, we have $\frac{p_\delta(z'|z,a)}{\delta} \underset{\delta \to 0+}{\longrightarrow} \mu(z', z, a)$ and $\frac{g_\delta(z,a)}{\delta} \underset{\delta \to 0+}{\longrightarrow} g(z, a)$, where $(z', z, a) \mapsto \mu(z', z, a) \in \mathbb{R}_+$ and $(z, a) \mapsto g(z, a) \in \mathbb{R}^N$ are functions that are defined for all distinct states $z' \neq z$ and action profiles $a \in A(z)$.

Next, we wish to define the "convergence" of the family $(\Gamma_\delta)_{\delta>0}$ as a convergence of the stochastic process of states and payoffs that is defined by the initial state and a strategy σ as $\delta \to 0+$. Obviously, in defining the convergence of the stochastic process of states and payoffs one has to take into account the stage duration δ. The state z_n in the play of the discrete-time stochastic game Γ_δ is interpreted as the state at time $n\delta$. Similarly, the sum $\sum_{j=0}^{n-1} g_\delta(z_j, a_j)$ of stage payoffs in stages $0 \leq j < n$, is interpreted as the cumulative payoff in the time interval $[0, n\delta]$.

In addition, a non-stationary strategy σ may represent completely different rules of behavior in the different games Γ_δ. For example, if σ is a Markov strategy, it depends on the stage, and stage n of the discrete-time game Γ_δ corresponds to time $n\delta$, which depends on the parameter δ. Therefore, we first define the convergence by imposing a condition on the dynamics defined by a stationary strategy. We say that Γ_δ *converges in stationary dynamics* if for all pure stationary strategies σ, states $z', z \in S$, times $t \geq 0$, and positive integers n_δ such that $n_\delta \delta \underset{\delta \to 0+}{\longrightarrow} t$, we have

$$P^z_{\delta,\sigma}(z_{n_\delta} = z') \underset{\delta \to 0+}{\longrightarrow} F^\sigma_{z,z'}(t)$$

and

$$E^z_{\delta,\sigma}\sum_{j=0}^{n_\delta} g_\delta(z_j,a_j) \underset{\delta\to 0+}{\longrightarrow} G_t(z,\sigma),$$

where $(\sigma,z',z,t)\mapsto F^\sigma_{z,z'}(t)\in\mathbb{R}$ and $(t,z,\sigma)\mapsto G_t(z,\sigma)\in\mathbb{R}^N$ are functions that are defined for all pure stationary strategies σ, states $z',z\in S$, and times $t\geq 0$.

The following proposition shows that the two convergence conditions are equivalent.

Proposition 5 *The following conditions are equivalent:*
(A) $(\Gamma_\delta)_{\delta>0}$ *converges in stationary dynamics.*
(B) $(\Gamma_\delta)_{\delta>0}$ *converges in data.*

Proof. (A) $\implies$ (B). Assume condition (A) holds. Obviously, $\sum_{z'\in S} P^z_{\delta,\sigma}(z_{n_\delta}=z')=1$. Therefore, $\sum_{z'\in S} F^\sigma_{z,z'}(t)=1$. Applying condition (A) to $n_\delta=0$ and $z'=z$, we have $F_{z,z}(0)=1$. Applying condition (A) to $t=0$ and all nonnegative integers n_δ with $\delta n_\delta \underset{\delta\to 0+}{\longrightarrow} 0$, we deduce that for every $\varepsilon>0$ there are $t_\varepsilon>0$ and $\delta_\varepsilon>0$ such that for every $0<\delta<\delta_\varepsilon$ and n with $n\delta\leq t_\varepsilon$, we have, $P^z_{\delta,\sigma}(z_n=z)>1-\varepsilon$ for all states $z\in S$ and pure stationary strategy profiles σ.

Fix $z\in S$ and $a\in A(z)$, set $K_\delta=\sum_{z'\neq z} p_\delta(z'\mid z,a)$, and let σ be a pure stationary strategy with $\sigma(z)=a$, and $n=n_\delta=[t_{1/3}/\delta]$ (where $[*]$ denotes the largest integer that is less than or equal to $*$). Then, for $\delta<\delta_{1/3}$, $1/3>P^z_{\delta,\sigma}(z_n\neq z)\geq\sum_{m=1}^n P^z_{\delta,\sigma}(\forall j<m\ z_j=z \text{ and } z\neq z_m=z_n)\geq \sum_{m=1}^n(1-K_\delta)^{m-1}K_\delta 2/3=(1-(1-K_\delta)^n)2/3$, which implies the inequality $(1-K_\delta)^n\geq 1/2$. Therefore, $\limsup_{\delta\to 0+}K_\delta/\delta<\infty$. Therefore, there is a positive constant K such that for all $\delta>0$, $z\in S$, and $a\in A(z)$, we have $\sum_{z'\neq z} p_\delta(z'\mid z,a)<K\delta$.

Next, we prove that if, for a pair of distinct states $z'\neq z$ and an action profile $a\in A(z)$ we have $\liminf_{\delta\to 0+} p_\delta(z'\mid z,a)/\delta<c$, then, for $t>0$ sufficiently small and a stationary strategy σ with $\sigma(z)=a$, we have $F^\sigma_{z,z'}(t)<ct$. Indeed, the set $\{z_n=z',\ z_0=z\}$ is the union of the disjoint sets $Y_{m,z''}=\{\forall 0\leq j<m,\ \ z_j=z_0, z_m=z'' \text{ and } z_n=z'\}$, where m ranges over the positive integers $1<m\leq n$ and z'' ranges over all states $z''\neq z$. Let $\varepsilon>0$ and set $n=n_\delta=[t_\varepsilon/\delta]$. Note that $P^z_{\delta,\sigma}(Y_{m,z''})\leq p_\delta(z'\mid z,a)$ for $z''=z'$ and $\sum_{m=1}^{n-1}\sum_{z\neq z''\neq z'} P^z_{\delta,\sigma}(Y_{m,z''})\leq\varepsilon K\delta n$ for δ sufficiently small. Therefore, if $\delta>0$ is sufficiently small so that, in addition, $p_\delta(z'\mid z,a)/\delta<c$

and for all $z'' \neq z$ we have $p_\delta(z'' \mid z_0 = z) \leq K\delta$, then $P^z_{\delta,\sigma}(z_n = z') \leq \sum_{m=1}^{n} P^z_{\delta,\sigma}(Y_{m,z'}) + \varepsilon K\delta n \leq (c + K\varepsilon)\delta n$. Therefore for $t > 0$ sufficiently small we have $F^\sigma_{z,z'}(t) < ct$.

Finally, we prove that if, for a pair of distinct states $z' \neq z$ and an action profile $a \in A(z)$ we have $\limsup_{\delta\to 0+} p_\delta(z' \mid z, a)/\delta > c$, then, for $t > 0$ sufficiently small and a stationary strategy σ with $\sigma(z) = a$, we have $F^\sigma_{z,'z}(t) > ct$. Indeed, the set $\{z_n = z',\ z_0 = z\}$ contains the disjoint sets $Y_{m,z'} = \{\forall 0 \leq j < m, \quad z_j = z_0, z_m = z' = z_n\}$, where m ranges over the positive integers $1 < m \leq n$. Let $\varepsilon > 0$ and set $n = n_\delta = [t_\varepsilon/\delta]$. Note that $P^z_{\delta,\sigma}(Y_{m,z'}) \geq (1-\varepsilon)^2 p_\delta(z' \mid z, a)$ for δ sufficiently small. Therefore, if $\delta > 0$ is sufficiently small so that, in addition, $p_\delta(z' \mid z, a)/\delta > c$, then $P^z_{\delta,\sigma}(z_n = z') \geq \sum_{m=1}^{n} P^z_{\delta,\sigma}(Y_{m,z'}) \geq n(1-\varepsilon)^2\delta c$. Therefore for $t > 0$ sufficiently small we have $F^\sigma_{z,z'}(t) > ct$.

We conclude that the $\limsup_{\delta\to 0+} p_\delta(z' \mid z, a)/\delta$ and the $\liminf_{\delta\to 0+} p_\delta(z' \mid z, a)/\delta$ coincide.

We will now prove that the second part of (B) holds. Fix a player $i \in N$ and assume that $\limsup_{\delta\to 0+} \|g^i_\delta\|/\delta < \infty$, where $\|g^i_\delta\| := \max_{z,a} |g^i_\delta(z, a)|$. For $t > 0$ let $\gamma_t(z, \sigma) = \frac{1}{t} G_t(z, \sigma)$. Then, for $\delta > 0$ sufficiently small, $g^i_\delta(z, \sigma(z))/\delta - 2\varepsilon\|g^i_\delta\|/\delta \leq \gamma^i_{t_\varepsilon}(z, \sigma) + \varepsilon$. Therefore

$$\limsup_{\delta\to 0+} g^i_\delta(z, \sigma(z))/\delta \leq \gamma^i_{t_\varepsilon}(z, \sigma) + \varepsilon + 2\varepsilon \limsup_{\delta\to 0+} \|g^i_\delta\|/\delta,$$

and therefore

$$\limsup_{\delta\to 0+} g^i_\delta(z, \sigma(z))/\delta \leq \liminf_{\varepsilon\to 0+} \gamma^i_{t_\varepsilon}(z, \sigma).$$

Similarly, for $\delta > 0$ sufficiently small, $\gamma^i_{t_\varepsilon}(z, \sigma) - \varepsilon \leq g^i_\delta(z, \sigma(z))/\delta + 2\varepsilon\|g^i_\delta\|/\delta$, and therefore $\limsup_{\varepsilon\to 0+} \gamma^z_{t_\varepsilon}(z, \sigma) \leq \liminf_{\delta\to 0+} g^i_\delta(z, \sigma(z))/\delta$. Given $a \in A(z)$ and applying these inequalities to a stationary strategy σ with $\sigma(z) = a$ we conclude that the $\liminf_{\delta\to 0+} g^i_\delta(z, a)/\delta$ and the $\limsup_{\varepsilon\to 0+} g^i_\delta(z, a)/\delta$ coincide.

Now, we prove that condition (A) implies that $\limsup_{\delta\to 0+} \|g^i_\delta\|/\delta < \infty$. For every $1 > \delta > 0$ let $z_\delta \in S$ and $a_\delta \in A(z)$ be such that $|g^i_\delta(z_\delta, a_\delta)| = \|g^i_\delta\|$. Let $\varepsilon > 0$, and let $\sigma = \sigma_\delta$ be a stationary strategy with $\sigma(z_\delta) = a_\delta$. Set $n = n_\delta = [t_\varepsilon/\delta]$ and $z_0 = z_\delta$. If $g^i(z_\delta, a_\delta) \geq 0$, then, for sufficiently small $\delta > 0$, we have $G^i_{t_\varepsilon}(z_\delta, \sigma) + t_\varepsilon/3 \geq E^{z_\delta}_\sigma \sum_{j=0}^{n-1} g^i_\delta(z_j, a_j) \geq (1-2\varepsilon) n g^i_\delta(z_\delta, a_\delta)$. Therefore, if $\varepsilon < 1/3$ we have $g^i_\delta(z_\delta, a_\delta)/\delta \leq 3|\gamma^i_{t_\varepsilon}(z, \sigma)| + 1$ for $\delta > 0$ sufficiently small. If $g^i(z_\delta, a_\delta) < 0$, then, for sufficiently small $\delta > 0$, we have $G^i_{t_\varepsilon}(z_\delta, \sigma) - t_\varepsilon/3 \leq E^{z_\delta}_\sigma \sum_{j=0}^{n-1} g^i_\delta(z_j, a_j) \leq (1-2\varepsilon) n g^i_\delta(z_\delta, a_\delta)$. Therefore, if $\varepsilon < 1/3$ we have $g^i_\delta(z_\delta,a_\delta)/\delta \geq -3|\gamma^i_{t_\varepsilon}(z,\sigma)|-1$. This proves that $\limsup_{\delta\to 0+}\|g^i_\delta\|/\delta \leq 3|\gamma^i_{t_\varepsilon}(z,\sigma)|+1<\infty$.

(B) $\implies$ (A). Define $\mu(z,z,a) = -\sum_{z'\neq z}\mu(z',z,a)$. Let σ be a stationary strategy and let Q be the $S\times S$ matrix, whose (z,z')-th entry is $Q_{z,z'} = \mu(z',z,\sigma(z))$. Note that for $\delta>0$ sufficiently small, $I+\delta Q$ is a transition matrix, where I stands for the identity matrix, and $\|I+\delta Q\| := \max_{z\in S}\sum_{z'\in S}|(I+\delta Q)_{z,z'}| = 1$. In addition, $e^{\delta Q}$ (which equals by definition the convergent sum $\sum_{j=0}^\infty \frac{\delta^j Q^j}{j!}$) is an $S\times S$ matrix, and $(e^{\delta Q})^n = e^{n\delta Q}$. Let P_δ be the $S\times S$ transition matrix, whose (z,z')-th entry is $(P_\delta)_{z,z'} = p_\delta(z'\mid z,\sigma(z))$. Therefore, if n is a positive integer, then $P^z_{\delta,\sigma}(z_n=z') = (P^n_\delta)_{z,z'}$. By the assumption on p_δ and the definitions of Q and $e^{\delta Q}$, we have $\|e^{\delta Q}-P_\delta\|\leq o(\delta)$ and $\|e^{\delta Q}-I-\delta Q\|\leq O(\delta^2)$.

For any two $S\times S$ matrices (or elements of a norm algebra) A and B we have $A^n-B^n = \sum_{k=1}^n A^{n-k}(A-B)B^{k-1}$, implying that $\|A^n-B^n\|\leq \|A-B\|\sum_{j=0}^{n-1}\|A\|^j\|B\|^{n-j}$. Therefore, $\|P^n_\delta - e^{n\delta Q}\|\leq \|P_\delta - e^{\delta Q}\|\sum_{j=0}^{n-1}\|e^{\delta Q}\|^j$. For $\delta>0$ sufficiently small, we have $\|e^{\delta Q}\|\leq 1+2\delta\|Q\|$, implying that $\|Q\|\|\sum_{j=0}^{n-1}\|e^{\delta Q}\|^j \leq e^{2n\delta\|Q\|}/(2\delta)$.

Therefore, $\|P^n_\delta - e^{n\delta Q}\|\leq o(1)e^{2n\delta\|Q\|}$. Therefore, $\|P^n_\delta - e^{tQ}\|\leq \|P^n_\delta - e^{n\delta Q}\| + \|e^{tQ}-e^{n\delta Q}\|\to 0$ as $n\delta\to t$. We conclude that $P^z_{\delta,\sigma}(z_n=z')\to F^\sigma_{z,z'}(t) = (e^{tQ})_{z,z'}\in\mathbb{R}$ as $\delta\to 0+$.

By assumption (B) we have $g_\delta(z,a) = \delta g(z,a)+o(\delta)$. Therefore, if $\delta\to 0+$ and $n_\delta\delta\to t>0$, then $|E^z_{\delta,\sigma}\sum_{j=0}^{n_\delta-1} g^i_\delta(z_j,a_j) - E^z_{\delta,\sigma}\sum_{j=0}^{n_\delta-1}\delta g^i(z_j,a_j)|\to 0$. If $\delta\to 0+$ and $n_\delta\delta\to t>0$, then, as shown earlier, $P^z_{\delta,\sigma}(z_n=z')\to F^\sigma_{z,z'}(t)$, and, therefore, $E^z_{\delta,\sigma}\sum_{j=0}^{n_\delta-1}\delta g^i(z_j,a_j)\to G_t(z,\sigma) = \int_0^t\sum_{z'\in S}F^\sigma_{z,z'}(s)g(z',\sigma(z'))\,ds$. Therefore, $E^z_{\delta,\sigma}\sum_{j=0}^{n_\delta-1} g^i_\delta(z_j,a_j)\to G_t(z,\sigma)$ as $\delta\to 0+$ and $n_\delta\delta\to t>0$. □

Remark that every continuous-time stochastic game $\Gamma = \langle N,S,A,\mu,g\rangle$ is a "data limit" of the discrete-time stochastic games $\Gamma_\delta = \langle N,S,A,p_\delta,g_\delta\rangle$, where $g_\delta(z,a) = \delta g(z,a)$ and $p_\delta(z'\mid z,a) = \delta\mu(z',z,a)$ for all pairs of distinct states $z'\neq z$ and action profile $a\in A(z)$.

The next proposition gives a sufficient condition for a family of Markov strategies σ_δ in Γ_δ to have a continuous-time limiting dynamics and payoffs as $\delta\to 0+$. In the formulas that follow, we view $\sigma_\delta(z,j)$ as a measure on $A(z)$; i.e., $\sigma_\delta(z,j)\in\Delta(A(z))$, and $\sigma_\delta(j) := (\sigma_\delta(z,j))_{z\in S}$ is an element of $\times_{z\in S}\Delta(A(z))$. Therefore, for any fixed $z\in S$, any linear combination of $\sigma_\delta(z,j)$ is a measure on $A(z)$. Similarly, if $\sigma: S\times\mathbb{R}_+\to\Delta(A)$ is measurable with $\sigma(z,t)\in\Delta(A(z))$, then, for any function $f\in L_1(\mathbb{R}_+)$, the integral

$\int_0^\infty f(t)\sigma(z,t)\,dt$ is well defined.

We say that the Markov strategies σ_δ in Γ_δ *converge* w* if for every continuous function $f:\mathbb{R}_+\to\mathbb{R}$ with bounded support the limit of $\sum_{j=0}^\infty f(j\delta)\delta\sigma_\delta(z,j)$ as $\delta\to 0+$ exists. In that case, there is a measurable function $\sigma: S\times\mathbb{R}_+\to\Delta(A)$ (with $\sigma(z,t)\in\Delta(A(z))$) such that for every $f\in L_1(\mathbb{R}_+)$ the limit of $\int_0^\infty f(t)\sigma_\delta(z,[t/\delta])\,dt$ as $\delta\to 0+$ exists and equals $\int_0^\infty f(t)\sigma(z,t)\,dt$, and we say that the discrete-time Markov strategies σ_δ *converge* w* *to* $\sigma: S\times\mathbb{R}_+\in\Delta(A)$. Whenever the conditional probability $P^{z_0}_{\delta,\sigma}(E_1\mid E_2)$ is independent of the initial state z_0, we suppress the superscript of the initial state z_0.

Proposition 6 *If the Markov strategies σ_δ in Γ_δ converge w* to $\sigma: S\times\mathbb{R}_+\to\Delta(A)$ and the family of discrete-time stochastic games Γ_δ obeys condition (A), then, for every $0\le s<t$, there are $S\times S$ transition matrices $F^\sigma(s,t)$ such that*

$$P_{\delta,\sigma_\delta}(z_n=z'\mid z_k=z)\to F^\sigma_{z,z'}(s,t)\ \textit{as}\ \delta\to 0+, k\delta\to s,\ \textit{and}\ n\delta\to t.$$

Proof. As the family of discrete-time stochastic games $(\Gamma_\delta)_{\delta>0}$ obeys condition (A), it satisfies condition (B). Therefore, there is a positive constant $K>0$ such that for every $z\in S$ we have $p_\delta(z\mid z,*)>1-K\delta$. Therefore, if $0\le n<k$, $|P_{\delta,\sigma_\delta}(z_n=z'\mid z_k=z)-I_{z,z'}|<1-(1-K\delta)^{k-n}\to 0$ as $k\delta-n\delta\to 0+$. Therefore, it suffices to prove that for every $s<t$ there are sequences $k_\delta<n_\delta$ such that $k_\delta\delta\to s$ and $n_\delta\delta\to t$ such that

$$P_{\delta,\sigma_\delta}(z_{n_\delta}=z'\mid z_{k_\delta}=z)\to F^\sigma_{z,z'}(s,t)\ \text{as}\ \delta\to 0+\,.$$

We will prove it for $n_\delta=[t/\delta]$ and $k_\delta=[s/\delta]+1$.

Assume that the Markov strategies σ_δ in Γ_δ converge w* to $\sigma: S\times\mathbb{R}_+\to\Delta(A)$. Let M be the space of all $S\times S$ matrices Q, let M_0 be the subset of all its matrices Q with $\sum_{z'\in S}Q_{z,z'}=0$ for every $z\in S$ and $Q_{z,z'}\ge 0$ for all $z\ne z'$, and let M_1 be the subset of M of all transition matrices. The identity matrix is denoted 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 M_1 is closed under multiplication. For an ordered list $F_1,\dots,F_j\in M$ we denote by $\prod_{i=1}^j F_i$ the matrix (ordered) product $F_1F_2\dots F_j$.

Let $Q:[0,\infty)\to M$ be defined by $Q_{z,z'}(u)=\mu(z',z,\sigma(z,u))$, and let $Q^\delta:[0,\infty)\to M$ be defined by $Q^\delta_{z,z'}(u)=(p_\delta(z'\mid z,\sigma_\delta(z,[u/\delta]))-I_{z,z'})/\delta$. As $(\Gamma_\delta)_{\delta>0}$ obeys condition (B), $Q^\delta_{z,z'}(u)=\mu(z',z,\sigma_\delta(z,[u/\delta]))+o(1)$ as $\delta\to 0+$. Therefore, $\int_s^t Q^\delta_{z,z'}(u)\,du=\mu(z',z,\int_s^t\sigma_\delta(z,[u/\delta]))\,du+o(1)$ as $\delta\to 0+$, where for a measure α on $A(z)$ we define $\mu(z', z, \alpha) := \sum_{a \in A(z)} \alpha(a)\mu(z', z, a)$. Therefore, as the Markov strategies σ_δ converge w* to σ, for every $s < t$ we have

$$\int_s^t Q^\delta(u)\, du \underset{\delta \to 0+}{\longrightarrow} \int_s^t Q(u)\, du.$$

Let G_j^δ be the transition matrix $(G_j^\delta)_{z,z'} = p_\delta(z' \mid z, \sigma_\delta(z, j))$, and given $0 \le s \le t$ we define $G^\delta(s,t)$ to be the transition matrix $\prod_{j=[s/\delta]+1}^{[t/\delta]-1} G_j^\delta$. It suffices to prove that $G^\delta(s,t)$ converges as $\delta \to 0+$.

Let $C = 2 \max_{z,a} |\mu(z,z,a)| < C'$. It follows that for every $t \ge 0$ we have $\|Q(t)\| \le C$, and for sufficiently small $\delta > 0$ we have $\|Q^\delta(t)\| < C'$.

Let $L_\delta(s,t) = [t/\delta] - [s/\delta] - 1$. Note that $t - s - 2\delta \le L_\delta(s,t)\delta \le t - s$. Using the triangle inequality and the fact that M is a Banach algebra, for all $0 \le t < s$, we have

$$\begin{aligned}
\|G^\delta(s,t) - I - \int_s^t Q^\delta(u)\, du\| &< \|G^\delta(s,t) - I - \int_{\delta\lceil s/\delta \rceil}^{[t/\delta]} Q^\delta(u)\, du\| + 2\delta C' \\
&\le 2\delta C' + \sum_{\ell=2}^{L_\delta(s,t)} \binom{L_\delta(s,t)}{\ell} (C'\delta)^\ell \\
&= 2\delta C' + (1 + C'\delta)^{L_\delta(s,t)} - 1 - L_\delta(s,t) C'\delta \\
&\le 4\delta C' + e^{(t-s)C'} - 1 - (t-s)C' \\
&\le (t-s)^2 C'^2
\end{aligned}$$

for $(t-s)C' \le 1$ and $\delta > 0$ sufficiently small. Therefore, for sufficiently small $\delta > 0$, we have

$$\|G^\delta(s,t) - I - \int_s^t Q^\delta(u)\, du\| \le (t-s)^2 C'^2.$$

For every sequence $s = t_0 < t_1 < \ldots < t_k = t$, set $A_j = G^\delta(t_{j-1}, t_j)$, $j = 1, \ldots, k$, $B_j = G^\delta_{[t_j/\delta]}$ for $j = 1, \ldots, k-1$, and $B_k = I$. Note that $G^\delta(t_0, t) = \prod_{j=1}^k (A_j B_j)$ and $\prod_{j=1}^k A_j - \prod_{j=1}^k (A_j B_j) = \sum_{i=1}^k (\prod_{j=1}^{i-1} A_j B_j) A_i (I - B_i) \prod_{j=i+1}^k A_j$, where a product over an empty set of indices is defined as the identity. Therefore, for a sufficiently large k, by setting $t_j = s + j(t-s)/k$ and $C_k = C' - 1/k$, there is $\delta(k) > 0$ (sufficiently small) such that for $0 < \delta < \delta(k)$, we have $\|I - B_j\| \le \frac{(t-s)^2 C_k^2}{k^2}$ (thus $\|B_j\| \le 1 + \frac{(t-s)^2 C_k^2}{k^2}$), and $\|A_j\| \le 1 + \frac{(t-s)^2 C_k^2}{k^2}$.

Therefore,

$$\|G^\delta(s,t)-\prod_{j=1}^k G^\delta(t_{j-1},t_j)\| \le \sum_{j=1}^k \left(1+\frac{(t-s)^2C_k^2}{k^2}\right)^{2k} \frac{(t-s)^2C_k^2}{k^2} \le 2(t-s)^2C_k^2/k.$$

Setting $F(t_{j-1},t_j) := I + \int_{t_{j-1}}^{t_j} Q(u)\,du$, for k sufficiently large and $\delta < \delta(k)$, we have

$$\|G^\delta(s,t) - \prod_{j=1}^k F(t_{j-1},t_j)\| \le 2(t-s)^2C'^2/k.$$

Therefore, $\sup_{0<\delta,\delta'<\delta(k)} \|G^\delta(s,t) - G^{\delta'}(s,t)\| \le 4(t-s)^2C'^2/k$, implying that $\lim_{k\to\infty}\sup_{0<\delta,\delta'<\delta(k)} \|G^\delta(s,t) - G^{\delta'}(s,t)\| = 0$. Therefore, $G^\delta(s,t)$ converges to a limit as $\delta \to 0+$. □

7 The continuous-time process

The proof of Proposition 6 implies the following classical[7] result on non-homogeneous continuous-time Markov processes.

Theorem 9 *Let $Q : [0,\infty) \to M$ (respectively, $\to M_0$) be measurable, with $\|Q(t)\| \le C$. Then there is a unique family $F(s,t) = F^Q(s,t)$, $t \ge s \ge 0$, of matrices in M (respectively, in M_1) such that for all $0 \le s \le t_1 \le t$ and $\delta > 0$ we have*

$$F(s,s) = I \tag{35}$$

$$\|F(t,t+\delta) - I - \int_t^{t+\delta} Q(x)\,dx)\| \le o(\delta) \tag{36}$$

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

$$F(s,t) = F(s,t_1)F(t_1,t). \tag{37}$$

This unique family satisfies

$$\frac{\partial F}{\partial t}(s,t) = F(s,t)Q(t) \quad \textit{for a.e. } t \tag{38}$$

[7]For a proof in the case where $Q(t)$ is continuous in t (and excluding inequality (39) and the w^* continuity), see, e.g., [25]. Variants of this result are stated in, e.g., [19, 14].

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

$$\|F(t,t+\delta) - I - \int_t^{t+\delta} Q(x)\,dx)\| \leq C\delta^2. \tag{39}$$

The implicitly defined map $Q \mapsto F(s,t)$ is w^* continuous.

An alternative proof of existence of transition matrices $F(s,t)$ is to construct a probability distribution on S-valued right-continuous functions on $[0,\infty)$ and prove that the transition matrices of this process satisfy the conditions of the theorem. This stochastic process is derived by defining the distribution of the random times $0 < t_1 < t_2 < \ldots$ of state changes together with the sequence of states z_{t_j}. This distribution P is given by defining the conditional probabilities

$$P(t_{j+1} \geq t_j + s \mid t_j, z = z_{t_j}) = e^{\int_{t_j}^{t_j+s} Q_{z,z}(u)\,du}$$

and, given t_j, z_{t_j}, for almost all t_{j+1},

$$P(z_{t_{j+1}} = z' \mid t_j, z = z_{t_j}, t_{j+1}) = -Q_{z,z'}(t_{j+1})/Q_{z,z}(t_{j+1}).$$

It remains to prove that a probability distribution on the right-continuous state processes is well defined by these conditional distributions and that the implied transition matrices satisfy (36), (38), and (39).

Lemma 6 *Assume that $S_0, S_1, \ldots$ are finite sets; $F(j)$ and $G(j)$, $j \geq 1$, are $S_{j-1} \times S_j$ transition matrices; $q, r \in \Delta(S_0)$; and q_k and r_k are the probability distributions on $\times_{j=0}^k S_j$ given by*

$$q_k(z_0, z_1, \ldots, z_k) = q(z_0) \prod_{j=1}^k F_{z_{j-1},z_j}(j)$$

and

$$r_k(z_0, z_1, \ldots, z_k) = r(z_0) \prod_{j=1}^k G_{z_{j-1},z_j}(j).$$

Then

$$\|q_k - r_k\| \leq \|q - r\| + E_{q_k} \sum_{j=1}^k \|F_{z_{j-1}}(j) - G_{z_{j-1}}(j)\|, \tag{40}$$

where $\|F_{z_{j-1}}(j)-G_{z_{j-1}}(j)\| := \sum_{z\in S_j} |F_{z_{j-1},z}(j)-G_{z_{j-1},z}(j)|$, *and (therefore)*

$$\|q_k - r_k\| \leq \|q-r\| + \sum_{j=1}^{k} \|F(j)-G(j)\|. \tag{41}$$

Proof. As $E_{q_k}\|F_{z_{j-1}}(j)-G_{z_{j-1}}(j)\| \leq \|F(j)-G(j)\|$, inequality (40) implies inequality (41). We prove inequality (40) by induction on k. For $k=1$,

$$\sum_{z_0,z_1} |q_1(z_0,z_1) - r_1(z_0,z_1)| =$$

$$\begin{aligned}
&= \sum_{z_0,z_1} |q(z_0)F_{z_0,z_1}(1) - r(z_0)G_{z_0,z_1}(1)| \\
&= \sum_{z_0,z_1} |q(z_0)(F_{z_0,z_1}(1) - G_{z_0,z_1}(1)) + (q(z_0)-r(z_0))G_{z_0,z_1}(1)| \\
&\leq \sum_{z_0,z_1} q(z_0)|F_{z_0,z_1}(1) - G_{z_0,z_1}(1)| + \sum_{z_0,z_1} |q(z_0)-r(z_0)|G_{z_0,z_1}(1) \\
&\leq \sum_{z_0} q(z_0)\sum_{z_1} |F_{z_0,z_1}(1) - G_{z_0,z_1}(1)| + \sum_{z_0} |q(z_0)-r(z_0)| \sum_{z_1} G_{z_0,z_1}(1) \\
&= E_q\|F_{z_0}(1) - G_{z_0}(1)\| + \sum_{z_0} |q(z_0)-r(z_0)| \\
&= E_q\|F_{z_0}(1) - G_{z_0}(1)\| + \|q-r\|.
\end{aligned}$$

Assume the result holds for $k-1$ and we prove it for k. For every finite set Y, the $S_{k-1}\times S_k$ transition matrices $F(k)$ and $G(k)$ define $(Y\times S_{k-1})\times S_k$ transition matrices $F^y(k)$ and $G^y(k)$ by $F^y_{(y,z_{k-1}),z_k}(k) = F_{z_{k-1},z_k}(k)$ and $G^y_{(y,z_{k-1}),z_k}(k) = G_{z_{k-1},z_k}(k)$. Obviously, $\|F^y_{y,z}(k) - G^y_{y,z}(k)\| = \|F_z(k) - G_z(k)\|$ for all $(y,z)\in Y\times S_{k-1}$, $q_k(z_0,z_1,\ldots,z_k) = q_{k-1}(z_0,z_1,\ldots,z_{k-1})F_{z_{k-1},z_k}(k)$ and $r_k(z_0,z_1,\ldots,z_k) = r_{k-1}(z_0,z_1,\ldots,z_{k-1})G_{z_{k-1},z_k}(k)$. The result for $k=1$ implies that

$$\|q_k - r_k\| \leq \|q_{k-1} - r_{k-1}\| + E_{q_{k-1}}\|F_{z_{k-1}}(k) - G_{z_{k-1}}(k)\|$$

and by the induction hypothesis

$$\|q_{k-1} - r_{k-1}\| \leq \|q-r\| + E_{q_{k-1}}\sum_{j=1}^{k-1} \|F_{z_{j-1}}(j) - G_{z_{j-1}}(j)(j)\|.$$

Summing the two inequalities (and observing that the projection of q_k on the first k coordinates $(z_0, \ldots, z_{k-1})$ is equal to q_{k-1}) proves (40). $\square$

Corollary 1 *Let k be a positive integer. Assume that $Q, Q' : [0, \infty) \to M_0$ are measurable, with $\|Q(t)\|, \|Q'(t)\| \leq C$. Then, for every $s > 0$ such that $ks \in \mathbb{N}$, the norm distance between the Q-distribution and the Q'-distribution of $z_0, z_{1/k}, \ldots, z_s$, is bounded by*

$$2sC/k + E_Q \sum_{j=1}^{sk} \| \int_{(j-1)/k}^{j/k} Q_{z_{(j-1)/k}}(t)\, dt - \int_{(j-1)/k}^{j/k} Q'_{z_{(j-1)/k}}(t)\, dt \|$$

$$\leq 2sC/k + \sum_{j=1}^{sk} \| \int_{(j-1)/k}^{j/k} Q(t)\, dt - \int_{(j-1)/k}^{j/k} Q'(t)\, dt \|.$$

Proof. Let $F(s,t)$ and $F'(s,t)$ be the transition matrices of the Q and Q' processes respectively. By the assumption on Q and Q' it follows from inequality (39) that $\|F_{z_{(j-1)/k}}((j-1)/k, j/k) - F'_{z_{(j-1)/k}}((j-1)/k, j/k)\|$ is $\leq 2C/k^2 + \| \int_{(j-1)/k}^{j/k} Q_{z_{(j-1)/k}}(t)\, dt - \int_{(j-1)/k}^{j/k} Q'_{z_{(j-1)/k}}(t)\, dt\|$. The result follows now from Lemma 6. $\square$

Corollary 2 *Assume that $Q, Q' : [0, \infty) \to M_0$ are (uniformly) bounded measurable functions. Then, for every $s \geq 0$, the norm distance between the Q-distribution and the Q'-distribution of $(z_t)_{0 \leq t \leq s}$ is bounded by*

$$E_Q \int_0^s \|Q_{z_t}(t) - Q'_{z_t}(t)\|\, dt \qquad (\leq \int_0^s \|Q(t) - Q'(t)\|\, dt).$$

Proof. Follows by going to the limit as $k \to \infty$ in the previous corollary. $\square$

Lemma 7 *Let σ be a correlated pure (mixed-action) Markov strategy profile. Then, for every subset of states $C \subset S$, $0 \leq s < \frac{1}{2\|\mu\|}$, and $z \in S$, we have*

$$\int_0^s \mu(C, z, \sigma(z,t))\, dt - s^2\|\mu\| \leq P_\sigma^z(z_s \in C) - \mathbb{I}(z \in C) \tag{42}$$

$$\leq \int_0^s \mu(C, z, \sigma(z,t))\, dt + s^2\|\mu\|, \tag{43}$$

where P^z_σ is the probability on plays defined by the initial state z and the strategy σ, and for every $[-1,1]$-valued measurable function f defined on $\cup_{z\in S}\{z\}\times\Delta(A(z))$ and every $z\in S$, we have

$$|E^z_\sigma \int_0^s f(z_t,\sigma(z_t,t))\,dt - \int_0^s f(z,\sigma(z,t))\,dt| \le s^2\|\mu\|, \tag{44}$$

where E^z_σ is the expectation with respect to the probability P^z_σ.

Proof. Apply Theorem 9 with $Q(t)$, $t\ge 0$, being the $S\times S$ matrix with $Q_{z,z'}(t)=\mu(z',z,\sigma(z,t))$. Then $\|Q(t)\|\le 2\|\mu\|$. For an $S\times S$ matrix F, $z\in S$, and $C\subset S$, define $F_{z,C}=\sum_{z'\in C}F_{z,z'}$. Note that for a matrix $F\in M_0$ we have $F_{z,C}\le \|F\|/2$ for every $z\in S$ and $C\subset S$. Therefore, inequalities (42) and (43) follow from (39) for $s\le 1/(2\|\mu\|)$.

Inequality (42) implies that for every $t\ge 0$ and every sequence $t_0=0<t_1<\ldots<t_k=t$ with $t_j-t_{j-1}\le 1/(2\|\mu\|)$, we have $P^z_\sigma(z_{t_j}=z_{t_{j-1}}\mid z_{t_{j-1}})\ge 1-(t_j-t_{j-1})\|\mu\|(1+t_j-t_{j-1})$. By setting $k=2^n$ and $t_j=jt/2^n$ we have $P^z_\sigma(z_{t_j}=z \quad \forall j\le 2^n)\ge (1-2^{-n}t(1+t/2^n)\|\mu\|)^{2^n}\to_{n\to\infty} e^{-t\|\mu\|}$. Therefore, $P^z_\sigma(z_s=z \quad \forall s\le t)\ge e^{-t\|\mu\|}$. In particular, for every $t\ge 0$ we have $P^z_\sigma(z_t=z)=E^z_\sigma \mathbb{I}(z_t=z)\ge e^{-t\|\mu\|}$.

$|f(z_t,\sigma(z_t,t))-f(z,\sigma(z,t))|\le 2\mathbb{I}(z_t\ne z)$. Therefore,

$$\begin{aligned}|E^z_\sigma \int_0^s f(z_t,\sigma(z_t,t))\,dt - \int_0^s f(z,\sigma(z,t))\,dt| &\le 2\int_0^s (1-e^{-t\|\mu\|})\,dt\\ &\le 2\int_0^s t\|\mu\|\,dt \le s^2\|\mu\|.\end{aligned}$$

□

An alternative (and complementary) proof of Lemma 7 is obtained by using the explicit law of the Markov process. The distribution of the first time of a state change is bounded stochastically from below by the exponential distribution with parameter $\|\mu\|$. Explicitly, let τ be a strategy and $t_1=\inf\{t>0: z_t\ne z_0\}$; then $P_\tau(t_1>s)\ge e^{-s\|\mu\|}$. If we set $t_{j+1}=\inf\{t>t_j: z_t\ne z_{t_j}\}$, we have $P_\tau(t_2>s)=P_\tau(t_1>s)+P_\tau(t_1<s<t_2)\ge e^{-\|\mu\|s}(1+\|\mu\|s)\ge 1-\|\mu\|^2s^2$. For a correlated Markov strategy σ, and $z\ne z_0$, the density of $t_1=t$ and $z_{t_1}=z$ is

$$\mu(z,z_0,\sigma(z_0,t))e^{\int_0^t \mu(z_0,z_0,\sigma(z_0,t))\,dt}.$$

For every subset $C \subset S$ we have $\mu(C, z, \sigma(z,t)) + \mu(S \setminus C, z, \sigma(z,t)) = 0$ and $P^z_\sigma(z_s \in C) - \mathbb{I}(z \in C) + P^z_\sigma(z_s \in S \setminus C) - \mathbb{I}(z \in S \setminus C) = 0$. Therefore, inequality (42) holds for $C = D$ if and only if inequality (43) holds for $C = S \setminus D$. Therefore, it suffices to prove inequalities (42) and (43) for $C \subset S \setminus \{z\}$. For every $s \geq t$ we have $P^z_\sigma(t_2 > s \mid t_1 = t) \geq e^{-(s-t)\|\mu\|}$ and $e^{-t(\mu(*))} \geq e^{-t\|\mu\|}$. Therefore, for $C \subset S \setminus \{z\}$, we have

$$\begin{aligned}
P^z_\sigma(z_s \in C) &\geq \sum_{z \in C} \int_0^s \mu(z, z, \sigma(z,t)) e^{\int_0^t \mu(z,z,\sigma(z,t))\,dt}\, e^{-(s-t)\|\mu\|}\,dt \\
&\geq \int_0^s \mu(C, z, \sigma(z,t))\,dt\, e^{-s\|\mu\|} \geq \int_0^s \mu(C, z, \sigma(z,t))\,dt - s^2\|\mu\|^2
\end{aligned}$$

and,

$$\begin{aligned}
P^z_\sigma(z_s \in C) &\leq \int_0^s \mu(C, z, \sigma(z,t)) e^{\int_0^t \mu(z,z,\sigma_u(z))\,du}\,dt + P^z_\sigma(t_2 \leq s) \\
&\leq \int_0^s \mu(C, z, \sigma(z,t))\,dt + s^2\|\mu\|^2.
\end{aligned}$$

This completes the proof of (42) and (43).

□

Lemma 8 *Let σ and τ be two strategies such that $\sigma(h_t) = \tau(h_t)$ on $z_u = z_0$ $\forall u \leq t$. Then*

$$|P^z_\sigma(z_s \in C) - P^z_\tau(z_s \in C)| \leq 2s^2\|\mu\|^2, \tag{45}$$

and for every function $f : \{(z,x) : z \in S \text{ and } x \in \Delta(A(z))\} \to [-1,1]$ and every probability measure ν on $[0,s]$, we have

$$|E^z_\sigma \int_0^s f(z_t, x_t)\,d\nu(t) - E^z_\sigma \int_0^s f(z_t, x_t)\,d\nu(t)| \leq 2s^2\|\mu\|. \tag{46}$$

Proof. Let $t_1 = s \wedge \inf\{t \geq 0 : z_t \neq z_0\}$ and $t_2 = \inf\{t \geq t_1 : z_t \neq z_{t_1}\}$. Note that $\{z_{t_1} \in C\} \setminus \{t_2 \leq s\} \subset \{z_s \in C\} \subset \{z_{t_1} \in C\} \cup \{t_2 \leq s\}$ and $P^z_\sigma(z_{t_1} \in C) = P^z_\tau(z_{t_1} \in C)$. Therefore, $P^z_\sigma(z_s \in C) \leq P^z_\sigma(z_{t_1} \in C) + P^z_\sigma(t_2 \leq$

$s) \leq P^z_\tau(z_{t_1} \in C) + 1 - e^{-s\|\mu\|}(1 + s\|\mu\|) \leq P^z_\tau(z_s \in C) + 2s^2\|\mu\|^2$.

$$
\begin{aligned}
E^z_\sigma \int_0^s f(z_t, x_t)\, d\nu(t) &\leq E^z_\sigma \int_0^s f(z_t, x_t)\mathbb{I}(t \leq t_1)\, d\nu(t) + E^z_\sigma \int_0^s \mathbb{I}(t > t_1)\, d\nu(t) \\
&\leq E^z_\tau \int_0^s f(z_t, x_t)\mathbb{I}(t \leq t_1)\, d\nu(t) + \int_0^s E^z_\sigma \mathbb{I}(t > t_1)\, d\nu(t) \\
&\leq E^z_\tau \int_0^s f(z_t, x_t)\, d\nu(t) + \int_0^s E^z_\tau \mathbb{I}(s > t_1)\, d\nu(t) \\
&\quad + \int_0^s E^z_\sigma \mathbb{I}(s > t_1)\, d\nu(t) \\
&\leq E^z_\tau \int_0^s f(z_t, x_t)\, d\nu(t) + 2\int_0^s 1 - e^{-s\|\mu\|}\, d\nu(t) \\
&\leq E^z_\tau \int_0^s f(z_t, x_t)\, d\nu(t) + 2s^2\|\mu\|.
\end{aligned}
$$

□

Lemma 9 *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 Markov strategy profile σ (or, more generally, for every strategy profile σ for which, for $s \geq t$, $\sigma_s(h)$ depends on only h_t, z_s, and s), we have*

$$P_\sigma(t < T \leq n \mid \mathcal{H}_t) = E_\sigma\left(\int_t^{T\wedge n} \mu(C, z_s, x_s)\, ds \mid \mathcal{H}_t\right) \tag{47}$$

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

Proof. Equality (47) holds trivially on $z_t \in C$. Assume that $z_t \notin C$. It suffices to prove the lemma for $t = 0$ and $n = 1$. Fix a (sufficiently large) positive integer $k > 1$. For every positive integer $1 \leq j \leq k$, $\mathbb{I}_j$ denotes the indicator of the event $(j-1)/k < T$, and $\mathbb{I}^*_j$ denotes the indicator of the event $(j-1)/k < T \leq j/k$. By Lemma 7, for every positive integer $1 \leq j \leq k$, $E_\sigma(\mathbb{I}^*_j \mid \mathcal{H}_{(j-1)/k}) = \mathbb{I}_j E_\sigma(\int_{(j-1)/k}^{j/k} \mu(C, z_s, x_s)\, dt + O(1/k^2) \mid \mathcal{H}_{(j-1)/k}))$.

Therefore,

$$\begin{aligned}
P_\sigma(T \leq 1 \mid \mathcal{H}_0) &= \sum_{j=1}^{k} E_\sigma(\mathbb{I}^*_j \mid \mathcal{H}_0) \\
&= \sum_{j=1}^{k} E_\sigma(E_\sigma(\mathbb{I}^*_j \mid \mathcal{H}_{(j-1)/k}) \mid \mathcal{H}_0) \\
&= \sum_{j=1}^{k} E_\sigma(\mathbb{I}_j E_\sigma(\int_{(j-1)/k}^{j/k} \mu(C, z_s, x_s)\, dt + O(1/k^2) \mid \mathcal{H}_{(j-1)/k}) \mid \mathcal{H}_0) \\
&= O(1/k) + E_\sigma(\int_0^{T\wedge 1} \mu(C, z_s, x_s)\, dt \mid \mathcal{H}_0).
\end{aligned}$$

As this holds for every k, equality (47) holds. □

The next two lemmas, which are of independent interest, are not used in the proof of the other results in this paper.

Lemma 10 *Let σ and τ be two pure (mixed-action) stationary correlated strategies of a continuous-time stochastic game with $\|\mu\| = 1$ and $0 \leq g^i \leq 1$. Then,*

$$d(D_\sigma, D_\tau) \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(z_t)\|\, dt), \tag{48}$$

where D_σ (respectively, D_τ) is the distribution of $(z_t)_{0\leq t\leq 1}$ w.r.t. P_σ (respectively, P_τ), and for every player i we have

$$|E_\sigma \int_0^1 g^i(z_t, \sigma(z_t))\, dt - E_\tau \int_0^1 g^i(z_t, \tau(z_t))\, dt| \leq 2E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\|\, dt). \tag{49}$$

Proof. Fix a positive integer k. Let $D^k_\sigma(j)$ (respectively, $D^k_\tau(j)$), $0 \leq j \leq k$, be the distribution of $z_0, z_{1/k}, \ldots z_{j/k}$ w.r.t. P_σ (respectively, P_τ). We prove by induction on j that the norm distance between $D^k_\sigma(j)$ and $D^k_\tau(j)$, is bounded by $E_\sigma(\int_{(j-1)/k}^{j/k} \|\sigma(z_t) - \tau(z_t)\| \, dt) + O(j/k^2)$.

$$\begin{aligned} d(D^k_\sigma(j), D^k_\tau(j)) &\leq E_\sigma(\int_0^{(j-1)/k} \|\sigma(z_t) - \tau(z_t)\| \, dt + O((j-1)/k^2) \\ &+ E_\sigma(\int_{(j-1)/k}^{j/k} \|\sigma(z_t) - \tau(z_t)| \, dt + O(1/k^2) \\ &= E_\sigma(\int_0^{j/k} \|\sigma(z_t) - \tau(z_t)\| \, dt + O(j/k^2). \end{aligned}$$

It follows that $d(D^k_\sigma(k), D^k_\tau(k)) \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\| \, dt) + O(1/k)$. As this holds for every k, and $d(D^k_\sigma(k), D^k_\tau(k)) \leq d(D^{\ell k}_\sigma(\ell k), D^{\ell k}_\tau(\ell k))$ for every positive integer k, we deduce that for every k we have $d(D^k_\sigma(k), D^k_\tau(k)) \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\| \, dt)$, and therefore $d(D_\sigma, D_\tau) \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\| \, dt)$.

Inequality (49) follows from (48) and the triangle inequality. Indeed, $|g(z_t, \sigma(z_t)) - g(z_t, \tau(z_t)| \leq \|\sigma(z_t) - \tau(x_t)\|$ and therefore

$$|E_\sigma \int_0^1 g^i(z_t, \sigma(z_t)) \, dt - E_\sigma \int_0^1 g^i(z_t, \tau(z_t)) \, dt| \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\| \, dt),$$

and by (49) and using $0 \leq g^i(z, a) \leq 1$, we have

$$|E_\sigma \int_0^1 g^i(z_t, \tau(z_t)) \, dt - E_\tau \int_0^1 g^i(z_t, \tau(z_t)) \, dt| \leq E_\sigma(\int_0^1 \|\sigma(z_t) - \tau(x_t)\| \, dt).$$

Summing the two inequalities we deduce (49). □

Lemma 11 *Assume $0 \leq g^i \leq 1$. Let k be a positive integer, let σ and τ be $1/k$-discretimized (correlated-action) strategy profiles, and let $s \geq 0$ with $ks \in \mathbb{N}$. Then, for every initial state $z_0 \in S$, we have*

$$\sum_{z \in S} |P^{z_0}_\sigma(z_s = z) - P^{z_0}_\tau(z_s = z)| \leq E^{z_0}_{P_\sigma} \int_0^s \|\sigma_t(h_t) - \tau_t(h_t)\| \, dt \tag{50}$$

and

$$|E^{z_0}_\sigma \int_0^s g^i_t \, dt - E^{z_0}_\tau \int_0^s g^i_t \, dt \,| \leq 2E^{z_0}_{P_\sigma} \int_0^s \|\sigma_t(h_t) - \tau_t(h_t)\| \, dt. \tag{51}$$

Proof. Inequalities (50) and (51) follow from Lemma 10. □

8 The discretimization approach

Continuous-time stochastic games can be viewed as a limit of continuous-time games where players have a short time delay in either the reaction to other players' actions, or the observation of other players' actions. This leads to our *discretimization approach*.

Given an increasing sequence $\mathcal{T} : t_0 = 0 < t_1 < t_2 < \dots$ with $t_k \to \infty$ as $k \to \infty$, we define a pure *$\mathcal{T}$-strategy* as a pure action strategy with $\sigma_t(h)$ being a function of $h_{t_{k-1}}$ and z_t whenever $t_k \le t < t_{k+1}$. The set of pure $\mathcal{T}$-strategies of player i is denoted by $\Sigma^i_{\mathcal{T}}$. We write $\mathcal{T} \subset \hat{\mathcal{T}}$ if $\{t_k : k \ge 0\} \subset \{\hat{t}_k : k \ge 0\}$. Note that $\Sigma^i_{\mathcal{T}} \subset \Sigma^i_{\hat{\mathcal{T}}}$ whenever $\mathcal{T} \subset \hat{\mathcal{T}}$. The diameter of $\mathcal{T}$, $d(\mathcal{T})$, is the supremum of $t_{k+1} - t_k$. A *$\mathcal{T}$-strategy* of player i is a probability distribution on $\Sigma^i_{\mathcal{T}}$.

A vector $V \in \mathbb{R}^S$ is the *discretimization value* of the ρ-discounted two-person zero-sum continuous time stochastic games if for every $\varepsilon > 0$ there is $\delta > 0$ such that if $\mathcal{T}^i \subset \hat{\mathcal{T}}^i$, $i = 1, 2$, with $d(\mathcal{T}^i) < \delta$, then there are $\mathcal{T}^i$-strategies σ^i, such that for every $\hat{\mathcal{T}}^i$-strategy τ^i we have

$$\varepsilon + E_{\sigma^1,\tau^2} \int_0^\infty e^{-\rho t} g(z_t, a_t)\, dt \ge V(z_0) \ge E_{\tau^1,\sigma^2} \int_0^\infty e^{-\rho t} g(z_t, a_t)\, dt - \varepsilon.$$

Similarly, $V \in \mathbb{R}^{N \times S}$ is a *discretimization equilibrium payoff* of the ρ-discounted non-zero-sum continuous-time stochastic game if for every $\varepsilon > 0$ there is $\delta > 0$ such that if $\mathcal{T}^i \subset \hat{\mathcal{T}}^i$, $i \in N$, with $d(\mathcal{T}^i) < \delta$, then there are $\mathcal{T}^i$-strategies σ^i, such that for every $\hat{\mathcal{T}}^i$-strategy τ^i we have

$$-\varepsilon + E_{\sigma^{-i},\tau^i} \int_0^\infty e^{-\rho t} g^i(z_t, a_t)\, dt \le V^i(z_0) \le E_\sigma \int_0^\infty e^{-\rho t} g^i(z_t, a_t)\, dt + \varepsilon.$$

A vector $v \in \mathbb{R}^S$ is the *discretimization uniform value* of the (undiscounted) two-person zero-sum continuous-time stochastic game if for every $\varepsilon > 0$ there is $s_\varepsilon > 0$ and $\delta > 0$ such that if $\mathcal{T}^i \subset \hat{\mathcal{T}}^i$, $i = 1, 2$, with $d(\mathcal{T}^i) < \delta$, then there are $\mathcal{T}^i$-strategies σ^i, such that for every $\hat{\mathcal{T}}^i$-strategy τ^i and every $s > s_\varepsilon$ we have

$$\varepsilon + \frac{1}{s} E_{\sigma^1,\tau^2} \int_0^s g(z_t, a_t)\, dt \ge v(z_0) \ge \frac{1}{s} E_{\tau^1,\sigma^2} \int_0^s g(z_t, a_t)\, dt - \varepsilon.$$

Similarly, $v \in \mathbb{R}^{N \times S}$ is a *discretimization uniform equilibrium payoff* of the (undiscounted) non-zero-sum continuous-time stochastic game, if for every

$\varepsilon > 0$ there is $\delta > 0$ such that if $\mathcal{T}^i \subset \hat{\mathcal{T}}^i$, $i \in N$, with $d(\mathcal{T}^i) < \delta$, then there are $\mathcal{T}^i$-strategies σ^i and $T > 0$, such that for every $\hat{\mathcal{T}}^i$-strategy τ^i and every $s > T$ we have

$$-\varepsilon + \frac{1}{s} E_{\sigma^{-i},\tau^i} \int_0^s g^i(z_t, a_t)\, dt \le V^i(z_0) \le \frac{1}{s} E_\sigma \int_0^s g^i(z_t, a_t)\, dt + \varepsilon.$$

Our proofs show that the value of the discounted two-person zero-sum game is a discretimization value, and the uniform value is a discretimization uniform value. Similarly, any equilibrium payoff of the discounted continuous-time game is a discretimization equilibrium payoff, and the undiscounted game has a discretimization uniform equilibrium payoff.

9 The asymptotic approach

Continuous-time stochastic games can be viewed as approximations of discrete-time stochastic games with frequent stages; see Section 6. This motivates our *asymptotic approach* that studies the strategic equilibrium of a sequence of discrete-time stochastic games that approximate the continuous-time stochastic game Γ.

A particular discrete-time approximation of a continuous-time stochastic game Γ that is described by the tuple $\langle N, S, A, g, \mu\rangle$ of the set of states S, action profiles A, payoff function g, and transition rates μ, is the classical (discrete-time) stochastic game $\Gamma_\delta = \langle N, S, A, g_\delta, p_\delta\rangle$ (with $\delta > 0$ sufficiently small so that $\delta\|\mu\| \le 1$) with state space S, action profiles A, stage payoff function g_δ defined by $g_\delta(z, a) := \delta g(z, a)$, and transition probabilities p_δ defined by $p_\delta(z' \mid z, a) := \delta\mu(z', z, a)$ for $z' \ne z$ and $p_\delta(z \mid z, a) := 1 - \delta\mu(z, a)$ where $\mu(z, a) := \sum_{z' \ne z} \mu(z', z, a)$. In this approximation δ stands for the duration of a stage. A play of the continuous-time stochastic game up to time $n\delta$ corresponds to the play of the n-stage discrete-time stochastic game Γ_δ, and the continuous-time stochastic game Γ with discount rate ρ corresponds to the discrete-time stochastic game Γ_δ with discount factor $1-\delta\rho$. We view Γ as the limit of the games Γ_δ as δ goes to 0, and our results on the continuous-time stochastic game translate to asymptotic results on the stochastic games Γ_δ.

A more general family of discrete-time stochastic games $\Gamma_\delta = \langle N, S, A, g_\delta, p_\delta\rangle$ that "converges" to the continuous time stochastic game $\Gamma_\delta = \langle N, S, A, g, \mu\rangle$ is a family $(\Gamma_\delta)_{\delta>0}$ such that for every $z \ne z'$ and $a \in A(z)$ we have

$g_\delta(z,a) = \delta g(z,a) + o(\delta)$ and $p_\delta(z' \mid, z, a) = \delta\mu(z', z, a)(1 + o(1))$ as δ goes to 0.

An *asymptotic limiting-average equilibrium payoff* of a family of discrete-time stochastic games $(\Gamma_\delta)_{\delta>0}$ is a vector payoff $u \in \mathbb{R}^{N\times S}$ such that for every $\varepsilon > 0$ there is $\delta_\varepsilon > 0$ such that for every $\delta < \delta_\varepsilon$ there is a profile of strategies σ_δ in Γ_δ such that for every strategy τ^i of player i in Γ_δ we have

$$\varepsilon + E^z_{\sigma_\delta} \liminf_{k\to\infty} \frac{1}{k\delta} \sum_{\ell=1}^{k} g^i_\delta(z_\ell, a_\ell) \geq u^i(z) \geq E^z_{\sigma^{-i}_\delta, \tau^i} \limsup_{k\to\infty} \frac{1}{k\delta} \sum_{\ell=1}^{k} g^i_\delta(z_\ell, a_\ell) - \varepsilon. \tag{52}$$

An *asymptotic uniform equilibrium payoff* of a family of discrete-time stochastic games $(\Gamma_\delta)_{\delta>0}$ is a vector payoff $u \in \mathbb{R}^{N\times S}$ such that for every $\varepsilon > 0$ there is $\delta_\varepsilon > 0$ and $s_\varepsilon > 0$ such that for every $\delta < \delta_\varepsilon$ there is a profile of strategies σ_δ in Γ_δ such that for every positive integer $k \geq s_\varepsilon/\delta$ and every strategy τ^i of player i in Γ_δ we have

$$\varepsilon + E^z_{\sigma_\delta} \frac{1}{k\delta} \sum_{\ell=1}^{k} g^i_\delta(z_\ell, a_\ell) \geq u^i(z) \geq E^z_{\sigma^{-i}_\delta, \tau^i} \frac{1}{k\delta} \sum_{\ell=1}^{k} g^i_\delta(z_\ell, a_\ell) - \varepsilon. \tag{53}$$

A *uniform-limiting equilibrium payoff* of a family of discrete-time stochastic games $(\Gamma_\delta)_{\delta>0}$ is a vector payoff $u \in \mathbb{R}^{N\times S}$ such that for every $\varepsilon > 0$ there is $\delta_\varepsilon > 0$ and $\sigma_\varepsilon > 0$ such that for every $\delta < \delta_\varepsilon$ there is a profile of strategies σ_δ in Γ_δ such that for every positive integer $k \geq s_\varepsilon/\delta$ and every strategy τ^i of player i in Γ_δ, inequalities (53) and (52) hold.

Note that if inequality (52) holds for σ_δ and τ^i, then inequality (53) holds for a sufficiently large k. However, there need not be s_ε such that inequality (53) holds for all $k \geq s_\varepsilon/\delta$. Therefore, the existence of a limiting-average equilibrium payoff does not imply the existence of a uniform equilibrium payoff. Moreover, there is a stochastic game with a single player (and without strategic choices, i.e., $|A(z)| = 1$) with countably many states that has a limiting-average equilibrium payoff but no uniform equilibrium payoff.

Similarly, existence of a uniform equilibrium payoff does not imply the existence of a limiting-average equilibrium payoff. There is a stochastic game with a single player with countably many states that has a uniform equilibrium payoff but no limiting-average equilibrium payoff.

Moreover, (it follows from the above that) there is a two-person zero-sum stochastic game with countably many states that has both a uniform equilibrium payoff and a limiting-average payoff that coincide, but has no uniform-limiting equilibrium payoff.

The present paper implies the following result.

Theorem 10 *Fix an N-player continuous-time stochastic game $\Gamma = \langle N, S, A, g, \mu \rangle$. Assume that the family $(\Gamma_\delta)_{\delta>0}$, $\Gamma = \langle N, S, A, g_\delta, p_\delta \rangle$, satisfies $g_\delta(z,a) = \delta g(z,a) + o(\delta)$ and $p_\delta(z' \mid z, a) = (\delta + o(\delta))\mu(z', z, a)$ for every $z' \neq z$ and $a \in A(z)$. Then the family of discrete-time stochastic games $(\Gamma_\delta)_{\delta>0}$ has a limiting-average equilibrium payoff.*

Stronger conclusions are implies by stronger asymptotic assumptions on p_δ, e.g., by assuming that $p_\delta = (\delta + o(\delta))\mu$ (where the function $o(\delta)$ is independent of the point in the domain of p_δ).

10 Related literature

10.1 Countinuous-time supergames and bargaining

Simon and Stinchcombe [28] 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) a 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 selects a constant action over a short time interval. In addition, the pure strategies choose a pure stage action, in contrast to our main model that enables us to choose a mixed action as a function of history.

Perry and Reny [24] 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.

10.2 The stochastic process of states

[14, 9] have shown[8] that for every correlated Markov (memoryless) strategy σ there is a unique transition probability $F(s,t,\sigma) = (F_{z',z}(s,t,\sigma))_{z,z' \in S}$ such that

$$P_\sigma(z_t = z' \mid z_s = z) = F_{z',z}(s,t,\sigma)$$

for every $t \geq s \geq 0$, and that it satisfies $F(s,s,\sigma) = I$ and the Kolmogorov forward differential equation:

$$\frac{\partial F(s,t,\sigma)}{\partial t} = F(s,t,\sigma)Q(t,\sigma) \text{ for a.e. } t \geq s \geq 0, \tag{54}$$

where $Q(t,\sigma)$ is the transition matrix $(Q_{z',z}(t,\sigma))_{z',z \in S}$ with $Q_{z',z}(t,\sigma) = \mu(z', z, \sigma_t(z))$.

10.3 Continuous-time MDP and stochastic games

Zachrisson [37] introduced two-player zero-sum continuous-time Markov games, called also continuous-time stochastic games, with only Markovian strategies, and proves that the discounted and finite horizon games have optimal (Markovian) strategies. Yehuda Levy [13] proved that a non-zero-sum stochastic game of fixed duration has a Markovian correlated equilibrium, but need not posses a Markovian equilibrium.

Jasso-Fuentes [8] 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 (respectively, two-person zero-sum Markov games), called the *uniformization technique*, appears also in [6] (respectively [37]).

Guo and Hernandez-Lerma [3] study two-person non-zero-sum continuous-time stochastic games with the discounted payoff criteria (and Borel actions spaces), and give conditions that ensure the existence of Nash equilibrium in stationary strategies.

[8] See, [14, p. 873] and [10, Equation (1.1) p. 919].

References

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

[2] 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.

[3] 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.

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

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

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

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

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

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

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

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

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

[13] Levy, Y. (2012), Continuous-time stochastic games of fixed duration. DP (forthcoming), Center for the study of Rationality, Hebrew University of Jerusalem.

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

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

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

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

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

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

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

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

[22] 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.

[23] 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.

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

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

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

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

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

[29] Solan, E. (2001), Characterization of correlated equilibrium in stochastic games, *International Journal of Game Theory*, 30, 259–277.

[30] Solan, E. (2003), Continuity of the value of competitive Markov decision processes, *Journal of Theoretical Probability*, 16, 831–845.

[31] Solan, E. and N. Vieille (2002), Correlated equilibrium in stochastic games, *Games and Economic Behavior*, 38, 362–399.

[32] Solan, E. and R. Vohra (2002), Correlated equilibrium and public signalling in absorbing games, *International Journal of Game Theory*, 31, 91–121.

[33] Sorin, S. (1986), Asymptotic properties of a non-zero-sum stochastic games, *International Journal of Game Theory*, 15, 101–107.

[34] Vigeral, G. (2012), A zero-sum stochastic game with compact action sets and no asymptotic value, CEREMADE, Universite Paris-Dauphine, preprint.

[35] Yushkevich, A.A. (1977), Controlled Markov models with countable state and continuous time, *Theory of Probability and its Applications* 22, 215-235.

[36] Yushkevich, A. A. and E. A. Feinberg (1979), On homogeneous Markov model with continuous time and finite or countable state space, *Theory of Probability and its Applications*, 24, 156-161.

[37] Zachrisson, L. E. (1964), Markov games, in *Advances in Game Theory*, Princeton University Press, Princeton, N. J.