# On Stochastic Hybrid Zero-Sum Games with non-linear slow dynamic*

Minh-Tuan NGUYEN † Eitan ALTMAN ‡ Vladimir GAITSGORY †

**Abstract**

A continuous time stochastic hybrid system, controlled by two players with opposite objectives (zero-sum game), is considered. The parameters of the system may jump at discrete moments of time according to a Markov Decision Process, i.e. a Markov chain that is directly controlled by both players, and has finite state and action spaces. Under assumption that the length of the intervals between the jumps is defined by a small parameter $\epsilon$, the value of this game is shown to have limit as the small parameter tends to zero. This limit is established to coincide with the viscosity solution of some Hamilton-Jacobi type equations.

**Key words:** Hybrid stochastic systems, stochastic games, small parameter, averaging, viscosity solutions, Hamilton-Jacobi-Isaacs equations, asymptotic optimality, non-linear dynamics.

## 1 Introduction and Statement of the problem

Consider the following hybrid stochastic controlled system. The state $Z_s \in \mathbb{R}^n$ evolves according to the following dynamics:

$$\frac{d}{ds}Z_s = f(Z_s, Y_s), \qquad s \in [t, T], \qquad Z_t = z \tag{1}$$

where $Y_s \in \mathbb{R}^k$ is the "control" and $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is a vector function. $Y_s$ is not chosen directly by the controllers, but is obtained as a result of controlling the following underlying stochastic discrete event system.

- Let $\epsilon$ be the basic time unit. Time is discretized, i.e. transitions occur at times $s = l\epsilon,\ l = 0, 1, 2, ..., \lfloor (T-t)\epsilon^{-1} \rfloor$, where $\lfloor x \rfloor$ stands for the greatest integer which is smaller than or equal to $x$.
- There is a finite state space $\mathbf{X}$ and two players having finite action spaces $\mathbf{A}_1$ and $\mathbf{A}_2$ respectively. Let $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2$.
- If the state is $v$ and actions $a = (a_1, a_2)$ are chosen by the players, then the next state is $w$ with probability $P_{vaw}$. Denote $\mathcal{P} = \{P_{vaw}\}$.

*This work is supported by the Australian Research Council (ARC).

†School of Mathematics – University of South Australia, The Levels Campus, Warrendi Road Mawson Lakes SA 5095 – Australia. Fax: +61 8 8320 5785. Email: *9208086Q@levels.unisa.edu.au; V.Gaitsgory@unisa.edu.au*

‡INRIA, BP93, 2004 Route des Lucioles, 06902 Sophia Antipolis Cedex – France. Email: *Eitan.Altman@sophia.inria.fr*

- A policy $u^i = \{u_0^i, u_1^i, ...\}$ in the set of policies $\mathcal{U}^i$ for player $i$, $i = 1, 2$ is a sequence of probability measures on $\mathbf{A}_i$ conditioned on the history of all previous states (the $\mathbf{X}$ component only) and actions of both players, as well as the current state. More precisely, define the set of histories:

$$\mathbf{H} := \bigcup_l \mathbf{H}_l, \quad \text{where } \mathbf{H}_l := \big\{(x_0, a_0^1, a_0^2, x_1, a_1^1, a_1^2, ..., x_l)\big\}$$

  are the sets of all sequences of $3l+1$ elements describing the possible samples of previous states and actions prior to $l$ as well as the currents state at stage $l$ (i.e. at time $l\epsilon$). (The range of $l$ will be either $l = 0, 1, ..., \lfloor (T-t)\epsilon^{-1} \rfloor$, or, in other contexts, all nonnegative integers, depending on whether we consider finite or infinite horizon problems). The policy at stage $l$ for player $i$, $u_l^i$, is a map from $\mathbf{H}_l$ to the set of probability measures over the action space $\mathbf{A}_i$. (Hence at each time $t = l\epsilon$, player $i$, observing the history $h_l$, chooses action $a_i$ with probability $p(a_i|h_l)$).

- Let $\mathcal{F}_l$ be the discrete $\sigma$-Algebra corresponding to $\mathbf{H}_l$. Each initial distribution $\xi$ and policy pair $u$ for the players uniquely define a probability measure $P_\xi^u$ over the space of samples $\mathbf{H}$ (equipped with the discrete $\sigma$-algebra), see e.g. [Hin70]. Denote by $E_\xi^u$ the corresponding expectation operator. On the above probability space are now defined the random processes $X_l$ and $A_l = (A_l^1, A_l^2)$, denoting the state and actions processes. When the initial distribution is concentrated on a single state $x$, we shall denote the corresponding probability measure and expectation by $P_x^u$ and $E_x^u$.

**Remark:** The reason that we do not include the $Z$ part of the state in the definition of the policies is that the trajectory of this component is fully determined by the trajectories of the $\mathbf{X}$ component together with the actions, for a fixed initial state $z$. The latter is assumed to be fixed and common knowledge for the players.

Let $g : \mathbf{X} \times \mathbf{A} \to \mathbb{R}^k$, be some given vector-valued bounded function and $X_n$ and $A_n = (A_n^1, A_n^2)$ denote the state and actions processes. Then $Y_s$ in (1) is given by

$$Y_s = g(X_{\lfloor s/\epsilon \rfloor}, A_{\lfloor s/\epsilon \rfloor}). \tag{2}$$

$Y_s$ and thus $Z_s$ are well defined stochastic processes, and are both $\mathcal{F}_{\lfloor (T-t)\epsilon^{-1} \rfloor}$ measurable.

We shall be especially interested in the following classes of policies.

- The stationary policies, denoted by $\mathcal{S}_1$, for player 1, and $\mathcal{S}_2$, for player 2. A policy $u$ is called stationary if $u_l$ depends only on the current state (the $\mathbf{X}$-component), and does not depend on previous states and actions nor on the time. Let $\mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2$.

- The Markov policies $\mathcal{M}_1, \mathcal{M}_2$: these are policies where $u_l^i$ depends only on the current $\mathbf{X}$ component of the state (at time $t = l\epsilon$) and on stage $l$, and does not depend on previous states and actions. Denote $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$.

Let us define the payoff of the game by the equation

$$J_\epsilon(t, z, x; u^1, u^2) = E_x^{(u^1,u^2)} \Big\{ \int_t^T F(Z_s, Y_s)\, ds + G(Z_T) \Big\},$$

when policies $u^1, u^2$ are used by the players, the initial state of the system is $z$, the initial state of the controlled Markov chain is $x$, $Z_s$ is obtained through (1) and $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^1$,

$G : \mathbb{R}^n \to \mathbb{R}^1$ are running cost and terminal cost function respectively.

In our dynamic game, player 1 wishes to maximize $J_\epsilon(t, z, x; u^1, u^2)$ and player 2 wants to minimize it. We define the upper and lower value functions of the hybrid game as

$$B_\epsilon^{up}(t, z, x) = \inf_{u^2 \in U^2} \sup_{u^1 \in U^1} J_\epsilon(t, z, x; u^1, u^2)$$
$$B_\epsilon^{lo}(t, z, x) = \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} J_\epsilon(t, z, x; u^1, u^2)$$

It can be shown (see Appendix) that the stochastic hybrid game has value $B_\epsilon(t, z, x)$. That is, for all $(t, z, x) \in [0, T] \times \mathbb{R}^n \times \mathbf{X}$.

$$B_\epsilon(t, z, x) \stackrel{\text{def}}{=} B_\epsilon^{up}(t, z, x) = B_\epsilon^{lo}(t, z, x).$$

Our model is characterized by the fact that $\epsilon$ is supposed to be a small parameter and our objective is to show that the value of the game has *a limit* as $\epsilon \to 0$ and this limit is a viscosity solution of some Hamilton-Jacobi type equations.

Notice that this result can be viewed as an extension of viscosity solutions for deterministic singularly perturbed zero-sum differential games (see [Gai96]) to the stochastic case under consideration.

This paper is a continuation and generalization of previous works [SAG97] which solves a hybrid problem restricted to a single controller and [AG94] which considers a linear hybrid game with linear cost. As in [SAG97, AG94], the fact that $\epsilon$ is small means that the variables $Y_s$ can be considered to be fast with respect to $Z_s$, since, by (2), they may have a finite (not tending with $\epsilon$ to zero) change at each interval of the length $\epsilon$. This along with the dynamic equation of the system (1) allow to decompose the game into stochastic sub-games on a sequence of intervals which are short with respect to the variables $Z_s$ (in the sense that $Z_s$ remain almost unchanged on these intervals) and which are long enough with respect to $Y_s$ (so that the corresponding stochastic sub-games show on these intervals there limit properties).

The type of model which we introduce is natural in the control of inventories or of production (see for example [SZ94]), where we deal with material whose quantity may change in a continuous way. Breakdowns, repairs and other control decisions yield the underlying controlled Markov chain. In particular, repair, or preventive maintenance decisions are typical actions of a player that minimizes costs. If there is some unknown parameter (disturbance) of the dynamics of the system (e.g. the probability of breakdowns) which may change in a way that depends on the current and past states in a way that is unknown and unpredictable by the minimizer, we may formulate this situation as a zero-sum game, where the minimizer wishes to guarantee the best performance (lowest expected cost) under the worst case behaviour of nature. Nature may then be modelled as the maximizing player.

Our model may also be used in the control of highly loaded queueing networks for which the fluid approximation holds (see Kleinrock [Kle76] p. 56). The quantities $Z_t$ may then represent the number of customers in the different queues whereas the underlying controlled Markov chain may correspond to routing, or flow control of, say, some on-off traffic, with again, nature controlling some disturbances in quantities such as service rates.

The remainder of this paper is organised as follows. In section 2 we give all imposed assumptions and then introduce the associated sub-game and the existence of the value of such game in section 3. The Limit Hamilton-Jacobi-Isaacs equations for the stochastic hybrid game is defined in section 4. The main results are presented in section 5. The most tedious proves are gathered in the last two sections.

## 2 Basic Assumptions

In our consideration, we use the following assumptions.

**Assumption 1.** *There exists a compact subset $D_1 \in \mathbb{R}^n$ which contains all solutions $Z_s$ of the system* (1) *obtained with different admissible controls $Y_s$ which are defined as piecewise constant function of time taking their values in a finite subset of $\mathbb{R}^k$. Denote this subset by $D_2$.*

**Assumption 2.** *All the functions used in the definitions of the stochastic hybrid and associated games are continuous on $D_1 \times D_2$ and also they satisfy the local Lipschitz conditions in $(z, y)$ with Lipschitz constant $L \geq 0$. That is, for any $(z_i, y_i) \in (D_1 \times D_2)$ $i = 1, 2$.*

$$\begin{aligned} \big\|f(z_1, y_1) - f(z_2, y_2)\big\| &\leq L\big\{\|z_1 - z_2\| - \|y_1 - y_2\|\big\}, \\ \big\|F(z_1, y_1) - F(z_2, y_2)\big\| &\leq L\big\{\|z_1 - z_2\| - \|y_1 - y_2\|\big\}, \\ \big\|G(z_1) - G(z_2)\big\| &\leq L\|z_1 - z_2\|. \end{aligned}$$

Notice that from Assumption 1 and Assumption 2, it follows that the functions $f$ and $F$ are bounded. That is, there exists a constant $M \geq 0$ such that for all $(z, y) \in D_1 \times D_2$

$$\|f(z, y)\| \leq M, \qquad \|F(z, y)\| \leq M \quad \text{and} \quad \|G(z)\| \leq M.$$

**Assumption 3.** *The stochastic process $\{X_n, A_n\}$ which is known as controlled Markov chain or Markov decision process has a unichain structure: under any pair of stationary policies for the two players, the state process constitutes a single ergodic class of states.*

## 3 $(z, \lambda)$-associated games

Consider a family of infinite horizon stochastic games, all with the same state and action spaces $\mathbf{X}$ and $\mathbf{A}$ as above and the same transition probabilities $\mathcal{P}$, parameterized by a pair $(z, \lambda) \in \mathbb{R}^n \times \mathbb{R}^n$. Let $r : \mathbb{R}^n \times \mathbb{R}^n \times \mathbf{X} \times \mathbf{A} \to \mathbb{R}$ be the immediate cost defined as follows

$$r(z, \lambda, x, a) = F\big(z, g(x, a)\big) + \lambda^T f\big(z, g(x, a)\big). \tag{3}$$

With the same definition of the set of policies $\mathcal{U} = (\mathcal{U}^1, \mathcal{U}^2)$ as above, let

$$\overline{\sigma}(z, \lambda, x, u) := \varliminf_{m \to \infty} \frac{1}{m} E_x^u \sum_{i=0}^{m-1} r(z, \lambda, X_i, A_i) \tag{4}$$

A policy pair $u_{z,\lambda} = (u^1_{z,\lambda}, u^2_{z,\lambda}) \in U$ is said to be a saddle point or an equilibrium policy pair for $(z, \lambda)$-associated game with infinite horizon expected average cost criterion, if for all $u^1 \in U^1, u^2 \in U^2$,

$$\overline{\sigma}(z, \lambda, x, u^1, u^2_{z,\lambda}) \leq \overline{\sigma}(z, \lambda, x, u^1_{z,\lambda}, u^2_{z,\lambda}) \leq \overline{\sigma}(z, \lambda, x, u^1_{z,\lambda}, u^2).$$

Let $\hat{f}_{z,\lambda} = (\hat{f}^1_{z,\lambda}, \hat{f}^2_{z,\lambda})$, where $\hat{f}^1_{z,\lambda}, \hat{f}^2_{z,\lambda}$ be some stationary equilibrium policy pair for the expected average problem. The existence of such stationary equilibrium policy pair is well known under our unichain assumption 3, see [Rog69, Sob71] (this extends to the countable case under Simultaneous Doeblin Condition, introduced in [Hor77] Section 11.1, with a communicating condition, or under contraction conditions, see e.g. [Fed78] and [AHS97], respectively). The function

$$\overline{\sigma}(z,\lambda) := \overline{\sigma}(z,\lambda,x,\hat{f}^1_{z,\lambda},\hat{f}^2_{z,\lambda}) \tag{5}$$

is then defined to be the value of the $(z,\lambda)$-associated game, and it is known to be independent on $x$ (which we shall thus omit from the notation). It can be computed using value iteration, (see e.g. [Wal81], chapter 13).

# 4 Limit Hamilton-Jacobi-Isaacs equations for the stochastic hybrid game.

Let us consider Hamilton-Jacobi equations

$$-\frac{\partial B(t,z)}{\partial t} + H\left(z, \frac{\partial B(t,z)}{\partial z}\right) = 0, \quad (t,z) \in [0,T) \times \mathbb{R}^n \tag{6}$$

with Hamiltonian $H(z,\lambda)$ being equal to $-\overline{\sigma}(z,\lambda)$ defined in (5). These equations will be referred to as *Limit Hamilton-Jacobi-Isaacs* (LHJI) equation for the stochastic hybrid game. Let us denote by $B(t,z)$ the viscosity solutions (see definition in the beginning of Section 6) of this equation which satisfy the boundary condition

$$B(T,z) = G(z), \quad \forall z \in \mathbb{R}^n\,. \tag{7}$$

In the following sections, it will be established that the value of our hybrid game converges to $B(t,z)$ as $\epsilon$ tends to zero.

As in stochastic hybrid optimal control problems e.g. [AG97, SAG97], the above results can be considered to be a justification of a decomposition of the stochastic hybrid game into the associated fast game allowing to describe an asymptotically optimal behaviour of the players if the slow parameters are fixed and the LHJI equations responsible for a "near-optimality" of the slow dynamics.

# 5 Main results

Our main result is now formulated as theorem below

**Theorem 1.** *Let Assumption 1–3 be true. Let equation* (6) *with* $H(z,\lambda) = -\overline{\sigma}(z,\lambda)$ *have the unique continuous viscosity solution* $B(t,z)$ *satisfying the boundary condition* (7). *Then the stochastic hybrid game have a value in the limit. That is,*

$$\lim_{\epsilon\to 0} B_\epsilon(t,z,x) \stackrel{def}{=} B(t,z), \tag{8}$$

*with the convergence being uniform on compact set* $[0,T] \times D_1 \times \mathbf{X}$.

We shall use the following property of the value function. It is an equi-continuous type and is crucial in our proof.

**Lemma 2.** *Corresponding to any compact set* $[0,T]\times D_1\times\mathbf{X}$ *there exists continuous functions* $\omega(\alpha), \mu(\alpha)$ *tending to zero as* $\alpha$ *tends to zero such that for any* $(t^i, z^i, x^i) \in [0,T]\times D_1\times\mathbf{X}, i = 1,2$

$$\left|B_\epsilon(t^1, z^1, x^1) - B_\epsilon(t^2, z^2, x^2)\right| \leq \omega(|t^1 - t^2| + |z^1 - z^2|) + \mu(\epsilon), \tag{9}$$

*with*

$$B_\epsilon(T, z, x) = G(z), \quad \forall (z,x) \in D_1 \times \mathbf{X}. \tag{10}$$

This lemma is established in the appendix.

Let us introduce the notation

$$V_\epsilon(t,z) \stackrel{\text{def}}{=} B_\epsilon(t, z, x^*)$$

where $x^*$ is some fixed (but arbitrary) state. By Lemma 2, if $(t,z,x)$ belongs to a compact set $[0,T] \times D_1 \times \mathbf{X}$, then

$$\left|B_\epsilon(t,z,x) - V_\epsilon(t,z)\right| \leq \mu(\epsilon). \tag{11}$$

Hence, to prove (8) it is sufficient to show that

$$\lim_{\epsilon \to 0} V_\epsilon(t,z) = B(t,z), \tag{12}$$

where the convergence is uniform with respect to $(t,z)$ from any compact subset of $[0,T]\times\mathbb{R}^n$. For the sake of brevity we shall refer to this sort of convergence as to $U$-convergence and the corresponding limits will be called $U$-limits.

From Lemma 2, it follows that for $(t^i, z^i) \in [0,T] \times D_1, i = 1,2$

$$\left|V_\epsilon(t^1, z^1) - V_\epsilon(t^2, z^2)\right| \leq \omega\big(|t^1 - t^2| + |z^1 - z^2|\big) + \mu(\epsilon). \tag{13}$$

**Lemma 3.** *Given any sequence* $\epsilon_i$ *tending to zero, one can find a subsequence* $\epsilon_{i_l} = \epsilon_l$ *of this sequence such that there exists the* $U$*-limit*

$$\lim_{\epsilon_l \to 0} V_{\epsilon_l}(t,z) \stackrel{\text{def}}{=} V(t,z). \tag{14}$$

The proof of the lemma is also given in the appendix.

—–

Let us show that any function obtained as $U$-limit in (14) coincides with $B(t,z)$. Notice that, by (13), any such function $V(t,z)$ is continuous on $[0,T] \times \mathbb{R}^n$ and, by (10) and (11), it satisfies the condition

$$V(T,z) = G(z), \quad \forall z \in \mathbb{R}^n.$$

Thus, to show that it coincides with $B(t,z)$ it is enough to show that it is a viscosity solution of (6) with $H(z,\lambda) = -\overline{\sigma}(z,\lambda)$.

# 6 Proof of Main Result

To begin this section, let us recall the definition of viscosity solutions.

**Definition.**

1. *A function $V(t,z)$ is called a viscosity sub-solution of* (6) *if*

$$-\frac{\partial v(\bar{t},\bar{z})}{\partial t} + H\Big(\bar{z}, \frac{\partial v(\bar{t},\bar{z})}{\partial z}\Big) \leq 0,$$

*for any $(\bar{t},\bar{z}) \in [0,T) \times \mathbb{R}^n$ and for each function $v(t,z)$ which has continuous partial derivatives on $[0,T) \times \mathbb{R}^n$ and satisfies the conditions: $v(\bar{t},\bar{z}) = V(\bar{t},\bar{z})$ and $v(t,z) \geq V(t,z)$ in some neighbourhood of $(\bar{t},\bar{z})$.*

2. *A function $V(t,z)$ is called a viscosity super-solution of* (6) *if*

$$-\frac{\partial v(\bar{t},\bar{z})}{\partial t} + H\Big(\bar{z}, \frac{\partial v(\bar{t},\bar{z})}{\partial z}\Big) \geq 0,$$

*for any $(\bar{t},\bar{z}) \in [0,T) \times \mathbb{R}^n$ and for each function $v(t,z)$ which has continuous partial derivatives on $[0,T) \times \mathbb{R}^n$ and which satisfies the conditions: $v(\bar{t},\bar{z}) = V(\bar{t},\bar{z})$ and $v(t,z) \leq V(t,z)$ in some neighbourhood of $(\bar{t},\bar{z})$.*

3. *A function $V(t,z)$ which is both viscosity sub- and super- solution is called a viscosity solution of equation* (6).

——-

**Proof of Theorem 1.** We first note that the hybrid game has the value (this is proved in the Appendix). It permits us to consider the value function $B_\epsilon(t,z,x)$ instead of its upper and lower value when dealing with dynamic programming approach.

Let $\Delta \stackrel{\text{def}}{=} \epsilon K(\epsilon)$ be a function of $\epsilon$ such that $K(\epsilon)$ takes integer value and

$$\lim_{\epsilon\to 0} \Delta = 0, \quad \lim_{\epsilon\to 0} K(\epsilon) = \infty.$$

Take $\bar{t} \in [0,T]$ with initial state of the system $Z_{\bar{t}} = \bar{z}$ and initial state of the Controlled Markov chain $X_{\lfloor \bar{t}/\epsilon \rfloor} = \bar{x}$. Then, it can be shown that

$$\begin{aligned} B_\epsilon^{up}(\bar{t},\bar{z},\bar{x}) &= \inf_{u^2\in U^2} \sup_{u^1\in U^1} E_x^{(u^1,u^2)}\Big\{\int_{\bar{t}}^{T} F(Z_s,Y_s)\,ds + G(Z_T)\Big\} \\ &= \inf_{u^2\in U^2} \sup_{u^1\in U^1} E_x^{(u^1,u^2)}\Big\{\int_{\bar{t}}^{\bar{t}+\Delta} F(Z_s,Y_s)\,ds + B_\epsilon\big(\bar{t}+\Delta, z(\bar{t}+\Delta), x(\bar{t}+\Delta)\big)\Big\}, \end{aligned} \tag{15}$$

where $z(\bar{t}+\Delta) = Z_{\bar{t}+\Delta}$ and $x(\bar{t}+\Delta) = X_{\lfloor(\bar{t}+\Delta)/\epsilon\rfloor}$.

Let $(\bar{t},\bar{z},\bar{x}) \in [0,T) \times D_1 \times \mathbf{X}$. Then, by Assumption 1, $\big(z(\bar{t}+\Delta), x(\bar{t}+\Delta)\big) \in D_1 \times \mathbf{X}$, where $D_1$ are compact sets in $\mathbb{R}^n$. Since the convergence in (14) is uniform with respect to $(t,z)$ from any compact subset of $[0,T] \times \mathbb{R}^n$, there exists a function $\tilde{\nu}(\epsilon_l)$,

$$\lim_{\epsilon_l\to 0} \tilde{\nu}(\epsilon_l) = 0, \tag{16}$$

such that

$$|V_{\epsilon_l}(t,z) - V(t,z)| \leq \tilde{\nu}(\epsilon_l), \ \forall (t,z) \in [0,T] \times D_1.$$

Using this and (11), one obtains from (15)

$$V(\bar{t},\bar{z}) = \inf_{u^2 \in U^2} \sup_{u^1 \in U^1} E_x^{(u^1,u^2)} \Big\{ \int_{\bar{t}}^{\bar{t}+\Delta} F(Z_s,Y_s)\,ds + V\big(\bar{t}+\Delta, z(\bar{t}+\Delta)\big) \Big\} + O(\tilde{\mu}(\epsilon_l)), \quad (17)$$

where

$$\tilde{\mu}(\epsilon_l) = \max\{\mu(\epsilon_l),\ \tilde{\nu}(\epsilon_l), \epsilon_l\}. \quad (18)$$

Let now $v(t,z)$ have continuous partial derivatives and satisfy the conditions: $v(\bar{t},\bar{z}) = V(\bar{t},\bar{z})$ and $v(t,z) \geq V(t,z)$ for $(t,z)$ in some neighbourhood of $(\bar{t},\bar{z})$. From (17) it follows then

$$v(\bar{t},\bar{z}) \leq \inf_{u^2 \in U^2} \sup_{u^1 \in U^1} E_x^{(u^1,u^2)} \Big\{ \int_{\bar{t}}^{\bar{t}+\Delta} F(Z_s,Y_s)\,ds + v\big(\bar{t}+\Delta, z(\bar{t}+\Delta)\big) \Big\} + O(\tilde{\mu}(\epsilon_l)). \quad (19)$$

By definition

$$z(\bar{t}+\Delta) = \bar{z} + \int_{\bar{t}}^{\bar{t}+\Delta} f(Z_s,Y_s)\,ds. \quad (20)$$

By Assumption 1 and 2, the function $f$ is continuous and its arguments belong to compact sets, the second term in the right hand side of (20) is of the order $O\big(\Delta(\epsilon_l)\big)$. Thus, substituting (20) into (19) and taking into account that $v(t,z)$ has continuous partial derivatives, one obtains

$$\frac{\partial v(\bar{t},\bar{z})}{\partial t} + \inf_{u^2 \in U^2} \sup_{u^1 \in U^1} E_x^{(u^1,u^2)} \left\{ \frac{1}{\Delta(\epsilon_l)} \int_{\bar{t}}^{\bar{t}+\Delta} \left[ F(Z_s,Y_s) + \left(\frac{\partial v(\bar{t},\bar{z})}{\partial z}\right)^T f(Z_s,Y_s) \right] ds \right\} + \frac{O\big(\tilde{\mu}(\epsilon_l)\big)}{\Delta(\epsilon_l)} + \frac{o\big(\Delta(\epsilon_l)\big)}{\Delta(\epsilon_l)} \geq 0. \quad (21)$$

Notice that for any $s \in [\bar{t}, \bar{t}+\Delta(\epsilon_l)]$

$$Z_s = \bar{z} + \int_{\bar{t}}^{\bar{t}+\Delta(\epsilon_l)} f(Z_s,Y_s)\,ds$$

Hence,

$$\big\|Z_s - \bar{z}\big\| \leq M\Delta(\epsilon_l).$$

This and Assumption 2 imply that

$$\frac{1}{\Delta(\epsilon_l)} \int_{\bar{t}}^{\bar{t}+\Delta} \left[ F(Z_s,Y_s) + \left(\frac{\partial v(\bar{t},\bar{z})}{\partial z}\right)^T f(Z_s,Y_s) \right] ds = \frac{1}{\Delta(\epsilon_l)} \int_{\bar{t}}^{\bar{t}+\Delta} \left[ F(\bar{z},Y_s) + \left(\frac{\partial v(\bar{t},\bar{z})}{\partial z}\right)^T f(\bar{z},Y_s) \right] ds + O\big(\Delta(\epsilon_l)\big) \quad (22)$$

Denote

$$m(\epsilon_l) = \left\lfloor \frac{\bar{t}+\Delta(\epsilon_l)}{\epsilon_l} \right\rfloor - \left\lfloor \frac{\bar{t}}{\epsilon_l} \right\rfloor$$

and note that

$$\left| m(\epsilon_l) - \frac{\Delta(\epsilon_l)}{\epsilon_l} \right| \leq 1,$$

then

$$\left| \frac{1}{m(\epsilon_l)} - \frac{\epsilon_l}{\Delta(\epsilon_l)} \right| \leq \frac{\epsilon_l^2}{\Delta^2(\epsilon_l)} \left( \frac{1}{1 - \epsilon_l/\Delta(\epsilon_l)} \right).$$

From this and (3), it follows that there exist positive constants $L_1$ and $L_2$ such that

$$\begin{aligned} \left\| \frac{1}{\Delta(\epsilon_l)} \int_{\bar{t}}^{\bar{t}+\Delta(\epsilon_l)} \Big[ F(\bar{z}, Y_s) + \lambda^T f(\bar{z}, Y_s) \Big] \, ds - \frac{\epsilon_l}{\Delta(\epsilon_l)} \sum_{i=\lfloor \bar{t}/\epsilon_l \rfloor + 1}^{\lfloor \bar{t}/\epsilon_l \rfloor + m(\epsilon_l)} r(\bar{z}, \lambda, X_i, A_i) \right\| &\leq L_1 \frac{\epsilon_l}{\Delta(\epsilon_l)} \\ \left\| \frac{\epsilon_l}{\Delta(\epsilon_l)} \sum_{i=\lfloor \bar{t}/\epsilon_l \rfloor + 1}^{\lfloor \bar{t}/\epsilon_l \rfloor + m(\epsilon_l)} r(\bar{z}, \lambda, X_i, A_i) - \frac{1}{m(\epsilon_l)} \sum_{i=\lfloor \bar{t}/\epsilon_l \rfloor + 1}^{\lfloor \bar{t}/\epsilon_l \rfloor + m(\epsilon_l)} r(\bar{z}, \lambda, X_i, A_i) \right\| &\leq L_2 \frac{\epsilon_l}{\Delta(\epsilon_l)}. \end{aligned} \tag{23}$$

Using (22) – (23), one may obtain from (21)

$$\frac{\partial v(\bar{t}, \bar{z})}{\partial t} + \inf_{u^2 \in U^2} \sup_{u^1 \in U^1} E_x^{(u^1, u^2)} \frac{1}{m(\epsilon_l)} \sum_{i=\lfloor \bar{t}/\epsilon_l \rfloor + 1}^{\lfloor \bar{t}/\epsilon_l \rfloor + m(\epsilon_l)} r\Big( \bar{z}, \frac{\partial x(\bar{t}, \bar{z})}{\partial z}, X_i, A_i \Big)$$
$$+ O\Big( \frac{\epsilon_l}{\Delta(\epsilon_l)} \Big) + O\big(\Delta(\epsilon_l)\big) + \frac{O\big(\tilde{\mu}(\epsilon_l)\big)}{\Delta(\epsilon_l)} + \frac{o\big(\Delta(\epsilon_l)\big)}{\Delta(\epsilon_l)} \geq 0. \tag{24}$$

Define now $\Delta(\epsilon_l)$ as follows

$$\Delta(\epsilon_l) = \sqrt{\tilde{\mu}(\epsilon_l)} \Rightarrow \frac{O\big(\tilde{\mu}(\epsilon_l)\big)}{\Delta(\epsilon_l)} = O\Big( \sqrt{\tilde{\mu}(\epsilon_l)} \Big).$$

Hence, passing to the limit in (24) as $\epsilon_l$ tends to zero and taking into account the associated game with the existence of stationary equilibrium policy pair as stated in (5), one obtains

$$\frac{\partial v(\bar{t}, \bar{z})}{dt} + \overline{\sigma} \left( \bar{z}, \frac{\partial v(\bar{t}, \bar{z})}{\partial z} \right) \geq 0 \Rightarrow -\frac{\partial v(\bar{t}, \bar{z})}{dt} + H \left( \bar{z}, \frac{\partial v(\bar{t}, \bar{z})}{\partial z} \right) \leq 0.$$

This establishes that $V(t, z)$ is a viscosity sub-solution of (6) on $[0, T) \times \mathbb{R}^n$.

Similarly, taking $v(t, z)$ having continuous partial derivatives and satisfying the conditions: $v(\bar{t}, \bar{z}) = V(\bar{t}, \bar{z})$ and $v(t, z) \leq V(t, z)$ in some neighbourhood of $(\bar{t}, \bar{z}) \in [0, T) \times \mathbb{R}^n$, one can obtain that

$$-\frac{\partial v(\bar{t}, \bar{z})}{dt} + H \left( \bar{z}, \frac{\partial v(\bar{t}, \bar{z})}{\partial z} \right) \geq 0$$

which means that $V(t, z)$ is a viscosity super-solution of (6) on $[0, t) \times \mathbb{R}^n$. Thus, $V(t, z)$ is a viscosity solution (6) on $[0, t) \times \mathbb{R}^n$ and, consequently, it coincides with $B(t, z)$.

This proves that $B_\epsilon(t, z, x)$ $U$-converges (as $\epsilon$ tends to zero) to $B(t, z)$ since, otherwise, by Lemma 3, one would be able to choose a subsequence $\epsilon_l$ tending to zero such that the $U$-limit (14) does not coincide with $B(t, z)$. □

# 7 Appendix

In this appendix we first present some general properties of the original game with fixed $\epsilon$, which allows us to obtain some properties of the limit game.

We first show that the original game is equivalent to a *stochastic game with finite state and action spaces.* This will allow us to use standard results to obtain the representation of the value and optimal policies.

## 7.1 An equivalent stochastic game

**Lemma 4.** *For a fixed $\epsilon$, the original hybrid game is equivalent to a finite-stage stochastic (Markov) game with finite state and action spaces and it has a* **value** *$B_\epsilon(t,z,x)$. That is*

$$B_\epsilon(t,z,x) = B_\epsilon^{up}(t,z,x) = B_\epsilon^{lo}(t,z,x).$$

In fact, introduce the following stochastic game:

- *State space*: consists of the *histories*

$$\mathcal{X} := \bigcup_{l=0}^{\lfloor (T-t)\epsilon^{-1} \rfloor} \mathbf{H}_l \quad \text{where } \mathbf{H}_l := \big\{(x_0, a_0^1, a_0^2, x_1, a_1^1, a_1^2, ..., x_l)\big\}.$$

  An element of the state space will be denoted by $h$; $n(h)$ will denote the length of the horizon.

- *Action spaces*: unchanged, i.e. $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2$.

- *Transition probabilities* these are obvious; for

$$h^1 = \big\{(x_0, a_0^1, a_0^2, x_1, a_1^1, a_1^2, ..., x_l)\big\}, h^2 = \big\{(y_0, b_0^1, b_0^2, y_1, b_1^1, b_1^2, ..., y_k)\big\},$$

  we have

$$\hat{\mathcal{P}}_{h^1,\alpha^1,\alpha^2,h^2} = \mathcal{P}_{x_l,\alpha^1,\alpha^2,y_k}$$

  for

$$k = l+1, x_0 = y_0, a_0^1 = b_0^1, a_0^2 = b_0^2, x_1 = y_1, ..., x_l = y_l, b_k^1 = \alpha^1, b_k^2 = \alpha^2,$$

  and zero 0 otherwise.

- *Immediate costs*:

$$c(t,z,h;a^1,a^2) = \int_{\epsilon n(h)}^{\epsilon n(h)+\epsilon} F(Z_s, Y_s)\, ds$$

  for $n(h) < \lfloor (T-t)\epsilon^{-1} \rfloor$, and

$$c(t,z,h;a^1,a^2) = \int_{\epsilon n(h)}^{T} F(Z_s, Y_s)\, ds + G(Z_T)$$

  for $n(h) = \lfloor (T-t)\epsilon^{-1} \rfloor$. Note that the immediate cost is parameterized by the initial $z$ and $t$. We did not write the immediate cost explicitly, however the random variables $Z_s, Y_s$ and $Z_T$ appearing in the immediate cost are fully determined by $h$ and the actions $a^1, a^2$.

Let us define the payoff of the new game: for any $h$ such that $n(h) \leq \lfloor (T-t)\epsilon^{-1} \rfloor$, we set

$$\hat{J}_\epsilon(t, z, h; v^1, v^2) = E_h^{(v^1,v^2)} \Big\{ \int_{t+\epsilon n(h)}^T F(Z_s, Y_s)\, ds + G(Z_T) \Big\}.$$

Note that each policy $u^i$ for player $i$ in the original game has an obvious equivalent Markov policy $v^i$ in this new game that achieves the same costs. It is thus simple to show that one may restrict to Markov policies in the new game (optimal Markov policies will depend of course on $z$ and $t$). The original policies generate the same costs in the original game as their equivalent new policies in the new game:

$$\hat{J}_\epsilon(t, z, h; v^1, v^2) = J_\epsilon(t, z, x; u^1, u^2),$$

where $h = x$.

Since the new game is a standard stochastic game with finite number of states and actions, it has a *value* (see e.g. Van Der Wal [Wal81], chapter 10). We conclude that the lower value and the upper value in the original game coincide and are equal to this value. Note also that dynamic programming can be used to characterize the value and optimal policies for both players. □

## 7.2 Proof of Lemma 2

It follows from arguments as in [Gai96] that there exists some real number $\omega$ such that for any policies $u^1$ and $u^2$ for the two players and any $z^1, z^2, t^1, t^2$ and $x$,

$$|J_\epsilon(t^1, z^1, x : u^1, u^2) - J_\epsilon(t^2, z^2, x : u^1, u^2)| \leq \omega(|t^1 - t^2| + |z^1 - z^2|) + O(\epsilon). \qquad (25)$$

This implies that (9) holds for the case where $x^1 = x^2$.

To conclude the proof, it thus suffices to show that for any $z$ and $t$,

$$|J_\epsilon(t, z, x^1; u^1, u^2) - J_\epsilon(t, z, x^2; u^1, u^2)| < \mu(\epsilon)$$

where $\mu$ is as in Lemma 2. We do this next. Choose some arbitrary $x^*$.

Denote $\eta = \inf\{n : X_n = x^*\}$. Then

$$\sup_{u^1,u^2,x} E_x^{u^1,u^2} \eta < \infty. \qquad (26)$$

Indeed, there exists a pure stationary pair $u^1, u^2$ that achieves this sup, since achieving the sup is equivalent to a problem of maximizing some total expected cost in a transient MDP with finite state and action spaces (see e.g. [Kal83] or [Hor77]). That the sup is finite follows from the unichain assumption 3.

Note that

$$\begin{aligned}
& B_\epsilon(t^1, z^1, x) - B_\epsilon(t^2, z^2, x) \\
&= \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} J_\epsilon(t^1, z^1, x; u^1, u^2) - \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} J_\epsilon(t^2, z^2, x; u^1, u^2) \\
&\leq \sup_{u^1 \in U^1} \left( \inf_{u^2 \in U^2} J_\epsilon(t^1, z^1, x; u^1, u^2) - \inf_{u^2 \in U^2} J_\epsilon(t^2, z^2, x; u^1, u^2) \right) \\
&\leq \sup_{u^1, u^2} |J_\epsilon(t^1, z^1, x; u^1, u^2) - J_\epsilon(t^2, z^2, x; u^1, u^2)|.
\end{aligned}$$

Since the same holds for $B_\epsilon(t^2, z^2, x) - B_\epsilon(t^1, z^1, x)$ we conclude from the last equation and from (25) that

$$|B_\epsilon(t^1, z^1, x) - B_\epsilon(t^2, z^2, x)| \le \omega(|t^1 - t^2| + |z^1 - z^2|) + O(\epsilon). \tag{27}$$

Denote $\tau = \min(T, t + \eta\epsilon)$ and $\sigma = t + \eta\epsilon$. Now, the optimality princile implies that

$$\begin{aligned} & B_\epsilon(t, z, x) \\ = & \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} E_x^{u^1,u^2} \left( \int_t^T F(Z_s, Y_s)\, ds + G(Z_T) \right) \\ = & \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} E_x^{u^1,u^2} \left( \int_t^\tau F(Z_s, Y_s)\, ds + \int_\tau^T F(Z_s, Y_s)\, ds + G(Z_T) \right) \\ = & \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} E_x^{u^1,u^2} \left( \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} + B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \right) \end{aligned}$$

Thus,

$$\begin{aligned} & B_\epsilon(t, z, x^1) - B_\epsilon(t, z, x^2) \\ = & \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} E_{x^1}^{u^1,u^2} \Big( \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} + B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \Big) \\ & - \sup_{u^1 \in U^1} \inf_{u^2 \in U^2} E_{x^2}^{u^1,u^2} \Big( \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} + B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \Big) \\ \le & \sup_{u^1,u^2} E_{x^1}^{u^1,u^2} \left| \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} \right| \\ & + \sup_{u^1,u^2} E_{x^2}^{u^1,u^2} \left| \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} \right| \\ & + \sup_{u^1,u^2} \left| E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \right|. \end{aligned}$$

Since the same bound holds also for $B_\epsilon(t, z, x^2) - B_\epsilon(t, z, x^1)$, we conclude that

$$\begin{aligned} & \left| B_\epsilon(t, z, x^1) - B_\epsilon(t, z, x^2) \right| & (28) \\ \le\ & 2 \sup_{u^1,u^2,x} \left| E_x^{u^1,u^2} \Big\{ \int_t^\tau F(Z_s, Y_s)\, ds + G(Z_T) 1\{\sigma > T\} \Big\} \right| & (29) \\ & + \sup_{u^1,u^2} \left| E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \right| & (30) \end{aligned}$$

The first term above is $O(\epsilon)$ since $F$ and $G$ are bounded, since $\tau \le t + \eta\epsilon$ and due to (26).

Next we bound the second term. We have

$$\begin{aligned} & \left| E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma \le T\} \right| \\ = & \Big| E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) \\ & - E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma > T\} + E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) 1\{\sigma > T\} \Big| \\ \le & \left| E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) \right| + \sup_{x,u^1,u^2} 2M(T - \tau + 1) P_x^{u^1,u^2}(\sigma > T). \end{aligned}$$

Due to (26), it follows that $P_x^{u^1,u^2}(\sigma > T)$ is of order of $O(\epsilon)$. It remains to estimate the first term in the right hand side of the above inequality.

Now, consider an arbitrary augmented probability space on which the two state and action trajectories are defined simultateously: those that start from initial states $x^1$ and $x^2$ respectively, and for which the marginal distribution of each trajectory separately is given by the corresponding probabilities $P_{x^1}^{u^1,u^2}$ and $P_{x^2}^{u^1,u^2}$, respectively. Let $\overline{P}_{x^1,x^2}^{u^1,u^2}$ be the probability measure governing the augmented probability space, and we denote by $\overline{E}_{x^1,x^2}^{u^1,u^2}$ the corresponding expectation.

Let $t^1$ and $t^2$ be the times at which we reach the state $x^*$ starting from states $x^1$ and state $x^2$, respectively. Let $z^1$ and $z^2$ be the value of the $z$ variable at those instants. Then we have by (27)

$$
\begin{aligned}
&|E_{x^1}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*) - E_{x^2}^{u^1,u^2} B_\epsilon(\tau, Z_\tau, x^*)| \\
= \;& |\overline{E}_{x^1,x^2}^{u^1,u^2}(B_\epsilon(t^1, z^1, x^*) - B_\epsilon(t^2, z^2, x^*))| \leq \omega \overline{E}_{x^1,x^2}^{u^1,u^2}(|t^1 - t^2| + |z^1 - z^2|).
\end{aligned}
$$

We have

$$
\sup_{x^1,x^2,u^1,u^2} \overline{E}_{x^1,x^2}^{u^1,u^2}|t^1-t^2| \leq \sup_{x^1,x^2,u^1,u^2} \overline{E}_{x^1,x^2}^{u^1,u^2}|t^1-t| + \sup_{x^1,x^2,u^1,u^2} \overline{E}_{x^1,x^2}^{u^1,u^2}|t^2-t| \leq 2\epsilon \sup_{u^1,u^2,x} E_x^{u^1,u^2}\eta.
$$

Due to the bounded function $f$, one may get

$$
\sup_{x^1,x^2,u^1,u^2} E_x^{u^1,u^2}|z^1 - z^2| \leq 2M\epsilon \sup_{u^1,u^2,x} E_x^{u^1,u^2}\eta.
$$

Thus, we see that (30) is bounded by $2\epsilon(M+1)\sup_{u^1,u^2,x} E_x^{u_1,u_2}\eta$. This concludes the proof. $\square$

Note that in the above proof we couple in some sense two systems that start in different initial states, in order to be able to compute expressions such as $E(t^1 - t^2)$. However, we did not have to make any particular assumption on the joint distribution between those two systems in order to obtain the required bounds.

## 7.3 Proof of Lemma 3

Firstly, one can establish that given a compact subset $[0,T] \times D$, $D \subset \mathbb{R}^n$, the following space

$$
\mathcal{V} = \left\{ V_\epsilon(t,z),\ \epsilon \to 0 \right\}
$$

where $V_\epsilon(t,z)$ are continuous functions on $[0,T] \times D$ satisfying the property (13) under sup metric, is complete and totally bounded. So, it is compact.

Hence, every sequence in $\mathcal{V}$ in the form $\left\{ V_{\epsilon_n},\ n \to \infty \right\}$ contains a subsequence $\left\{ V_{\epsilon_l} \right\}$ such that the convergence (14) will be uniform with respect to $(t,z) \in [0,T] \times D$.

Notice that our claim is stronger than that since U-convergence is uniform with respect to $(t,z)$ from *any compact subset* of $[0,T] \times \mathbb{R}^n$. So, we just remove the dependence of compact set $D$ by choosing a sequence of subsets $[0,T] \times D^l$, where $D^l \subset \mathbb{R}^n$ is the closed ball centered at zero and having the radius $l$, $l = 1, 2, \ldots$, and then, using a diagolization procedure one can construct a subsequence providing the required $U$-convergence. $\square$

## References


[AG94] E. Altman and V. Gaitsgory. "A Hybrid (differential-stochastic) zero-sum Game with fast stochastic part." In *Preprint volume of the sixth International Symposium on Dynamic Games and Applications*, pp. 475 – 480, St-Jovite, Quebec, Canada, July 13-15 1994. The International Society of Dynamic Games.

[AG97] E. Altman and V. Gaitsgory. "Asymptotic Optimization of a nonlinear Hybrid System Governed by a Markov Decision Process." *SIAM J. Control and Optimization*, **35**(6):2070–2085, November 1997.

[AHS97] E. Altman, A. Hordijk, and F. M. Spieksma. "Contraction Conditions for Average and $\alpha$-Discount Optimality in Countable State Markov Games with Unbounded Rewards." *MOR*, **22**:588–618, August 1997.

[Fed78] A. Federgruen. "On N-person Stochastic Games with Denumerable State Space." *Adv. Appl. Prob.*, **10**:452–471, 1978.

[Gai96] V. Gaitsgory. "Limit Hamilton-Jacobi-Isaacs equations for singularly perturbed zero-sum differential games." *Mathematical Analysis and Applications*, **202**(0352):862–899, 1996.

[Hin70] K. Hinderer. *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, volume **33** of *Lecture Notes in Operations Research and Mathematical Systems*. Springer-Verlag, 1970.

[Hor77] A. Hordijk. *Dynamic Programming and Markov Potential Theory*. Mathematical Centre Tracts 51. Mathematisch Centrum, Amsterdam, second edition, 1977.

[Kal83] L. C. M. Kallenberg. *Linear Programming and Finite Markovian Control Problems*. Mathematical Centre Tracts 148. Mathematical Centre, Amsterdam, 1983.

[Kle76] L. Kleinrock. *Queuing systems, Volume II: Computer applications*. John Wiley, New York, 1976.

[Rog69] P. Rogers. "Nonzero-Sum Stochastic Games." Technical Report ORC 69-8, Berkeley, University of California, Operation Research Centre, 1969.

[SAG97] P. Shi, E. Altman, and V. Gaitsgory. "On asymptotic optimization for A Class of Nonlinear Stochastic Hybrid Systems." *Mathematical Methods of Operations Research*, **46**(1):1228–1249, 1997.

[Sob71] M. Sobel. "Noncooperative Stochastic Games." *Annal of Mathematical Statistics*, **42**:1930–1935, 1971.

[SZ94] S. P. Sethi and Q. Zhang. *Hierarchical decision making in stochastic manufacturing systems*. Birkhauser, Boston-Basel, Berlin, 1994.

[Wal81] J. Van Der Wal. *Stochastic Dynamic Programming: Successive Approximations and Nearly Optimal Strategies for Markov Decision Processes and Markov Games*. Mathematical Centre Tracts **139**. Mathematisch Centrum, Amsterdam, 1981.