Mathematical Methods of Operations Research (1998) 47:289–315

On Asymptotic Optimization of a Class of Nonlinear Stochastic Hybrid Systems[1]

Peng Shi

Centre for Industrial and Applied Mathematics, School of Mathematics, The University of South Australia, SA 5095, Australia
e-mail: matsp@zarniwoop. levels.unisa.edu.au

Eitan Altman

INRIA, BP93, 2004 Route des Lucioles, 06902 Sophia Antipolis Cedex, France
e-mail: eitan. altman@sophia.inria.fr

Vladimir Gaitsgory

Centre for Industrial and Applied Mathematics, School of Mathematics, The University of South Australia, SA 5095, Australia
e-mail: mavg@lux.levels.unisa.edu.au

Abstract: We consider the problem of control for continuous time stochastic hybrid systems in finite time horizon. The systems considered are nonlinear: the state evolution is a nonlinear function of both the control and the state. The control parameters change at discrete times according to an underlying controlled Markov chain which has finite state and action spaces. The objective is to design a controller which would minimize an expected nonlinear cost of the state trajectory. We show using an averaging procedure, that the above minimization problem can be approximated by the solution of some deterministic optimal control problem. This paper generalizes our previous results obtained for systems whose state evolution is linear in the control.

Key Words: Hybrid stochastic systems, Markov decision processes, nonlinear systems.

1 Introduction

We consider in this paper a controlled hybrid system: control actions are taken periodically at discrete times, and they influence in some probabilistic sense the parameters of a system that evolves in continuous time. More precisely, the state of the continuous part of system is described by some nonlinear differential equation; its dynamics is parameterized by some vector which may take a finite number of different values. The value of these param-

[1] This work is supported by the Australian Research Council. All correspondence should be directed to the first author.

eters are functions of a controlled Markov chain, that has jumps at some fixed moments of time.

As an example for such a system, consider the admission control into a telecommunications network. The state of the continuous time system may be taken to be the amount of workload (i.e. the transmission time required by the information packets) in the different nodes. The dynamics of this state is determined by the number, the routes and the type of sessions that are present in the network. These can be described by a Markov chain that takes a finite number of states. This Markov chain has transitions corresponding to the end of sessions, or to the beginning of new sessions. The latter, however, may be controlled by the network; the actions available are thus to accept or reject a new coming call, having some requirements for its routing, bandwidth, and duration.

The objective is to minimize an expected nonlinear cost of the state trajectory over a finite horizon problem. We consider the behavior of the hybrid system when the time between transitions of the Markov chain is small. Using an averaging procedure, we show that the above minimization problem can be approximated by the solution of some deterministic optimal control problem. This paper generalizes our previous results obtained for systems whose state evolution is linear in the control [2, 3].

Note that the problem we deal with here is closely related with singularly perturbed stochastic optimal control problems which were intensively studied in the literature, see, e.g. [4, 5, 6, 7, 9, 10, 19, 20, 21, 22]. Our result differs, however, from the ones obtained in the references above and it can be regarded as an extension of the averaging technique developed for deterministic singularly perturbed optimal control problems (see, e.g. [16]).

Notation: Throughout this paper $\mathbf{R}^n$ and $\mathbf{R}^{n\times m}$ denote, respectively, the n dimensional Euclidean space and the set of all $n \times m$ real matrices. P_x^u and E_x^u are, respectively, the probability measure and mathematical expectation corresponding to an initial distribution x and a policy u (which will be specified later). $\|\cdot\|$ will refer to the L_1 norm in the finite dimensional space. That is, for $q \in \mathbf{R}^k$ and $A \in \mathbf{R}^{n\times k}$, $\|q\| = \max_{i=1,2,\ldots,k}|q_i|$ and $\|A\| = \max_{\|q\|=1}\|Aq\|$. $\lfloor x \rfloor$ stands for the greatest integer which is smaller than or equal to x. To distinguish the variables, we use capital letters for stochastic variables, i.e. Z and H, and small letters for deterministic variables, i.e. z and h.

The remainder of this paper proceeds as follows: In Section 2, we describe the class of systems under consideration and formulate the problem. Also some preliminaries are recalled. Section 3 presents the main result about an approximation of the problem of optimal control of the hybrid stochastic system by a deterministic optimal control problem. Approaches that allow to characterize the solution of the deterministic optimal control problem and to use this solution to obtain an asymptotically optimal policy for the hybrid stochastic system are discussed. The fundamental lemmas which are used to achieve our main result are proved in Appendices.

2 Problem Formulation and Preliminaries

Consider the following hybrid stochastic control system

$$\dot{Z}(t) = f(Z(t), Y(t)) \ , \quad Z(0) = z_0, \ t \in [0,1] \ , \tag{2.1}$$

where $Z(t)$ is the state, z_0 is the initial state, $Y(t)$ is a control, $f(\cdot,\cdot) :\rightarrow \mathbf{R}^n$ is a function. The controls $Y(t)$ will be defined below as a piecewise constant function of time taking there values in a finite subset of $\mathbf{R}^k$. We shall denote this subset as D_2. By D_1 we shall denote a compact subset of $\mathbf{R}^n$ which will be assumed to contain the solutions $Z(t)$ of (2.1) obtained with the admissible controls.

$Y(t)$ is not chosen directly by the controller, but is obtained as a result of controlling the following underlying stochastic discrete event system. Let ε be the basic time unit. Time is discretized, i.e. transitions occur at times $t = n\varepsilon, n = 0, 1, \ldots, \lfloor \varepsilon^{-1} \rfloor$. There is a finite state space $\mathbf{X} = \{1, 2, \ldots, N\}$ and a finite action space $\mathbf{A}$. If a state is v and an action a is chosen then the next state is w with the probability P_{vaw}. A policy $u = \{u_0, u_1, \ldots\}$ in the set of policy $\mathbf{U}$ is a sequence of probability measures on $\mathbf{A}$; at each time $t = n\varepsilon$ the controller chooses u_n based on the history of all previous states and actions, as well as the present state. In this paper, our attention is concentrated on the following classes of policies:

- Markov policies, denoted by $\mathscr{M}$, i.e. policies for which u_n depends only on the current state, and does not depend on previous states and actions.
- Stationary policies, denoted by $\mathscr{S}$, i.e. policies for which u_n depends only on the current state, and does not depend on previous states and actions nor on the time.

The stochastic process $\{X_n, A_n\}$ is known as a controlled Markov chain, or Markov decision process (MDP), and well studied by researchers in the past three decades, see, e.g. [13] and the references therein. We assume throughout the paper that under any stationary policy, the state space forms an aperiodic Markov chain such that all states communicate (regular Markov chain). The results of the paper hold in fact under weaker ergodicity assumptions, however the restricted assumption makes the presentation clearer.

We make the following assumption on the nonlinear function $f(Z(t), Y(t))$.

Assumption 2.1: There exist positive numbers C and M such that

$$\|f(z_1, y) - f(z_2, y)\| \leq C\|z_1 - z_2\| \ , \quad \forall z_i \in D_1, \ i = 1, 2; \ y \in D_2$$

$$\|f(z, y)\| \leq M \ , \quad \forall (z, y) \in D_1 \times D_2$$

For convenience, we denote by $\mathbf{H}$ the set of all possible states-actions histories which can be observed till time $\lfloor \varepsilon^{-1} \rfloor$:

$$\mathbf{H} = \bigcup \{h\} , \quad h = \{(x_n, a_n), n = 0, 1, \ldots, \lfloor \varepsilon^{-1} \rfloor\} .$$

Let $\mathscr{F}$ be the σ-algebra of all subsets of $\mathbf{H}$. Each policy u and initial state x determine a probability measure on $\mathscr{F}$, on which the stochastic state and action process $H = \{X_n, A_n, n = 0, 1, \ldots, \lfloor \varepsilon^{-1} \rfloor\}$ is defined.

Let $g : \mathbf{X} \times \mathbf{A} \to \mathbf{R}^k$, be some given vector-valued function. Then, $Y(t)$ in (2.1) is given by

$$Y(t) = g(X_{\lfloor t/\varepsilon \rfloor}, A_{\lfloor t/\varepsilon \rfloor}) . \tag{2.2}$$

The set D_2 is defined as the set of all possible values of g on $\mathbf{X} \times \mathbf{A}$. The system (2.1) with $Y(t)$ defined in (2.2) is a hybrid system. First, $Y(t)$ changes its values via some random jumps whereas $Z(t)$ is smooth (differentiable) function of time. Secondly, $Y(t)$ being controlled "stochastically" through controlling the transition probabilities plays by itself the role of a "direct" control with respect to $Z(t)$.

Remark 2.1: It should be noted that the system (2.1) *can be used to represent many important physical systems. It is natural in the control of inventories or of production, where we deal with material whose quantity may "slowly" change in a continuous way. Breakdowns, repairs and other control decisions yield the underlying MDP. Our model may also be used in the control of highly loaded queuing networks for which the fluid approximation holds, see, e.g.* [18]. *The slow variables Z_t may then represent the number of customers in the different queues whereas the underlying MDP may correspond to routing, or admission of some calls.* □

For fixed ε and $X_0 = x$, the control problem considered in this paper is as follows: find a policy u that achieves

$$\mathbf{Q}_\varepsilon : F_x^\varepsilon(z_0) = \inf_{u \in U} E_x^u G(Z_1) , \tag{2.3}$$

where Z_1 is the solution of (2.1), and the cost function G: $\mathbf{R}^n \to \mathbf{R}$, satisfies the following assumption:

Assumption 2.2: There exists a positive number C_G such that

$$\|G(z_1) - G(z_2)\| \leq C_G \|z_1 - z_2\|$$

for any z_1 and $z_2 \in D_1$.

Our objective is to construct a policy u_ε which is asymptotically optimal for $\mathbf{Q}_\varepsilon$. More precisely, the difference between the cost under this policy and $F_x^\varepsilon(z_0)$ converges to zero as $\varepsilon \to 0$.

In the reminder of this section, we recall a general result on MDPs that establishes the uniform convergence of the state-action frequencies to their limits.

Let m and K be arbitrary integer numbers and

$$h = \{x_{m+1}, a_{m+1}, \ldots, x_{m+K}, a_{m+K}\}$$

be a state-action trajectory of the length K. Let

$$\psi_m^K(h; w, a) \triangleq \frac{1}{K} \sum_{n=m+1}^{m+K} 1\{x_n = w, a_n = a\} , \quad \psi_m^K(h) = \{\psi_m^k(h; w, a)\} ,$$

where $1\{x_n = w, x_n = a\}$ is the indicator function (that is, it is equal to one when $x_n = w, a_n = a$, and it is equal to zero otherwise).

If H is a random realization of h, then we denote

$$\psi_m^K(w, a) = \psi_m^K(H; w, a)$$

$$\psi_m^K \triangleq \{\psi_m^K(w, a)\}_{w,a} .$$

Let $\eta(u) = \{\eta(u; w, a)\}$ be the vector of steady state probabilities of state-action, (w, a), pairs obtained when using a stationary policy u, i.e.

$$\eta(u; w, a) = \lim_{n \to \infty} P_x^u(X_n = w, A_n = a) . \tag{2.4}$$

Notice that due to the ergodicity assumption on our model, $\eta(u; w, a)$ does not depend on the initial distribution.

Let

$$W \stackrel{\text{def}}{=} \bigcup_{u \in S} \{\eta(u)\} \tag{2.5}$$

and denote

$$d_K^1 = dist\{\psi_0^K, W\} = \inf_{\eta \in W} \|\psi_0^K - \eta\} \ .$$

It follows from [12] and [1] that for any policy u and initial distribution ξ,

$$\lim_{K \to \infty} d_K^1 = 0 \ , \quad P_\xi^u a.s.$$

This implies by the bounded convergence theorem that

$$\lim_{K \to \infty} E_\xi^u d_K^1 = 0 \ .$$

For any policy u for which the limit

$$\psi_0 := \lim_{K \to \infty} \psi_0^K$$

exists $P_\xi^u - a.s.$ (in particular, for any policy $u \in S$, the above limit exists and $\psi_0(w, a) = \eta(u; w, a)$ independently on the initial distribution ξ), define

$$d_K^2 = \|\psi_0^K - \psi_0\| \ .$$

Lemma 2.1: [3] The following holds

$$\lim_{K \to \infty} \sup_{\xi} \sup_{u \in \mathbf{U}} E_\xi^u d_K^1 = 0$$

$$\lim_{K \to \infty} \sup_{\xi} \sup_{u \in S} E_\xi^u d_K^2 = 0$$

Before ending this section, we recall an inequality which will be used in the proof of our main results.

Lemma 2.2: [16] Let $N(\varepsilon)$ be a function of ε with its values being natural numbers tending to infinity as ε tends to zero. Then, if there exist a scalar $L > 0$ and a function of ε, $\phi(\varepsilon) \geq 0$, such that the nonnegative numbers Δ_l satisfy the inequality

$$\Delta_{l+1} \leq \Delta_l + LN(\varepsilon)^{-1}\Delta_l + \phi(\varepsilon)N(\varepsilon)^{-1} , \quad \Delta_0 = 0, \, l = 0, 1, \ldots, k < N(\varepsilon)$$

Δ_l also satisfy the inequality

$$\Delta_l \leq \phi(\varepsilon)L^{-1}e^L , \quad l = 1, \ldots, k+1 .$$

3 Main Results

Let us define the point-to-set mapping $V(z)$: $D_1 \to 2^{R^n}$

$$\begin{aligned} V(z) &= \bigcup_{u \in S} \sum_{(w,a)} \eta(u; w, a) f(z, g(w, a)) \\ &= \bigcup_{\eta \in W} \sum_{w,a} \eta(w, a) f(z, g(w, a)) . \end{aligned} \tag{3.1}$$

Notice that $V(z)$ is compact and convex (and even polyhedral) since W has all these properties (see [12]).

Consider the differential inclusion

$$\dot{z}(t) \in V(z(t)) , \quad z(0) = z_0 . \tag{3.2}$$

Lemma 3.1: Corresponding to any solution $\bar{z}(t)$ of (3.2), there exists a Markov policy $u_\varepsilon(\bar{z})$ such that the random trajectory $Z(t)$ of (2.1) obtained with this policy $u_\varepsilon(\bar{z})$ satisfies the inequality

$$\max_{t \in [0,1]} E_x^{u_\varepsilon(\bar{z})} \| Z(t) - \bar{z}(t) \| \leq \gamma(\varepsilon) , \tag{3.3}$$

where $\gamma(\varepsilon)$ satisfies $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$.

Proof: See Appendix A. ■

Lemma 3.1 shows that the solutions of differential inclusion (3.2) are approximated by random trajectories of (2.1).

Lemma 3.2: There exists a function $\tilde{z}_\varepsilon(t,h)$ *such that*

i) *for a fixed* $h \in \mathbf{H}$, *it is a solution of* (3.2);
ii) *for any policy* u, $\tilde{z}_\varepsilon(t) = \tilde{z}_\varepsilon(t,H)$ *satisfies*

$$\max_{t\in[0,1]} E_x^u \|Z(t) - \tilde{z}_\varepsilon(t)\| \le \gamma(\varepsilon) \ , \tag{3.4}$$

where $Z(t)$ *is the solution of* (2.1) *and* $\gamma(\varepsilon)$ *is as in Lemma* 3.1.

Proof: See Appendix B. ■

Remark 3.1: When $f(z,y)$ *in* (2.1) *is linear in the second argument (for each value of the first argument), i.e.* $f(z,y) = f_1(z) + f_2(z)y$, *lemmas* 3.1 *and* 3.2 *reduce to the corresponding results in* [3]. □

Now we are ready to present our main result in this paper.

Define the "deterministic" optimal control problem $\mathbf{Q_0}$ as follows:
$\mathbf{Q_0}$: Find a solution $z(t)$ of (3.2) which minimizes the cost function

$$F_x^0(z_0) \triangleq \inf_z G(z(1)) \tag{3.5}$$

over the trajectories z of system (3.2).

The following theorem about an approximation of $\mathbf{Q}_\varepsilon$ by $\mathbf{Q_0}$ can be easily established on the basis of lemmas 3.1 and 3.2.

Theorem 3.1: The values $F_x^\varepsilon(z_0)$ *of the original problem* $\mathbf{Q}_\varepsilon$ *converge to the value* $F_x^0(z_0)$ *of the problem* $\mathbf{Q_0}$, *as* $\varepsilon \to 0$. *More precisely,*

$$|F_x^\varepsilon(z_0) - F_x^0(z_0)| \le C_G \gamma(\varepsilon) \ ,$$

where $\gamma(\varepsilon)$ *is as in Lemmas* 3.1 *and* 3.2 *and* C_G *is a constant from Assumption* 2.2. *If* z^* *is the solution of* (3.2) *such that* (3.5) *is minimized, then the Markov*

policy $u_\varepsilon(z^*)$ *mentioned in Lemma* 3.1 *satisfies*

$$|E_x^{u_\varepsilon(z^*)}G(Z(1)) - F_x^\varepsilon(z_0)| \leq C_G\gamma(\varepsilon) .$$

That is, $u_\varepsilon(z^*)$ *is asymptotically optimal for* $\mathbf{Q}_\varepsilon$.

Proof: Let u be an arbitrary policy, $Z(t)$ be the solution of (2.1) and $\tilde{z}^\varepsilon(t)$ be a solution of (3.2) satisfying (3.4). Then, by Assumption 2.2 and Lemma 3.2,

$$|E_x^u G(Z(1)) - E_x^u G(\tilde{z}^\varepsilon(1))| \leq C_G E_x^u \|Z(1) - \tilde{z}^\varepsilon(1)\| \leq C_G\gamma(\varepsilon) . \tag{3.6}$$

Note that

$$G(\tilde{z}^\varepsilon(1)) \geq F_x^0(z_0) ,$$

which implies

$$E_x^u G(\tilde{z}^\varepsilon(1)) \geq F_x^0(z_0)$$

for any policy u. Combining the above inequality with (3.6), it can be shown that

$$E_x^u G(Z(1)) \geq F_x^0(z_0) - C_G\gamma(\varepsilon) .$$

Hence

$$F_x^\varepsilon(z_0) = \inf_u E_x^u G(Z(1)) \geq F_x^0(z_0) - C_G\gamma(\varepsilon) . \tag{3.7}$$

Let now z^* be a solution which minimizes $\mathbf{Q}_0$. We have that from Lemma 3.1

$$\begin{aligned} |E_x^{u_\varepsilon(z^*)}G(Z(1)) - F_x^0(z_0)| &= |E_x^{u_\varepsilon(z^*)}G(Z(1)) - G(z^*(1))| \\ &\leq C_G E_x^{u_\varepsilon(z^*)}\|Z(1) - z^*(1)\| \leq C_G\gamma(\varepsilon) . \end{aligned}$$

Therefore

$$E_x^{u_\varepsilon(z^*)} G(Z(1)) \leq F_x^0(z_0) + C_G \gamma(\varepsilon) \ . \tag{3.8}$$

Since $E_x^{u_\varepsilon(z^*)} G(Z(1)) \geq F_x^\varepsilon(z_0)$, the inequalities (3.7) and (3.8) conclude the proof of the theorem. ■

Remark 3.2: In [2], *the linear case (where* $f(\cdot,\cdot)$ *appearing in* (2.1) *is linear) was analyzed. In that case, it follows from the analysis in* [2] *that* γ *can be chosen such that*

$$\lim_{\varepsilon \to 0} \varepsilon^{-(1/2)} \gamma(\varepsilon) = 0 \ .$$

Hence, for the linear case, simple bounds on the rate of convergence are available for Lemma 3.1 *and* 3.2 *as well as for Theorem* 3.1. □

4 Construction of an Asymptotically Optimal Policy

Let $\bar{z}(t)$ be an arbitrary solution for $\mathbf{Q_0}$. We show below how to construct the policy $u_\varepsilon(\bar{z})$ (appearing in Lemma 3.1 and in Theorem 3.1). Choose a function $\Delta = \Delta(\varepsilon)$ in such a way that

$$\lim_{\varepsilon \to 0} \Delta(\varepsilon) = 0 \ , \quad \lim_{\varepsilon \to 0} \frac{\Delta(\varepsilon)}{\varepsilon} = \infty \tag{4.1}$$

and set

$$\tau_l = \tau(l, \varepsilon) \stackrel{\Delta}{=} l\Delta(\varepsilon) \ , \quad l = 0, 1, 2, \ldots, \ell(\varepsilon) \ , \quad \ell(\varepsilon) \stackrel{\Delta}{=} \lfloor \Delta(\varepsilon)^{-1} \rfloor \ , \quad \tau_{\ell(\varepsilon)+1} = 1 \ . \tag{4.2}$$

Let v_l be the projection (see Appendix A) of the vector

$$(\Delta(\varepsilon))^{-1} \int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t) dt$$

onto the set $V(\bar{z}(\tau_l))$, $l = 0, 1, \ldots, \ell(\varepsilon) - 1$.

By definition there exists a stationary policy s_l such that

$$v_l = \sum_{w,a} \eta(s_l; w, a) f(\bar{z}(\tau_l), g(w, a)) \ .$$

Construct now $u_\varepsilon(z)$ as the Markov policy obtained by applying s_l during $n = \lfloor \tau_l/\varepsilon \rfloor, \lfloor \tau_l/\varepsilon \rfloor + 1, \ldots, \lfloor \tau_{l+1}/\varepsilon \rfloor - 1$, where $l = 0, 1, \ldots, \ell(\varepsilon) - 1$, and by applying an arbitrary stationary policy during $\lfloor \tau_{\ell(\varepsilon)}/\varepsilon \rfloor$, $\lfloor \tau_{\ell(\varepsilon)}/\varepsilon \rfloor + 1, \ldots,$ $\lfloor \varepsilon^{-1} \rfloor$. In the proof of Lemma 3.1 it is established that the policy $u_\varepsilon(z)$ thus constructed satisfies inequality (3.3).

By Theorem 3.1, the policy $u_\varepsilon(z^*)$ constructed on the basis of an optimal solution of the deterministic problem $\mathbf{Q}_0$ is asymptotically optimal for the problem $\mathbf{Q}_\varepsilon$. Let us now consider some ways of characterization of the optimal trajectory $z^*(t)$ of $\mathbf{Q}_0$.

4.1 Hamilton-Jacoby-Bellman (HJB) Equation for Q_0

One way of characterization of the optimal control in the problem $\mathbf{Q}_0$ is related to the HJB equation. Define the Hamiltonian of (3.2) as

$$\mathscr{H}(z, \lambda) = \max\{\lambda^T v | v \in V(z)\} \ . \tag{4.3}$$

As follows from (3.1), $\mathscr{H}(z, \lambda)$ is equal to the optimal value of the following linear programming (LP) problem

$$\mathscr{H}(z, \lambda) = \max_{\eta} \left\{ \sum_{w,a} \lambda^T f(z, g(w, a)) \eta(w, a) | \eta = \{\eta(w, a)\} \in W \right\} , \tag{4.4}$$

where the characterization of W as a set of linear constrains was given explicitly in [11]. The HJB equation of $\mathbf{Q}_0$ is written in the form

$$-\frac{\partial B_0(t, z)}{\partial t} + \mathscr{H}\left(z, -\frac{\partial B_0(t, z)}{\partial z}\right) = 0 \ , \quad B_0(1, z) = G(z) \ . \tag{4.5}$$

The HJB equation allows us to construct both necessary and sufficient conditions of optimality for $\mathbf{Q}_0$ and, in particular, to verify whether a given $z(t)$, solution of (3.2), is optimal in $\mathbf{Q}_0$ (see details in [8]). On the other hand, the

viscosity solution of (4.5) (see e.g. [15]) defines the optimal value of the problem $\mathbf{Q}_0$ on the interval $[s, 1]$ subject to the initial condition $z_s = z$ which provides an approximation for the optimal value $B_\varepsilon(s, z, x)$ of the problem $\mathbf{Q}_\varepsilon$ on the same interval $[s, 1]$ subject to the same initial condition $z_s = z$ and with the initial state of the MDP being x. More precisely, since, by definition, $B_\varepsilon(0, z, x) = G_\varepsilon(z)$ and $B_0(0, z) = G_0(z)$, from Theorem 3.1 it follows that

$$\lim_{\varepsilon \to 0} B_\varepsilon(0, z, x) = B_0(0, z) \ .$$

Similarly to this theorem, one can establish also that

$$\lim_{\varepsilon \to 0} B_\varepsilon(s, z, x) = B_0(s, z) \ ,$$

with the convergence being uniform with respect to $s \in [0, 1]$, $x \in \mathbf{X}$ and $z \in Z$, where Z is a compact subset of $\mathbf{R}^n$.

The Hamiltonian $\mathscr{H}(z, \lambda)$ allows also the following representation.

Let us introduce a parameterized set $\mathscr{L} = \{L(z, \lambda)\}$ of MDPs, $(z, \lambda) \in \mathbb{R}^n \times \mathbb{R}^n$, all of which have $\mathbf{X}$ and $\mathbf{A}$ as state and action spaces, and $\mathscr{P} = \{P_{vaw}, v, w \in \mathbf{X}, a \in \mathbf{A}\}$ as transition probabilities. They differ by the immediate cost, which is given by

$$r(z, \lambda; w, a) = \lambda^T f(z, g(w, a)) \ .$$

Consider the problem of maximization of the infinite horizon expected average cost related to an initial distribution ξ over $\mathbf{X}$

$$\mathscr{H}_\xi(z, \lambda) \stackrel{\Delta}{=} \sup_u \mathscr{H}_\xi(z, \lambda; u) \ , \quad \mathscr{H}_\xi(z, \lambda; u) \stackrel{\Delta}{=} \overline{\lim_{m \to \infty}} \frac{1}{m+1} E_\xi^u \sum_{j=0}^{m} r(z, \lambda; X_j, A_j) \ . \tag{4.6}$$

This problem is well known to be equivalent to the linear programming problem (4.4) (see, i.e. [12]). Namely:

a) The optimal value of the above problem does not depend on the initial distribution ξ and it is equal to the optimal value of (4.4)

$$\mathscr{H}_\xi(z, \lambda) = \mathscr{H}(z, \lambda) \ .$$

b) There is a one-to-one correspondence between optimal stationary policies of $L(z, \lambda)$ and the optimal solutions of (4.4).

The approximate solution of Q_ε via Q_0 has, thus, a decomposition structure. One fixes first the slow parameters z and λ and finds optimal stationary policy for "fast" MDP (4.6) and then finds an optimal (or near optimal) regime of changing the slow parameters via solution of HJB (4.3).

4.2 Reduction to Bolza Problem

Another, in a sense dual decomposition procedure for an approximate solution of Q_ε can be constructed in case the objective function is given in the integral form.

Assume that instead of (2.3), we consider the following objective function:

$$\inf E \int_0^1 \varphi(Z(t), Y(t))dt \,, \tag{4.7}$$

where $\varphi(\cdot, \cdot)$ satisfies the same assumptions as $f(\cdot, \cdot)$.

Define $\Phi(z, v)$ to be the optimal value of the following linear programming problem

$$\Phi(z, v) \triangleq \inf_{\eta} \left\{ \sum_{w,a} \varphi(z, g(w, a))\eta(w, a) \Big| \sum_{w,a} f(z, g(w, a))\eta(w, a) = v \in V(z),\ \eta \in W \right\}. \tag{4.8}$$

For $v \notin V(z)$, we take $\Phi(z, v)$ to be equal to $+\infty$.

Similar to the proof of Theorem 1 in [17], it can be shown that $\mathbf{Q}_0$ is equivalent to the Bolza problem

$$\inf \int_0^1 \Phi(z(t), \dot{z}(t))dt \,. \tag{4.9}$$

Notice that the LP problem (4.8) is equivalent to the following MDP (see, e.g. [1]) restricted to stationary policies:

$$\inf_u \varliminf_{m\to\infty} \left\{ \frac{1}{m+1} \sum_{j=0}^{m} \varphi(z, g(X_j, A_j)) \right\} \tag{4.10}$$

subject to

$$\underline{\lim}_{m\to\infty}\left\{\frac{1}{m+1}\sum_{j=0}^{m} f(z, g(X_j, A_j)) = v | v \in V(z)\right\} .$$

The equivalence is understood in the sense that they have the same value, and that based on the solution of (4.8), one can construct an optimal policy for the MDP and vice versa. Thus again with fixed slow parameters (this time, z and v), one finds optimal stationary strategies for fast MDP with constraint (4.10) and then consider slow deterministic Bolza problem (4.9).

5 Conclusion

The system we dealt with in this paper includes parameters jumping at discrete times governed by a controlled Markov chain with finite state and action spaces. We considered the problem of optimal control of this nonlinear stochastic hybrid system under the condition that the intervals between the jumps are small. We showed that an asymptotically optimal policy for this problem can be found on the basis of solution of some specially constructed deterministic optimal control problem.

Appendices

A Proof of Lemma 3.1

Proof: From (3.1), we have that by Assumption 2.1

$$\max_{t\in[0,1]}\{\|\eta\| : \eta \in V(z(t))\} \leq M , \quad \forall z(t) , \tag{A.1}$$

and

$$\rho(V(z_1), V(z_2)) \leq C\|z_1 - z_2\| , \quad \forall z_1, z_2 , \tag{A.2}$$

where $\rho(\cdot,\cdot)$ is the Hausdorff metric, which is defined as follows:
for any two arbitrary bounded sets B and D,

$$\rho(B,D) \triangleq \max\left\{\sup_{\xi\in B} dist\{\xi,D\}, \quad \sup_{\xi\in D} dist\{\xi,B\}\right\}$$

and for any set V

$$dist\{\xi,V\} \triangleq \inf_{v\in V}\|\xi - v\| \ .$$

Let $\bar{z}(t)$ be an arbitrary solution to the differential inclusion (3.2) and let $\Delta(\varepsilon)$ be as in (4.1). By (A.1),

$$\|z(t) - z(\tau_l)\| \le M\Delta(\varepsilon) \ , \quad \forall t \in [\tau_l, \tau_{l+1}] \ , \tag{A.3}$$

where τ_l, $l = 0, 1, \ldots, \ell(\varepsilon) - 1$ are defined in (4.2).

By virtue of (A.2) and (A.3), we have

$$\begin{aligned}\dot{\bar{z}}(t) \in V(\bar{z}(t)) &\subset V(\bar{z}(\tau_l)) + C\|\bar{z}(t) - \bar{z}(\tau_l)\|\bar{B} \\ &\subset V(\bar{z}(\tau_l)) + CM\Delta(\varepsilon)\bar{B} \ ,\end{aligned} \tag{A.4}$$

where $\bar{B}$ is the closed ball in $\mathbf{R}^n$ with the center in the origin and with the unit radius.

From (A.4) and the fact that $V(z)$ is convex it follows

$$\Delta^{-1}(\varepsilon)\int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt \in V(\bar{z}(\tau_l)) + CM\Delta(\varepsilon)\bar{B}$$

which implies that

$$dist\left(\Delta^{-1}(\varepsilon)\int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt, V(\bar{z}(\tau_l))\right) \le CM\Delta(\varepsilon) \ . \tag{A.5}$$

Define the vectors v_l, $l = 0, 1, \ldots, \ell(\varepsilon) - 1$, as the projections of the vectors

$$\Delta^{-1}(\varepsilon)\int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt$$

onto the sets $V(\bar{z}(\tau_l))$, i.e.

$$v_l \triangleq argmin\left\{ \left\| \Delta^{-1}(\varepsilon) \int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt - v \right\| | v \in V(\bar{z}(\tau_l)) \right\} .$$

As noticed in Section 4, there exists a policy s_l such that

$$v_l = \sum_{w,a} \eta(s_l, w, a) f(\bar{z}(\tau_l), g(w, a)) . \tag{A.6}$$

Define a Markov policy $u_\varepsilon(\bar{z})$ as indicated in Section 4. That is, define it as one following the stationary policy s_l during $n = \left\lfloor \frac{\tau_l}{\varepsilon} \right\rfloor, \left\lfloor \frac{\tau_l}{\varepsilon} + 1 \right\rfloor, \left\lfloor \frac{\tau_{l+1}}{\varepsilon} - 1 \right\rfloor$.

Let us define the sequence of the vectors ξ_l as the solution to the equation

$$\zeta_{l+1} = \zeta_l + \Delta(\varepsilon) v_l , \quad l = 0, 1, \ldots, \ell(\varepsilon) - 1 \; ; \; \zeta_0 = z_0 . \tag{A.7}$$

Since

$$\bar{z}(\tau_{l+1}) = \bar{z}(\tau_l) + \int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt \tag{A.8}$$

by virtue of (A.5), subtracting (A.7) from (A.8), we have that

$$\begin{aligned} \|\bar{z}(\tau_{l+1}) - \zeta_{l+1}\| &\leq \|\bar{z}(\tau_l) - \zeta_l\| + \Delta(\varepsilon) \left\| \Delta^{-1}(\varepsilon) \int_{\tau_l}^{\tau_{l+1}} \dot{\bar{z}}(t)dt - v_l \right\| \\ &\leq \|\bar{z}(\tau_l) - \zeta_l\| + CM\Delta^2(\varepsilon) \end{aligned}$$

which implies (see (4.2)) that

$$\|\bar{z}(\tau_l) - \zeta_l\| \leq \ell(\varepsilon) CM\Delta^2(\varepsilon) \leq CM\Delta(\varepsilon) , \quad l = 0, 1, \ldots, \ell(\varepsilon) . \tag{A.9}$$

Now let us define a sequence of random vectors Z_l, $l = 0, 1, \ldots, \ell(\varepsilon)$ accord-

ing to the following relations

$$Z_{l+1} = Z_l + \int_{\tau_l}^{\tau_{l+1}} f(Z_l, Y(t))dt \ , \quad Z_0 = z_0 \ , \tag{A.10}$$

where $Y(t)$ is defined by (2.2) with the policy $u_\varepsilon(\bar{z})$.

Subtracting (A.10) from (A.7), one obtains

$$\begin{aligned}
&E_x^{u_\varepsilon(\bar{z})}\|\zeta_{l+1} - Z_{l+1}\| \\
&\quad \le E_x^{u_\varepsilon(\bar{z})}\|\zeta_l - Z_l\| + \Delta(\varepsilon)E_x^{u_\varepsilon(\bar{z})}\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(Z_l, Y(t))dt - v_l\right\| \\
&\quad \le E_x^{u_\varepsilon(\bar{z})}\|\zeta_l - Z_l\| + \Delta(\varepsilon)E_x^{u_\varepsilon(\bar{z})} \\
&\qquad \times \left\{ \left\| \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(Z_l, Y(t))dt - \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t))dt \right\| \right. \\
&\qquad \left. + \left\| \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t))dt - v_l \right\| \right\} .
\end{aligned} \tag{A.11}$$

By Assumption 2.1, we have that from (A.9)

$$\begin{aligned}
&\left\| \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(Z_l, Y(t))dt - \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t))dt \right\| \\
&\quad \le \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} C\|Z_l - \bar{z}(\tau_l)\|dt \\
&\quad \le \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} C(\|Z_l - \zeta_l\| + \|\zeta_l - \bar{z}(\tau_l)\|)dt \\
&\quad \le C(\|Z_l - \zeta_l\| + \|\zeta_l - \bar{z}(\tau_l)\|) \\
&\quad \le C\|Z_l - \zeta_l\| + C^2 M\Delta(\varepsilon) \ .
\end{aligned} \tag{A.12}$$

Now, substitute (A.12) into (A.11), and take into account (A.6), we have

$$E_x^{u_\varepsilon(\bar{z})}\|\zeta_{l+1} - Z_{l+1}\|$$

$$\leq E_x^{u_\varepsilon(\bar{z})}\|\zeta_l - Z_l\| + C\Delta(\varepsilon)E_x^{u_\varepsilon(\bar{z})}\|Z_l - \zeta_l\| + C^2M\Delta^2(\varepsilon)$$

$$+ \Delta(\varepsilon)E_x^{u_\varepsilon(\bar{z})}\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t))dt - \sum_{(w,a)} \eta(s_l, w, a)f(\bar{z}(\tau_l), y(w, a))\right\| . \tag{A.13}$$

Consider the state-action frequencies ψ_m^K corresponding to the realization H. It follows from Lemma 2.1 that there exists some μ: $\mathbf{N} \to \mathbf{R}$ with

$$\lim_{K\to\infty} \mu(K) = 0$$

such that for any stationary policy s applied during $n = m+1, \ldots, m+K$, and any distribution ξ over X_m

$$E_\xi^s\left(\max_{(w,a)} |\psi_m^K(w, a) - \eta(s, w, a)|\right) \leq \mu(K) . \tag{A.14}$$

Denote

$$K(\varepsilon) = \min_{l=0,1,\ldots,\ell(\varepsilon)-1} (\lfloor \tau_{l+1}/\varepsilon \rfloor - \lfloor \tau_l/\varepsilon \rfloor)$$

and notice that

$$2 \geq \lfloor \tau_{l+1}/\varepsilon \rfloor - \lfloor \tau_l/\varepsilon \rfloor - K(\varepsilon) \geq 0 , \quad \left|K(\varepsilon) - \frac{\Delta(\varepsilon)}{\varepsilon}\right| \leq 1$$

$$\Rightarrow \quad \left|\frac{1}{K(\varepsilon)} - \frac{\varepsilon}{\Delta(\varepsilon)}\right| \leq \frac{\varepsilon^2}{\Delta^2(\varepsilon)}\left(\frac{1}{1 - \varepsilon/\Delta(\varepsilon)}\right) . \tag{A.15}$$

From (A.15) it follows that there exist positive constants L_1 and L_2 such that

$$\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t))dt - \frac{\varepsilon}{\Delta(\varepsilon)} \sum_{n=\lfloor \tau_l/\varepsilon \rfloor + 1}^{\lfloor \tau_l/\varepsilon \rfloor + K(\varepsilon)} f(\bar{z}(\tau_l), y(X_n, A_n))\right\| \leq L_1 \frac{\varepsilon}{\Delta(\varepsilon)} \tag{A.16}$$

$$\left\| \frac{\varepsilon}{\Delta(\varepsilon)} \sum_{n=\lfloor \tau_l/\varepsilon \rfloor+1}^{\lfloor \tau_l/\varepsilon \rfloor+K(\varepsilon)} f(\bar{z}(\tau_l), y(X_n, A_n)) - \frac{1}{K(\varepsilon)} \sum_{n=\lfloor \tau_l/\varepsilon \rfloor+1}^{\lfloor \tau_l/\varepsilon \rfloor+K(\varepsilon)} f(\bar{z}(\tau_l), y(X_n, A_n)) \right\|$$

$$\leq L_2 \frac{\varepsilon}{\Delta(\varepsilon)} \ . \tag{A.17}$$

Since

$$\frac{1}{K(\varepsilon)} \sum_{n=\lfloor \tau_l/\varepsilon \rfloor+1}^{\lfloor \tau_l/\varepsilon \rfloor+K(\varepsilon)} f(\bar{z}(\tau_l), y(X_n, A_n)) = \sum_{w,a} \psi^{K(\varepsilon)}_{\lfloor \tau_l/\varepsilon \rfloor}(H; W, a) f(\bar{z}(\tau_l), y(w, a)) \ ,$$

it is easy to show that using (A.14), (A.16)–(A.17)

$$E_x^{u_\varepsilon(\bar{z})} \left\| \frac{1}{\Delta(\varepsilon)} \int_{\tau_l}^{\tau_{l+1}} f(\bar{z}(\tau_l), Y(t)) dt - \sum_{w,a} \eta(s^l; w, a) f(\bar{z}(\tau_l), y(w, a)) \right\|$$

$$\leq (L_1 + L_2) \frac{\varepsilon}{\Delta(\varepsilon)} + E_x^{u_\varepsilon(\bar{z})}$$

$$\times \left\| \sum_{n=\lfloor \tau_l/\varepsilon \rfloor+1}^{\lfloor \tau_l/\varepsilon \rfloor+K(\varepsilon)} \psi^{K(\varepsilon)}_{\lfloor \tau_l/\varepsilon \rfloor}(H; w, a) f(\bar{z}(\tau_l), y(w, a)) - \sum_{w,a} \eta(s^l; w, a) f(\bar{z}(\tau_l), y(w, a)) \right\|$$

$$\leq (L_1 + L_2) \frac{\varepsilon}{\Delta(\varepsilon)}$$

$$+ E_x^{u_\varepsilon(\bar{z})} \left\{ E^{s^l}_{X_{\lfloor \tau_l/\varepsilon \rfloor}} \sum_{w,a} (|\psi^{K(\varepsilon)}_{\lfloor \tau_l/\varepsilon \rfloor}(H; w, a) - \eta(s^l; w, a)| \, \| f(\bar{z}(\tau_l), y(w, a)) \|) \right\}$$

$$\leq (L_1 + L_2) \frac{\varepsilon}{\Delta(\varepsilon)} + L_3 \mu(K(\varepsilon)) \ ,$$

where

$$L_3 = \max_{(z,y) \in D_1 \times D_2} \| f(z, y) \| \ .$$

Now, substitute (A.18) into (A.13), we have that

$$E_x^{u_\varepsilon(\bar{z})} \| \zeta_{l+l} - Z_{l+1} \|$$

$$\leq E_x^{u_\varepsilon(\bar{z})} \| \zeta_l - Z_l \| + C\Delta(\varepsilon) E_x^{u_\varepsilon(\bar{z})} \| Z_l - \zeta_l \| + C^2 M \Delta^2(\varepsilon) \tag{A.19}$$

$$+ (L_1 + L_2)\varepsilon + L_3 \Delta(\varepsilon) \mu(K(\varepsilon)) \ .$$

Applying Lemma 2.2 to (A.19), we have

$$E_x^{u_\varepsilon(\bar{z})}\|\zeta_l - Z_l\| \leq \nu(\varepsilon) , \tag{A.20}$$

where $\nu(\varepsilon) \to 0$ as $\varepsilon \to 0$.

On the other hand, we know that

$$Z(\tau_{l+1}) = Z(\tau_l) + \int_{\tau_l}^{\tau_{l+1}} f(Z(t), Y(t))dt . \tag{A.21}$$

Subtract (A.21) from (A.10) we have

$$\begin{aligned} &E_x^{u_\varepsilon(\bar{z})}\|Z(\tau_{l+1}) - Z_{l+1}\| \\ &\quad \leq E_x^{u_\varepsilon(\bar{z})}\left[\|Z(\tau_l) - Z_l\| + \int_{\tau_l}^{\tau_{l+1}} C\|Z(t) - Z_l\|dt\right] . \end{aligned} \tag{A.22}$$

Note that

$$E_x^{u_\varepsilon(\bar{z})}\|Z(t) - Z(\tau_l)\| \leq M\Delta(\varepsilon) . \tag{A.23}$$

Substituting (A.23) into (A.22), we have that

$$\begin{aligned} &E_x^{u_\varepsilon(\bar{z})}\|Z(\tau_{l+1}) - Z_{l+1}\| \\ &\quad \leq E_x^{u_\varepsilon(\bar{z})}\left[\|Z(\tau_l) - Z_l\| + \int_{\tau_l}^{\tau_{l+1}} C\|Z(\tau_l) - Z_l\|dt\right] + CM\Delta^2(\varepsilon) \\ &\quad = E_x^{u_\varepsilon(\bar{z})}\|Z(\tau_l) - Z_l\| + C\Delta(\varepsilon)E_x^{u_\varepsilon(\bar{z})}\|Z(\tau_l) - Z_l\| + CM\Delta^2(\varepsilon) \end{aligned}$$

which implies by Lemma 2.2 that

$$E_x^{u_\varepsilon(\bar{z})}\|Z(\tau_l) - Z_l\| \leq C_1\Delta(\varepsilon) , \tag{A.24}$$

where C_1 is a positive constant.

Finally, combining (A.9), (A.20) and (A.24), we conclude that there exists a

positive constant C_2 such that

$$
\begin{aligned}
&E_x^{u_\varepsilon(\bar{z})}\|\bar{z}(\tau_l) - Z(\tau_l)\| \\
&\quad \le E_x^{u_\varepsilon(\bar{z})}\{\|\bar{z}(\tau_l) - \zeta_l\| + \|\zeta_l - Z_l\| + \|Z_l - Z(\tau_l)\|\} \\
&\quad \le \nu(\varepsilon) + C_2\Delta(\varepsilon) \ , \quad l = 0, 1, \ldots, \ell(\varepsilon) \ ,
\end{aligned}
$$

where $C_2 = CM + C_1$. This together with (A.23) and (A.3) complete our proof. ■

B Proof of Lemma 3.2

Proof: Let $h = \{x_0, a_0, \ldots, x_{\lfloor \varepsilon^{-1} \rfloor}, a_{\lfloor \varepsilon^{-1} \rfloor}\} \in \mathbf{H}$ be some state-action trajectory and

$$
y(t, h) \stackrel{\Delta}{=} g(x_{\lfloor t/\varepsilon \rfloor}, a_{\lfloor t/\varepsilon \rfloor}) \ , \tag{B.1}
$$

where $g(\cdot, \cdot)$ is defined in (2.2).

Denote by $\sigma_l(h)$ the projection of $\psi^{K(\varepsilon)}_{\lfloor \tau_l/\varepsilon \rfloor}(h)$ on W, i.e. $\sigma_l(h) := \{\sigma_l(h; w, a)\}_{w,a}$ is the solution of

$$
\min_{\eta}\{\|\psi^{K(\varepsilon)}_{\lfloor \tau_l/\varepsilon \rfloor}(h) - \eta\| \,\big|\, \eta \in W\} \ .
$$

It follows from Lemma 2.1 that there exists a function $\nu(K)$

$$
\lim_{K \to \infty} \nu(K) = 0
$$

such that for any policy u,

$$
E_x^u dist\{\psi_m^K(H), W\} \le \nu(K) \ .
$$

Hence,

$$E_x^u\left\{\max_{w,a}|\psi_{[\tau_l/\varepsilon]}^{K(\varepsilon)}(H;w,a)-\sigma_l(H;w,a)|\right\}\le \nu(K(\varepsilon))\ . \tag{B.2}$$

Let

$$v_0(h)=\sum_{w,a}\sigma_0(h;w,a)f(z_0,y(w,a))$$

$$\zeta_1(h)=z_0+\Delta(\varepsilon)v_0(h)\ .$$

Now, we can define the vectors $v_1(h),v_2(h),\ldots,v_l(h),\zeta_1(h),\zeta_2(h),\ldots,\zeta_{l+1}(h)$, as follows:

$$v_l(h)=\sum_{w,a}\sigma_l(h;w,a)f(\zeta_l(h),y(w,a))\ ,\quad l=0,1,\ldots,\ell(\varepsilon)-1\ , \tag{B.3}$$

and

$$\zeta_{l+1}(h)=\zeta_l(h)+\Delta(\varepsilon)v_l(h)\ ,\quad l=0,1,\ldots,\ell(\varepsilon)-1,\zeta_0(h)=z_0\ . \tag{B.4}$$

By definition, it is easy to see that

$$v_l(h)\in V(\zeta_l(h))\ ,\quad \forall l=0,1,\ldots,\ell(\varepsilon)-1\ .$$

Following the procedure in the proof of Lemma 3.1 from (A.15) to (A.18), it can be shown that there exists a function $\nu_1(\varepsilon)$ which tends to zero as ε tends to zero such that

$$E_x^u\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}}f(\zeta_l(H),y(t,H))dt-v_l(H)\right\|\le \nu_1(\varepsilon)\ . \tag{B.5}$$

Define the piecewise linear function $\zeta(t,h)$ according to the formula

$$\zeta(t,h)=\begin{cases}\zeta_l(h)+(t-\tau_l)v_l(h), & t\in[\tau_l,\tau_{l+1}],\ l=0,1,\ldots,\ell(\varepsilon)-2\\ \zeta_{\ell(\varepsilon)-1}(h)+(t-\tau_{\ell(\varepsilon)-1})v_{\ell(\varepsilon)-1}(h), & t\in[\tau_{\ell(\varepsilon)-1},1]\ .\end{cases}$$

Along the same line as that in the proof of Lemma 3.1 and [16], we have the following inequalities

$$\begin{aligned} dist(\dot{\zeta}(t,h), V(\zeta(t,h))) &= dist(v_l(h), V(\zeta(t,h))) \\ &\leq dist(v_l(h), V(\zeta(\tau_l,h))) + \rho(V(\zeta(\tau_l,h)), V(\zeta(t,h))) \\ &\leq dist(v_l(h), V(\zeta_l(h))) + MC\Delta(\varepsilon) \\ &= MC\Delta(\varepsilon) \ , \quad \forall t \in (\tau_l, \tau_{l+1}) \ , \end{aligned} \tag{B.6}$$

where $\rho(\cdot,\cdot)$ and $dist(\cdot,\cdot)$ are defined as in Lemma 3.1. From Filippov Theorem [14] and (B.6), it follows that there exists a solution $z(t,h)$ of the differential inclusion

$$\dot{z}(t,h) \in V(z(t,h)) \ , z(0,h) = z_0$$

such that

$$\max_{t\in[0,1]} \|z(t,h) - \zeta(t,h)\| \leq C_2\Delta(\varepsilon) \ , \tag{B.7}$$

where C_2 is a positive number, which can be chosen to be the same for all $h \in \mathbf{H}$.

Let $\tilde{z}(t,h)$ be the solution of the equation

$$\dot{\tilde{z}}(t,h) = f(\tilde{z}(t,h), y(t,h)) \ , \quad \tilde{z}(0,h) = z_0 \ .$$

Notice that if u is a policy and H is a random realization of states-actions history, then

$$\tilde{z}(t,H) = Z(t) \ . \tag{B.8}$$

By definition, we have

$$\tilde{z}(\tau_{l+1},h) = \tilde{z}(\tau_l,h) + \int_{\tau_l}^{\tau_{l+1}} f(\tilde{z}(t,h), y(t,h))dt \ , \quad \tilde{z}(0,h) = z_0 \ , \tag{B.9}$$

where $y(t,h)$ is defined by (B.1).

Let $\tilde{z}_l(h), l = 0, 1, \ldots, \ell(\varepsilon)$ be the solution of the difference equation

$$\tilde{z}_{l+1}(h) = \tilde{z}_l(h) + \int_{\tau_l}^{\tau_{l+1}} f(\tilde{z}_l(h), y(t,h))dt \ , \quad \tilde{z}_0(h) = z_0 \ , \tag{B.10}$$

with z_0 being the same as the initial condition in (2.1).

Now, subtracting (B.10) from (B.9) we get

$$\begin{aligned} &\|\tilde{z}(\tau_{l+1}, h) - \tilde{z}_{l+1}(h)\| \\ &\quad \le \|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\| + \int_{\tau_l}^{\tau_{l+1}} C\|\tilde{z}(t,h) - \tilde{z}_l(h)\|dt \ . \end{aligned} \tag{B.11}$$

Note that

$$\|\tilde{z}(t,h) - \tilde{z}(\tau_l, h)\| \le M\Delta(\varepsilon) \ , \quad t \in [\tau_l, \tau_{l+1}] \ . \tag{B.12}$$

Substitute (B.12) into (B.11), we have that

$$\begin{aligned} &\|\tilde{z}(\tau_{l+1}, h) - \tilde{z}_{l+1}(h)\| \\ &\quad \le \|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\| + \int_{\tau_l}^{\tau_{l+1}} C\|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\|dt + CM\Delta^2(\varepsilon) \\ &\quad = \|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\| + C\Delta(\varepsilon)\|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\| + CM\Delta^2(\varepsilon) \ . \end{aligned}$$

By Lemma 2.2, the above inequality implies that there exists a positive number C_1 such that

$$\|\tilde{z}(\tau_l, h) - \tilde{z}_l(h)\| \le C_1\Delta(\varepsilon) \ . \tag{B.13}$$

Let u be an arbitrary control policy and H a realization in **H**. Subtract (B.10) from (B.4) and take the mathematical expectation, we have

$$\begin{aligned} &E_x^u\|\tilde{z}_{l+1}(H) - \zeta_{l+1}(H)\| \\ &\quad \le E_x^u\|\tilde{z}_l(H) - \zeta_l(H)\| + \Delta(\varepsilon)E_x^u\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\tilde{z}_l(H), y(t,H))dt - v_l(H)\right\| \end{aligned}$$

$$\leq E_x^u\|\tilde{z}_l(H) - \zeta_l(H)\|$$

$$+ \Delta(\varepsilon)E_x^u\left\{\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\tilde{z}_l(H), y(t,H))dt - \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\zeta_l(H), y(t,H))dt\right\|\right.$$

$$\left. + \left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\zeta_l(H), y(t,H))dt - v_l(H)\right\|\right\} . \tag{B.14}$$

By Assumption 2.1, we know that

$$E_x^u\left\|\frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\tilde{z}_l(H), y(t,H))dt - \frac{1}{\Delta(\varepsilon)}\int_{\tau_l}^{\tau_{l+1}} f(\zeta_l(H), y(t,H))dt\right\|$$

$$\leq CE_x^u\|\tilde{z}_l(H) - \zeta_l(H)\| . \tag{B.15}$$

Substituting (B.15) and (B.5) into (B.14), we have

$$E_x^u\|\tilde{z}_{l+1}(H) - \zeta_{l+1}(H)\|$$

$$\leq E_x^u\|\tilde{z}_l(H) - \zeta_l(H)\| + C\Delta(\varepsilon)E_x^u\|\tilde{z}_l(H) - \zeta_l(H)\| + \Delta(\varepsilon)v_1(\varepsilon) . \tag{B.16}$$

By Lemma 2.2, from (B.16) it follows that

$$E_x^u\|\tilde{z}_l(H) - \zeta_l(H)\| \leq v_2(\varepsilon) , \tag{B.17}$$

where $v_2(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Finally, taking into account (B.8), (B.12), (B.13), (B.17) and (B.7), we have

$$E_x^u\|Z(t) - z(t,H)\|$$

$$= E_x^u\|\tilde{z}(t,H) - z(t,H)\|$$

$$\leq E_x^u\|\tilde{z}(\tau_l,H) - z(\tau_l,H)\| + 2M\Delta(\varepsilon)$$

$$\leq E_x^u[\|\tilde{z}(\tau_l, H) - \tilde{z}_l(H)\| + \|\tilde{z}_l(H) - \zeta_l(H)\|$$

$$+ \|\zeta(\tau_l, H) - z(\tau_l, H)\|] + 2M\Delta(\varepsilon)$$

$$\leq C_1\Delta(\varepsilon) + \nu_2(\varepsilon) + C_2\Delta(\varepsilon) + 2M\Delta(\varepsilon) , \quad \forall t \in [\tau_l, \tau_{l+1}] .$$

The right hand side of last inequality tends to zero as ε tends to zero, which establishes our desired result. ■

References

[1] Altman E, Shwartz H (1991) Adaptive control of constrained Markov chains: Criteria and policies. Annals of Operations Research 28(1):101–134

[2] Altman E, Gaitsgory V (1993) Control of a hybrid stochastic system. System & Control Letters 20(4):307–314

[3] Altman E, Gaitsgory V (1994) Asymptotic optimization of a nonlinear hybrid system governed by a Markov decision process. Technical report, Preprints

[4] Bensoussan A, Blankenship GL (1987) Singular perturbations in stochastic control. In: Kokotovic PV, Bensoussan A, Blankenship GL (eds.) Singular perturbations and asymptotic analysis in control systems, Lecture Notes in Control and Information Sciences, Springer-Verlag, New York, pp. 171–260.

[5] Bensoussan A (1989) Perturbation methods in optimal control problems. John Wiley, New York

[6] Bielecki D, Filar JA (1991) Singular perturbed Markov control problem: Limiting average cost. Annals of Operations Research 28(1):153–168

[7] Boukas EK, Zhang Q, Yin G (1995) Robust production and maintenance planning in stochastic manufacturing systems. IEEE Trans. Automat. Control 40(6):1098–1102

[8] Clarck FH (1983) Optimization and nonsmooth analysis. John Wiley, New York

[9] Delebecque F, Quadrat J (1978) Contribution of stochastic control singular perturbation averaging and team theories to an example of large scale systems: Management and hydropwer production. IEEE Trans. Automat. Control 23(2):209–222

[10] Delebecque F, Quadrat J (1979) Optimal control of Markov chains admitting strong and weak interactions. Automatica 17(2):281–296

[11] Derman C, Strauch RE (1966) A note on memoryless rules for controlling sequential control processes. Annals of Math. and Stat. 37(2):276–278

[12] Derman C (1970) Finite state Markovian decision processes. Academic Press, New York

[13] Filar J, Vrieze K (1996) Competitive Markov decision processes. Springer-Verlag, Berlin

[14] Filippov AF (1959) To some questions of the theory of optimal control. Vestnik Moskov Univ. Ser. I Mat. Mekh. (2):25–32, In Russian

[15] Fleming WH, Soner HM (1993) Controlled Markov processes and viscosity solutions. Springer-Verlag, New York

[16] Gaitsgory V (1992) Suboptimization of singularly perturbed control systems. SIAM J. Contr. & Optimiz. 30(5):1228–1249

[17] Gaitsgory V (1993) Suboptimal control of singularly perturbed systems and periodic optimization. IEEE Trans. Automat. Control 38(6):888–903

[18] Kleinrock L (1976) Queuing systems, Volume II: Computer applications. John Wiley, New York
[19] Kushner H (1990) Weak convergence and singularly perturbed stochastic control and filtering problems. Birkhauser, Boston-Basel, Berlin
[20] Pervozvansky AA, Gaitsgory V (1988) Theory of suboptimal decisions. Kluwer Academic Publishers, Dordrecht, The Netherlands
[21] Philips RG, Kokotovic PV (1981) A singular perturbation approach to modeling and control of Markov chains. IEEE Trans. Automat. Control 26(9):1087–1094
[22] Sethi SP, Zhang Q (1994) Hierarchical decision making in stochastic manufacturing systems. Birkhauser, Boston-Basel, Berlin

Received: June 1996