CONTROL OF A HYBRID STOCHASTIC SYSTEM

Eitan ALTMAN*
INRIA
Centre Sophia Antipolis
06565 Valbonne Cedex, France

Vladimir Gaitsgory
Department of Economics
Bar-Ilan Univ., Ramat-Gan, Israel

November 1992

Abstract

We consider in this paper a continuous time stochastic hybrid control system with a finite time horizon. The objective is to minimize a linear function of the expected state trajectory. The state evolves according to a linear dynamics. However, the parameters of the state evolution equation may change at discrete times according to a controlled Markov chain which has finite state and action spaces. We use a procedure similar in form to the maximum principle; this determines a control strategy which is asymptotically optimal as the number of transitions during the finite time horizon grows to infinity.

Keywords: Hybrid stochastic systems, asymptotic optimality, linear dynamics, Markov decision processes, finite horizon.

1 Introduction and statement of the problem.

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

$$\frac{d}{dt} Z_t = AZ_t + BY_t, \qquad t \in [0,1], \qquad Z_0 = z \tag{1}$$

where $Y_t \in \mathbb{R}^k$ is the "control" and $A(n \times n)$ and $B(n \times k)$ are matrices of real numbers. 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,

*The work of this author was supported by the Chateaubriand fellowship from the French embassy in Israel

i.e. transitions occur at times $t = n\epsilon$, $n = 0, 1, 2, ..., \lfloor \epsilon^{-1} \rfloor$, where $\lfloor x \rfloor$ stands for the greatest integer which is smaller or equal to x. There is a finite state space $\mathbf{X} = \{1, ..., N\}$ and a finite action space $\mathbf{A}$. If the state is v and action a is chosen then the next state is w with probability P_{vaw}. A policy $u = \{u_0, u_1, ...\}$ in the set of policies U is a sequence of probability measures on $\mathbf{A}$; at each time $t = n\epsilon$ the controller chooses u_n based on the history of all previous states and actions, as well as the present state. Let X_n and A_n, $n = 0, 1, ...$ denote the resulting state and action processes. Denote by P_x^u and E_x^u the probability measure and expectation that correspond to an initial state $X_0 = x$ and a policy u. Sometimes we shall assume an initial distribution ξ on X_0, instead of a fixed initial state. In that case P_ξ^u denotes the corresponding probability measure. Let $y^j : \mathbf{X} \times \mathbf{A} \to \mathbb{R}$, $j = 1, ..., k$ be some given functions. Then Y_t in (1) is given by

$$Y_t = y(X_{\lfloor t/\epsilon \rfloor}, A_{\lfloor t/\epsilon \rfloor}). \tag{2}$$

The underlying stochastic process $\{X_n, A_n\}$ is known as a Controlled Markov Chain, or Markov Decision Process (MDP) see c.f. Derman [4] pp. 2-4. We shall be especially interested in stationary policies, denoted by $\mathcal{S}$, i.e. policies for which u_t depends only on the current state, and does not depend on previous states and actions nor on the time. 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.

Let c be an n-dimensional vector representing the (linear) operating cost related to the process Z_t. We consider the control problem with ϵ and x fixed:

$\mathbf{Q1}^\epsilon$: find a policy u that achieves $F_x^\epsilon = \inf_{u \in U} E_x^u c^T Z_1$, where Z_1 is obtained through (1).

Remarks:

(i) $\mathbf{Q1}^\epsilon$ is equivalent to the problem: find a policy u that achieves $\inf_{u \in U} c^T \overline{Z}_1$, where $\overline{Z}_t \in \mathbb{R}_n$ is given by

$$\frac{d}{dt}\overline{Z}_t = A\overline{Z}_t + BE_x^u Y_t, \qquad t \in [0, 1], \qquad \overline{Z}_0 = z \tag{3}$$

(ii) By solving the problem $\mathbf{Q1}^\epsilon$, one can also solve a problem with an integral cost function, i.e. to find a policy u that achieves $\inf_{u \in U} E_x^u \int_0^1 c^T Z_t dt$. This is obtained by using a new variable R_t defined by $dR_t/dt = c^T Z_t$.

Note that the controller does not require knowledge of the initial value z of Z_0, which may be assumed to be zero. More precisely, due to the linearity of the system (1), if a control strategy is optimal (or δ-optimal) for a given Z_0, then it is optimal (or δ-optimal, respectively) for any other value of Z_0.

The type of model which we introduce is natural in the control of inventories or of production, where we deal with material whose quantity may change in a continuous (linear) way. Breakdowns, repairs and other control decisions yield the underlying MDP. Our model may also be used in the control of highly loaded queueing networks for which the fluid approximation holds (see Kleinrock [8] p. 56). The quantities Z_t may then represent the number of customers in the different queues whereas the underlying MDP may correspond to routing, or flow control of, say, some long on-off traffic.

Our model is characterized by the fact that ϵ is to be a small parameter. We construct a set of policies $u^*(\epsilon)$ such that $u^*(\epsilon)$ is "almost" optimal for $\mathbf{Q1}^\epsilon$ for all ϵ small enough. More precisely, the difference between the cost under $u^*(\epsilon)$ and F_x^ϵ converges to zero as $\epsilon \to 0$.

The fact that ϵ is chosen to be small means that the variables Y_t along with the MDP X_t can be considered to be fast with respect to the time scale t in which Z_t evolves; Indeed, Y_t and X_t may have large jumps between $t = m\epsilon$ and $t = (m+1)\epsilon$, whereas the corresponding change in Z_t in that period is of order of ϵ. The problem is, thus, close in nature to singular perturbed control problems. The main idea in dealing with this category of problems is to optimize slow motions, assuming that the fast ones are approximated by their quasi stationary states obtained with "frozen" slow variables and controls, see [3, 9, 10, 11], and references therein. In our case, these quasi stationary states are related to the steady state distribution of the (controlled) Markov chains obtained when using stationary policies.

In Section 2 we present the main result; we construct a non-stationary policy for the hybrid control problem $\mathbf{Q1}^\epsilon$. We prove in Section 3 that the policy introduced in Section 2 is indeed asymptotically optimal as ϵ tends to zero.

Below, B^T will denote the transpose of a matrix (or of a column vector) B, and $\|B\|$ will denote the sum of absolute values of the components of B.

2 Construction of a sub-optimal policy

Consider a parameterized set $\mathcal{L} = \{L(\Lambda)\}$ of MDPs, $\Lambda \in \mathbb{R}^n$, all of which have $\mathbf{X}$ and $\mathbf{A}$ as state and action spaces, and $\mathcal{P} = \{P_{vaw}\}$ as transition probabilities. The immediate cost function in $L(\Lambda)$ is $r(v, a) = \Lambda^T B y(v, a)$. Let $u(\Lambda)$ be an optimal stationary deterministic policy for MDP

$L(\Lambda)$, which is thus a policy that achieves

$$J(\Lambda) := \sup_{u \in U} \left\{ \underline{\lim}_{m\to\infty} \frac{1}{m+1} E_\xi^u \sum_{j=0}^m \Lambda^T B y(X_j, A_j) \right\} \tag{4}$$

It follows from [4] p. 79 that $J(\Lambda)$ does not depend on the initial distribution ξ, and that $u(\Lambda)$ indeed exists; it can be computed using well known methods such as the value iteration, policy improvement or Linear Programming, see c.f. Derman [4] Section 6. Let $u(\Lambda, \delta)$ denote some δ-optimal stationary policy for the MDP $L(\Lambda)$, i.e. $u(\Lambda, \delta)$ satisfies

$$J(\Lambda) - \delta \leq \lim_{m\to\infty} \frac{1}{m+1} E_\xi^{u(\Lambda,\delta)} \sum_{j=0}^m \Lambda^T B y(X_j, A_j) \tag{5}$$

for any initial distribution ξ.

Let $\lambda_t \in \mathbb{R}^n$, $t \in [0,1]$. Consider the following equation:

$$\frac{d}{dt}\lambda_t = -A^T \lambda_t, \quad \lambda_1 = -c \tag{6}$$

Choose some function $\Delta(\epsilon)$ such that

$$\lim_{\epsilon\to 0} \Delta(\epsilon) = 0, \qquad \lim_{\epsilon \to 0} \frac{\Delta(\epsilon)}{\epsilon} = \infty \tag{7}$$

Let $\tau_l = \tau(l,\epsilon) := l\Delta(\epsilon),\ l = 0, 1, 2, ..., \lfloor \Delta(\epsilon)^{-1} \rfloor$.

Let $u^*(\epsilon, \delta)$ be the policy obtained by applying $u(\lambda_{\tau_l}, \delta)$ (that is $u(\Lambda, \delta)$ with $\Lambda = \lambda_{\tau_l}$) during $n = \lfloor \tau_l/\epsilon \rfloor, \lfloor \tau_l/\epsilon \rfloor + 1, ..., \lfloor \tau_{l+1}/\epsilon \rfloor - 1$. Note that $n(\epsilon) := (\lfloor \tau_{l+1}/\epsilon \rfloor - \lfloor \tau_l/\epsilon \rfloor - 1) \to \infty$ as $\epsilon \to 0$ and that

$$|\epsilon n(\epsilon) - \Delta(\epsilon)| \leq 2\epsilon. \tag{8}$$

In the next Section we prove that

Theorem 1 *For any $\delta \geq 0$, $u^*(\epsilon, \delta)$ is asymptotically δ-optimal for problem* $\mathbf{Q1}^\epsilon$ *as $\epsilon \to 0$, i.e.*

$$\overline{\lim_{\epsilon\to 0}} \left(E_x^{u^*(\epsilon,\delta)} c^T Z_1 - F_x^\epsilon \right) \leq \delta$$

Note that the fact that the suboptimal policy is chosen as the result of the optimization of the problem (4) with Λs defined by the solution of the conjugate system (6), allows one to consider the procedure as an analog to the maximum principle which, as it is well known, is not only a necessary but also a sufficient condition of optimality for linear deterministic systems. (See also Lemma 2 below).

3 Asymptotic optimality of the stationary policy

In order to establish Theorem 1 we shall need two Lemmas.

Let $\hat{V}$ be some compact set in $\mathbb{R}^k$. Let $\mathcal{K}(\hat{V})$ be the set of measurable functions $[0,1] \to \hat{V}$. Consider the following linear deterministic control system **Q2**: find a control $\hat{Y} \in \mathcal{K}(\hat{V})$ that achieves

$$J^* := \inf_{Y \in \mathcal{K}(\hat{V})} c^T \hat{Z}_1 \tag{9}$$

subject to

$$\frac{d}{dt}\hat{Z}_t = A\hat{Z}_t + B\hat{Y}_t, \qquad t \in [0,1], \qquad \hat{Z}_t \in \mathbb{R}^n, \qquad \hat{Z}_0 = \hat{z}_0 \tag{10}$$

where A, B and c are as in $\mathbf{Q1}^\epsilon$. Let $\lambda_t \in R^n$, $t \in [0,1]$.

Lemma 2 *Consider the system, conjugate to the linear system (10) (which obviously coincides with (6)):*

$$\frac{d}{dt}\lambda_t = -A^T\lambda_t, \quad \lambda_1 = -c \tag{11}$$

Then (i) $\hat{Y}^0$ is optimal for **Q2** *(achieves J^*) if and only if*

$$\hat{Y}_t^0 = argmax\left\{\lambda_t^T By | y \in \hat{V}\right\} \tag{12}$$

almost every where.
(ii) Choose some $\delta \geq 0$, and let $\hat{Y}_t^\delta$ be such that

$$\int_0^1 \lambda_t^T B\hat{Y}_t^\delta dt \geq \int_0^1 \max_y \left\{\lambda_t^T By | y \in \hat{V}\right\} dt - \delta, \ \forall t \in [0,1]. \tag{13}$$

Then

$$c^T\hat{Z}_1 \leq J^* + \delta \tag{14}$$

where $\hat{Z}_1$ corresponds to the choice of $\hat{Y}_t^\delta$ (through (10)).

Proof: Denote $\lambda_\tau = -\left(exp\{A(1-\tau)\}\right)^T c$. Obviously, λ satisfies (11). Since $\hat{Z}_t$ clearly satisfies

$$\hat{Z}_t = \int_0^t exp\{A(t-\tau)\}B\hat{Y}_\tau d\tau + \hat{z}_0$$

we get

$$c^T \hat{Z}_1 = -\int_0^1 \lambda_\tau^T B \hat{Y}_\tau d\tau + c^T \hat{z}_0 \tag{15}$$

Assume that (12) holds. Then (15) implies that $\hat{Y}^0$ is optimal. Similarly, it is easily seen that (13) implies (14). Finally, if (12) does not hold then it follows from (15) that $\hat{Y}^0$ is not optimal. ■

Let $\eta(u)$ be the steady state probability on the pairs of state and actions when using a stationary policy u, i.e.

$$\eta(u; v, a) = \lim_{n\to\infty} P_x^u(X_n = v, A_n = a). \tag{16}$$

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

$$W \stackrel{\text{def}}{=} \cup_{u\in\mathcal{S}} \eta(u), \qquad V \stackrel{\text{def}}{=} \cup_{u\in\mathcal{S}} \sum_{v,a} \eta(u; v, a) y(v, a) \tag{17}$$

($y(v, a)$ are defined above (2)). Let $\overline{f}^n(\xi, u)$ be the matrix whose elements are given by $\overline{f}^n(\xi, u; v, a) \stackrel{\text{def}}{=} (n+1)^{-1} \sum_{j=0}^n P_\xi^u(X_j = v, A_j = a)$ and let

$$M_\xi^n \stackrel{\text{def}}{=} \cup_{u\in S} \overline{f}^n(\xi, u), \qquad L_\xi^n \stackrel{\text{def}}{=} \cup_{u\in U} \overline{f}^n(\xi, u)$$

Recall that the Hausdorff metric (distance) between two arbitrary bounded sets B and D is defined as follows

$$dist\{B, D\} \stackrel{\text{def}}{=} \max\{\sup_{\zeta\in B} dist\{\zeta, D\}, \sup_{\zeta\in D} dist\{\zeta, B\}\}$$

and the distance between a point ζ and an arbitrary set V is

$$dist\{\zeta, V\} \stackrel{\text{def}}{=} \inf_{\nu\in V} \|\zeta - \nu\|.$$

Here and throughout the paper $\|\cdot\|$ stands for the Euclidean norm. From (16) it follows that M_ξ^n converges to W in the Hausdorff metric sense. Moreover, the rate of convergence in (16) is geometric and hence the distance between $\overline{f}^n(\xi, u)$ and $\eta(u)$ (for stationary u) is of order n^{-1} (see e.g. [2]). Hence one may easily obtain that

$$dist\{M_\xi^n, W\} = O(n^{-1}) \tag{18}$$

where the estimate is uniform with respect to $\xi \in \mathbf{X}$.

We show below that L_ξ^n also converges to W in the Hausdorff metric sense and that the rate of convergence is also the same $O(n^{-1})$ Note that this is stronger than the well known result that $\overline{\lim}_{n\to\infty}\{\overline{f}^n(\xi,u)\} \subset W$, $\forall u \in U$, $\forall \xi$, see [4] Section 7, Theorem 2,3).

Theorem 3 *There exists a nonnegative function $\hat{\gamma} : \mathbb{N} \to \mathbb{R}$, satisfying $\overline{\lim}_{n\to\infty} \hat{\gamma}(n) \times n < \infty$, such that*

$$dist\{L_\xi^n, W\} < \hat{\gamma}(n),$$

uniformly in the initial distribution ξ on $\mathbf{X}$.

Proof: Define the transition operator G, as follows. Given a function $\zeta = \{\zeta(\nu,a)\}$, $\nu \in \mathbf{X}, a \in \mathbf{A}$, let $G : \mathbb{R}^{|\mathbf{X}\times\mathbf{A}|} \to \mathbb{R}^{|\mathbf{X}|}$ be given by $G^\nu(\zeta) := \sum_{a\in\mathbf{A}} \zeta(\nu,a) - \sum_{a\in\mathbf{A}}\sum_{v\in\mathbf{X}} \zeta(v,a)P_{va\nu}$. For $\zeta(v,a) = P_\xi^u\{X_n = v, A_n = a\}$ we get

$$\begin{aligned} G^\nu(\zeta) &= P_\xi^u(X_n=\nu) - \sum_{a\in\mathbf{A}}\sum_{w\in\mathbf{X}} P_\xi^u(X_n = w, A_n = a)P_{wa\nu} \\ &= P_\xi^u(X_n=\nu) - P_\xi^u(X_{n+1}=\nu) \end{aligned}$$

It then follows that for any ξ and $u \in U$,

$$G^\nu(\overline{f}^n(\xi,u)) = \frac{1}{n+1}\sum_{s=0}^{n}[P_\xi^u(X_s=\nu) - P_\xi^u(X_{s+1}=\nu)] = \frac{1}{n+1}\Big(\xi(\nu) - P_\xi^u(X_{n+1}=\nu)\Big),$$

and $\overline{f}^n(\xi,u)$ satisfies $\sum_{\nu,a}\overline{f}^n(\xi,u;\nu,a) = 1$, $\overline{f}^n(\xi,u) \geq 0$. Hence L_ξ^n is contained in the set

$$H_n := \left\{\zeta \in \mathbb{R}^{|\mathbf{X}|\times|\mathbf{A}|} : \zeta \geq 0; G^\nu(\zeta) \leq I(n+1)^{-1}, -G^\nu(\zeta) \leq I(n+1)^{-1}, \forall \nu \in \mathbf{X}; \sum_{\nu,a}\zeta(\nu,a) = 1\right\} \tag{19}$$

where each component of $I \in \mathbb{R}^{|\mathbf{X}|}$ is equal to one. Next we relate H_n to W. W is a convex and compact polyhedral set given by

$$W = \left\{\zeta \in \mathbb{R}^{|\mathbf{X}|\times|\mathbf{A}|} : G(\zeta) = 0, \zeta \geq 0, \sum_{\nu\in\mathbf{X},a\in\mathbf{A}}\zeta(\nu,a) = 1\right\} \tag{20}$$

(see [4] p. 95). From [7] it follows that the Hausdorff distance between H_n and W is of order n^{-1}. Moreover, it is uniform with respect to ξ since both H_n and W do not depend on ξ. Hence

$L^n_\xi \subset W + O(n^{-1})B$, where B is the closed ball with the radius one and with the center in the origin and $O(n^{-1})$ is estimated uniformly with respect to ξ.

From (18) it follows that $W \subset M^n_\xi + O(n^{-1})B$. On the other hand by the definition $M^n_\xi \subset L^n_\xi$. Thus $W \subset L^n_\xi + O(n^{-1})B$ which completes the proof of the theorem.

■

Lemma 4 *There exists a nonnegative function $\gamma : \mathbb{N} \to \mathbb{R}$ satisfying $\overline{\lim}_{n\to\infty} \gamma(n) \times n < \infty$ such that*
(i) For each $u \in U$ there exists a $\zeta^u \in V$ (where V is defined in (17)) such that

$$\left\| \frac{1}{n+1} \sum_{j=0}^{n} E^u_x y(X_j, A_j) - \zeta^u \right\| < \gamma(n), \tag{21}$$

with $\gamma(n)$ independent of u and x and with ζ^u independent of x.
(ii) In particular for $u \in \mathcal{S}$,

$$\left\| \frac{1}{n+1} \sum_{j=0}^{n} E^u_x y(X_j, A_j) - \sum_{v,a} \eta(u; v, a) y(v, a) \right\| < \gamma(n), \tag{22}$$

and

$$\left| \frac{1}{n+1} \sum_{j=0}^{n} \lambda^T_r B E^u_x y(X_j, A_j) - \sum_{v,a} \eta(u; v, a) \lambda^T_r B y(v, a) \right| < \gamma(n), \tag{23}$$

for any $r \in [0, 1]$, where λ_t is given in (11).
(iii) V is convex and compact.

Proof: From (17) and Theorem 3 it follows that

$$\inf_{\zeta\in V} \left\| \frac{1}{n+1} \sum_{j=0}^{n} E^u_\xi y(X_j, A_j) - \zeta \right\| \le \|y\| \inf_{\beta\in W} \left\| \overline{f}^n(x, u) - \beta \right\| < \|y\| \, \hat{\gamma}(n) \tag{24}$$

where $\|y\| := (\sum_{v,a}(y(v, a))^2)^{1/2}$. This establishes (i) with $\gamma(n) = \|y\| \, \hat{\gamma}(n)$. Let $u \in \mathcal{S}$. We note that

$$\left\| \frac{1}{n+1} \sum_{j=0}^{n} E^u_x y(X_j, A_j) - \sum_{v,a} \eta(u; v, a) y(v, a) \right\| \le \frac{1}{n+1} \sum_{j=0}^{n} \|P^u_x(X_j = \bullet, A_j = \bullet) - \eta(u; \bullet, \bullet)\| \, \|y\| \tag{25}$$

and

$$\left| \frac{1}{n+1} \sum_{j=0}^{n} \lambda_r^T B E_x^u y(X_j, A_j) - \sum_{v,a} \eta(u; v, a) \lambda_r^T B y(v, a) \right| \leq$$

$$\frac{1}{n+1} \sum_{j=0}^{n} \| P_x^u (X_j = \bullet, A_j = \bullet) - \eta(u; \bullet, \bullet) \| \max_{t \in [0,1]} \|\lambda_t\| \, \|B\| \, \|y\|$$

(ii.2) then follows since there exist a constants ρ and $\alpha < 1$ such that for any initial state x and stationary policy u,

$$\| P_x^u (X_n = \bullet, A_n = \bullet) - \eta(u; \bullet, \bullet) \| \leq \left\| P_x^u (X_n = \bullet) - \sum_a \eta(u; \bullet, a) \right\| < \rho \alpha^n$$

(see e.g. [1] Lemma 4.6).
(iii) follows immediately from (20) and the definition of V. ■

Proof of Theorem 1: The proof is based on relating between the stochastic problem $\mathbf{Q1}^\epsilon$ introduced in Section 1, and the deterministic problem **Q2** introduced in Section 3.

Choose an arbitrary $u \in U$. Let $\overline{Y}_t^{x,u} := E_x^u Y_t$; let $\overline{Z}_t$ be the solution of (3) (the stochastic problem) that corresponds to that choice of u, and to x given in $\mathbf{Q1}^\epsilon$.

Recall the definition of V (above Theorem 3); it will be used as the compact set $\hat{V}$ of the admissible controls for the deterministic problem **Q2**. We first show that there exists a control $\tilde{Y} \in \mathcal{K}(V)$ for the deterministic problem such that the corresponding solution $\tilde{Z}$ of (10), satisfies

$$|c^T \overline{Z}_1 - c^T \tilde{Z}_1| \leq \left\| c^T \right\| \left\| \overline{Z}_1 - \tilde{Z}_1 \right\| \leq \kappa(\epsilon) \tag{26}$$

where $\kappa : \mathbb{R} \to \mathbb{R}$ does not depend on u and satisfies $\lim_{\epsilon \to 0} \kappa(\epsilon) = 0$. Hence

$$F_x^\epsilon \geq J^\star - \kappa(\epsilon). \tag{27}$$

where F_x^ϵ, $J^\star$ are the optimal values of $\mathbf{Q1}^\epsilon$ and **Q2** respectively. The proof will then follow by showing that

$$\overline{\lim_{\epsilon \to 0}} E_x^{u^*(\epsilon,\delta)} c^T Z_1 \leq J^\star + \delta \tag{28}$$

since by combining (27) with (28) we would then get indeed

$$\overline{\lim_{\epsilon \to 0}} (E_x^{u^*(\epsilon,\delta)} c^T Z_1 - F_x^\epsilon) \leq \overline{\lim_{\epsilon \to 0}} E_x^{u^*(\epsilon,\delta)} c^T Z_1 - J^* \leq \delta \tag{29}$$

We begin thus by establishing (26). On the basis of Lemma 4 (i) and (8) one can verify that corresponding to any vector: $\int_{\tau(l,\epsilon)}^{\tau(l+1,\epsilon)} \overline{Y}_t^{x,u} dt$, there exists a $\tilde{y}_l \in V$ such that

$$\left\| \int_{\tau_l}^{\tau_{l+1}} (\overline{Y}_t^{x,u} - \tilde{y}_l) dt \right\| \leq M_1 \Delta(\epsilon) \gamma(n(\epsilon)) + M_2 \epsilon$$

where M_1 and M_2 are some positive constants. Let $\tilde{Z}_t$ be the solution of the system (10) where $\hat{Y}_t = \tilde{y}_l$ for $\tau_l \leq t < \tau_{l+1}$. We get

$$\tilde{Z}_{\tau_{l+1}} = \tilde{Z}_{\tau_l} + \int_{\tau_l}^{\tau_{l+1}} \left(A\tilde{Z}_t + B\tilde{y}_l \right) dt. \tag{30}$$

On the other hand from (3) it follows that

$$\overline{Z}_{\tau_{l+1}} = \overline{Z}_{\tau_l} + \int_{\tau_l}^{\tau_{l+1}} \left(A\overline{Z}_t + B\overline{Y}_t^{x,u} \right) dt.$$

Hence after the subtraction we get

$$\begin{aligned}
&\left\| \overline{Z}_{\tau_{l+1}} - \tilde{Z}_{\tau_{l+1}} \right\| \\
\leq \;& \left\| \overline{Z}_{\tau_l} - \tilde{Z}_{\tau_l} \right\| + \left\| \int_{\tau_l}^{\tau_{l+1}} A(\overline{Z}_t - \overline{Z}_{\tau_l}) dt \right\| + \left\| \int_{\tau_l}^{\tau_{l+1}} A(\tilde{Z}_t - \tilde{Z}_{\tau_l}) dt \right\| \\
& + \|A\| \, \Delta(\epsilon) \left\| \overline{Z}_{\tau_l} - \tilde{Z}_{\tau_l} \right\| + M_1 \Delta(\epsilon) \, \|B\| \, \gamma(n(\epsilon)) + M_2 \, \|B\| \, \epsilon \\
\leq \;& \left\| \overline{Z}_{\tau_l} - \tilde{Z}_{\tau_l} \right\| + \|A\| \, \Delta(\epsilon) \left\| \overline{Z}_{\tau_l} - \tilde{Z}_{\tau_l} \right\| + \Delta(\epsilon) \left[2M_3 \Delta(\epsilon) \, \|A\| + M_1 \, \|B\| \, \gamma(n(\epsilon)) + M_2 \epsilon \Delta(\epsilon)^{-1} \right]
\end{aligned}$$

where M_3 is a positive constant such that

$$\left\| \tilde{Z}_t - \tilde{Z}_{\tau_l} \right\| \leq M_3 \Delta(\epsilon), \quad \left\| \overline{Z}_t - \overline{Z}_{\tau_l} \right\| \leq M_3 \Delta(\epsilon) \quad \forall t \in [\tau_l, \tau_{l+1}].$$

Now it follows from Gaitsgory [5] Proposition 5.1 (whith $N_\epsilon = \Delta(\epsilon)^{-1}$) that there exists a function $C : \mathbb{R} \to \mathbb{R}$ such that for any $l = 0, 1, 2, ..., \lfloor \Delta(\epsilon)^{-1} \rfloor$.

$$\left\| \overline{Z}_{\tau_l} - \tilde{Z}_{\tau_l} \right\| \leq C(\epsilon) \tag{31}$$

and $\lim_{\epsilon \to 0} C(\epsilon) = 0$. This implies (26).

It remains to establish (28). For the sake of breavity let us use the following notations: $u(\lambda_{\tau_l}, \delta) = u_l$, $E_x^{u^\star(\epsilon,\delta)} y(X_{\lfloor t/\epsilon \rfloor}, A_{\lfloor t/\epsilon \rfloor}) = \overline{Y}_t$. Since by definition, $u^\star(\epsilon, \delta) = u_l$ during $j =$

$\lfloor \tau_l/\epsilon \rfloor, j = \lfloor \tau_l/\epsilon \rfloor + 1, ..., j = \lfloor \tau_{l+1}/\epsilon \rfloor - 1$, one may write

$$\left| \int_{\tau_l}^{\tau_{l+1}} \lambda_{\tau_l}^T B \overline{Y}_s ds - \epsilon E_x^{u^\star(\epsilon,\delta)} \sum_{j=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_{l+1}/\epsilon \rfloor - 1} \lambda_{\tau_l}^T B E_{\tilde{x}}^{u_l} y(X_j, A_j) \right| \leq M\epsilon$$

where M is some positive constant and $\tilde{x} = X_{\lfloor \tau_l/\epsilon \rfloor}$ is the random state at the moment $\lfloor \tau_l/\epsilon \rfloor$ generated by policy $u^\star(\epsilon, \delta)$ with the initial state $X_0 = x$. By Lemma 4 (ii),

$$\left| n(\epsilon)^{-1} \sum_{j=\lfloor \tau_l/\epsilon \rfloor}^{\lfloor \tau_{l+1}/\epsilon \rfloor - 1} \lambda_{\tau_l}^T B E_{\tilde{x}}^{u_l} y(X_j, A_j) - \sum_{v,a} \eta(u_l, v, a) \lambda_{\tau_l}^T B y(v, a) \right| \leq \gamma(n(\epsilon)).$$

Hence, by (8),

$$\left| \int_{\tau_l}^{\tau_{l+1}} \lambda_{\tau_l}^T B \overline{Y}_s ds - \Delta(\epsilon) \sum_{v,a} \eta(u_l, v, a) \lambda_{\tau_l}^T B y(v, a) \right| \leq M_4 \epsilon + \Delta(\epsilon) \gamma(n(\epsilon))$$

where M_4 is a positive constant. Using the definition of $u(\lambda_{\tau_l}, \delta) = u_l$ as a stationary policy which provides the fulfillment of (5) with $\Lambda = \lambda_{\tau_l}$ and taking also into account (16), (17) one obtaines

$$\sum_{v,a} \eta(u_l, v, a) \lambda_{\tau_l}^T B y(v, a) \geq \max\{\sum_{v,a} \eta(u, v, a) \lambda_{\tau_l}^T B y(v, a) | u \in \mathcal{S}\} - \delta$$

$$= \max\{\lambda_{\tau_l}^T B y | y \in V\} - \delta = h_{\tau_l} - \delta$$

where $h_t := \max\{\lambda_t^T B y | y \in V\}$. From the last two inequalities it follows it follows that

$$\int_{\tau_l}^{\tau_{l+1}} \lambda_{\tau_l}^T B \overline{Y}_s ds \geq \Delta(\epsilon)(h_{\tau_l} - \delta) - M_4 \epsilon - \Delta(\epsilon) \gamma(n(\epsilon)). \tag{32}$$

The functions $\lambda(t)$ and $h(t)$ are uniformly continuous on $[0, 1]$. This and (32) allows us to conclude that there exists a function $\hat{C}(\epsilon)$ such that

$$\int_0^1 \lambda_s^T B \overline{Y}_s ds \geq \int_0^1 h(s) ds - \delta - \hat{C}(\epsilon), \quad \lim_{\epsilon \to 0} \hat{C}(\epsilon) = 0. \tag{33}$$

Recalling now that $\overline{Z}_t$ is the solution of (3) with $E_x^u Y_t = E_x^{u^\star(\epsilon,\delta)} y(X_{\lfloor t/\epsilon \rfloor}, A_{\lfloor t/\epsilon \rfloor}) = \overline{Y}_t$ and using Lemma 2 (ii), one obtains (28) and thus completes the proof. ■

References

[1] E. Altman and A. Shwartz, Adaptive control of constrained Markov chains: criteria and policies, *Annals of Oper. Res.*, **28**, special issue on Markov Decision Processes, 1991, pp. 101-134.

[2] E. Altman and O. Zeitouni, "Rate of convergence of empirical measures and costs in controlled Markov chains and transient optimality", EE Pub. 759, Technion, May 1990, submited to *Mathematics of Operations Research.*

[3] A. Bensoussann, *Perturbation methods in optimal control problems*, John Wiley, New York, 1989.

[4] C. Derman, *Finite State Markovian Decision Processes,* Academic Press, 1970.

[5] V. Gaitsgory, Suboptimization of singularly perturbed control systems, *SIAM j. opt. control*, **30:5**, 1992, pp 1228-1249.

[6] V. Gaitsgory, Use of the averaging method in control problems, *Differential Equations* (translated from Russian), **22**, 1986, pp. 1290-1299.

[7] A. J. Hoffman, On approximate solutions of systems of linear inequalities, *J. Research National Bureau of Standards*, **49**, No. 2, 1959.

[8] L. Kleinrock, *Queuing Systems, Volume II: Computer Applications*, John Wiley, New York, 1976.

[9] P. V. Kokotovic, Applications of singular perturbations techniques to control problems, *SIAM Review*, **26**, 1984, pp. 501-550.

[10] P. V. Kokotovic, H. Khalil and J. O'Reilly, *Singular perturbations in control analysis and design*, Academic Press, New-York, 1986.

[11] R. E. O'Malley, *Introduction to singular perturbations*, Academic Press, New-York, 1974.