SIAM J. CONTROL OPTIM.
Vol. 43, No. 1, pp. 325–340

ON A REPRESENTATION OF THE LIMIT OCCUPATIONAL MEASURES SET OF A CONTROL SYSTEM WITH APPLICATIONS TO SINGULARLY PERTURBED CONTROL SYSTEMS*

VLADIMIR GAITSGORY†

Abstract. A representation of the limit occupational measures set of a control system in terms of the vector function defining the system's dynamics is established. Applications in averaging of singularly perturbed control systems are demonstrated.

Key words. singularly perturbed control systems, occupational measures, averaging method, limit occupational measures sets, approximation of slow motions

AMS subject classifications. 34E15, 34C29, 34A60, 93C70

DOI. 10.1137/S0363012903424186

1. Introduction and preliminaries.

Under certain conditions, the set of occupational measures generated by admissible controls and corresponding solutions of a control system converges (as the length of the time interval tends to infinity) to a limit set. If this limit set is independent of the initial conditions within some subset of the state space, it is called a limit occupational measures set (LOMS) of the system.

Criteria for the existence of the LOMS were discussed in [24], [25], where it was used as a tool for analysis of singularly perturbed control systems (SPCS).

In this paper, we give a natural representation for the LOMS which enhances its applications in SPCS. Our main results are stated in Theorem 2.1. We establish that, under the assumptions made, the convex hull of a union of occupational measures sets converges to a convex and compact set of probability measures defined in (2.6) (Theorem 2.1(i)) and that the LOMS of the control system is equal to this set if it exists (Theorem 2.1(ii)). We also give necessary and sufficient conditions for the existence of the LOMS (Theorem 2.1(iii)).

The paper consists of six sections. Theorem 2.1 is stated in section 2 and proved in sections 4–6. Applications in SPCS are discussed in section 3.

Singularly perturbed problems of control and optimization have been studied intensively in both deterministic and stochastic settings (see [1], [2], [3], [4], [5], [6], [7], [8], [11], [12], [14], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [36], [37], [38], [39], [40], [41], [42], [45], [46], [48] and the references therein).

Originally, the most common approaches to SPCS, especially in the deterministic case, were related to an approximation of the SPCS by the systems obtained via equating the singular perturbations parameter to zero (with further application of the boundary layer method (see [37], [44]) for an asymptotical description of the fast dynamics). This type of approach was successfully applied to a number of important classes of problems (see [30], [31], [38] and also [17], [36], [39], [45] for some recent results obtained in this direction).

*Received by the editors March 10, 2003; accepted for publication (in revised form) December 1, 2003; published electronically June 25, 2004. This work was supported by Australian Research Council Discovery-Project grant DP0346099 and IREX grant X00106494.
http://www.siam.org/journals/sicon/43-1/42418.html

†School of Mathematics, University of South Australia, The Mawson Lakes Campus, Mawson Lakes SA 5095, Australia (v.gaitsgory@unisa.edu.au).

Various averaging type approaches allowing a consideration of more general classes of SPCS, in which the equating of the small parameter to zero may not lead to a right approximation, were developed in [2], [3], [4], [5], [6], [7], [11], [20], [21], [22], [23], [24], [25], [26], [27], [41], [46].

In [24] and [25], in particular (see also [3], [5], [46] for related results), the slow trajectories were approximated by the solutions of the averaged system in which the controls are measure-valued and take their values in the LOMS of the associated system (that is, the system that would describe the fast dynamics if the slow state variables were "frozen"). The paper continues this line of research by establishing that the LOMS allows a representation in terms of the vector function defining the right-hand side of the associated system.

Let us introduce some notation and definitions which are used throughout the paper. Given a compact metric space W, $\mathcal{B}(W)$ will stand for the σ-algebra of its Borel subsets and $\mathcal{P}(W)$ will denote the set of probability measures defined on $\mathcal{B}(W)$. The set $\mathcal{P}(W)$ will be treated as a compact metric space with a metric ρ, which is consistent with its weak convergence topology (see, e.g., [13]). A sequence $\gamma^k \in \mathcal{P}(W)$ converges to $\gamma \in \mathcal{P}(W)$ in this metric if and only if

$$\lim_{k\to\infty} \int_W q(w)\gamma^k(dw) \;=\; \int_W q(w)\gamma(dw) \tag{1.1}$$

for any continuous $q(w) : W \to \mathbb{R}^1$. There are many ways of defining such a metric ρ. In this paper, we will use the following definition: $\forall \gamma', \gamma'' \in \mathcal{P}(W)$,

$$\rho(\gamma', \gamma'') \stackrel{\text{def}}{=} \sum_{l=1}^{\infty} \frac{1}{2^l} \left| \int_W q_l(w)\gamma'(dw) - \int_W q_l(w)\gamma''(dw) \right|, \tag{1.2}$$

where $q_l(\cdot)$, $l = 1, 2, \ldots$, is a sequence of Lipschitz continuous functions which is dense in the unit ball of $C(W)$ (the space of continuous functions on W). Using the metric ρ, one can define the Hausdorff metric ρ_H on the set of subsets of $\mathcal{P}(W)$ as follows: $\forall\, \Gamma_i \subset \mathcal{P}(W)$, $i = 1, 2$,

$$\rho_H(\Gamma_1, \Gamma_2) \stackrel{\text{def}}{=} \max \left\{ \sup_{\gamma\in\Gamma_1} \rho(\gamma, \Gamma_2), \sup_{\gamma\in\Gamma_2} \rho(\gamma, \Gamma_1) \right\}, \tag{1.3}$$

where $\rho(\gamma, \Gamma_i) \stackrel{\text{def}}{=} \inf_{\gamma'\in\Gamma_i} \rho(\gamma, \gamma')$. It can be verified (see, e.g., Lemma Π2.4, p. 205 in [22]) that, with the definition of the metric ρ as in (1.2),

$$\rho_H(co\Gamma_1, co\Gamma_2) \le \rho_H(\Gamma_1, \Gamma_2), \tag{1.4}$$

where *co* stands for the convex hull of the corresponding set.

In what follows, we will deal with the convergence in the Hausdorff metric of sets in $\mathcal{P}(W)$ defined as unions of occupational measures. Given a measurable function $w(t) : [0, S] \to W$, the occupational measure $p^{w(\cdot)} \in \mathcal{P}(W)$ generated by this function is defined by taking

$$p^{w(\cdot)}(Q) \stackrel{\text{def}}{=} \frac{1}{S} \text{meas} \Big\{ t \mid w(t) \in Q \Big\} \qquad \forall Q \in \mathcal{B}(W),$$

where meas $\{\cdot\}$ stands for the Lebesgue measure on $[0, S]$.

2. Main theorem. Consider a control system

$$\dot{y}(\tau) \ = \ f(u(\tau), y(\tau)), \quad \tau \in [0, S], \tag{2.1}$$

where the function $f(u, y) : U \times \mathbb{R}^m \to \mathbb{R}^m$ is continuous in (u, y) and satisfies Lipschitz conditions in y, U is a compact metric space, and the controls are Lebesgue measurable functions $u(\tau) : [0, S] \to U$.

Let Y be a compact subset of $\mathbb{R}^m$ and $Y^\delta \stackrel{\text{def}}{=} Y + \delta B$, where δ is a positive number and B is the closed unit ball in $\mathbb{R}^m$, and let us introduce the following definition and assumptions.

DEFINITION. *A pair $(u(\tau), y(\tau))$ is called admissible (δ-admissible) for system* (2.1) *on the interval* $[0, S]$ *if $u(\tau)$ is a control, $y(\tau)$ is the corresponding solution of* (2.1)*, and $y(\tau) \in Y$ ($y(\tau) \in Y^\delta$) $\forall \tau \in [0, S]$.*

Assumption I. For any initial condition $y(0) \in Y$, there exists a control $u(\tau)$ such that the corresponding solution of (2.1) does not leave Y on $[0, S]$ for any $S > 0$.

Assumption II. For any Lipschitz continuous function $g(u, y) : U \times \mathbb{R}^m \to \mathbb{R}^1$,

$$\left| \frac{1}{S} \sup_{(u(\cdot), y(\cdot))} \int_0^S g(u(\tau), y(\tau)) d\tau \ - \frac{1}{S} \sup_{(u^\delta(\cdot), y^\delta(\cdot))} \int_0^S g(u^\delta(\tau), y^\delta(\tau)) d\tau \right| \stackrel{\text{def}}{=} \mu_g(\delta, S) \ \to \ 0 \tag{2.2}$$

as $\delta \to 0$ and $S \to \infty$, where the *sup*s in the above expression are, respectively, over all admissible pairs and over all δ-admissible pairs which satisfy the condition $y^\delta(0) \in Y$.

Assumption I is equivalent to the assumption that the viability kernel of Y is equal to Y. It is satisfied, for example, if, for any $y \in Y$, there exists $u \in U$ such that $f(u, y) = 0$. More general sufficient (and necessary) conditions for this assumption to be satisfied can be found in [9], [10].

Note that if Assumption I is replaced by a stronger assumption that Y is a forward invariant set—that is, all solutions of (2.1) obtained with the measurable controls $u(\tau) : [0, S] \to U$ $\forall S > 0$ do not leave Y—then Assumption II is satisfied automatically. In this case, all δ-admissible pairs satisfying the condition $y^\delta(0) \in Y$ are admissible and (2.2) is valid with $\mu_g(\delta, S) \equiv 0$.

Let $(u(\tau), y(\tau)) : [0, S] \to U \times Y$ be an admissible pair and let $p^{(u(\cdot), y(\cdot))} \in \mathcal{P}(U \times Y)$ be the occupational measure generated by this pair. Denote by $\Gamma(S, y)$ and $\Gamma(S, Y)$ the sets of occupational measures defined by the equations

$$\Gamma(S, y) \stackrel{\text{def}}{=} \bigcup_{(u(\tau),\ y(\tau))} \left\{ p^{(u(\cdot), y(\cdot))} \right\}, \qquad \Gamma(S, Y) \stackrel{\text{def}}{=} \bigcup_{y \in Y} \left\{ \Gamma(S, y) \right\}, \tag{2.3}$$

where the first union is over all admissible pairs of (2.1) satisfying the initial conditions $y(0) = y$ and the second is over all initial conditions

$$y(0) \ = \ y \ \in \ Y. \tag{2.4}$$

DEFINITION. *A convex and compact set $\Gamma \subset \mathcal{P}(U \times Y)$ is called the LOMS of system* (2.1) *on Y if there exists a function $\nu(S)$, $\lim_{S \to \infty} \nu(S) = 0$, such that*

$$\rho_H(\Gamma(S, y), \Gamma) \ \leq \ \nu(S) \quad \forall y \ \in \ Y. \tag{2.5}$$

In Theorem 2.1 below, we relate the LOMS of system (2.1) on Y to the set $W \subset \mathcal{P}(U \times Y)$ defined by the equation

$$W \stackrel{\text{def}}{=} \left\{ \gamma \mid \gamma \in \mathcal{P}(U \times Y); \ \int_{U \times Y} (\phi'(y))^T f(u, y) \gamma(du, dy) = 0 \ \forall \phi(\cdot) \in C^1 \right\}, \tag{2.6}$$

where C^1 is the space of continuously differentiable functions $\phi(y) : \mathbb{R}^m \to \mathbb{R}^1$ and $\phi'(y)$ is the vector column of partial derivatives (the gradient) of $\phi(y)$. Note that, as can be easily verified, the set W is convex and compact in the weak convergence topology of $\mathcal{P}(U \times Y)$.

THEOREM 2.1. *Let Assumptions* I *and* II *be satisfied. Then the following hold:*

(i) *The estimate*

$$\rho_H(co\Gamma(S,Y),W) \;\le\; \bar{\nu}(S) \tag{2.7}$$

is valid for some $\bar{\nu}(S)$, $\lim_{S\to\infty}\bar{\nu}(S) = 0$.

(ii) *If the LOMS* Γ *of system* (2.1) *on* Y *exists, it is equal to the set* W*:*

$$\Gamma \;=\; W. \tag{2.8}$$

(iii) *The LOMS* Γ *of system* (2.1) *on* Y *exists if and only if*

$$\rho_H(\Gamma(S,y'),\Gamma(S,y'')) \;\le\; \hat{\nu}(S) \quad \forall y', y'' \in Y, \tag{2.9}$$

for some $\hat{\nu}(S)$, $\lim_{S\to\infty}\hat{\nu}(S) = 0$.

Proof. Statements (i) and (ii) of the theorem are proved in sections 5–6. Statement (iii) is proved in section 4. □

3. On applications in SPCS. Let us consider an SPCS defined on the interval $[0,T]$ $(T > 0)$ by the equations

$$\epsilon\dot{y}(t) = f(u(t), y(t), z(t)), \qquad y(0) = y_0, \tag{3.1}$$

$$\dot{z}(t) = g(u(t), y(t), z(t)), \qquad z(0) = z_0, \tag{3.2}$$

where $\epsilon > 0$ is a small parameter; $f : U \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$, $g : U \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ are continuous vector functions satisfying Lipschitz conditions in z and y; U is a compact metric space, and the controls are measurable functions satisfying the inclusion $u(t) \in U$.

Along with (3.1)–(3.2), let us consider the associated system

$$\dot{y}(\tau) = f(u(\tau), y(\tau), z), \quad \tau \in [0,S], \tag{3.3}$$

in which (in contrast to (3.1)) z is a vector of fixed parameters: z = const. The controls in (3.3) are measurable functions satisfying the inclusion $u(\tau) \in U$.

Assume that, $\forall z$ from a sufficiently large area $Z \subset \mathbb{R}^n$, the solutions of the associated system (3.3) with the initial conditions in a compact set $Y \subset \mathbb{R}^m$ do not leave this set $\forall \tau \ge 0$ (that is, Y is forward invariant with respect to the solutions of (3.3)). Assume also that the solutions of the SPCS (3.1)–(3.2) do not leave $Y \times Z'$ for $t \in [0,T]$, where Z' is a compact set belonging to the interior of Z.

Let $u(\tau) \in U$ be a control and $y(\tau) \in Y$ be the solution of the associated system (3.3), obtained with this control and the initial condition $y(0) = y$. Denote by $\gamma^{(u(\cdot),y(\cdot))} \in \mathcal{P}(U \times Y)$ the occupational measure generated by the pair $(u(\cdot), y(\cdot)) : [0,S] \to U \times Y$ and denote by $\Gamma(z,S,y) \subset \mathcal{P}(U \times Y)$ the union of all such occupational measures.

Assume that the LOMS $\Gamma(z)$ of the associated system exists, that is,

$$\lim_{S\to\infty} \rho_H(\Gamma(z,S,y), \Gamma(z)) = 0,$$

with the convergence being uniform with respect to $(y, z) \in Y \times Z$. Note that, by Theorem 2.1(ii), $\Gamma(z) = W(z)$ (the latter is defined in (2.6), with the dependence on z being due to the fact that the vector function $f(\cdot)$ includes the dependence on z).

Define $\tilde{g}(\gamma, z) : \mathcal{P}(U \times Y) \to \mathbb{R}^n$ by the equation

$$\tilde{g}(\gamma, z) \stackrel{\text{def}}{=} \int_{U \times Y} g(u, y, z)\gamma(du, dy)$$

and consider the *averaged* system

$$\dot{z}(t) = \tilde{g}(\gamma(t), z(t)), \qquad z(0) = z_0, \tag{3.4}$$

in which the controls are Lebesgue measurable functions $\gamma(\cdot) : [0, T] \to \mathcal{P}(U \times Y)$ satisfying the inclusion

$$\gamma(t) \in W(z(t)). \tag{3.5}$$

The following result is a corollary of Theorem 2.1(ii) and Theorem 4.2 in [24].

COROLLARY 3.1. *Let the assumptions made above be satisfied. Also let the multivalued map* $V_g(\cdot) : Z \to 2^{R^n}$,

$$V_g(z) \stackrel{\text{def}}{=} \bigcup_{\gamma \in W(z)} \{\tilde{g}(\gamma, z)\},$$

be Lipschitz continuous. Then the following hold:

(i) *Corresponding to any solution* $(z^{sp}(t), y^{sp}(t))$ *of* (3.1)–(3.2) *there exists a solution* $z^{av}(t)$ *of* (3.4) *such that*

$$\max_{t \in [0,T]} ||z^{sp}(t) - z^{av}(t)|| \le \mu(\epsilon), \qquad \lim_{\epsilon \to 0} \mu(\epsilon) = 0. \tag{3.6}$$

(ii) *Corresponding to any solution* $z^{av}(t)$ *of* (3.4) *there exists a solution* $(z^{sp}(t), y^{sp}(t))$ *of* (3.1)–(3.2) *which satisfies* (3.6).

Proof. The proof follows from Theorem 4.2 in [24] with the replacement of $\Gamma(z)$ by $W(z)$ (see also Theorem 2.6 in [25] and related results in [3]). □

Sufficient conditions for the assumptions used in Corollary 3.1 to be valid have been discussed in [24], [25], where it was noticed, in particular, that these assumptions (including the existence of the LOMS and the Lipschitz continuity of $V_g(z)$) are satisfied if there exist positive definite matrices C and D such that, for any $u \in U$, any $y^1, y^2 \in R^m$, and any $z \in Z$,

$$(f(u, y^1, z) - f(u, y^2, z))^T C(y^1 - y^2) \le -(y^1 - y^2)^T D(y^1 - y^2).$$

The existence of such C and D can be guaranteed, for example, if $f(u, y, z) = A(z)y + B(z)u$, where $A(z)$ and $B(z)$ are matrices functions of the corresponding dimensions and the eigenvalues of $A(z)$ have negative real parts $\forall z \in Z$.

Let $G(\cdot) : \mathbb{R}^n \to \mathbb{R}^1$ be a continuous function. From Corollary 3.1 it follows that the optimal value of the problem

$$\inf_{(z^{sp}(\cdot), y^{sp}(\cdot))} G(z^{sp}(T)) \stackrel{\text{def}}{=} G^{\star}_{\epsilon}, \tag{3.7}$$

where inf is over the solutions of (3.1)–(3.2), converges (as ϵ tends to zero) to the optimal value of the problem

$$\inf_{z^{av}(\cdot)} G(z^{av}(T)) \stackrel{\text{def}}{=} G^{\star}, \tag{3.8}$$

where inf is over the solutions of (3.4). That is,

$$\lim_{\epsilon\to 0} G^\star_\epsilon = G^\star. \tag{3.9}$$

Also, it can be shown that a near optimal solution of (3.7) can be constructed on the basis of the optimal (or near optimal) solution of (3.8) (see details in [23], [24], [25], [27]).

Note that statement (i) of Corollary 3.1 remains valid even in the event the assumption about the existence of the LOMS is not satisfied, with (3.9) being replaced in this case by a weaker statement that

$$\liminf_{\epsilon\to 0} G^\star_\epsilon \geq G^\star.$$

The validity of this can be established via a straightforward extension of the averaging techniques used in [23], [24], [25], [27] in combination with Theorem 2.1(i).

In conclusion, let us observe that a numerical analysis of the solutions of (3.4) satisfying the inclusion (3.5) can be based on the fact that the set $W(z)$ allows the representation in the form of a countable system of equations:

$$W(z) = \left\{ \gamma \mid \gamma \in \mathcal{P}(U\times Y); \int_{U\times Y} (\phi_i'(y))^T f(u,y,z)\gamma(du,dy) = 0\ ,\ i=1,2,\dots \right\}\ . \tag{3.10}$$

Here, $\{\phi_i(\cdot)\}$ is a sequence of continuously differentiable functions such that any function $\phi(\cdot)\in C^1$ and its gradient $\phi'(\cdot)$ can be simultaneously approximated on Y by linear combinations of functions from $\{\phi_i\}$ and their corresponding gradients. (An example of such a sequence is the sequence of the monomials $y_1^{i_1}\dots y_m^{i_m}$, $i_1,\dots,i_m = 0,1,\dots$, where y_j $(j=1,\dots,m)$ stands for the jth component of y; see, e.g., [33].) To numerically approximate the solutions of the averaged system (3.4)–(3.5) one may need to truncate the system of equations in (3.10) and subsequently approximate the resulting set by the set of measures supported on a grid. The details of such a procedure will be studied in a different paper.

4. Sets of time averages and proof of Theorem 2.1(iii). Let $\delta_0 > 0$ be fixed and $q_l(u,y) : U\times Y^{\delta_0} \to \mathbb{R}^1$, $l = 1,2,\dots$, be a sequence of Lipschitz continuous functions which is dense in the space of continuous functions on $U\times Y^{\delta_0}$. Let

$$h(u,y)\ =\ (q_1(u,y),\dots,q_j(u,y)),\ j=1,2,\dots, \tag{4.1}$$

and let $V_h(S,y)$ be the set of time averages defined by the equation

$$V_h(S,y) \overset{\text{def}}{=} \bigcup_{(u(\cdot),\ y(\cdot))} \left\{ \frac{1}{S}\int_0^S h\big(u(\tau),y(\tau)\big)d\tau \right\}, \tag{4.2}$$

where the union is over all admissible pairs of (2.1) satisfying the initial conditions (2.4).

The proof of Theorem 2.1(iii) is based on the following proposition.

Proposition 4.1. *Let Assumption* I *be satisfied. Then the following hold:*

(i) *The LOMS* Γ *of system* (2.1) *on* Y *exists if and only if, for every vector function* $h(u,y)$ *defined in* (4.1)*, there exist a convex and compact set* V_h *and a function* $\nu_h(S)$, $\lim_{S\to\infty}\nu_h(S) = 0$*, such that*

$$d_H(V_h(S,y),V_h)\ \leq\ \nu_h(S)\quad \forall y\ \in\ Y, \tag{4.3}$$

where $d_H(\cdot,\cdot)$ is the Hausdorff metric defined on the bounded subsets of the corresponding finite dimensional space by the Euclidean norm.

(ii) *For a given $h(u,y)$ as in* (4.1)*, a convex and compact set V_h satisfying* (4.3) *exists if and only if*

$$d_H(V_h(S,y'),V_h(S,y'')) \le \hat{\nu}_h(S) \quad \forall y',y'' \in Y \tag{4.4}$$

for some $\hat{\nu}_h(S)$ tending to zero as S tends to infinity.

Proof. The proof of Proposition 4.1(i) is similar to that of Theorem 3.1 in [24] (see also Corollary 3.7 in [25] and the more general result in [7]). The proof of the "if" statement in Proposition 4.1(ii) follows exactly the same steps as that of Proposition 3.2 in [26] (see also [21], [22], and [7]), where this statement was proved for the case when Y is forward invariant with respect to the system (2.1). The proof of the "only if" statement is obvious. □

Proof of Theorem 2.1(iii). If (2.9) is valid, then the estimate (4.4) is true for every $h(u,y)$ defined in (4.1). Hence, by Proposition 4.1(ii), for every such $h(u,y)$, there exists a convex and compact set V_h satisfying (4.3). This, by Proposition 4.1(i), implies the existence of the LOMS. Thus, the "if" statement in Theorem 2.1(iii) is proved. The proof of the "only if" statement is obvious. □

To conclude this section, let us show that Assumption II can be equivalently reformulated in terms of convergence to zero of the Hausdorff metric between the sets of time averages defined below. For $0 < \delta \le \delta_0$, let

$$V_h^\delta(S,y) \stackrel{\text{def}}{=} \bigcup_{(u^\delta(\cdot),\ y^\delta(\cdot))} \left\{ \frac{1}{S}\int_0^S h\big(u^\delta(\tau),y^\delta(\tau)\big)d\tau \right\}, \tag{4.5}$$

where, in contrast to (4.2), the union is over all δ-admissible pairs of (2.1) satisfying the initial conditions (2.4). Denote

$$V_h(S,Y) \stackrel{\text{def}}{=} \bigcup_{y\in Y} \{V_h(S,y)\}, \qquad V_h^\delta(S,Y) \stackrel{\text{def}}{=} \bigcup_{y\in Y} \{V_h^\delta(S,y)\}.$$

The following lemma is used in the proof of Theorem 2.1(i) (see section 5 below).

LEMMA 4.2. *Assumption* II *is equivalent to that, for any $h(\cdot)$ as in* (4.1)*,*

$$d_H(coV_h(S,Y),coV_h^\delta(S,Y)) \stackrel{\text{def}}{=} \bar{\nu}_h(\delta,S) \to 0 \tag{4.6}$$

as $\delta \to 0$ and $S \to \infty$.

Proof. Let $\Psi_V(\cdot)$ stand for the support function of a set $V \subset R^j$. That is, for any $\eta \in R^j$, $\Psi_V(\eta) \stackrel{\text{def}}{=} \sup_{v\in V} \eta^T v$. If Assumption II is satisfied, then, taking $g(u,y) = \eta^T h(u,y)$ in (2.2), one can obtain that the function $\mu_g(\delta,S)$ defined by the equation

$$|\Psi_{coV_h(S,Y)}(\eta) - \Psi_{coV_h^\delta(S,Y)}(\eta)| = |\Psi_{V_h(S,Y)}(\eta) - \Psi_{V_h^\delta(S,Y)}(\eta)| \stackrel{\text{def}}{=} \mu_g(\delta,S) \tag{4.7}$$

tends to zero as $\delta \to 0$ and $S \to \infty$. Using a standard argument based on the separability of convex sets, one can verify that (4.7) implies (4.6). Thus, the validity of (4.6) is implied by the validity of Assumption II.

Now let (4.6) be satisfied for any $h(\cdot)$ constructed as in (4.1). Then $\mu_g(\delta,S)$ in (4.7) tends to zero as δ tends to zero and S tends to infinity. By taking all but one component of η to be equal to zero in (4.7), one can verify that (2.2) is valid for any $g(u,y) = q_l(u,y)$, $l = 1,2,\ldots$. Since the sequence $q_l(u,y)$, $l = 1,2,\ldots$, is dense in $C(U \times Y^\delta)$, it implies that (2.2) is valid for any continuous (and, in particular, Lipschitz continuous) $g(\cdot)$. This completes the proof of the lemma. □

5. Proofs of Theorem 2.1(i) and Theorem 2.1(ii).

Proof of Theorem 2.1(i). To prove the required statement, one needs to establish the validity of the following two inequalities:

$$\sup_{\gamma \in W} \rho(\gamma, co\Gamma(S,Y)) \leq \bar{\nu}(S), \tag{5.1}$$

$$\sup_{\gamma \in co\Gamma(S,Y)} \rho(\gamma, W) \leq \bar{\nu}(S). \tag{5.2}$$

Let us first prove the validity of (5.2). It is straightforward to verify that from the convexity of W it follows that

$$\sup_{\gamma \in co\Gamma(S,Y)} \rho(\gamma, W) = \sup_{\gamma \in \Gamma(S,Y)} \rho(\gamma, W).$$

Hence, to prove (5.2) it is enough to show that the function $\bar{\nu}(S)$ defined by the equation

$$\bar{\nu}(S) \stackrel{\text{def}}{=} \sup_{\gamma \in \Gamma(S,Y)} \rho(\gamma, W) \tag{5.3}$$

tends to zero as S tends to infinity. Assume this is not the case. Then there exist a positive number δ, a sequence $S^k \to \infty$, and sequences $y^k \in Y$ and $\gamma^k \in \Gamma(S^k, y^k)$ such that $\rho(\gamma^k, W) \geq \delta$, $k = 1, 2, \ldots$. Without loss of generality one may assume that there exists $\lim_{k\to\infty} \gamma^k \stackrel{\text{def}}{=} \gamma \in \mathcal{P}(U \times Y)$ (since $\mathcal{P}(U \times Y)$ is compact). From the continuity of the metric it follows that

$$\rho(\gamma, W) \geq \delta. \tag{5.4}$$

By the definition of the convergence in $\mathcal{P}(U \times Y)$ (see (1.1)),

$$\lim_{k\to\infty} \int_{U\times Y} (\phi\prime(y))^T f(u,y)\gamma^k(du,dy) \;=\; \int_{U\times Y} (\phi\prime(y))^T f(u,y)\gamma(du,dy) \tag{5.5}$$

for any $\phi \in C^1$. Also, from the fact that $\gamma^k \in \Gamma(S^k, y^k)$, it follows that there exists an admissible pair $(u^k(\tau), y^k(\tau))$ (for system (2.1) on the interval $[0, S^k]$) such that

$$\int_{U\times Y} (\phi\prime(y))^T f(u,y)\gamma^k(du,dy) \;=\; \frac{1}{S^k}\int_0^{S^k} (\phi\prime(y^k(\tau)))^T f(u^k(\tau), y^k(\tau))d\tau.$$

The second integral is apparently equal to

$$\frac{\phi(y^k(S^k)) - \phi(y^k(0))}{S^k}$$

and tends to zero as S^k tends to infinity (since $y^k(\tau) \in Y \;\forall\; \tau \in [0, S^k]$ and Y is a compact set). This and (5.5) imply that

$$\int_{U\times Y} (\phi\prime(y))^T f(u,y)\gamma(du,dy) \;=\; 0 \;\; \forall \phi \in C^1 \quad \Rightarrow \quad \gamma \;\in\; W.$$

The latter contradicts (5.4) and, hence, $\bar{\nu}(s)$ defined in (5.3) tends to zero as S tends to infinity. This proves (5.2).

Let us now prove the validity of inequality (5.1). Let $\hat{r}$ be a positive number such that Y is contained in the interior of $\hat{r}B \stackrel{\text{def}}{=} \hat{B}$ (as above, B is the closed unit ball in $\mathbb{R}^m$) and let $\psi(y) : \mathbb{R}^m \to [0,1]$ be a continuously differentiable function such that

$$\psi(y) \ = \ 1 \quad \forall y \ \in Y, \qquad \psi(y) \ = \ 0 \quad \forall y \ \in \mathbb{R}^m / \hat{B}. \tag{5.6}$$

Define the function $F(u,y) : U \times \mathbb{R}^m \to \mathbb{R}^m$ by the equation

$$F(u,y) \ \stackrel{\text{def}}{=} \ \psi(y) f(u,y). \tag{5.7}$$

Note that, by (5.6),

$$F(u,y) \ = \ f(u,y) \quad \forall (u,y) \ \in U \times Y, \qquad F(u,y) = 0 \quad \forall (u,y) \ \in U \times \mathbb{R}^m / \hat{B}. \tag{5.8}$$

Let $\bar{C}(U \times \mathbb{R}^m)$ be the space of bounded continuous functions on $U \times \mathbb{R}^m$ taking values in $\mathbb{R}^1$, and let $\hat{C}^1$ be the space of continuously differentiable functions on $\mathbb{R}^m$ taking values in $\mathbb{R}^1$ and vanishing at infinity. Note that, since Y is a compact set, one can replace C^1 by $\hat{C}^1 \subset C^1$ in (2.6) without adding new elements to the set W. Define a linear operator $A : \hat{C}^1 \to \bar{C}(U \times \mathbb{R}^m)$ by the equation

$$(A\phi)(u,y) \ \stackrel{\text{def}}{=} \ (\phi'(y))^T F(u,y) \quad \forall \phi \ \in \hat{C}^1.$$

This operator satisfies the conditions of Theorem 4.1 in [43], namely, the following:

(i) $\hat{C}^1$, the domain of A, is an algebra and is dense in the space $\hat{C}$ of continuous functions on $\mathbb{R}^m$ which vanish at infinity (this is an immediate consequence of the Stone–Weierstrass theorem; see, e.g., Theorem IV.6.16 in [15]).

(ii) For each $\phi \in \hat{C}^1$ and $u \in U$, $(A\phi)(u,\cdot) \stackrel{\text{def}}{=} (\phi'(\cdot))^T F(u,\cdot) \in \hat{C}$.

(iii) For each $\phi \in \hat{C}^1$,

$$\lim_{||y|| \to \infty} \max_{u \in U} (A\phi)(u,y) \ = \ \lim_{||y|| \to \infty} \max_{u \in U} (\phi'(y))^T F(u,y) \ = \ 0.$$

(iv) For each $u \in U$, the operator $A_u \phi \stackrel{\text{def}}{=} (A\phi)(u,\cdot)$ satisfies the positive maximum principle, i.e., if $\phi(y^*) = \sup_y \phi(y) > 0$, then

$$(A_u \phi)(y^*) \ = \ (\phi'(y^*))^T F(u,y^*) \ \leq \ 0.$$

Note that (ii) and (iii) are satisfied because of (5.8) and that (iv) follows from the fact that $\phi'(y^*) = 0$.

Let us consider now an arbitrary $\gamma \in W$ and extend its definition to the Borel subsets of $U \times \mathbb{R}^m$ by taking $\gamma(Q) \stackrel{\text{def}}{=} \gamma(Q \cap (U \times Y))$ $\forall Q \in \mathcal{B}(U \times \mathbb{R}^m)$. By (5.8) and (2.6),

$$\int_{U \times \mathbb{R}^m} (A\phi)(u,y) \gamma(du,dy) \ = \ \int_{U \times \mathbb{R}^m} (\phi'(y))^T F(u,y) \gamma(du,dy)$$

$$= \ \int_{U \times Y} (\phi'(y))^T F(u,y) \gamma(du,dy) \ = \ \int_{U \times Y} (\phi'(y))^T f(u,y) \gamma(du,dy) \ = \ 0$$

$\forall \phi \in \hat{C}^1$. From Theorem 4.1 in [43] it follows that there exist a probability space $(\Omega, \mathcal{F}, P)$, a filtration $\{\mathcal{F}_\tau\}$ of σ-subalgebras of $\mathcal{F}$, and a $\mathcal{P}(U) \times \mathbb{R}^m$-valued random process $(\lambda(\tau), y(\tau)) \ = \ (\lambda(\tau,\omega), y(\tau,\omega))$ which satisfies the following conditions:

(a) $(\lambda(\cdot), y(\cdot))$ is $\{\mathcal{F}_\tau\}$-progressive and stationary with

$$E[\lambda(\tau)(D_1)\chi_{D_2}(y(\tau))] = \gamma(D_1 \times D_2) \quad \forall \tau \ \geq \ 0, \tag{5.9}$$

where D_1 and D_2 are arbitrary Borel subsets of U and $\mathbb{R}^m$, respectively, and $\chi_{D_2}(\cdot)$ is the indicator function of D_2.

(b) For every $\phi \in \hat{C}^1$, $\phi(y(\tau)) - \int_0^\tau \int_U (\phi'(y(s)))^T F(u, y(s))\lambda(s)(du)ds$ is an $\{\mathcal{F}_\tau\}$-martingale.

A further characterization of the pair $(\lambda(\tau,\omega), y(\tau,\omega))$ is given by the following lemma.

Lemma 5.1. *There exists a subset Δ of Ω such that $P(\Delta) = 0$ and such that, $\forall \omega \in \Omega/\Delta$, the pair $(\lambda(\tau,\omega), y(\tau,\omega))$ satisfies the equation*

$$\dot{y}(\tau,\omega) \ = \ \bar{F}(\lambda(\tau,\omega), y(\tau,\omega)) \tag{5.10}$$

for almost all $\tau \in [0,S]$ $(\forall S > 0)$, where

$$\bar{F}(\lambda, y) \ \stackrel{\text{def}}{=} \int_U F(u,y)\lambda(du). \tag{5.11}$$

Proof. Proof of the lemma is given in section 6 below. □

Let τ_i, $i = 1, 2, \ldots$, stand for a sequence of all rational numbers belonging to the interval $[0,S]$. By (5.9),

$$P\{\omega \mid y(\tau_i,\omega) \ \in Y\} \ = \ E[\chi_Y(y(\tau_i))] \ = \ E[\lambda(\tau_i)(U)\chi_Y(y(\tau_i))] \ = \ \gamma(U \times Y) \ = \ 1.$$

That is, for every i there exists a subset Δ_i of Ω such that $P(\Delta_i) = 0$ and such that

$$y(\tau_i,\omega) \in Y \ \forall \omega \in \Omega/\Delta_i \quad \Rightarrow \quad y(\tau_i,\omega) \in Y, \ i = 1,2,\ldots, \ \forall \omega \in \Omega/(\cup_i \Delta_i). \tag{5.12}$$

From the fact that $y(\tau,\omega)$ satisfies (5.10) it follows that $y(\cdot,\omega)$ is continuous (in fact, absolutely continuous) for $\omega \in \Omega/\Delta$. Thus, the inclusions (5.12) and the fact that Y is compact imply that

$$y(\tau,\omega) \ \in Y \quad \forall \tau \in [0,S], \quad \forall \omega \in \Omega/\bar{\Delta}, \tag{5.13}$$

where $\bar{\Delta} \stackrel{\text{def}}{=} \Delta \cup (\cup_i \Delta_i)$, with

$$P(\bar{\Delta}) \ = \ 0. \tag{5.14}$$

Note that, by (5.8) and (5.11), equation (5.10) is equivalent to

$$\dot{y}(\tau,\omega) \ = \ \bar{f}(\lambda(\tau,\omega), y(\tau,\omega)) \tag{5.15}$$

for $\omega \in \Omega/\bar{\Delta}$, where

$$\bar{f}(\lambda, y) \ \stackrel{\text{def}}{=} \int_U f(u,y)\lambda(du). \tag{5.16}$$

Consider the control system

$$\dot{y}(\tau) \ = \ \bar{f}(\lambda(\tau), y(\tau)), \tag{5.17}$$

where $\lambda(\tau)$ is a relaxed control, that is, a Lebesgue measurable function $\lambda(\tau) : [0,S] \to \mathcal{P}(U)$ (see [47]). A pair $(\lambda(\tau), y(\tau))$ will be called admissible (for system (5.17) on the interval $[0,S]$) if $\lambda(\tau)$ is a relaxed control, $y(\tau)$ is the corresponding solution of (5.17), and $y(\tau) \in Y \ \forall \tau \in [0,S]$. Let $h(u,y)$ be as in (4.1) and

$$\bar{h}(\lambda, y) \stackrel{\text{def}}{=} \int_U h(u,y)\lambda(du). \tag{5.18}$$

Consider the set of the time averages

$$\bar{V}_h(S,y) \stackrel{\text{def}}{=} \bigcup_{(\lambda(\cdot),\ y(\cdot))} \left\{ \frac{1}{S}\int_0^S \bar{h}\big(\lambda(\tau), y(\tau)\big)d\tau \right\}, \tag{5.19}$$

where the union is over all admissible pairs of (5.17) satisfying the initial conditions (2.4).

By the relaxation theorem (see, e.g., Theorem 10.4.4, p. 402 in [10]),

$$\bar{V}_h(S,y) \ \subset \ \mathrm{cl} V_h^{\delta}(S,y) \ \ \forall \delta \ > \ 0, \tag{5.20}$$

where cl stands for the closure of the set and $V_h^{\delta}(S,y)$ is defined in (4.5). Since $y(\tau,\omega)$ satisfies (5.13) and (5.15), from (5.20) and (4.6) it follows that, $\forall \omega \in \Omega/\bar{\Delta}$,

$$\begin{aligned} &\frac{1}{S}\int_0^S \bar{h}\big(\lambda(\tau,\omega), y(\tau,\omega)\big)d\tau \ \in \ \bar{V}_h(S, y(0,\omega)) \\ &\subset \ \mathrm{cl}V_h^{\delta}(S,Y) \ \subset \bar{co}V_h(S,Y) \ + \ \bar{\nu}_h(\delta,S)B_j, \end{aligned}$$

where $\bar{co}$ is the closed convex hull of the corresponding set and B_j is the closed unit ball in $\mathbb{R}^j$ (j is the dimension of the vector function $h(\cdot)$). Using now (5.14), one obtains from here that

$$\frac{1}{S}\int_0^S E[\bar{h}\big(\lambda(\tau,\omega), y(\tau,\omega)\big)]d\tau \ \in \ \bar{co}V_h(S,Y) \ + \ \bar{\nu}_h(\delta,S)B_j. \tag{5.21}$$

From (5.9), however, it follows that

$$E[\bar{h}\big(\lambda(\tau,\omega), y(\tau,\omega)\big)] \ = \ \int_{U\times Y} h(u,y)\gamma(du,dy). \tag{5.22}$$

Consequently, by (5.21),

$$\begin{aligned} &\int_{U\times Y} h(u,y)\gamma(du,dy) \ \in \ \bar{co}V_h(S,Y) \ + \ \bar{\nu}_h(\delta,S)B_j \\ \Rightarrow \quad &\int_{U\times Y} h(u,y)\gamma(du,dy) \ \in \ \bar{co}V_h(S,Y) \ + \ \bar{\nu}_h(S)B_j, \end{aligned}$$

where

$$\bar{\nu}_h(S) \stackrel{\text{def}}{=} \limsup_{\delta\to 0} \bar{\nu}_h(\delta,S).$$

Note that from the fact that $\bar{\nu}_h(\delta,S) \to 0$ as $\delta \to 0$ and $S \to \infty$ it follows that $\bar{\nu}_h(S) \to 0$ as $S \to \infty$. Since γ is an arbitrary element of W, one can conclude that

$$\bigcup_{\gamma\in W} \left\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \right\} \ \subset \ \bar{co}V_h(S,Y) \ + \ \bar{\nu}_h(S)B_j. \tag{5.23}$$

The set $V_h(S,Y)$ allows the representation

$$V_h(S,Y) \;=\; \bigcup_{\gamma\in\Gamma(S,Y)} \Big\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \Big\}.$$

Hence,

$$\bar{co} V_h(S,Y) = \bar{co} \bigcup_{\gamma\in\Gamma(S,Y)} \Big\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \Big\}$$

$$= \bigcup_{\gamma\in\bar{co}\Gamma(S,Y)} \Big\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \Big\}.$$

That is, (5.23) can be rewritten in the form

$$\bigcup_{\gamma\in W} \Big\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \Big\} \;\subset\; \bigcup_{\gamma\in\bar{co}\Gamma(S,Y)} \Big\{ \int_{U\times Y} h(u,y)\gamma(du,dy) \Big\} \;+\; \bar{\nu}_h(S)B_j.$$

Since it is true for any $h(u,y)$ as in (4.1), the validity of (5.1), with some $\bar{\nu}(S)$ tending to zero as S tends to infinity, follows from Lemma 3.5 in [25]. This completes the proof of Theorem 2.1(i). □

Proof of Theorem 2.1(ii). If the LOMS Γ exists, then, by (1.4),

$$\rho_H(co\Gamma(S,Y),\Gamma) \;=\; \rho_H(co\Gamma(S,Y),co\Gamma) \;\leq\; \rho_H(\Gamma(S,Y),\Gamma) \;\leq\; \nu(S),$$

where $\nu(S)$ is from (2.5). This and (2.7) imply equation (2.8) (because both Γ and W are compact). □

6. Proof of Lemma 5.1. The proof is divided into three steps. First, it is established that the random processes

$$J_i(\tau) \;\stackrel{\text{def}}{=}\; y_i(\tau) - \int_0^\tau \bar{F}_i(\lambda(s),y(s))ds\;, \qquad K_i(\tau) \;\stackrel{\text{def}}{=}\; y_i^2(\tau) - 2\int_0^\tau y_i(s)\bar{F}_i(\lambda(s),y(s))ds, \tag{6.1}$$

$i=1,\ldots,m$, are $\{\mathcal{F}_\tau\}$-martingales, where $y_i(\cdot), \bar{F}_i(\cdot)$, are the ith components of $y(\cdot)$ and $\bar{F}(\cdot)$, respectively. That is, for $\tau > \sigma \geq 0$,

$$E[\; J_i(\tau) \mid \mathcal{F}_\sigma \;] = J_i(\sigma), \qquad E[\; K_i(\tau)|\mathcal{F}_\sigma \;] = K_i(\sigma). \tag{6.2}$$

Second, it is shown that the processes $J_i^2(\tau)$, $i = 1,\ldots,m$, are $\{\mathcal{F}_\tau\}$-martingales. That is, for $\tau > \sigma \geq 0$,

$$E[\; J_i^2(\tau) \mid \mathcal{F}_\sigma \;] = J_i^2(\sigma), \quad i=1,\ldots,m. \tag{6.3}$$

Finally, the statement of the lemma is proved on the basis of (6.3).

Let us verify the validity of (6.2). Let $N > 0$ and $\psi_N(\theta) : [0,\infty) \to [0,\infty)$ be a continuously differentiable function such that

$$\psi_N(\theta) = 1 \quad \forall\theta\in[0,N^2]\;, \qquad \psi_N(\theta) = 0 \quad \forall\theta\in[N^2+1\;,\infty),$$

and $\psi_N(\theta)\in[0,1]$ $\forall\theta\in(N^2,N^2+1)$. Define $\phi_{i,N}(\cdot)\in\hat{C}^1$ by the equation $\phi_{i,N}(y) \stackrel{\text{def}}{=} y_i\psi_N(||y||^2)$, where $||y||$ is the Euclidean norm of y. Note that $|\phi_{i,N}(y)| \leq |y_i|$ and that $\phi_{i,N}(y) = y_i$ for $||y|| \leq N$.

By condition (b), the process

$$J_{i,N}(\tau) \stackrel{\text{def}}{=} \phi_{i,N}(y(\tau)) - \int_0^\tau \phi'_{i,N}(y(s))^T \bar{F}_i(\lambda(s), y(s))ds$$

is an $\{\mathcal{F}_\tau\}$-martingale. Hence,

$$E[\ J_{i,N}(\tau) \mid \mathcal{F}_\sigma\ ] = J_{i,N}(\sigma), \quad \tau > \sigma. \tag{6.4}$$

It can be seen, however, that for N large enough,

$$E[\,|J_{i,N}(\tau) - J_i(\tau)|\,] = E[\,|\phi_{i,N}(y(\tau)) - y_i(\tau)|\,] \leq 2E[\,|y_i(\tau)|\chi_{Q_N}(y(\tau))\,]$$
$$\leq 2\sqrt{E[\ |y_i(\tau)|^2]}\sqrt{E[\ \chi_{Q_N}(y(\tau))\ ]} \leq 2\sqrt{E[\ |y_i(\tau)|^2]}\sqrt{\gamma(U \times Q_N)} = 0,$$

where $\chi_{Q_N}(\cdot)$ is the indicator function of the set $Q_N \stackrel{\text{def}}{=} \{y \mid ||y|| > N\}$ and (5.8), (5.11) as well as (5.9) and the fact that $\gamma(U \times Y) = 1$ have been used. Thus, for sufficiently large N, $J_{i,N}(\tau) = J_i(\tau)$ a.s. and, hence, (6.4) implies the validity of the first equation in (6.2). The validity of the second equation in (6.2) is verified in a similar way (by using the test functions $\phi_{i,N}(y) \stackrel{\text{def}}{=} y_i^2\psi_N(||y||^2)$).

Let us now prove (6.3). Using the second equation in (6.2), one can write

$$E[J_i^2(\tau) \mid \mathcal{F}_\sigma] - J_i^2(\sigma) = E[J_i^2(\tau) - K_i(\tau) \mid \mathcal{F}_\sigma] - (J_i^2(\sigma) - K_i(\sigma))$$
$$= E\left[-2\int_0^\tau (y_i(\tau) - y_i(s))\bar{F}_i(\lambda(s), y(s))ds + \left(\int_0^\tau \bar{F}_i(\lambda(s), y(s))ds\right)^2 \mid \mathcal{F}_\sigma\right]$$
$$- \left(-2\int_0^\sigma (y_i(\sigma) - y_i(s))\bar{F}_i(\lambda(s), y(s))ds + \left(\int_0^\sigma \bar{F}_i(\lambda(s), y(s))ds\right)^2\right)$$
$$= E\left[-2(y_i(\tau) - y_i(\sigma))\int_0^\sigma \bar{F}_i(\lambda(s), y(s))ds - 2\int_\sigma^\tau (y_i(\tau) - y_i(s))\bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right]$$
$$+ E\left[2\int_\sigma^\tau \bar{F}_i(\lambda(s), y(s))ds \int_0^\sigma \bar{F}_i(\lambda(s), y(s))ds + \left(\int_\sigma^\tau \bar{F}_i(\lambda(s), y(s))ds\right)^2 \mid \mathcal{F}_\sigma\right].$$

Note that, by the first equation in (6.2),

$$E\left[(y_i(\tau) - y_i(\sigma))\int_0^\sigma \bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right]$$
$$- E\left[\int_\sigma^\tau \bar{F}_i(\lambda(s), y(s))ds \int_0^\sigma \bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right] = 0.$$

Hence, to complete the proof of (6.3), it is now sufficient to show that

$$E\left[\int_\sigma^\tau (y_i(\tau) - y_i(s))\bar{F}_i(\lambda(s), y(s))ds|\mathcal{F}_\sigma\right] = \frac{1}{2}E\left[\left(\int_\sigma^\tau \bar{F}_i(\lambda(s), y(s))ds\right)^2 |\mathcal{F}_\sigma\right]. \tag{6.5}$$

Using again the fact that $J_i(\cdot)$ is a martingale, one can obtain that

$$E\left[\int_\sigma^\tau (y_i(\tau) - y_i(s))\bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right]$$
$$= \int_\sigma^\tau E[\,(y_i(\tau) - y_i(s))\bar{F}_i(\lambda(s), y(s)) \mid \mathcal{F}_\sigma]ds$$
$$= \int_\sigma^\tau E[\,E[\,(y_i(\tau) - y_i(s)) \mid \mathcal{F}_s]\bar{F}_i(\lambda(s), y(s)) \mid \mathcal{F}_\sigma]ds$$
$$= \int_\sigma^\tau E\left[\left(\int_s^\tau \bar{F}_i(\lambda(s'), y(s'))ds'\right)\bar{F}_i(\lambda(s), y(s)) \mid \mathcal{F}_\sigma\right]ds$$
$$= E\left[\int_\sigma^\tau \left(\int_s^\tau \bar{F}_i(\lambda(s'), y(s'))ds'\right)\bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right].$$

Since

$$E\left[\int_\sigma^\tau \left(\int_s^\tau \bar{F}_i(\lambda(s'), y(s'))ds'\right)\bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right]$$
$$= E\left[\int_\sigma^\tau \left(\int_\sigma^s \bar{F}_i(\lambda(s'), y(s'))ds'\right)\bar{F}_i(\lambda(s), y(s))ds \mid \mathcal{F}_\sigma\right]$$

and the sum of the left- and the right-hand sides in the above equation is equal to $E[\,(\int_\sigma^\tau \bar{F}_i(\lambda(s), y(s))ds)^2 \mid \mathcal{F}_\sigma]$, it follows that (6.5) is valid and, thus, (6.3) is established.

From (6.3) and the first equation in (6.2), it follows that, for $\tau > \sigma \geq 0$,

$$E[\,(J_i(\tau) - J_i(\sigma))^2 \mid \mathcal{F}_\sigma] = 0 \quad \Rightarrow \quad E[\,(J_i(\tau) - J_i(\sigma))^2] = 0, \tag{6.6}$$

$i = 1, \ldots, m$. By Kolmogorov's continuity theorem (see, e.g., Theorem 1.10, p. 23 in [35]), there exists a continuous version of the $\mathbb{R}^m$-valued process $J(\cdot) \stackrel{\text{def}}{=} \{J_i(\cdot)\}$, $i = 1, \ldots, m$. For this version, (6.6) implies that

$$J(\tau) = J(0) \quad \forall \tau \in [0, S], \quad \forall S > 0.$$

In accordance with our notation (see (6.1)), the latter is equivalent to

$$y(\tau) = y(0) + \int_0^\tau \bar{F}(\lambda(s), y(s))ds,$$

which, in turn, is equivalent to (5.10). This completes the proof of the lemma. ∎

Acknowledgment. The author wants to express his gratitude to J.-B. Lasserre for useful discussions enhancing the progress in writing this paper and also to V. Borkar for referring the author to the result of R. H. Stockbridge (see [43]) as well as to some other results in the theory of control martingale problems, which were used in the proof of the main theorem.

REFERENCES

[1] E. Altman and V. Gaitsgory, *Asymptotic optimization of a nonlinear hybrid system governed by a Markov decision process*, SIAM J. Control Optim., 35 (1997), pp. 2070–2085.
[2] O. Alvarez and M. Bardi, *Viscosity solutions methods for singular perturbations in deterministic and stochastic control*, SIAM J. Control Optim., 40 (2001), pp. 1159–1188.

[3] Z. Artstein, *The chattering limit of singularly perturbed optimal control problems*, in Proceedings of CDC-2000, Control and Decision Conference, Sydney, 2000, pp. 564–569.
[4] Z. Artstein, *An occupational measure solution to a singularly perturbed optimal control problem*, Control Cybernet., 31 (2002), pp. 623–642.
[5] Z. Artstein and V. Gaitsgory, *Tracking fast trajectories along a slow dynamics: A singular perturbations approach*, SIAM J. Control Optim., 35 (1997), pp. 1487–1507.
[6] Z. Artstein and V. Gaitsgory, *The value function of singularly perturbed control systems*, Appl. Math. Optim., 41 (2000), pp. 425–445.
[7] Z. Artstein and V. Gaitsgory, *Convergence to convex compact sets in infinite dimensions*, J. Math. Anal. Appl., 284 (2003), pp. 471–480.
[8] Z. Artstein and A. Leizarowitz, *Singularly perturbed control systems with one-dimensional fast dynamics*, SIAM J. Control Optim., 41 (2002), pp. 641–658.
[9] J.-P. Aubin, *Viability Theory*, Birkhäuser Boston, Boston, 1991.
[10] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser Boston, Boston, 1990.
[11] F. Bagagiolo and M. Bardi, *Singular perturbation of a finite horizon problem with state-space constraints*, SIAM J. Control Optim., 36 (1998), pp. 2040–2060.
[12] A. Bensoussan, *Perturbation Methods in Optimal Control*, John Wiley, New York, 1989.
[13] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
[14] F. Colonius and R. Fabri, *Controllability for systems with slowly varying parameters*, ESAIM Control Optim. Calc. Var., 9 (2003), pp. 207–216.
[15] N. Danford and J.T. Schwartz, *Linear Operators*, Interscience, New York, 1958.
[16] A. Dontchev, *Time-scale decomposition of the reachable set of constrained linear systems*, Math. Control Signals Systems, 5 (1992), pp. 327–340.
[17] A. Dontchev, T. Donchev, and I. Slavov, *A Tichonov-type theorem for singularly perturbed differential inclusion*, Nonlinear Anal., 26 (1996), pp. 1547–1554.
[18] T. Donchev and I. Slavov, *Averaging method for one-sided Lipschitz differential inclusions with generalized solutions*, SIAM J. Control Optim., 37 (1999), pp. 1600–1613.
[19] J.A. Filar, V. Gaitsgory, and A.B. Haurie, *Control of singularly perturbed hybrid stochastic systems*, IEEE Trans. Automat. Control, 46 (2001), pp. 179–190.
[20] O.P. Filatov and M.M. Hapaev, *Averaging of Systems of Differential Inclusions*, Moscow University Publishing House, Moscow, 1998 (in Russian).
[21] V. Gaitsgory, *Use of the averaging method in control problems*, Differential Equations, 22 (1986), pp. 1290–1299 (translated from Russian).
[22] V. Gaitsgory, *Control of Systems with Fast and Slow Motions*, Nauka, Moscow, 1991 (in Russian).
[23] V. Gaitsgory, *Suboptimization of singularly perturbed control systems*, SIAM J. Control Optim., 30 (1992), pp. 1228–1249.
[24] V. Gaitsgory and A. Leizarowitz, *Limit occupational measures set for a control system and averaging of singularly perturbed control systems*, J. Math. Anal. Appl., 233 (1999), pp. 461–475.
[25] V. Gaitsgory and M.-T. Nguyen, *Multiscale singularly perturbed control systems: Limit occupational measures sets and averaging*, SIAM J. Control Optim., 41 (2002), pp. 954–974.
[26] G. Grammel, *Averaging of singularly perturbed systems*, Nonlinear Anal., 28 (1997), pp. 1851–1865.
[27] G. Grammel, *Periodic near optimal control*, J. Math. Anal. Appl., 248 (2000), pp. 124–144.
[28] Y. Kabanov and S. Pergamenshikov, *On convergence of attainability sets for controlled two-scale stochastic linear systems*, SIAM J. Control Optim., 35 (1997), pp. 134–159.
[29] Y. Kabanov and S. Pergamenshchikov, *Two-Scale Stochastic Systems*, Springer-Verlag, Berlin, 2003.
[30] P.V. Kokotovíc, *Applications of singular perturbation techniques to control problems*, SIAM Rev., 26 (1984), pp. 501–550.
[31] P.V. Kokotovíc, H.K. Khalil, and J. O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*, Academic Press, New York, 1986.
[32] H.J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser Boston, Boston, 1990.
[33] J.G. Llavona, *Approximation of Continuously Differentiable Functions*, North–Holland Math. Stud. 130, North–Holland, Amsterdam, 1986.
[34] A. Leizarowitz, *Order reduction is invalid for singularly perturbed control problems with a vector fast variable*, Math. Control Signals Systems, 15 (2002), pp. 101–119.
[35] R.S. Liptser and A.N. Shiryev, *Statistic of Random Processes* I. *General Theory*, Springer-Verlag, New York, 1977.

[36] S.D. Naidu, *Singular perturbations and time scales in control theory and applications: An overview*, Dynam. Contin. Discrete Impuls. Systems, Series B: Applications and Algorithms, 9 (2002), pp. 233–278.
[37] R.E. O'Malley, Jr., *Introduction to Singular Perturbations*, Academic Press, New York, 1974.
[38] R.E. O'Malley, Jr., *Singular perturbations and optimal control*, in Mathematical Control Theory, W.A. Copel, ed., Lecture Notes in Math. 680, Springer-Verlag, Berlin, 1978, pp. 170–218.
[39] M. Quincampoix and H. Zhang, *Singular perturbations in non-linear optimal control systems*, Differential Integral Equations, 8 (1995), pp. 931–944.
[40] A.A. Pervozvansky and V. Gaitsgory, *Theory of Suboptimal Decisions*, Kluwer Academic, Dordrecht, The Netherlands, 1988.
[41] V.A. Plotnikov, A.V. Plotnikov, and A.N. Vituk, *Differential Equations with Multivalued Right-Hand Sides: Asymptotic Methods*, AstroPrint, Odessa, Russia, 1999 (in Russian).
[42] S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser Boston, Boston, 1994.
[43] R.H. Stockbridge, *Time-average control of a martingale problem. Existence of a stationary solution*, Ann. Probab., 18 (1990), pp. 190–205.
[44] A.B. Vasil'eva and V.F. Butuzov, *Asymptotic Expansions of Solutions of Singularly Perturbed Equations*, Nauka, Moscow, 1973 (in Russian).
[45] V. Veliov, *A generalization of Tichonov theorem for singularly perturbed differential inclusions*, J. Dynam. Control Systems, 3 (1997), pp. 1–28.
[46] A. Vigodner, *Limits of singularly perturbed control problems with statistical limits of fast motions*, SIAM J. Control Optim., 35 (1997), pp. 1–28.
[47] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.
[48] G.G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications. A Singular Perturbation Approach*, Springer-Verlag, New York, 1997.