SIAM J. CONTROL OPTIM.
Vol. 44, No. 4, pp. 1436–1473

ON EXISTENCE OF LIMIT OCCUPATIONAL MEASURES SET OF A CONTROLLED STOCHASTIC DIFFERENTIAL EQUATION*

VIVEK BORKAR† AND VLADIMIR GAITSGORY‡

Abstract. We establish that, under certain conditions, the set of occupational measures as well as the set of mathematical expectations of occupational measures generated by the admissible controls and the corresponding solutions of a controlled stochastic differential equation (CSDE) converge (with the time horizon tending to infinity) to a set called limit occupational measures set (LOMS) and we show that this limit set coincides with the set of stationary marginal distributions of the CSDE. We also demonstrate the applicability of our results for averaging of singularly perturbed CSDE.

Key words. singularly perturbed controlled stochastic differential equations, occupational measures, averaging method, limit occupational measures sets, approximation of slow motions

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

DOI. 10.1137/S0363012904443476

1. Introduction.

In this paper we establish that, under certain conditions, the set of occupational measures as well as the set of mathematical expectations of occupational measures generated by the admissible controls and the corresponding solutions of a controlled stochastic differential equation (CSDE) converge (with the time horizon tending to infinity) to a set called limit occupational measures set (LOMS) and we show that this limit set coincides with the set of stationary marginal distributions of the CSDE.

The motivation for our study is the applicability of results to averaging of singularly perturbed CSDE. We show that, given a singularly perturbed CSDE, the slow components of its state variables are approximated by the solutions of the averaged system in which the controls take values in the LOMS of the system describing the fast dynamics. In the deterministic control setting, a similar approach was used in [4], [5], [6], [7], [8], [26], [27], [28] (see also [18], [19], [24], [25], [30], [46], [51] for related results). The current paper is based on a combination of ideas developed in the deterministic setting and also on results of [11], [13], and [49] which describe the set of stationary marginal distributions of the CSDE.

Note that singularly perturbed problems of control and optimization have been considered in both deterministic and stochastic literature (see [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [12], [17], [18], [19], [20], [21], [22], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [35], [36], [37], [38], [39], [41], [42], [43], [44], [45], [46], [48], [50], [51], [53] and references therein). Singularly perturbed CSDE, in particular, have been studied in [2], [3], [12], [32], [33], and [38], where earlier references can also be found. In [2], [3], and [12] the Hamilton–Jacobi–Bellman (HJB) equations corresponding to

*Received by the editors May 5, 2004; accepted for publication (in revised form) February 8, 2005; published electronically November 4, 2005.

http://www.siam.org/journals/sicon/44-4/44347.html

†School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Rd., Mumbai 400005, India (borkar@tifr.res.in). This project was partly supported by grant III.5(157)/99-ET from the Department of Science and Technology, Government of India.

‡School of Mathematics, University of South Australia, Mawson Lakes Campus, Mawson Lakes SA, 5095, Australia (v.gaitsgory@unisa.edu.au). This project was supported by the Australian Research Council Grant DP0346099.

singularly perturbed CSDE were analyzed. In [2], in particular, it was shown that the optimal value function of the problem of optimal control of singularly perturbed CSDE with a fairly general structure (the only structural constraint was the periodicity in fast variables) converges to a viscosity solution of the HJB equation in which the fast variables are averaged out. In [12], results concerning asymptotic behavior of singularly perturbed CSDE with nongenerate diffusion were obtained. In [32] and [33], singularly perturbed CSDE linear in fast variables were studied and the limit behavior of the attainability sets was described. In [38], weak convergence methods were used to establish a number of important results concerning mainly the case when the fast dynamics are not controlled.

The results obtained in this paper can be used for an approximation of the slow dynamics of singularly perturbed CSDE having a general structure (that is, in particular, nonlinear and nonperiodic in fast variables, and having a controlled fast dynamics). It allows one to treat stochastic nondegenerate and degenerate diffusion cases (as well as a purely deterministic case) in a similar manner and also to deal with the situation when the classical approach, based on equating the singular perturbation parameter to zero, may not lead to a correct approximation of the slow dynamics.

The paper is organized as follows: In section 2 we introduce some notations and define the LOMS of the CSDE as the limit towards which converges the set of mathematical expectations of occupational measures generated by the controls and solutions of the CSDE. In section 3 we identify necessary and sufficient conditions for the LOMS to exist (Theorems 3.2 and 3.3) and also sufficient conditions for every element of the LOMS to be asymptotically approximated (in mean) by an occupational measure obtained with some admissible control (Theorem 3.4). In section 4 we establish that if the LOMS exists, it coincides with the set of marginal stationary distributions of the CSDE (Theorem 4.1) and show that every occupational measure converges to this set in mean (Theorem 4.2). The proofs for sections 3 and 4 are contained in sections 6 and 7.

In section 5 we demonstrate the applicability of above mentioned results to averaging of singularly perturbed CSDE (Theorem 5.1). The proofs for section 5 are contained in section 8.

2. Preliminaries. For a compact set U and m dimensional Euclidean space R^m, $\mathcal{P}(U\times R^m)$ and $\mathcal{P}(U\times \bar{R}^m)$ will stand for the spaces of probability measures defined on the σ- algebras of Borel subsets of $U\times R^m$ and $U\times \bar{R}^m$, respectively, with $\bar{R}^m$ being the one point compactification of R^m (see, e.g., [23, p. 126]). Note that any probability measure μ on $U \times R^m$ may be identified with the unique probability measure on $U \times \bar{R}^m$ that restricts to μ on $U \times R^m$ and perforce assigns zero probability to its complement. Conversely, any probability measure μ on $U\times \bar{R}^m$, assigning probability one to $U \times R^m$, defines a unique probability measure on $U \times R^m$. Thus, $\mathcal{P}(U \times R^m)$ can be considered as a subset of $\mathcal{P}(U \times \bar{R}^m)$ consisting of the probability measures μ on $U \times \bar{R}^m$ with $\mu(U \times R^m) = 1$.

The set $\mathcal{P}(U \times \bar{R}^m)$ will be treated as a compact metric space with a metric $\rho(\cdot,\cdot)$ consistent with its weak convergence topology which is metrizable and compact. There are many ways of how $\rho(\cdot,\cdot)$ can be introduced. In this paper the following definition will be used (in most of the cases): for any $\mu',\mu'' \in \mathcal{P}(U \times \bar{R}^m)$,

$$\rho(\mu',\mu'') \stackrel{def}{=} \sum_{i=1}^{\infty} 2^{-i} \left| \int f_i(u,y)\mu'(du,dy) - \int f_i(u,y)\mu''(du,dy)\right|, \tag{2.1}$$

where $f_i(u,y), i = 1,2,\ldots,$ is the sequence of Lipschitz continuous functions which is dense in the unit ball of $C(U \times \bar{R}^m)$ (the space of continuous functions defined on $U \times \bar{R}^m$). Using the metric ρ, one can define the Hausdorff metric ρ_H on the set of subsets of $\mathcal{P}(\bar{R}^m \times U)$ as follows: $\forall \mathcal{M}_i \subset \mathcal{P}(U \times \bar{R}^m), i = 1, 2,$

$$\rho_H(\mathcal{M}_1, \mathcal{M}_2) \stackrel{def}{=} \max \left\{ \sup_{\mu \in \mathcal{M}_1} \rho(\mu, \mathcal{M}_2), \sup_{\mu \in \mathcal{M}_2} \rho(\mu, \mathcal{M}_1) \right\}, \tag{2.2}$$

where (here and in what follows)

$$\rho(\mu, \mathcal{M}_i) \stackrel{\text{def}}{=} \inf_{\mu' \in \mathcal{M}_i} \rho(\mu, \mu'). \tag{2.3}$$

Remark 1. Note that, if $\mathcal{M}_1$ and/or $\mathcal{M}_2$ are not closed, then from the fact that $\rho_H(\mathcal{M}_1, \mathcal{M}_2) = 0$ it does not follow that $\mathcal{M}_1 = \mathcal{M}_2$. That is, $\rho_H(\cdot, \cdot)$ is, in fact, a semimetric. By some abuse of terminology we still will refer to it as to a metric keeping in mind that its equality to zero is equivalent to the equality of the closures of the corresponding sets.

We will be dealing with a CSDE

$$dy(\tau) = a(u(\tau), y(\tau))d\tau + b(y(\tau))dW(\tau) \tag{2.4}$$

with the initial conditions

$$y(0) = y_0, \tag{2.5}$$

where:

- the functions $a(u, y) : U \times R^m \to R^m$ and $b(y) : R^m \to R^{m\times m}$ are continuous and satisfy Lipschitz conditions in y, with $a(u, y)$ satisfying it uniformly with respect to $u \in U$;
- U is a compact metric space;
- $W(\cdot)$ is an R^m-valued standard Brownian motion;
- y_0 is an R^m-valued random variable independent of $W(\cdot)$;
- *admissible controls* $u(\cdot)$ are $U-$valued random processes progressively measurable with respect to a right continuous and complete filtration $\{\mathcal{F}_\tau\} \subset \mathcal{F}$ of σ-fields (with $(\Omega, \mathcal{F}, \mathcal{P})$ being a given probability space) such that:
  - $\{y_0$ and $W(\theta); \theta \le \tau\}$ is measurable with respect to $\mathcal{F}_\tau$ for $\tau \ge 0$,
  - For $\tau' \ge \tau \ge 0,\ W(\tau') - W(\tau)$ is independent of $\mathcal{F}_\tau$.

Let $S > 0$, $u(\cdot)$ be an admissible control and $y(\cdot)$ be the corresponding solution of the CSDE (2.4) on the interval $[0, S]$. Define the occupational measure $\mu_S^{u(\cdot),y(\cdot)}$ generated by the pair $(u(\cdot), y(\cdot))$ on this interval by taking

$$\mu_S^{u(\cdot),y(\cdot)}(Q) \stackrel{def}{=} \frac{1}{S} meas\{\tau \ : \ (u(\tau), y(\tau)) \in Q\} \tag{2.6}$$

for any Borel subset Q of $U \times \bar{R}^m$, with *meas* standing for the Lebesgue measure on $[0, S]$. Note that $\mu_S^{u(\cdot),y(\cdot)}$ is uniquely defined by

$$\int f_i(u, y) \mu_S^{u(\cdot),y(\cdot)}(du, dy) = \frac{1}{S} \int_0^S f_i(u(\tau), y(\tau))d\tau, \quad i = 1, 2, \ldots, \tag{2.7}$$

where $f_i(\cdot)$ are as in (2.1). From (2.7) it follows that, for any fixed $\mu \in \mathcal{P}(U \times \bar{R}^m)$, the value of the metric $\rho(\mu, \mu_S^{u(\cdot),y(\cdot)})$ is a random variable, which allows one to easily verify that $\mu_S^{u(\cdot),y(\cdot)}$ is a $\mathcal{P}(U \times \bar{R}^m)$- valued random variable. Define also the mathematical expectation $E[\mu_S^{u(\cdot),y(\cdot)}]$ of $\mu_S^{u(\cdot),y(\cdot)}$ as the probability measure on $\bar{R}^m \times U$ such that

$$E[\mu_S^{u(\cdot),y(\cdot)}](Q) \stackrel{def}{=} \frac{1}{S} E[meas\{\tau \ : \ (y(\tau), u(\tau)) \in Q\}] \tag{2.8}$$

for any Borel subset Q of $U \times \bar{R}^m$. By (2.7) and (2.8),

$$\int f_i(u,y) E[\mu_S^{u(\cdot),y(\cdot)}](du,dy) = E\left[\int_0^S f_i(u,y)\mu_S^{u(\cdot),y(\cdot)}(du,dy)\right]$$

$$= E\left[\frac{1}{S}\int_0^S f_i(u(\tau),y(\tau))d\tau\right], \qquad i = 1,2,\dots, \tag{2.9}$$

with $E[\mu_S^{u(\cdot),y(\cdot)}]$ being uniquely defined by these equations.

Denote by $\mathcal{M}(S,y_0)$ and $E[\mathcal{M}(S,y_0)]$ the collections of the occupational measures and their mathematical expectations:

$$\mathcal{M}(S,y_0) \stackrel{\rm def}{=} \bigcup_{(u(\cdot),\ y(\cdot))} \{\mu_S^{u(\cdot),y(\cdot)}\}, \quad E[\mathcal{M}(S,y_0)] \stackrel{\rm def}{=} \bigcup_{(u(\cdot),\ y(\cdot))} \{E[\mu_S^{u(\cdot),y(\cdot)}]\}, \tag{2.10}$$

where the unions are over all admissible controls and corresponding solutions of (2.4) with the initial condition (2.5).

By analogy with the deterministic setting (see [26], [27], and [28]), we introduce the following definition.

DEFINITION. *A convex and compact set $\mathcal{M} \subset \mathcal{P}(U \times R^m)$ will be called LOMS of the CSDE* (2.4) *with respect to the initial conditions having probability distributions from a given class $\mathcal{C}$ if, for any initial conditions with the distribution from this class,*

$$\rho_H(E[\mathcal{M}(S,y_0)],\mathcal{M}) \leq \ \nu^{\mathcal{C}}(S), \qquad \lim_{S\to\infty} \nu^{\mathcal{C}}(S) = 0. \tag{2.11}$$

The following assumption about the solutions of (2.4) will be used throughout the paper.

Assumption 1. There exists $\alpha > 0$ such that any solution of (2.4) obtained with an admissible control satisfies the inequality

$$\sup_{\tau,u(\cdot)} E[||y(\tau)||^{\alpha}] \leq c_1(E[||y_0||^{\alpha}] + 1), \ c_1 \ = \text{const.} \tag{2.12}$$

As an example let us consider the case when the CSDE (2.4) is linear. That is,

$$a(u,y) \stackrel{\rm def}{=} A_1 y + A_2 u, \qquad b(y) \stackrel{\rm def}{=} A_3, \tag{2.13}$$

where U is a compact subset of R^s (for some natural s) and $A_i, i = 1,2,3$, are matrices of the corresponding dimensions. In this case the solution of (2.4) can be presented in the form

$$y(\tau) = e^{A_1\tau}y(0) + \int_0^{\tau} e^{A_1(\tau-\tau')}A_2 u(\tau')d\tau' + \int_0^{\tau} e^{A_1(\tau-\tau')}A_3 dW(\tau') \tag{2.14}$$

and it is easy to verify that Assumption 1 will be valid with $\alpha = 2$ if the eigenvalues of A_1 have negative real parts. Note that, for general nonlinear systems, sufficient conditions for Assumption 1 to be valid can be derived from the existence of the corresponding Liapunov functions (see, e.g., [11], [16], and [34] for classical results on the uncontrolled case).

3. Strong and weak h-approximation conditions. Let $h(u,y) : U \times \bar{R}^m \to R^j$ be defined by

$$h(u,y) \stackrel{\text{def}}{=} (f_1(u,y), f_2(u,y), \dots, f_j(u,y)), \tag{3.1}$$

where, as above, $f_i(\cdot)$ are as in the definition of the metric $\rho(\cdot,\cdot)$ (see (2.1)). In some instances (e.g., in the definitions below or in Lemma 3.1 and Theorems 3.2, 3.4(i)) we will consider j, and hence $h(\cdot)$, as being fixed. In other cases (e.g., in Theorems 3.3 and 3.4(ii)), the reference "*for every* $h(u,y)$ *as in* (3.1)" will be used in order to indicate that j can be any positive integer: $j = 1, 2, \dots$.

Definition. *We shall say that the CSDE* (2.4) *satisfies* strong h-approximation condition (S-h-AC) *if, for any initial condition* y_0' *and admissible control* $u'(\cdot)$, *corresponding to any other initial condition* y_0'' *there exists an admissible control* $u''(\cdot)$ *such that the solutions* $y'(\cdot)$ *and* $y''(\cdot)$ *of the CSDE* (2.4) *(obtained with* y_0', $u'(\cdot)$ *and* y_0'', $u''(\cdot)$, *respectively) satisfy the inequality*

$$E\left[\left\|\frac{1}{S}\int_0^S h(u'(\tau),y'(\tau))d\tau - \frac{1}{S}\int_0^S h(u''(\tau),y''(\tau))d\tau\right\|\right]$$

$$\le \nu_h(S)(1 + E[||y_0'||^\alpha] + E[||y_0''||^\alpha]) \tag{3.2}$$

for some monotone decreasing $\nu_h(\cdot) : [0,\infty) \to [0,\infty)$ *such that* $\lim_{S\to\infty} \nu_h(S) = 0$ *(*α *is the same as in Assumption* 1*).*

Definition. *We shall say that the CSDE* (2.4) *satisfies* weak h-approximation condition (W-h-AC) *if*

$$\left\|E\left[\frac{1}{S}\int_0^S h(u'(\tau),y'(\tau))d\tau\right] - E\left[\frac{1}{S}\int_0^S h(u''(\tau),y''(\tau))d\tau\right]\right\|$$

$$\le \nu_h(S)(1 + E[||y_0'||^\alpha] + E[||y_0''||^\alpha]), \tag{3.3}$$

where y_0', y_0'', $u'(\cdot), u''(\cdot)$, $y'(\cdot), y''(\cdot)$, $\nu_h(\cdot)$ *and* α *are as above.*

Note that, in the linear case (2.13), one can take $u''(\cdot) = u'(\cdot)$ and obtain (see (2.14)) that

$$E[||y'(\tau) - y''(\tau)||] = E[||e^{A_1\tau}(y_0' - y_0'')||] \le ||e^{A_1\tau}||(E[||y_0'||] + E[||y_0''||]). \tag{3.4}$$

Since $h(\cdot)$ satisfies Lipschitz conditions, the validity of S-h-AC will follow from (3.4) (with $\nu_h(S) = O(\frac{1}{S})$ and $\alpha \ge 1$) if the eigenvalues of A_1 have negative real parts, in which case $||e^{A_1\tau}|| \le \beta_1 e^{-\beta_2\tau}$, with β_1, β_2 being positive constants. Note that A_3 in (2.13) can be degenerate or, in fact, it can be zero (the deterministic case). Note also that a Liapunov-type stability condition which leads to the validity of a similar estimate (and, thus, leads to the fulfillment of S-h-AC with $u''(\cdot) = u'(\cdot)$) for a nonlinear CSDE can be found in [11].

Remark 2. Note that W-h-AC is an auxiliary condition which is introduced in order to simplify our consideration. It is obvious that it is implied by S-h-AC, but we were unable to construct an example in which W-h-AC is satisfied while S-h-AC is not. We leave the question of whether it is possible to construct such an example (or whether W-h-AC and S-h-AC are equivalent) open. Note that, in case of the uncontrolled dynamics (U consists only of one point; say, $U = \{\bar{u}\}$), S-h-AC is implied by W-h-AC and, hence, W-h-AC and S-h-AC are equivalent. In fact, as is noticed later (see Remark 4 on page 10), if W-h-AC is satisfied, then there exists a nonrandom vector $\tilde{h}$ such that

$$E\left[\left\|\frac{1}{S}\int_0^S h(\bar{u},y(\tau))d\tau - \tilde{h}\right\|\right] \leq \bar{\nu}^{(C,\alpha)}(S), \qquad \lim_{S\to\infty} \bar{\nu}^{(C,\alpha)}(S) = 0 \tag{3.5}$$

for any solution $y(\cdot)$ of (2.4) which has the initial condition satisfying the inequality $E[||y(0)||^\alpha] \leq C = \text{const}$. It follows that, for any two solutions $y'(\cdot)$ and $y''(\cdot)$ of (2.4) with the initial conditions satisfying a similar inequality,

$$E\left[\left\|\frac{1}{S}\int_0^S h(\bar{u},y'(\tau))d\tau - \frac{1}{S}\int_0^S h(\bar{u},y''(\tau))d\tau\right\|\right] \leq 2\bar{\nu}^{(C,\alpha)}(S).$$

Therefore, S-h-AC is satisfied.

For $h(u,y)$ as in (3.1), let $V_h(S,y_0)$ stand for the collection of random variables

$$V_h(S,y_0) \stackrel{\text{def}}{=} \bigcup_{(u(\cdot),\ y(\cdot))} \left\{\frac{1}{S}\int_0^S h(u(\tau),y(\tau))d\tau\right\} = \bigcup_{\mu\in\mathcal{M}(S,y_0)} \left\{\int h(u,y)\mu(du,dy)\right\} \tag{3.6}$$

and $E[V_h(S,y_0)]$ stand for the set of the corresponding mathematical expectations

$$\begin{aligned} E[V_h(S,y_0)] &\stackrel{\text{def}}{=} \bigcup_{(u(\cdot),\ y(\cdot))} \left\{E\left[\frac{1}{S}\int_0^S h(u(\tau),y(\tau))d\tau\right]\right\} \\ &= \bigcup_{\mu\in E[\mathcal{M}(S,y_0)]} \left\{\int h(u,y)\mu(du,dy)\right\}, \end{aligned} \tag{3.7}$$

where, in both (3.6) and (3.7), the first unions are over the admissible controls and corresponding solutions of (2.4) with the initial conditions (2.5).

Next, we introduce the Hausdorff metric $d_H^E(\cdot,\cdot)$ on collections of random variables as follows.

Definition. *Let V_1 and V_2 be two collections of integrable random variables defined on the same probability space and taking values in R^j. Then*

$$d_H^E(V_1,V_2) \stackrel{def}{=} \max\left\{\sup_{\zeta\in V_1} d^E(\zeta,V_2),\ \sup_{\zeta\in V_2} d^E(\zeta,V_1)\right\}, \tag{3.8}$$

with

$$d^E(\zeta,V_2) \stackrel{\text{def}}{=} \inf_{\zeta'\in V_2} E[||\zeta-\zeta'||]\ \ \forall\zeta\in V_1, \quad d^E(\zeta,V_1) \stackrel{\text{def}}{=} \inf_{\zeta'\in V_1} E[||\zeta-\zeta'||]\ \ \forall\zeta\in V_2, \tag{3.9}$$

where (here and in what follows) $||\cdot||$ is the Euclidean norm in R^j.

It is easy to see that d_H^E is nonnegative, symmetric, and satisfies the triangle inequality. For the constant valued collections of random variables, which can be viewed as just subsets of R^j, the definition above is reduced to the "standard" definition of the Hausdorff metric (semimetric) in R^j:

$$d_H(V_1,V_2) = \max\left\{\sup_{\zeta\in V_1} d(\zeta,V_2), \sup_{\zeta\in V_2} d(\zeta,V_1)\right\}, \qquad d(\zeta,V_i) = \inf_{\zeta'\in V_i} ||\zeta-\zeta'||, \ \ i=1,2. \tag{3.10}$$

Note that, as in the case with $\rho_H(\cdot,\cdot)$ (see Remark 1 on page 3), the equality $d_H(\cdot,\cdot) = 0$ is equivalent to the fact that the closures of the corresponding subsets of R^j are equal.

LEMMA 3.1. *S-h-AC is equivalent to the fulfillment of the inequality*

$$d_H^E(V_h(S,y_0'),V_h(S,y_0'')) \leq \nu_h(S)(1+E[||y_0'||^\alpha]+E[||y_0''||^\alpha]), \tag{3.11}$$

and W-h-AC is equivalent to the fulfillment of the inequality

$$d_H(E[V_h(S,y_0')],E[V_h(S,y_0'')]) \leq \nu_h(S)(1+E[||y_0'||^\alpha]+E[||y_0''||^\alpha]) \tag{3.12}$$

for any initial conditions y_0' and y_0''.

Proof. The proof is obvious. □

DEFINITION. *We shall say that the initial condition* (2.5) *has a probability distribution belonging to the class* (C,α) *if*

$$E[||y_0||^\alpha] \leq C = \text{const.} \tag{3.13}$$

THEOREM 3.2. *Let Assumption* 1 *be valid. If the CSDE* (2.4) *satisfies W-h-AC, then there exists a convex and compact set $V_h \subset R^j$ such that, for any initial condition y_0 with the probability distribution from the class (C,α),*

$$d_H(E[V_h(S,y_0)],V_h) \leq \nu_h^{C,\alpha}(S), \qquad \lim_{S\to\infty}\nu_h^{C,\alpha}(S) = 0. \tag{3.14}$$

Conversely, if there exists V_h such that (3.14) *is valid for any initial condition y_0 with the probability distribution from the class (C,α), then W-h-AC is satisfied for any initial conditions y_0', y_0'' with the probability distributions from this class.*

Proof. The fact that the validity of (3.14) implies W-h-AC is obvious since from (3.14) it follows that

$$d_H(E[V_h(S,y_0')],E[V_h(S,y_0'')]) \leq d_H(E[V_h(S,y_0')],V_h) + d_H(V_h,E[V_h(S,y_0'')])$$
$$\leq 2\nu_h^{C,\alpha}(S),$$

which, by (3.12), leads to the fulfillment of W-h-AC. The proof of the fact that W-h-AC implies the existence of a convex and compact set V_h which satisfies (3.14) for any initial condition y_0 with the probability distribution from the class (C,α) is given in section 6. □

THEOREM 3.3. *Let Assumption* 1 *be valid. If the CSDE* (2.4) *satisfies W-h-AC for any vector function $h(u,y)$ as in* (3.1). *Then the LOMS $\mathcal{M}$ of the CSDE* (2.4) *with respect to the initial conditions having the probability distribution from the class (C,α) exists. That is, for any initial condition y_0 with the probability distribution from this class, the estimate is valid:*

$$\rho_H(E[\mathcal{M}(S,y_0)],\mathcal{M}) \leq \nu^{C,\alpha}(S), \qquad \lim_{S\to\infty}\nu^{C,\alpha}(S) = 0. \tag{3.15}$$

Also, the LOMS $\mathcal{M}$ allows the representation

$$\mathcal{M} \stackrel{\text{def}}{=} \{\mu \in \mathcal{P}(U\times R^m) \mid \int h(u,y)\mu(du,dy) \in V_h \quad \forall h(u,y) \text{ as in } (3.1)\}, \tag{3.16}$$

where V_h are convex and compact sets the existence of which (for every $h(u,y)$ as in (3.1)) *is established by Theorem* 3.2.

Conversely, if there exists a convex and compact set $\mathcal{M} \subset \mathcal{P}(U \times R^m)$ which satisfies (3.15) *with any initial condition y_0 having the probability distribution from the class (C, α), then W-h-AC is satisfied for any vector function $h(\cdot)$ as in* (3.1) *and any initial conditions y_0', y_0'' with the probability distributions from this class. Also, for any $h(\cdot)$ as in* (3.1), *the estimate* (3.14) *is valid with*

$$V_h = \bigcup_{\mu \in \mathcal{M}} \left\{ \int h(u, y)\mu(du, dy) \right\}. \tag{3.17}$$

Proof of Theorem 3.3 is in the end of section 6.

THEOREM 3.4. *Let Assumption* 1 *be valid.* (i) *If the CSDE* (2.4) *satisfies S-h-AC, then, for any initial condition y_0 with the probability distribution from the class (C, α),*

$$\sup_{\zeta \in V_h} d^E(\zeta, V_h(S, y_0)) \leq \tilde{\nu}_h^{(C,\alpha)}(S), \qquad \lim_{S \to \infty} \tilde{\nu}_h^{(C,\alpha)}(S) = 0, \tag{3.18}$$

where V_h is as in (3.14) *and $d^E(\cdot, \cdot)$ is defined by* (3.9).

(ii) *If the CSDE* (2.4) *satisfies S-h-AC for any vector function $h(\cdot)$ as in* (3.1), *then, for any initial condition y_0 with the probability distribution from the class (C, α),*

$$\sup_{\mu \in \mathcal{M}} \rho^E(\mu, \mathcal{M}(S, y_0)) \leq \tilde{\nu}^{(C,\alpha)}(S), \qquad \lim_{S \to \infty} \tilde{\nu}^{(C,\alpha)}(S) = 0, \tag{3.19}$$

where

$$\rho^E(\mu, \mathcal{M}(S, y_0)) \stackrel{\text{def}}{=} \inf_{\mu' \in \mathcal{M}(S, y_0)} E[\rho(\mu, \mu')]. \tag{3.20}$$

Proof of Theorem 3.4 is in section 6.

In conclusion of this section let us consider the following simple result which is used in the proof of Theorem 3.4.

PROPOSITION 3.5. *Let V_i, $i = 1, \dots, k$ be collections of random variables defined on the same probability space such that any element $\zeta_i \in V_i$ is independent from any element $\zeta_j \in V_j$ for $i \neq j$. Assume also that*

$$E[||\zeta_i||^2] \leq \bar{c} = \text{const} \qquad \forall \zeta_i \in V_i, \qquad i = 1, 2 \dots k.$$

Then

$$d_H^E \left(\frac{1}{k} \sum_1^k V_i, \ \frac{1}{k} \sum_1^k E[V_i] \right) \leq \sqrt{\frac{\bar{c}}{k}}, \tag{3.21}$$

where $E[V_i]$ stands for the set of mathematical expectations of the elements of V_i and $d_H^E(\cdot, \cdot)$ is defined in (3.8).

Proof. Take an arbitrary element $\zeta \in \frac{1}{k} \sum_1^k V_i$. By definition it is presented in the form $\zeta = \frac{1}{k} \sum_1^k \zeta_i$, where $\zeta_i \in V_i$. Consider

$$\bar{\zeta} \stackrel{\text{def}}{=} E[\zeta] = \frac{1}{k} \sum_1^k E[\zeta_i] \in \frac{1}{k} \sum_1^k E[V_i].$$

Due to the independence of ζ_i, $i = 1, \ldots, k$,

$$E[||\zeta - \bar{\zeta}||] \leq \sqrt{E[||\zeta - \bar{\zeta}||^2]} = \sqrt{\frac{1}{k^2}\sum_1^k E[||\zeta_i - E[\zeta_i]||^2]} \leq \sqrt{\frac{\bar{c}}{k}}. \tag{3.22}$$

Now take an arbitrary $\bar{\zeta} \in \frac{1}{k}\sum_1^k E[V_i]$. By definition, there exist $\zeta_i \in V_i$ such that $\bar{\zeta} = \frac{1}{k}\sum_1^k E[\zeta_i]$. Define $\zeta \stackrel{\text{def}}{=} \frac{1}{k}\sum_1^k \zeta_i \in \frac{1}{k}\sum_1^k V_i$. Similarly to (3.22), one can establish that $E[||\zeta - \bar{\zeta}||] \leq \sqrt{\frac{\bar{c}}{k}}$. This completes the proof of the proposition. ◻

4. Representation of the limit occupational measures set. Let $C_0^2(R^m)$ be the space of twice continuously differentiable functions $f(y) : R^m \to R^1$ which vanish at infinity along with their first and second derivatives and let $\mathcal{D}$ be a countable dense set in $C_0^2(R^m)$. Let $L : C_0^2(R^m) \to C_b(U \times R^m)$ be the operator defined as follows:

$$(Lf)(y,u) \;=\; \frac{1}{2}tr(b(y)b^T(y)\nabla^2 f(y)) + \langle \nabla f(y), a(u,y)\rangle \quad \forall f \in C_0^2(R^m). \tag{4.1}$$

Define the set of probability measures $D \subset \mathcal{P}(R^s \times A)$ by

$$D \;=\; \{\mu \in \mathcal{P}(U \times R^m) : \int (Lf)(u,y)\mu(du,dy) = 0 \quad \forall f \in \mathcal{D}\}. \tag{4.2}$$

and introduce the following assumption.

Assumption 2. For some $\alpha > 0$,

$$\int ||y||^\alpha \mu(du,dy) \leq c_2 = \text{const} \qquad \forall \mu \in D. \tag{4.3}$$

Note that the set D is convex and it is easy to verify that it is compact if Assumption 2 is satisfied. In fact, from this assumption it follows that D is tight and, hence, by Prohorov's theorem (see, e.g., Theorem 2.3.1, p. 25 in [15]), it is relatively compact in $\mathcal{P}(U \times R^m)$. Also, D is closed. This implied the compactness.

In [13] and [49] it was shown that, under some mild conditions, the set D represents the set of marginal distributions of stationary relaxed solutions of (2.4). In the following theorem it is established that, if W-h-AC is satisfied for any $h(u,y)$ as in (3.1)), then the LOMS of the CSDE (2.4) exists (the existence being implied by Theorem 3.3) and coincides with D.

THEOREM 4.1. *Let Assumptions* 1 *and* 2 *be satisfied with* $\alpha \geq 2$. *Then,*

(i) *The estimate*

$$\rho_H\left(\bigcup_{\{y_0\}\in(C,\alpha)} \{E[\mathcal{M}(S,y_0)]\},\ D\right) \leq \bar{\nu}^{(C,\alpha)}(S), \qquad \lim_{S\to\infty} \bar{\nu}^{(C,\alpha)}(S) = 0, \tag{4.4}$$

is valid, where the union is over all initial conditions with the probability distribution from the class (C,α).

(ii) *If W-h-AC is satisfied for any* $h(u,y)$ *as in* (3.1), *then the LOMS* $\mathcal{M}$ *of the CSDE* (2.4) *with respect to the initial conditions having the probability distribution from the class* (C,α) *exists and is equal to* D*:*

$$\mathcal{M} = D. \tag{4.5}$$

Proof. The statement (i) of the theorem is proved in section 7 on the basis of Theorem 4.2 stated below. The validity of (ii) is proved on the basis of (i) as follows. By (3.15),

$$\rho_H\left(\bigcup_{\{y_0\}\in(C,\alpha)}\{E[\mathcal{M}(S,y_0)]\},\ \mathcal{M}\right)\le \nu^{C,\alpha}(S).$$

If now one assumes that (4.4) is valid, it will follow that

$$\rho_H(\mathcal{M},D)\le \nu^{C,\alpha}(S)+\bar{\nu}^{(C,\alpha)}(S) \quad\Rightarrow\quad \rho_H(\mathcal{M},D)=0.$$

The latter implies (4.5) since both $\mathcal{M}$ and D are compact. □

THEOREM 4.2. *Let Assumption* 1 *be satisfied with* $\alpha\ge 2$. *Then, for any initial condition* y_0 *with the probability distribution from the class* (C,α),

$$\sup_{\mu\in\mathcal{M}(S,y_0)} E[\rho(\mu,D)]\le\bar{\nu}^{(C,\alpha)}(S),\qquad \lim_{S\to\infty}\bar{\nu}^{(C,\alpha)}(S)=0, \tag{4.6}$$

where $\rho(\mu,D)$ *is defined as in* (2.3).

Proof of Theorem 4.2 is in section 7.

COROLLARY 4.3. *Let Assumptions* 1 *and* 2 *be valid with* $\alpha\ge 2$. *If the CSDE* (2.4) *satisfies W-h-AC for any vector function* $h(\cdot)$ *as in* (3.1), *then, for any initial condition* y_0 *with the probability distribution from the class* (C,α),

$$\rho_H(E[\mathcal{M}(S,y_0)],D)\le\ \nu^{C,\alpha}(S),\qquad \lim_{S\to\infty}\nu^{(C,\alpha)}(S)=0. \tag{4.7}$$

If the CSDE (2.4) *satisfies S-h-AC for any vector function* $h(\cdot)$ *as in* (3.1), *then, for any initial condition* y_0 *with the probability distribution from the class* (C,α),

$$\sup_{\mu\in D}\rho^E(\mu,\mathcal{M}(S,y_0))\le\tilde{\nu}^{(C,\alpha)}(S),\qquad \lim_{S\to\infty}\tilde{\nu}^{(C,\alpha)}(S)=0. \tag{4.8}$$

Proof. The proof follows immediately from Theorems 3.3, 3.4, and 4.1(ii). □

Remark 3. The estimate (4.8) is, in fact, equivalent to the validity of S-h-AC for any vector function $h(\cdot)$. Let us show that if (4.8) is satisfied, then, for any $y_0', u'(\cdot)$ and any y_0'', there exists $u''(\cdot)$ such that (3.2) is valid (with y_0', y_0'' being assumed to be from the class (C,α)). Let $\mu_S^{u'(\cdot)y'(\cdot)}$ be the occupational measure generated by the pair $(u'(\cdot),y'(\cdot))$ on the interval $[0,S]$. By (4.6), there exists $\mu_S'\in D$ such that $E[\rho(\mu_S^{u'(\cdot)y'(\cdot)},\mu_S')]\le\bar{\nu}^{(C,\alpha)}(S)$. From (4.8) it follows, in turn, that the estimate $E[\rho(\mu_S',\mu_S'')]\le\tilde{\nu}^{(C,\alpha)}(S)$ is valid for some $\mu_S''\in\mathcal{M}(S,y_0'')$. By definition, μ_S'' is an occupational measure generated on the interval $[0,S]$ by some admissible control $u''(\cdot)$ and the corresponding solution $y''(\cdot)$ of the CSDE (2.4) which satisfies the initial condition $y''(0)=y_0''$. That is, $\mu_S''=\mu_S^{u''(\cdot)y''(\cdot)}$ and $E[\rho(\mu_S^{u'(\cdot)y'(\cdot)},\mu_S^{u''(\cdot)y''(\cdot)})]\le\bar{\nu}^{(C,\alpha)}(S)+\tilde{\nu}^{(C,\alpha)}(S)$. The latter estimate implies the validity of S-h-AC for any $h(\cdot)$ as in (3.1).

Remark 4. Under Assumptions 1 and 2, one can show that, corresponding to any extreme point μ of D, there exists an admissible control $u_\mu(\cdot)$ and the corresponding solution $y_\mu(\cdot)$ of the CSDE (2.4) such that, for any $h(\cdot)$ as in (3.1), there almost surely exists the limit

$$\lim_{S\to\infty}\frac{1}{S}\int_0^S h(u_\mu(\tau),y_\mu(\tau))d\tau=\int h(u,y)\mu(du,dy). \tag{4.9}$$

We do not give the proof of this statement in the paper (it is based on results of [13], [49], and the ergodic theory). Let us note only that, for the uncontrolled case mentioned in Remark 2 on page 5, it follows that, if W-h-AC is satisfied, then the value of the integral on the right-hand side of (4.9) is the same for any extreme points μ of D and, hence, it is the same for all elements of D. Denoting this value as $\tilde{h}$ and using (4.6), one can easily verify the validity of (3.5).

Let $g(u, y) : U \times R^m \to R^n$ be continuous and satisfy Lipschitz conditions in y. Define the collection of R^n-valued random variables $V_g(S, y_0)$ similarly to (3.6) with the replacement of $h(\cdot)$ by $g(\cdot)$. That is,

$$V_g(S, y_0) \stackrel{\text{def}}{=} \bigcup_{(u(\cdot),y(\cdot))} \left\{ \frac{1}{S} \int_0^S g(u(\tau), y(\tau))d\tau \right\} = \bigcup_{\mu \in \mathcal{M}(S,y_0)} \left\{ \int g(u, y)\mu(du, dy) \right\}, \tag{4.10}$$

where, as in (3.6), the first union is over all admissible controls and corresponding solutions of the CSDE (2.4). Define also the set $V_g \subset R^n$ by

$$V_g \stackrel{\text{def}}{=} \bigcup_{\mu \in D} \left\{ \int g(u, y)\mu(du, dy) \right\}, \tag{4.11}$$

where the union is over elements of D. Note that from the convexity of D it follows that V_g is convex. Also, the following two corollaries used in averaging of singularly perturbed CSDE (see section 5 below) are valid.

Corollary 4.4. *Let Assumptions* 1 and 2 *be satisfied with* $\alpha \geq 2$. *Then the set* V_g *is compact and there exists a function* $\bar{\nu}_g^{(C,\alpha)}(S)$, *tending to zero as* S *tends to infinity, such that for any initial condition* y_0 *with the distribution from the class* (C, α),

$$\sup_{v \in V_g(S,y_0)} E[d(v, V_g)] \leq \bar{\nu}_g^{(C,\alpha)}(S). \tag{4.12}$$

Corollary 4.5. *Let Assumptions* 1 and 2 *be valid with* $\alpha \geq 2$ *and let the CSDE* (2.4) *satisfies S-h-AC for any vector function* $h(u, y)$ *as in* (3.1). *Then there exists a function* $\tilde{\nu}_g^{(C,\alpha)}(S)$, *tending to zero as* S *tends to infinity, such that for any initial condition* y_0 *with the probability distribution from the class* (C, α),

$$\sup_{v \in V_g} d^E(v, V_g(S, y_0)) \leq \tilde{\nu}_g^{(C,\alpha)}(S). \tag{4.13}$$

Proofs of Corollaries 4.4 *and* 4.5. The proofs follow from Theorems 4.1(ii), (4.2), and Theorems 4.1(ii), (3.4), respectively, if $g(\cdot) = h(\cdot)$ (with $h(\cdot)$ being as in (3.1)). In the general case, the corollaries are proved in section 7. □

5. Application in averaging of singularly perturbed controlled stochastic differential equations.

Consider the following singularly perturbed CSDE:

$$dy^\epsilon(t) = \frac{1}{\epsilon} a(u(t), y^\epsilon(t))dt + \frac{1}{\sqrt{\epsilon}} b(y^\epsilon(t))dB_1(t). \tag{5.1}$$

$$dz^\epsilon(t) = g(u(t), y^\epsilon(t), z^\epsilon(t))dt + \sigma(z^\epsilon(t))dB_2(t), \tag{5.2}$$

with the initial conditions

$$y^\epsilon(0) = y_0, \qquad z^\epsilon(0) = z_0, \tag{5.3}$$

where:

- ϵ is a small positive parameter;
- the functions $b(\cdot) : R^m \to R^{m\times m}$ and $\sigma(\cdot) : R^n \to R^{n\times n}$ satisfy Lipschitz conditions; the functions $a(u,y) : U \times R^m \to R^m$ and $g(u,y,z) : U \times R^m \times R^n \to R^n$ are continuous and satisfy Lipschitz conditions, respectively, in y and (y,z) uniformly with respect to $u \in U$;
- U is a compact metric space;
- $B_1(\cdot)$ and $B_2(\cdot)$ are R^m- and R^n-valued independent standard Brownian motions;
- y_0 and z_0 are R^m- and R^n-valued random variables which have bounded fourth moments and are independent of $B_1(\cdot), B_2(\cdot)$;
- *admissible controls* $u(\cdot)$ are U-valued random processes progressively measurable with respect to a right continuous and complete filtration $\{\hat{\mathcal{F}}_t\}$ of σ-fields such that
  - $\{y_0, z_0 \ \& \ B_1(\theta), B_2(\theta); \theta \le t\}$ is measurable with respect to $\hat{\mathcal{F}}_t$ for $t \ge 0$,
  - For $t' \ge t \ge 0$, $B_1(t') - B_1(t)$ and $B_2(t') - B_2(t)$ are independent of $\hat{\mathcal{F}}_t$.

Let us set $\tau = \frac{t}{\epsilon}, y(\tau) = y^\epsilon(\epsilon\tau), u'(\tau) = u^\epsilon(\epsilon\tau), W(\tau) = \frac{1}{\sqrt{\epsilon}}B_1(\epsilon\tau)$ and $\{\mathcal{F}_\tau\} = \{\hat{\mathcal{F}}_{\epsilon\tau}\}$. The subsystem (5.1) takes then the form of the CSDE (2.4) and is called *the associated system*. Assuming that the associated system satisfies Assumptions 1 and 2, let us define the averaged CSDE by

$$dz(t) = \tilde{g}(\mu(t), z(t))dt + \sigma(z(t))dB_2(t), \tag{5.4}$$

where:

- $\tilde{g}(\mu, z) \stackrel{\text{def}}{=} \int_{R^m\times U} g(u,y,z)\mu(du,dy) : \mathcal{P}(R^m \times U) \times R^n \to R^n$;
- the Brownian motion $B_2(\cdot)$ and the initial condition $z(0) = z_0$ are the same as in (5.2);
- *admissible controls* $\mu(\cdot)$ are $\{\hat{\mathcal{F}}_t\}$-progressive D-valued random processes (D being defined in (4.2)).

Let $G(\cdot) : R^n \to R$ be Lipschitz continuous and $T > 0$. Consider the problem of optimal control

$$\inf_{u(\cdot),y^\epsilon(\cdot),z^\epsilon(\cdot)} E[G(z^\epsilon(T))] \stackrel{\text{def}}{=} G^*_\epsilon, \tag{5.5}$$

where *inf* is over the admissible controls and the corresponding solutions of the singularly perturbed CSDE (5.1) and (5.2). Consider also the problem

$$\inf_{\mu(\cdot),z(\cdot)} E[G(z(T))] \stackrel{\text{def}}{=} G^*_{av}, \tag{5.6}$$

where *inf* is over the admissible controls and the corresponding solutions of the averaged CSDE (5.4). Note that, if $\sigma(\cdot) \equiv 0$, then the averaged system (5.4) is purely deterministic and the minimization in (5.6) can be restricted to open loop controls (a similar phenomenon was dealt with in [1], where the fast dynamics were defined by a Markov decision process).

THEOREM 5.1. (i) *Let the associated system satisfy Assumptions* 1 *and* 2 *with* $\alpha = 4$. *Then, corresponding to any admissible solution* $(y^\epsilon(\cdot), z^\epsilon(\cdot))$ *of the singularly perturbed CSDE* (5.1) and (5.2), *there exists an admissible solution* $z(\cdot)$ *of the averaged CSDE* (5.4) *such that*

$$\max_{t\in[0,T]} E[||z^\epsilon(t) - z(t)||^2] \le \tilde{\nu}(\epsilon), \qquad \lim_{\epsilon\to 0} \tilde{\nu}(\epsilon) = 0 \tag{5.7}$$

and the following inequality is valid:

$$\liminf_{\epsilon\to 0} G^*_\epsilon \geq G^*_{av}. \tag{5.8}$$

(ii) *If, in addition, the associated system satisfies S-h-AC for any vector function* $h(\cdot)$ *as in* (3.1), *then, corresponding to an arbitrary admissible solution* $z(\cdot)$ *of the averaged CSDE* (5.4), *there exists an admissible solution* $(y^\epsilon(\cdot), z^\epsilon(\cdot))$ *of the singularly perturbed CSDE* (5.1) and (5.2) *such that* (5.7) *is valid and*

$$\lim_{\epsilon\to 0} G^*_\epsilon \; = \; G^*_{av}. \tag{5.9}$$

Proof. The proof's details are outlined in section 8. □

Remark 5. The approximation of the z-components of the state variables of the CSDE (5.1) and (5.2) by the solutions of the averaged system (5.4) stated in Theorem 5.1 has many similarities with the classical relaxed control setting (see [52]). In contrast to the latter, however, the approximation established in Theorem 5.1 is asymptotic (that is valid when the small parameter tends to zero) and also the controls used in (5.4) take values in the LOMS (and not in the space of all probability measures defined on the control set).

Remark 6. Note that the conditions of Theorem 5.1 can be relaxed. Namely, the theorem remains valid if Assumptions 1 and 2 are satisfied with $\alpha > 2$ and also if they are satisfied with $\alpha = 2$ (to prove the result in the latter case, one needs to impose some additional conditions; in particular, one needs to assume that there exists an integrable random variable η such that the solution of the associated system satisfy the inequality $||y(\tau)||^2 \leq \eta \; \forall \tau \geq 0$). Note also that a statement similar to Theorem 5.1 is valid for singularly perturbed CSDE in which the fast subsystem may depend on the slow state variables. The proof of such a statement is in many ways similar to one outlined in section 8 but it is more technically involved and we do not include it in the paper.

Let us consider a special case when $b(y) = 0$. That is, the associated system is deterministic and it can be written in the form

$$\frac{dy(\tau)}{d\tau} = a(u(\tau), y(\tau)). \tag{5.10}$$

Assume that there exist positive definite matrices F_1 and F_2 such that, for any y', y'' and any $u \in U$,

$$(a(u,y') - a(u,y''))^T F_1(y' - y'') \leq -(y' - y'')^T F_2(y' - y''), \tag{5.11}$$

Note that (5.11) is satisfied if $a(u,y) = A_1 y + A_2 u$ (as in (2.13)), with the eigenvalues of A_1 having negative real parts. Taking $y^T F_1 y$ as a Liapunov function, one can easily verify (see, e.g., [27]) that solutions $y'(\tau)$ and $y''(\tau)$ of (5.10) obtained with the same control and with initial conditions $y'(0) = y'_0, \; y''(0) = y''_0$, satisfy the inequality

$$||y'(\tau) - y''(\tau)|| \leq \beta_1 e^{-\beta_2 \tau} ||y'_0 - y''_0|| \quad \forall \tau \geq 0, \tag{5.12}$$

where β_1 and β_2 are some positive constants. This implies the validity of S-h-AC. From (5.12) it follows (see Theorem 3.1(ii) in [25]) that there exists a compact set $Y \subset R^m$ such that any solution $y(\cdot)$ of (5.10) satisfies the inequality

$$\min_{y\in Y} ||y(\tau) - y|| \leq \beta_1 e^{-\beta_2 \tau} \min_{y\in Y} ||y(0) - y|| \qquad \forall \tau \geq 0, \tag{5.13}$$

where β_1, β_2 are as in (5.12) (that is, Y is forward invariant with respect to the solutions of (5.10) and is a global attractor for these solutions). Using the inequality (5.13), it is straightforward to verify that Assumption 1 is satisfied with an arbitrary positive α. Also, using this inequality, one can establish that $\mu(U \times Y) = 1 \ \forall \mu \in \mathcal{M}$, where, as above, $\mathcal{M}$ is the limit occupational measures set. Hence, this set allows the representation (see (4.2) and Theorem 4.1(ii))

$$\mathcal{M} = D = \{\mu \in \mathcal{P}(U \times Y) : \int_{U \times Y} \langle \nabla f(y), a(u,y) \rangle \mu(du, dy) = 0 \ \forall \ f \in \mathcal{D}\}, \tag{5.14}$$

where $\mathcal{P}(U \times Y)$ is the space of probability measures defined on Borel subsets of $U \times Y$. Note that the representation of the LOMS in the form (5.14) is consistent with one obtained for the deterministic case in [26] and that Assumption 2 is satisfied automatically in this case.

As an example, let us consider the singularly perturbed CSDE (5.1) and (5.2), in which: $y = (y_1, y_2)$ (that is, $y \in R^2$); U is a square in R^2: $U = \{(u_1, u_2) : \ |u_i| \leq 1, \ i = 1, 2\}$;

$$b(y) = 0, \qquad a(u,y) = (-y_1 + u_1, \ -y_2 + u_2); \tag{5.15}$$

and the slow dynamics are one dimensional ($z \in R^1$) with $z_0 = 0$ (zero initial condition) and with

$$\sigma(z) = \sigma = \text{const}, \qquad g(u,y,z) = g(u,y) \stackrel{\text{def}}{=} y_2 u_1 - y_1 u_2. \tag{5.16}$$

Consider the optimal control problem (5.5) with $G(z) = z$. Using (5.15) and (5.16), it is easy to verify that, if the fast subsystem (5.1) is multiplied by ϵ and, then ϵ is formally equated to zero, the resulting slow dynamics become uncontrolled and the value of the objective function is equal to zero. The limit of the optimal value of (5.5) is, however, strictly less than zero: $\lim_{\epsilon \to 0} G^*_\epsilon < 0$. This is evidenced by the fact that, if the rapidly oscillating controls $u_1(t) = cos(\frac{t}{\epsilon})$, $u_2(t) = sin(\frac{t}{\epsilon})$ are used, then the value of the objective function can be verified to be equal to $-0.5T + O(\epsilon) < 0$. Thus, the classical approach based on the equating of the singular perturbation parameter to zero is not applicable in the given example. The averaged problem is equivalent in this case to the infinite dimensional linear program

$$\min_{\mu \in D} \int_{U \times Y} g(u,y) \mu(du, dy) \stackrel{\text{def}}{=} g^*, \tag{5.17}$$

with $G^*_{av} = g^* T$, where D is as in (5.14) and $g(u,y)$ is defined in (5.16). Note that the solution of the problem (5.17) has been found numerically by approximating the problem with finite dimensional linear programs (using the approach proposed in [29]). The optimal value g^* of (5.17), in particular, was found to be approximately equal to -0.7679. One may conclude, therefore (by (5.9)), that $\lim_{\epsilon \to 0} G^*_\epsilon \approx -0.7679T$.

In some cases the averaged system can be equivalent to the system obtained via equating of the singular parameter to zero. To illustrate that, let us assume that the associated system is linear (that is, (2.13) is true, with eigenvalues of A_1 having negative real parts). Let us assume also that the slow subsystem (5.2) is linear in y and u. That is,

$$g(u,y,z) \stackrel{\text{def}}{=} A_4(z) y + A_5(z) u, \tag{5.18}$$

with $A_4(z), A_5(z)$ being matrices functions of the corresponding dimensions, then the averaged system becomes equivalent to

$$dz(t) = A_4(z(t))\bar{y}(t) + A_5(z(t))\bar{u}(t) + \sigma(z(t))dB_2(t), \tag{5.19}$$

where

$$(\bar{u}(t), \bar{y}(t)) \in \Omega \stackrel{\text{def}}{=} \left\{(\bar{u}, \bar{y}) \mid (\bar{u}, \bar{y}) = \int (u, y)\mu(du, dy),\ \mu \in D\right\}. \tag{5.20}$$

Under the assumption that U is convex, it can be shown (although, we do not do it in this paper) that the set Ω defined above can be represented in the form

$$\Omega = \{(\bar{u}, \bar{y}) \mid \bar{y} = -A_1^{-1}A_2\bar{u},\ \bar{u} \in U\}. \tag{5.21}$$

and that (5.19) is equivalent to the system

$$dz(t) = (-A_4(z(t))A_1^{-1}A_2 + A_5(z(t)))\bar{u}(t) + \sigma(z(t))dB_2(t), \qquad \bar{u}(t) \in U. \tag{5.22}$$

Note that the system (5.22) can be obtained via multiplying (5.1) by ϵ, then formally equating ϵ to zero and expressing y as a function of u, and then substituting the result into (5.2).

6. Proofs for section 3. Proofs of Theorems 3.2 and 3.4 are based on a number of lemmas stated below.

LEMMA 6.1. *Let a function $\psi(S) : (0, \infty) \to R^1$ be such that, for some monotone decreasing function $\nu(S)$, $lim_{S\to\infty}\nu(S) = 0$, the following inequalities are valid:*

$$|\psi(S) - \psi(kS)| \le \nu(S), \quad k = 1, 2, \dots. \tag{6.1}$$

Let also

$$|\psi(S') - \psi(S'')| \le \frac{\alpha|S'' - S'|}{\max\{S', S''\}} \qquad \forall S', S'' > 0, \qquad \alpha = \text{const}. \tag{6.2}$$

Then there exists a limit

$$lim_{S\to\infty}\psi(S) \stackrel{\text{def}}{=} A \tag{6.3}$$

and the estimate

$$|\psi(S) - A| \le \nu(S) \qquad \forall S > 0 \tag{6.4}$$

is valid.

Proof. To establish the existence of the limit it is sufficient to show that, corresponding to any $\delta > 0$, there exists $S_\delta > 0$ such that, for any $S'' \ge S' \ge S_\delta$,

$$|\psi(S'') - \psi(S')| \le \delta. \tag{6.5}$$

Note that from (6.1) it follows that, for any $k_2 \ge k_1 \ge 1$,

$$|\psi(S) - \psi\left(\frac{k_2}{k_1}S\right)| \le |\psi(S) - \psi(k_2S)| + |\psi(k_2S) - \psi\left(\frac{k_2}{k_1}S\right)| \le \nu(S) + \nu\left(\frac{k_2}{k_1}S\right) \le 2\nu(S).$$

Choose integer $k_2 \geq k_1 \geq 1$ in such a way that

$$0 \leq \frac{S''}{S'} - \frac{k_2}{k_1} \leq \frac{\delta}{2\alpha} \Rightarrow 0 \leq S'' - \frac{k_2}{k_1}S' \leq \frac{\delta}{2\alpha}S'.$$

Then, by (6.2),

$$|\psi(S'')-\psi(S')| \leq |\psi(S'')-\psi\left(\frac{k_2}{k_1}S'\right)|+|\psi\left(\frac{k_2}{k_1}S'\right)-\psi(S')| \leq \frac{\alpha|S''-\frac{k_2}{k_1}S'|}{\frac{k_2}{k_1}S'}+2\nu(S')$$

$$\leq \frac{\delta}{2} + 2\nu(S).$$

Choosing S_δ to be such that $\nu(S_\delta) = \frac{\delta}{2}$, one establishes (6.5). Thus the limit (6.3) exists. Passing to the limit as $k \to \infty$ in (6.1), one obtains the estimate (6.4). □

In the following lemmas, it is always supposed that Assumption 1 is satisfied.

LEMMA 6.2. *Let $h(\cdot)$ be as in (3.1) and the constant c_h be defined by*

$$c_h \stackrel{\text{def}}{=} \max_{(u,y)\in U\times\bar{R}^m} ||h(u,y)||. \tag{6.6}$$

Then the following estimates are valid:

$$\sup_{\zeta\in V_h(S,y_0)} E[||\zeta||] \leq c_h \qquad \Rightarrow \qquad \sup_{\zeta\in E[V_h(S,y_0)]} ||\zeta|| \leq c_h; \tag{6.7}$$

$$d_H^E(V_h(S',y_0),V_h(S'',y_0)) \leq \frac{2c_h|S''-S'|}{\max\{S',S''\}}; \tag{6.8}$$

$$d_H(E[V_h(S',y_0)],E[V_h(S'',y_0)]) \leq \frac{2c_h|S''-S'|}{\max\{S',S''\}}. \tag{6.9}$$

Proof. Note that (6.7) is obvious and that (6.9) follows from (6.8) since

$$d_H(E[V_1],E[V_2]) \leq d_H^E(V_1,V_2)$$

for any collections of random variables V_1 and V_2 such that $E[||\zeta||] < \infty \ \ \forall\zeta \in V_i, i = 1,2$. Let us prove (6.8). Assume that $S'' \geq S'$. Then, by (6.7), for any admissible control $u(\cdot)$ and corresponding solution $y(\cdot)$ of the CSDE (2.4),

$$E\left[\left\|\frac{1}{S'}\int_0^{S'} h(u(\tau),y(\tau))d\tau - \frac{1}{S''}\int_0^{S''} h(u(\tau),y(\tau))d\tau\right\|\right]$$

$$\leq E\left[\left\|\left(\frac{1}{S'}-\frac{1}{S''}\right)\int_0^{S'} h(u(\tau),y(\tau))d\tau\right\|\right] + E\left[\left\|\frac{1}{S''}\int_{S'}^{S''} h(u(\tau),y(\tau))d\tau\right\|\right]$$

$$\leq \frac{2c_h(S''-S')}{S''}.$$

This implies (6.8). □

LEMMA 6.3. *Let y_0 be fixed (nonrandom) and $\Psi(p,S,y_0)$ be the support function of the set $E[V_h(S,y_0)]$:*

$$\Psi_h(p,S,y_0) \stackrel{\text{def}}{=} \sup_{v\in E[V_h(S,y_0)]} \{p^T v\}.$$

If the CSDE (2.4) *satisfies W-h-AC, then there exists a convex, positively homogeneous and Lipschitz continuous function* $\Psi_h(p)$ *such that*

$$|\Psi_h(p, S, y_0) - \Psi_h(p)| \leq c\nu_h(S)||p||(1 + ||y_0||^\alpha), \tag{6.10}$$

where $\nu_h(S)$ *is the function introduced in* (3.2) *and* $c = 1 + c_1 (c_1$ *is the constants from* (2.12)).

Proof. Note, first, that $\Psi_h(p, S, y_0)$ allows also the representation

$$\Psi_h(p, S, y_0) = \frac{1}{S} \sup_{(u(\cdot),y(\cdot))} \left\{ E\left[\int_0^S p^T h(u(\tau), y(\tau)) d\tau | y(0) = y_0 \right] \right\},$$

where the sup is over all admissible controls and corresponding solutions of (2.4).

From (6.7) it follows that

$$|\Psi_h(p, S, y_0)| \leq c_h||p||, \qquad |\Psi_h(p', S, y_0) - \Psi_h(p'', S, y_0)| \leq c_h||p' - p''|| \tag{6.11}$$

and from (6.9) it follows that

$$|\Psi_h(p, S', y_0) - \Psi_h(p, S'', y_0)| \leq \frac{2c_h||p|||S'' - S'|}{\max\{S', S''\}}. \tag{6.12}$$

Also, by (3.12),

$$|\Psi_h(p, S, y_0') - \Psi_h(p, S, y_0'')| \leq \nu_h(S)||p||(1 + ||y_0'||^\alpha + ||y_0''||^\alpha). \tag{6.13}$$

Note that if (6.10) is established, then the fact that $\Psi_h(p)$ is convex, positively homogeneous and Lipschitz continuous will follow from the fact that $\Psi_h(p, S, y_0)$ is convex, positively homogeneous and Lipschitz continuous in p (see (6.11)).

By Lemma 6.1, to establish (6.10), it is sufficient to verify the validity of the following estimates:

$$|\Psi(p, kS, y_0) - \Psi(p, S, y_0)| \leq c\nu_h(S)||p||(1 + ||y_0||^\alpha), k = 1, 2, \dots. \tag{6.14}$$

For $k = 1$ it is obvious. Assume that

$$|\Psi(p, (k-1)S, y_0) - \Psi(p, S, y_0)| \leq c\nu_h(S)||p||(1 + ||y_0||^\alpha) \tag{6.15}$$

and show the validity of (6.14) using the induction. Define the collection of random variables $W_h(S, y_0)$ as follows:

$$W_h(S, y_0) \stackrel{\text{def}}{=} \bigcup_{(u(\cdot),y(\cdot))} \left\{ \left(\frac{1}{S} \int_0^S h(u(\tau), y(\tau)) d\tau,\ y(S) \right) \right\}, \tag{6.16}$$

where the union is over all admissible controls and corresponding solutions of the CSDE (2.4). Using dynamic programming, one can obtain

$$\Psi_h(p, kS, y_0) = \sup_{(\zeta,\eta) \in W_h((k-1)S, y_0)} \left\{ \frac{k-1}{k} E[p^T \zeta] + \frac{1}{k} E[\Psi_h(p, S, \eta)] \right\}. \tag{6.17}$$

By (6.13) and (2.12), for any η such that $(\zeta, \eta) \in W_h((k-1)S, y_0)$,

$$|E[\Psi_h(p, S, \eta)] - \Psi_h(p, S, y_0)\ | \leq E[|\Psi_h(p, S, \eta) - \Psi_h(p, S, y_0)|]$$

$$\leq \nu_h(S)||p||(1+E[||\eta||^\alpha]+||y_0||^\alpha) \leq \nu_h(S)||p||(1+c_1(1+||y_0||^\alpha)+||y_0||^\alpha)$$

$$= c\nu_h(S)||p||(1+||y_0||^\alpha),$$

with the constant c being as defined in the statement of the lemma. Hence, using (6.17), one can obtain

$$\left|\Psi_h(p,kS,y_0) - \left(\frac{k-1}{k}\Psi_h(p,(k-1)S,y_0) + \frac{1}{k}\Psi_h(p,S,y_0)\right)\right|$$

$$= \left|\Psi_h(p,kS,y_0) - \sup_{(\zeta,\eta)\in W_h((k-1)S,y_0)} \left\{\frac{k-1}{k}E[p^T\zeta] + \frac{1}{k}\Psi_h(p,S,y_0)\right\}\right|$$

$$\leq \frac{1}{k}\sup_{(\zeta,\eta)\in W_h((k-1)S,y_0)} \{|E[\Psi_h(p,S,\eta)] - \Psi_h(p,S,y_0)|\} \leq \left(\frac{1}{k}\right) c\nu_h(S)||p||(1+||y_0||^\alpha).$$

From (6.15) it follows, on the other hand, that

$$\left|\frac{k-1}{k}\Psi_h(p,(k-1)S,y_0) - \frac{k-1}{k}\Psi_h(p,S,y_0))\right| \leq \left(\frac{k-1}{k}\right) c\nu_h(S)||p||(1+||y_0||^\alpha)$$

$$\Rightarrow \qquad |\Psi_h(p,kS,y_0) - \Psi_h(p,S,y_0)|$$

$$\leq \left|\Psi_h(p,kS,y_0) - \left(\frac{k-1}{k}\Psi_h(p,(k-1)S,y_0\right) + \frac{1}{k}\Psi_h(p,S,y_0))\right|$$

$$+ \left|\frac{k-1}{k}\Psi_h(p,(k-1)S,y_0) - \frac{k-1}{k}\Psi_h(p,S,y_0)\right| \leq \left(\frac{1}{k} + \frac{k-1}{k}\right) c\nu_h(S)||p||(1+||y_0||^\alpha).$$

The latter implies (6.14). □

LEMMA 6.4. *Let the CSDE* (2.4) *satisfy W-h-AC and let V_h be a convex and compact subset of R^j defined by*

$$V_h \stackrel{\text{def}}{=} \{v \mid p^T v \leq \Psi_h(p)\ \forall p \in R^j\}. \tag{6.18}$$

Then, for any y_0 with the probability distribution from the class (C,α),

$$d_H(coE[V_h(S,y_0)], V_h) \leq c(1+C)\nu_h(S), \tag{6.19}$$

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

Comment. The notation (6.18) anticipates the fact that this set will coincide with the set V_h, the existence of which is claimed in Theorem 3.2.

Proof. Note that the fact that the set V_h is convex and compact follows from its definition in the form (6.18) and from the continuity of $\Psi_h(p)$.

Let y_0 be random. The support functions of both $E[V_h(S,y_0)]$ and $coE[V_h(S,y_0)]$ are equal to $E[\Psi_h(p,S,y_0)]$:

$$\sup_{v\in coE[V_h(S,y_0)]} \{p^T v\} = \sup_{v\in E[V_h(S,y_0)]} \{p^T v\} = E[\Psi_h(p,S,y_0)].$$

The support function for V_h is $\Psi_h(p)$ (see Corollary 13.2.1 in [47]). Hence (see, e.g., Lemma П2.9, p. 207 in [24]),

$$d_H(coE[V_h(S,y_0)],V_h) \leq \sup_{||p||\leq 1} |E[\Psi_h(p,S,y_0)] - \Psi_h(p)|. \tag{6.20}$$

By (6.10),

$$| E[\Psi_h(p,S,y_0)] - \Psi_h(p) | \leq c\nu_h(S)||p||(1+E[||y_0||^\alpha]) \leq c\nu_h(S)||p||(1+C) \tag{6.21}$$

for any y_0 with the probability distribution from the class (C,α). This and (6.20) imply (6.19). □

LEMMA 6.5. *For any $S > 0$ and $k = 1, 2, \ldots$, there exists a collection of random variables $V_h'(kS, y_0)$ such that*

$$V_h'(kS,y_0) \subset V_h(kS,y_0) \qquad \Rightarrow \qquad E[V_h'(kS,y_0)] \subset E[V_h(kS,y_0)] \tag{6.22}$$

and such that: (i) *The estimate*

$$d_H(E[V_h'(kS,y_0)], coE[V_h(S,y_0)]) \leq \frac{\bar{c}_h}{k} + c\nu_h(S)(1+E[||y_0||^\alpha]), \tag{6.23}$$

is valid if W-h-AC is satisfied; and (ii) *The estimate*

$$d_H^E(V_h'(kS,y_0), coE[V_h(S,y_0)] \leq \frac{c_h}{\sqrt{k}} + \frac{\bar{c}_h}{k} + c\nu_h(S)(1+E[||y_0||^\alpha]), \tag{6.24}$$

is valid if S-h-AC is satisfied ($c_h, \bar{c}_h$ and c being constants).

Proof. The following three parts detail the proof.

Part I: *Construction of $V_h'(kS,y_0)$*. Consider the CSDE (2.4) on the interval $[0,kS]$ $(k = 1,2,\ldots)$. Denote by $\{u(\cdot)\}^{0,kS}$ the family of admissible controls on the interval $[0,kS]$ such that the restriction of any control from this family to the interval $((i-1)S, iS]$ $(k \geq i \geq 1)$ is conditionally independent on $\mathcal{F}_{(i-1)S}$ conditioned on $y((i-1)S)$.

Define $V'(kS,y_0)$ as the collection of random variables

$$V_h'(kS,y_0) \stackrel{\text{def}}{=} \bigcup_{\{u(\cdot),y(\cdot)\}^{0,kS}} \left\{ \frac{1}{kS}\int_0^{kS} h(u(\tau),y(\tau))d\tau \right\} \tag{6.25}$$

and $E[V'(kS,y_0)]$ as the set of corresponding mathematical expectations

$$E[V_h'(kS,y_0)] \stackrel{\text{def}}{=} \bigcup_{\{u(\cdot),y(\cdot)\}^{0,kS}} \left\{ E\left[\frac{1}{kS}\int_0^{kS} h(u(\tau),y(\tau))d\tau \right] \right\}, \tag{6.26}$$

where, in both cases, the union is over the controls from $\{u(\cdot)\}^{0,kS}$ and the corresponding solutions of the CSDE (2.4) on the interval $[0,kS]$. Note that, by definition,

$$V_h'(S,y_0) = V_h(S,y_0), \qquad E[V_h'(S,y_0)] = E[V_h(S,y_0)] \tag{6.27}$$

and that the inclusions (6.22) are valid for $k = 2,3,\ldots$.

Let $\{u(\cdot),y(\cdot)\}_\eta^{(i-1)S,iS}$ be the family of restrictions to the interval $((i-1)S,iS]$ of the controls $\{u(\cdot)\}^{0,kS}$ and the solutions of the CSDE (2.4) which are obtained with these controls and satisfy the initial conditions $y((i-1)S) \stackrel{\text{def}}{=} \eta$. Define the collection of random variables $V'_h((i-1)S, iS, \eta)$ $(i = 1, \ldots, k)$:

$$V_h((i-1)S,\ iS,\ \eta) \stackrel{\text{def}}{=} \bigcup_{\{u(\cdot),y(\cdot)\}_\eta^{(i-1)S,iS}} \left\{ \frac{1}{S} \int_{(i-1)S}^{iS} h(u(\tau), y(\tau)) d\tau \right\} \tag{6.28}$$

and the set of corresponding mathematical expectations $E[V_h((i-1)S, iS, \eta)]$:

$$E[V_h((i-1)S, iS, \eta)] \stackrel{\text{def}}{=} \bigcup_{\{u(\cdot),y(\cdot)\}_\eta^{(i-1)S,iS}} \left\{ E\left[\frac{1}{S} \int_{(i-1)S}^{iS} h(u(\tau), y(\tau)) d\tau \right] \right\}. \tag{6.29}$$

It is easy to verify that

$$V'_h(kS, y_0) = \bigcup_{(\zeta,\eta) \in W'_h((k-1)S, y_0)} \left\{ \frac{k-1}{k}\zeta + \frac{1}{k} V_h((k-1)S,\ kS,\ \eta) \right\} \tag{6.30}$$

$$\Rightarrow\ E[V'_h(kS, y_0)] = \bigcup_{(\zeta,\eta) \in W'_h((k-1)S, y_0)} \left\{ \frac{k-1}{k} E[\zeta] + \frac{1}{k} E[V_h((k-1)S,\ kS,\ \eta)] \right\}, \tag{6.31}$$

where

$$W'_h((k-1)S, y_0) \stackrel{\text{def}}{=} \bigcup_{\{u(\cdot),y(\cdot)\}^{0,(k-1)S}} \left\{ \left(\frac{1}{(k-1)S} \int_0^{(k-1)S} h(u(\tau), y(\tau)) d\tau,\ y((k-1)S) \right) \right\}, \tag{6.32}$$

the union being over the controls from the family of restrictions of the elements of $\{u(\cdot)\}^{0,kS}$ to the interval $[0, (k-1)S]$ and corresponding solutions of the CSDE (2.4).

Part II: *Proof of Lemma* 6.5(i). Using induction, let us show that

$$d_H(E[V'_h(kS, y_0)],\ \frac{1}{k} \sum_1^k E[V_h((i-1)S,\ iS,\ y_0)] \leq c\nu_h(S)(1+E[||y_0||^\alpha]), \quad k = 1, 2, \ldots. \tag{6.33}$$

For $k = 1$ it is immediate since, by definition, $E[V_h(S, y_0)] = E[V_h(0,\ S,\ y_0)]$. Assume that the estimate

$$d_H(E[V'_h((k-1)S,\ y_0)], \frac{1}{k-1} \sum_1^{k-1} E[V_h((i-1)S,\ iS,\ y_0)]) \leq c\nu_h(S)(1+E[||y_0||^\alpha]) \tag{6.34}$$

is valid. From Assumption 1 and W-h-AC (see (3.12)) it follows that, for any η such that $(\zeta, \eta) \in W'_h((k-1)S, y_0)$,

$$d_H(E[V_h((k-1)S,\ kS,\ \eta)],\ E[V_h((k-1)S,\ kS,\ y_0)]) \leq \nu_h(S)(1+E[||\eta||^\alpha]+E[||y_0||^\alpha])$$

$$\leq \nu_h(S)(1 + c_1(1 + E[||y_0||^\alpha]) + E[||y_0||^\alpha]) = c\nu_h(S)(1 + E[||y_0||^\alpha]).$$

This and (6.31) lead to the estimate

$$d_H(E[V'(kS, y_0)],\ \frac{k-1}{k} E[V'((k-1)S,\ y_0)] + \frac{1}{k} E[V_h((k-1)S,\ kS,\ y_0)])$$

$$= d_H\left(E[V'(kS,y_0)], \bigcup_{(\zeta,\eta)\in W'_h((k-1)S,y_0)}\left\{\frac{k-1}{k}E[\zeta]+\frac{1}{k}E[V_h((k-1)S,\ kS,\ y_0)]\right\}\right)$$
$$\leq \left(\frac{1}{k}\right)c\nu_h(S)(1+E[||y_0||^\alpha]).$$

Using the estimate above and (6.34), one can further obtain that

$$d_H(E[V'(kS,y_0)],\ \frac{1}{k}\sum_1^k E[V_h((i-1)S,\ iS,\ y_0)])$$
$$\leq d_H\left(\frac{k-1}{k}E[V'((k-1)S,\ y_0)]+\frac{1}{k}E[V_h((k-1)S,\ kS,\ y_0)]\right),$$
$$\frac{1}{k}\sum_1^k E[V_h((i-1)S,\ iS,\ y_0)])\ +\ \left(\frac{1}{k}\right)c\nu_h(S)(1+E[||y_0||^\alpha])$$
$$\leq\ d_H\left(\frac{k-1}{k}E[V'((k-1)S,\ y_0)],\frac{k-1}{k}\frac{1}{k-1}\sum_1^{k-1}E[V_h((i-1)S,\ iS,\ y_0)]\right)$$
$$+\left(\frac{1}{k}\right)c\nu_h(S)(1+E[||y_0||^\alpha])\leq\left(\frac{k-1}{k}\right)c\nu_h(S)(1+E[||y_0||^\alpha])+\left(\frac{1}{k}\right)c\nu_h(S)(1+E[||y_0||^\alpha])$$
$$= c\nu_h(S)(1+E[||y_0||^\alpha]).$$

Thus, (6.33) is established.

Since

$$E[V_h((i-1)S,\ iS,\ y_0)] = E[V_h(S,y_0)] \quad \forall i=1,\dots,k,$$

(6.33) is equivalent to

$$d_H(E[V'_h(kS,y_0)],\ \frac{1}{k}\sum_1^k E[V_h((S,y_0)]) \leq c\nu_h(S)(1+E[||y_0||^\alpha]).$$

By Shapley–Folkman's theorem (see, e.g., [24, p. 204])

$$d_H\left(\frac{1}{k}\sum_1^k E[V_h(S,y_0)],\ coE[V_h(S,y_0)]\right)\leq\frac{2(j+1)c_h}{k}, \tag{6.35}$$

where c_h is as in (6.6) and j is the dimension of the Euclidean space containing the subsets above. These imply (6.23) with $\bar{c}_h \stackrel{\text{def}}{=} 2(j+1)c_h$.

Part III: *Proof of Lemma* 6.5(ii). Let $y_0^0 \stackrel{\text{def}}{=} y_0$ and let y_0^i $i=1,\dots,k-1$ have the same probability distribution as y_0 and be independent of y_0 and among themselves (and also independent of $W(\cdot)$). Using induction, let us show that

$$d_H^E\left(V'_h(kS,y_0),\ \frac{1}{k}\sum_1^k V_h((i-1)S,\ iS,\ y_0^{i-1})\right)\leq c\nu_h(S)(1+E[||y_0||^\alpha]),\quad k=1,2,\dots. \tag{6.36}$$

For $k = 1$, the above expression is obviously true. Assume that

$$d_H^E\left(V_h'((k-1)S, y_0),\ \frac{1}{k-1}\sum_1^{k-1} V_h((i-1)S,\ iS,\ y_0^{i-1})\right) \leq c\nu_h(S)(1+E[||y_0||^\alpha]). \tag{6.37}$$

From Assumption 1 and S-h-AC (see (3.11)) it follows that, for any η such that $(\zeta, \eta) \in W_h'((k-1)S, y_0)$,

$$d_H^E\left(V_h((k-1)S,\ kS,\ \eta),\ V_h((k-1)S,\ kS,\ y_0^{k-1})\right) \leq \nu_h(S)(1+E[||\eta||^\alpha]+E[||y_0||^\alpha])$$

$$\leq \nu_h(S)(1+c_1(1+E[||y_0||^\alpha])+E[||y_0||^\alpha]) = c\nu_h(S)(1+E[||y_0||^\alpha]).$$

Hence, by (6.30),

$$d_H^E\left(V'(kS, y_0),\ \frac{k-1}{k}V'((k-1)S,\ y_0) + \frac{1}{k}V_h((k-1)S,\ kS,\ y_0^{k-1})\right)$$

$$= d_H^E\left(V'(kS, y_0), \bigcup_{(\zeta,\eta)\in W_h'((k-1)S,y_0)} \left\{\frac{k-1}{k}\zeta + \frac{1}{k}V_h((k-1)S,\ kS,\ y_0^{k-1})\right\}\right)$$

$$\leq \left(\frac{1}{k}\right) \sup_{(\zeta,\eta)\in W_h'((k-1)S,y_0)} d_H^E(V_h((k-1)S,\ kS,\ \eta),\ V_h((k-1)S,\ kS,\ y_0^{k-1}))$$

$$\leq \left(\frac{1}{k}\right) c\nu_h(S)(1+E[||y_0||^\alpha]).$$

Using this estimate and (6.37), one obtains

$$d_H^E\left(V'(kS, y_0),\ \frac{1}{k}\sum_1^k V_h((i-1)S,\ iS,\ y_0^{i-1})\right)$$

$$\leq d_H^E\left(\frac{k-1}{k}V'((k-1)S,\ y_0) + \frac{1}{k}V_h((k-1)S,\ kS,\ y_0^{k-1}),\ \frac{1}{k}\sum_1^k V_h((i-1)S,\ iS,\ y_0^{i-1})\right)$$

$$+ \left(\frac{1}{k}\right) c\nu_h(S)(1+E[||y_0||^\alpha])$$

$$\leq d_H^E\left(\frac{k-1}{k}V'((k-1)S,\ y_0),\ \frac{k-1}{k}\frac{1}{k-1}\sum_1^{k-1} V_h((i-1)S,\ iS,\ y_0^{i-1})\right)$$

$$+ \left(\frac{1}{k}\right) c\nu_h(S)(1+E[||y_0||^\alpha]) \leq \left(\frac{k-1}{k}\right) c\nu_h(S)(1+E[||y_0||^\alpha]) + \left(\frac{1}{k}\right) c\nu_h(S)(1+E[||y_0||^\alpha])$$

$$= c\nu_h(S)(1+E[||y_0||^\alpha]).$$

This proves the validity of the estimate (6.36).

The elements of $V_h((i-1)S,\ iS,\ y_0^{i-1}), i = 1, 2, \ldots, k$, are mutually independent and, by (6.6),

$$E[||\zeta||^2] \leq c_h^2 \qquad \forall \zeta \in V_h((i-1)S,\ iS,\ y_0^{i-1}), \quad i = 1, 2, \ldots, k.$$

Hence, one can use Proposition 3.5 and the fact that $E[V_h((i-1)S,\ iS,\ y_0^{i-1})] = E[V_h(S, y_0)]$ to obtain

$$d_H^E\left(\frac{1}{k}\sum_1^k V_h((i-1)S,\ iS,\ y_0^{i-1}),\ \frac{1}{k}\sum_1^k E[V_h((i-1)S,\ iS,\ y_0^{i-1})]\right)$$
$$= d_H^E\left(\frac{1}{k}\sum_1^k V_h((i-1)S,\ iS,\ y_0^{i-1}),\ \frac{1}{k}\sum_1^k E[V_h(S,y_0)]\right) \le \frac{c_h}{\sqrt{k}}.$$

The last estimate along with (6.35) and (6.33) imply (6.24). □

Corollary 6.6. *For any y_0 with the probability distribution from the class (C,α),*

$$d_H(E[V_h'(kS,y_0)], V_h) \le \frac{\bar{c}_h}{k} + 2c(1+C)\nu_h(S) \tag{6.38}$$

if W-h-AC is satisfied and

$$d_H^E(V_h'(kS,y_0), V_h) \le \frac{c_h}{\sqrt{k}} + \frac{\bar{c}_h}{k} + 2c(1+C)\nu_h(S) \tag{6.39}$$

if S-h-AC is satisfied.

Proof. The estimates follow from (6.19), (6.23) and (6.19), (6.24), respectively. □

Lemma 6.7. *For any y_0 with the probability distribution from the class (C,α),*

$$\sup_{\zeta\in V_h} d(\zeta,\ E[V_h(S,y_0)]) \le \nu_h^1(S), \qquad \lim_{S\to\infty} \nu_h^1(S) = 0 \tag{6.40}$$

if W-h-AC is satisfied, and

$$\sup_{\zeta\in V_h} d^E(\zeta,\ V_h(S,y_0)) \le \nu_h^2(S), \qquad \lim_{S\to\infty} \nu_h^2(S) = 0 \tag{6.41}$$

if S-h-AC is satisfied.

Proof. Using (6.9) with $S'' = S$ and $S' = \lfloor S^{\frac12}\rfloor S^{\frac12}$, one can obtain

$$d_H(E[V_h(\lfloor S^{\frac12}\rfloor S^{\frac12}, y_0)],\ E[V_h(S,y_0)]) \le \frac{2c_h}{S^{\frac12}}, \tag{6.42}$$

where $\lfloor S^{\frac12}\rfloor$ stands for the integer part of $S^{\frac12}$. Hence,

$$\sup_{\zeta\in V_h} d(\zeta,\ E[V_h(S,y_0)]) \le \sup_{\zeta\in V_h} d(\zeta,\ E[V_h(\lfloor S^{\frac12}\rfloor S^{\frac12}, y_0)]) + \frac{2c_h}{S^{\frac12}}. \tag{6.43}$$

From (6.22) and (6.38) (with the replacement of S by $S^{\frac12}$ and the replacement of k by $\lfloor S^{\frac12}\rfloor$) it follows, on the other hand, that

$$\sup_{\zeta\in V_h} d(\zeta,\ E[V_h(\lfloor S^{\frac12}\rfloor S^{\frac12}, y_0)]) \le \sup_{\zeta\in V_h} d(\zeta,\ E[V_h'(\lfloor S^{\frac12}\rfloor S^{\frac12}, y_0)])$$
$$\le d_H(E[V_h'(\lfloor S^{\frac12}\rfloor S^{\frac12}, y_0)],\ V_h) \le \frac{\bar{c}_h}{\lfloor S^{\frac12}\rfloor} + 2c(1+C)\nu_h(S^{\frac12})$$

This and (6.43) imply (6.40) with $\nu_h^1(S) \stackrel{\text{def}}{=} \frac{2c_h}{S^{\frac12}} + \frac{\bar{c}_h}{\lfloor S^{\frac12}\rfloor} + 2c(1+C)\nu_h(S^{\frac12})$.

To establish (6.41), one can use (6.8) and obtain, similarly to (6.42), that

$$d_H^E(V_h(\lfloor S^{\frac{1}{2}} \rfloor S^{\frac{1}{2}}, y_0),\ V_h(S, y_0)) \leq \frac{2c_h}{S^{\frac{1}{2}}},$$

$$(6.44)\qquad \Rightarrow \quad \sup_{\zeta \in V_h} d^E(\zeta,\ V_h(S, y_0)) \leq \sup_{\zeta \in V_h} d^E(\zeta,\ V_h(\lfloor S^{\frac{1}{2}} \rfloor S^{\frac{1}{2}}, y_0)]) + \frac{2c_h}{S^{\frac{1}{2}}}.$$

By (6.22) and (6.39) (with S being replaced by $S^{\frac{1}{2}}$ and k by $\lfloor S^{\frac{1}{2}} \rfloor$ as above),

$$\sup_{\zeta \in V_h} d^E(\zeta,\ V_h(\lfloor S^{\frac{1}{2}} \rfloor S^{\frac{1}{2}}, y_0)) \leq \sup_{\zeta \in V_h} d^E(\zeta,\ V_h'(\lfloor S^{\frac{1}{2}} \rfloor S^{\frac{1}{2}}, y_0))$$

$$\leq d_H^E(V_h'(\lfloor S^{\frac{1}{2}} \rfloor S^{\frac{1}{2}}, y_0),\ V_h) \leq \frac{c_h}{\sqrt{\lfloor S^{\frac{1}{2}} \rfloor}} + \frac{\bar{c}_h}{\lfloor S^{\frac{1}{2}} \rfloor} + 2c(1 + C)\nu_h(S^{\frac{1}{2}}).$$

This and (6.44) imply (6.41) with $\nu_h^2(S)$ being equal to the sum of $\frac{2c_h}{S^{\frac{1}{2}}}$ and the right-hand side of the last inequality. □

Proof of Theorem 3.2. By (6.19),

$$sup_{\zeta \in E[V_h(S,y_0)]} d(\zeta, V_h) \leq sup_{\zeta \in coE[V_h(S,y_0)]} d(\zeta, V_h) \leq d_H(coE[V_h(S, y_0)], V_h)$$

$$\leq c(1 + C)\nu_h(S).$$

Comparing this estimate and (6.40), one obtains (3.14) with $\nu_h^{C,\alpha}(S) \stackrel{\text{def}}{=} \max\{c(1 + C)\nu_h(S), \nu_h^1(S)\}$. □

Proof of Theorem 3.4. The statement (i) of the theorem is established by (6.41), with $\tilde{\nu}_h^{C,\alpha}(S) \stackrel{\text{def}}{=} \nu_h^2(S)$. Let us prove the statement (ii). Assume it is not true. Then there exists a number $\delta > 0$, and sequences $\mu_l \in \mathcal{M}$ and $S_l, l = 1, 2, \dots,$ ($S_l \to \infty$ as $l \to \infty$) such that, for any $\mu' \in \mathcal{M}(S_l, y_0)$,

$$(6.45)\quad E[\rho(\mu_l, \mu')] = \sum_{i=1}^{\infty} 2^{-i} E\left[\left|\int f_i(u, y)\mu_l(du, dy) - \int f_i(u, y)\mu'(du, dy)\right|\right] \ \geq\ \delta$$

$$\Rightarrow \sum_{i=1}^{N} 2^{-i} E[|\int f_i(u, y)\mu_l(du, dy) - \int f_i(u, y)\mu'(du, dy)|] \ \geq\ \frac{\delta}{2} \quad \forall \mu' \in \mathcal{M}(S_l, y_0)$$

for N large enough. Hence, for any $\mu' \in \mathcal{M}(S_l, y_0)$,

$$E\left[\sqrt{\sum_{i=1}^{N}\left|\int f_i(u, y)\mu_l(du, dy) - \int f_i(u, y)\mu'(du, dy)\right|^2}\,\right] \ \geq\ \frac{c_N \delta}{2}, \quad c_N = \text{const.}$$

(6.46)

Let $h(u, y)$ be defined by (3.1) with $j = N$ and let $\zeta_l \stackrel{\text{def}}{=} \int h(u, y)\mu_l(du, dy)$; by (3.17), $\zeta_l \in V_h$. Also, by (3.6), $V_h(S_l, y_0)$ is the union of all $\zeta' \stackrel{\text{def}}{=} \int h(u, y)\mu'(du, dy)$ over $\mu' \in \mathcal{M}(S_l, y_0)$. The estimate (6.46) is equivalent, thus, to

$$E[||\zeta_l - \zeta'||] \geq \frac{c_N \delta}{2} \quad \forall \zeta' \in V_h(S_l, y_0) \qquad \Leftrightarrow \qquad d^E(\zeta_l, V_h(S_l, y_0)) \geq \frac{c_N \delta}{2}.$$

The latter contradicts (6.41). □

Proof of Theorem 3.3. Let Assumption 1 and W-h-AC be satisfied for any $h(\cdot)$ as in (3.1). Then, by Theorem 3.2, for any such $h(\cdot)$ and any initial condition y_0 with the probability distribution from the class (C, α), there exists a convex and compact set V_h such that (3.14) is satisfied. From Corollary 3.7 in [28] (see also Theorem 3.1(i) in [27] and more general results in [7]) it follows that

$$\rho_H(E[\mathcal{M}(S, y_0)], \bar{\mathcal{M}}) \leq \nu^{C,\alpha}(S) \tag{6.47}$$

for some $\nu^{C,\alpha}(S)$ tending to zero as S tends to infinity, where

$$\bar{\mathcal{M}} \stackrel{\text{def}}{=} \{\mu \in \mathcal{P}(U \times \bar{R}^m) \mid \int h(u,y)\mu(du,dy) \in V_h \quad \forall h(u,y) \;\; \textit{as in } (3.1)\}.$$

It is easy to verify that the set $\bar{\mathcal{M}}$ is convex and compact. Hence, the estimate (3.15) will be established if one shows that $\mathcal{M} = \bar{\mathcal{M}}$. Since, by definition, $\mathcal{M} \subset \bar{\mathcal{M}}$, it is enough to show that $\bar{\mathcal{M}} \subset \mathcal{M}$.

For $N = 1, 2, \ldots$, let $Y_N \stackrel{\text{def}}{=} \{y \in R^m \mid ||y|| \leq N\}$ and $Y_N^c \stackrel{\text{def}}{=} \{y \in R^m \mid ||y|| > N\}$. By (2.8) and (2.10), for any $\mu \in E[\mathcal{M}(S, y_0)]$, there exists an admissible control $u(\cdot)$ and the corresponding solution $y(\cdot)$ of the CSDE (2.4) such that

$$\mu(U \times Y_N^c) = \frac{1}{S}\int_0^S E[\chi_{Y_N^c}(y(\tau))]d\tau \;\leq\; \frac{1}{N^\alpha}\frac{1}{S}\int_0^S E[\chi_{Y_N^c}(y(\tau))\; ||y(\tau)||^\alpha]d\tau,$$

where $\chi_{Y_N^c}(y(\tau))$ is the indicator function of Y_N^c. Hence, using Assumption I and the fact that y_0 has the probability distribution from the class (C, α), one obtains that, for any $\mu \in E[\mathcal{M}(S, y_0)]$,

$$\mu(U \times Y_N^c) \leq \frac{1}{N^\alpha}\frac{1}{S}\int_0^S E[||y(\tau)||^\alpha]d\tau \leq \frac{1}{N^\alpha}c_1(C+1)$$
$$\Rightarrow \qquad \mu(U \times Y_N) \geq 1 - \frac{1}{N^\alpha}c_1(C+1).$$

Take an arbitrary $\mu \in \bar{\mathcal{M}}$. By (6.47), there exists a sequence $\mu_i \in E[\mathcal{M}(S_i, y_0)]$ (S_i tends to infinity as i tends to infinity) such that

$$\lim_{i\to\infty}\rho(\mu_i, \mu) = 0 \quad \Rightarrow \quad \mu(U \times Y_N) \geq \limsup_{i\to\infty}\mu_i(U \times Y_N) \geq 1 - \frac{1}{N^\alpha}c_1(C+1)$$
$$\Rightarrow \qquad \mu(U \times R^m) = 1.$$

The latter implies that $\mu \in \mathcal{M}$ and, hence, $\bar{\mathcal{M}} \subset \mathcal{M}$.

Using the second representation for $E[V_h(S, y_0)]$ in (3.7), it is straightforward to verify that the validity of (3.15) implies the validity of (3.14) with V_h as in (3.17) for any $h(\cdot)$ as in (3.1). The fact that W-h-AC is satisfied for any such $h(\cdot)$ follows now from Theorem 3.2. ☐

7. Proofs for sections 4. *Proof of Theorem* 4.2. Assume that (4.6) is not valid. Then there exist a number $\delta > 0$ and initial conditions y_0 with the probability distribution from the class (C, α) such that, for some S_i, $\lim_{i\to\infty} S_i = \infty$, and some $\mu_i \in \mathcal{M}(S_i, y_0)$,

$$E[\inf_{\mu' \in D}\rho(\mu_i, \mu')] \geq \delta \quad i = 1, 2, \ldots, \qquad \Rightarrow \qquad E[\inf_{\mu' \in D}\rho(\mu^*, \mu')] \geq \delta, \tag{7.1}$$

where it is assumed (without loss of generality) that $\mu_i \to \mu^*$ in law as $i \to \infty$. From the fact that $\mu_i \in \mathcal{M}(S_i, y_0)$ it follows that there exist an admissible control $u^i(\cdot)$ and the corresponding solution $y^i(\cdot)$ of the CSDE (2.4) (with the initial conditions $y^i(0) = y_0$) such that μ_i is the occupational measure of the pair $(u^i(\cdot), y^i(\cdot))$ on the interval $[0, S_i]$ (see (2.6)). Hence, for any $f \in \mathcal{D}$

$$\frac{1}{S_i}(f(y^i(S) - f(y(0))) = \int_{U\times \bar{R}^m} (Lf)(u,y)\mu_i(du,dy)$$

$$+\frac{1}{S_i}\int_0^{S_i} \langle \nabla f(y^i(\tau)), b(y^i(\tau))dW(\tau)\rangle,$$

where, in order for the integration in the first term of the right-hand side to be legitimate, the definition of $(Lf) : U\times R^m \to R^1$ (see (4.1) above) is formally extended to $(Lf) : U \times \bar{R}^m \to R^1$ by setting $(Lf)(\cdot,\infty) \stackrel{\text{def}}{=} 0$.

The left-hand side and the variance of the second term on the right-hand side of the above expression tend to zero as $S_i \to \infty$ (this can be easily derived from the fact that the probability distribution of y_0 belongs to the class (C, α) and from that Assumption 1 is satisfied with $\alpha \geq 2$). Hence, the first term on the right-hand side tends to zero in law.

By Skorohod's theorem (see, e.g., [15, p. 23]), there exist $\mathcal{P}(U \times \bar{R}^m)$-valued random variables $\tilde{\mu}_i$ and $\tilde{\mu}^*$ defined on a common probability space such that they agree in law with μ_i and μ^*, respectively, and such that

$$\lim_{i\to\infty} \rho(\tilde{\mu}_i, \tilde{\mu}^*) = 0 \quad a.s. \tag{7.2}$$

$$\Rightarrow \quad \lim_{i\to\infty} \int_{U\times\bar{R}^m} (Lf)(u,y)\tilde{\mu}_i(dxdu) = \int_{U\times\bar{R}^m} (Lf)(u,y)\tilde{\mu}^*(du,dy) \quad a.s.$$

$$\Rightarrow \quad \int_{U\times\bar{R}^m} (Lf)(u,y)\tilde{\mu}^*(du,dy) = 0 \ \ a.s. \quad \Rightarrow \quad \int_{U\times\bar{R}^m} (Lf)(u,y)\mu^*(du,dy) = 0 \ \ a.s.$$

Since $\mathcal{D}$ is countable, the last expression is valid for all $f \in \mathcal{D}$ outside a common zero probability set. Hence, if one establishes that $\mu^* \in \mathcal{P}(U \times R^m)$ a.s., it will follow that $\mu^* \in D$ a.s. and, thus, it will contradict (7.1).

To complete the proof of the theorem, one needs to show now that $\mu^* \in \mathcal{P}(U\times R^m)$ a.s. That is, one needs to show that

$$\mu^*(U \times R^m) = 1 \qquad a.s. \tag{7.3}$$

From Assumption 1 it follows that, for any $\delta > 0$, there exists a compact set $Y_\delta \subset R^m$ such that

$$E[\tilde{\mu}_i(U \times Y_\delta)] = E[\mu_i(U \times Y_\delta)] \geq 1 - \delta,$$

where it is also taken into account that $\tilde{\mu}_i$ and μ_i agree in law. By (7.2),

$$\tilde{\mu}^*(U \times Y_\delta) \geq \limsup_{i\to\infty} \tilde{\mu}_i(U \times Y_\delta) \quad a.s.$$

$$\Rightarrow \quad E\left[\tilde{\mu}^*(U \times Y_\delta)\right] \geq E[\limsup_{i\to\infty} \tilde{\mu}_i(U \times Y_\delta)] \geq \limsup_{i\to\infty} E[\tilde{\mu}_i(U \times Y_\delta)] \geq 1 - \delta.$$

Since μ^* and $\tilde{\mu}^*$ agree in law,

$$E[\tilde{\mu}^*(U \times Y_\delta)] = E[\mu^*(U \times Y_\delta)] \quad \Rightarrow \quad E[\mu^*(U \times Y_\delta)] \geq 1 - \delta.$$

Since δ can be arbitrary small, the latter implies that $E[\mu^*(U \times R^m)] = 1$ which, in turn, implies the validity of (7.3). This completes the proof of the theorem. ∎

Proof of Theorem 4.1(i). It can be easily verified that $\rho(\mu, D)$ is a convex function of μ. Hence, by (4.6),

$$\rho(E[\mu], D) \leq E[\rho(\mu, D)] \leq \bar{\nu}^{(C,\alpha)}(S) \quad \forall \mu \in \mathcal{M}(S, y_0)$$
$$\Rightarrow \sup_{\mu \in E[\mathcal{M}(S,y_0)]} \rho(\mu, D) \leq \bar{\nu}^{(C,\alpha)}(S).$$

Since the above estimate is uniform with respect to the initial conditions y_0 which have the probability distribution from the class (C, α), it follows that

$$\sup_{\{y_0\} \in (C,\alpha)} \left\{ \sup_{\mu \in E[\mathcal{M}(S,y_0)]} \rho(\mu, D) \right\} \leq \bar{\nu}^{(C,\alpha)}(S).$$

In Lemma 7.1 below it is shown that

$$\sup_{\mu \in D} \rho \left(\mu, \bigcup_{\{y_0\} \in (C,\alpha)} E[\mathcal{M}(S, y_0)] \right) = 0. \tag{7.4}$$

These imply (4.4). ∎

LEMMA 7.1. *The equation* (7.4) *is valid if the conditions of Theorem* 4.1 *are satisfied.*

Proof. For any $h(\cdot)$ as in (3.1), define the set D_h by

$$D_h \stackrel{\text{def}}{=} \bigcup_{\mu \in D} \left\{ \int h(u, y)\mu(du, dy) \right\}. \tag{7.5}$$

As follows from Lemma 3.5 in [28], the validity of (7.4) will be established if one shows that

$$\sup_{v \in D_h} d \left(v, \bigcup_{\{y_0\} \in (C,\alpha)} \{E[V_h(S, y_0)]\} \right) = 0. \tag{7.6}$$

Take an arbitrary element $v \in D_h$. By definition, there exists $\mu \in D$ such that $v = \int h(u, y)\mu(du, dy)$. From results in [13] and [49] it follows that there exists m-dimensional standard Brownian motion $W'(\cdot)$ and a stationary $\mathcal{P}(U) \times R^m$ - valued random process $(\lambda'(\tau), y'(\tau))$ such that

$$dy'(\tau) = \tilde{a}(\lambda'(\tau), y'(\tau))dt + b(y'(\tau))dW'(\tau), \qquad \tilde{a}(\lambda, y) \stackrel{\text{def}}{=} \int a(u, y)\lambda(du), \tag{7.7}$$

and

$$E\left[\int h(u, y'(\tau))\lambda'(\tau)(du) \right] = \int h(u, y)\mu(du, dy) = v \quad \forall \tau \geq 0, \tag{7.8}$$

$$E[||y'(\tau)||^\alpha] = \int ||y||^\alpha \mu(du,dy) \le c_2 \quad \forall \tau \ge 0, \tag{7.9}$$

with $W'(\cdot)$ being independent of $y'(0)$ and $\lambda'(\cdot)$ being nonanticipative (i.e., for $\bar\tau > \tau, W'(\bar\tau) - W'(\tau)$ is independent of $\{y'(0)$ and $W'(\theta), \lambda'(\theta), \theta \le \tau\}$); ($c_2$ is the constant from Assumption 2). Using Filippov type chattering lemma for CSDE (see, e.g., [14, p. 15]), one can establish that there exists a sequence of admissible controls $u^i(\cdot)$ and the corresponding sequence of solutions $y^i(\cdot)$ of the CSDE (2.4) (considered with $W'(\cdot)$ instead of $W(\cdot)$) such that $y^i(0) = y'(0)$ and such that

$$\lim_{i\to\infty} E\left[\left\|\frac{1}{S}\int_0^S h(u^i(\tau), y^i(\tau))d\tau - \frac{1}{S}\int_0^S \int h(u, y'(\tau))\lambda'(\tau)(du)d\tau\right\|\right] = 0. \tag{7.10}$$

From (7.8) and (7.10) it follows that

$$\lim_{i\to\infty} \left\| E\left[\frac{1}{S}\int_0^S h(u^i(\tau), y^i(\tau))d\tau\right] - v\right\| = 0. \tag{7.11}$$

Since

$$E\left[\frac{1}{S}\int_0^S h(u^i(\tau), y^i(\tau))d\tau\right] \in E[V_h(S, y'(0))] \subset \bigcup_{\{y_0\}\in(C,\alpha)} \{E[V_h(S, y_0)]\}$$

(the last inclusion being due to the fact that, by (7.9), $y'(0)$ has the probability distribution from the class (C, α) with $C \ge c_2$), one can use (7.11) to obtain that

$$dist\left(v, \bigcup_{\{y_0\}\in(C,\alpha)} \{E[V_h(S, y_0)]\}\right) = 0.$$

As v is an arbitrary element of D_h, this implies (7.6). □

The proofs of Corollaries 4.4 and 4.5 are based on the following result.

LEMMA 7.2. *A sequence $\mu^k \in \mathcal{P}(U \times R^m), k = 1, 2, \dots,$ converges to $\mu \in \mathcal{P}(U \times R^m)$ in the metric ρ defined in* (2.1) *(that is, $\lim_{k\to\infty} \rho(\mu^k, \mu) = 0$) if and only if*

$$\lim_{k\to\infty} \int f(u,y)\mu^k(du,dy) = \int f(u,y)\mu(du,dy)$$

for any bounded continuous function $f(u, y) : U \times R^m \to R^1$.

Proof. follows from Theorem 2.1.1 in [15]. □

Proof of Corollary 4.4. By Assumption 2 (see (4.3)), for any $\mu \in D$ and $N \ge 1$,

$$N^{\alpha-1}\int_{||y||\ge N} ||y||\mu(du,dy) \le \int_{||y||\ge N} ||y||^\alpha \mu(du,dy) \le \int ||y||^\alpha \mu(du,dy) \le c_2$$

$$\int_{||y||\ge N} ||y||\mu(du,dy) \le \frac{c_2}{N^{\alpha-1}} \qquad \forall \mu \in D. \tag{7.12}$$

Let $\xi_N(\theta) : [0,\infty) \to [0,1]$ be a continuous function such that $\xi_N(\theta) = 1$ for $\theta \in [0, N]$ and such that $\xi_N(\theta) = 0$ for $\theta \in [N+1, \infty)$. Let $g_N(u, y) \stackrel{\text{def}}{=} g(u, y)\xi_N(||y||)$. According to these definitions, $g_N(u, y) = g(u, y)$ for $||y|| \le N$ and also

$$||g_N(u,y)|| \le ||g(u,y)||, \quad ||g_N(u,y)|| \le \max_{u\in U, ||y||\le N} ||g(u,y)|| \qquad \forall (u,y) \in U \times Y.$$

Due to the Lipschitz continuity of $g(u,y)$,

$$||g(u,y)|| \leq a_1 + a_2||y|| \quad \Rightarrow \quad ||g_N(u,y)|| \leq a_1 + a_2||y|| \qquad \forall y \in R^m, \tag{7.13}$$

where $a_i = \text{const}, i = 1, 2$. By (7.12), it is implied that

$$\left\| \int g(u,y)\mu(du,dy) - \int g_N(u,y)\mu(du,dy) \right\| \leq \frac{a_3}{N^{\alpha-1}} \;\; \forall \mu \in D, \quad a_3 = \text{const}. \tag{7.14}$$

From (7.12) and (7.13) it follows that the set V_g is bounded. Let us prove that it is closed by showing that, if $\mu_k \in D$ and the limit $\lim_{k\to\infty} \int g(u,y)\mu_k(du,dy)$ exists, then this limit belongs to V_g. Assume the above limit does exist. Due to the fact that D is compact, one may also assume (without loss of generality) that $\lim_{k\to\infty} \rho(\mu_k,\mu) = 0$ for some $\mu \in D$. By virtue of Lemma 7.2 and since $g_N(u,y)$ is bounded, the latter leads to the equality $\lim_{k\to\infty} \int g_N(u,y)\mu_k(du,dy) = \int g_N(u,y)\mu(du,dy)$, which, in turn, implies that

$$\begin{aligned}
&\lim_{k\to\infty} \left\| \int g(u,y)\mu_k(du,dy) - \int g(u,y)\mu(du,dy) \right\| \\
&\leq \limsup_{k\to\infty} \left\| \int g(u,y)\mu_k(du,dy) - \int g_N(u,y)\mu_k(du,dy) \right\| \\
&+ \lim_{k\to\infty} \left\| \int g_N(u,y)\mu_k(du,dy) - \int g_N(u,y)\mu(du,dy) \right\| \\
&+ \left\| \int g_N(u,y)\mu(du,dy) - \int g(u,y)\mu(du,dy) \right\| \\
&\leq \frac{2a_3}{N^{\alpha-1}} \quad \Rightarrow \quad \lim_{k\to\infty} \int g(u,y)\mu_k(du,dy) = \int g(u,y)\mu(du,dy) \in V_g.
\end{aligned}$$

This proves that V_g is compact.

Let $V_{g_N}(S,y_0)$ and V_{g_N} be defined by (4.10) and (4.11) with the replacement of $g(\cdot)$ by $g_N(\cdot)$. By (7.14),

$$d_H(V_{g_N}, V_g) \leq \frac{a_3}{N^{\alpha-1}}. \tag{7.15}$$

Similarly to (7.12), from Assumption 1 it follows that, for any y_0 having a probability distribution from the class (C,α),

$$E\left[\int_{||y||\geq N} ||y||\mu(du,dy) \right] \leq \frac{a_4}{N^{\alpha-1}} \qquad \forall \mu \in \mathcal{M}(S,y_0) \tag{7.16}$$

$$\Rightarrow \; E\left[\left\| \int g(u,y)\mu(du,dy) - \int g_N(u,y)\mu(du,dy) \right\| \right] \leq \frac{a_5}{N^{\alpha-1}} \;\; \forall \mu \in \mathcal{M}(S,y_0) \tag{7.17}$$

$$\Rightarrow \qquad d_H^E(V_{g_N}(S,y_0), V_g(S,y_0)) \leq \frac{a_5}{N^{\alpha-1}}, \tag{7.18}$$

where a_4, a_5 are positive constants. From (7.15) and (7.18) it follows that, to prove (4.12), it is enough to prove that

$$\sup_{v\in V_{g_N}(S,y_0)} E[d(v, V_{g_N})] \leq \nu_{g_N}(S), \qquad \lim_{S\to\infty} \nu_{g_N}(S) = 0. \tag{7.19}$$

Assume it is not true. Then there exists $\delta > 0$ and sequences $S_i, \lim_{i\to\infty} S_i = 0$, and $\mu_i \in \mathcal{M}(S_i, y_0$ such that

$$E\left[d\left(\int g_N(u,y)\mu_i(du,dy), V_{g_N}\right)\right] \geq \delta, \tag{7.20}$$

with $\mu_i \to \mu^*$ in law as $i \to \infty$. Like in the proof of Theorem 4.2, let $\tilde{\mu}_i$ and $\tilde{\mu}^*$ be $\mathcal{P}(U \times R^m)$-valued random variables defined on a common probability space such that they agree in law with μ_i and μ^*, respectively, and such that (7.2) is satisfied. From (7.2) and Lemma 7.2 it follows that

$$\lim_{i\to\infty} \int g_N(u,y)\tilde{\mu}_i(du,dy) = \int g_N(u,y)\tilde{\mu}^*(du,dy) \in V_{g_N} \quad a.s.,$$

the last inclusion being implied by the fact that $\tilde{\mu}^* \in D$ (which is established similarly to the proof of Theorem 4.2). Hence, $E[d(\int g_N(u,y)\tilde{\mu}^*(du,dy), V_{g_N})] = 0$. This contradicts the following inequalities resulting from (7.20) and the fact that $\tilde{\mu}_i$ and μ_i agree in law:

$$E\left[d\left(\int g_N(u,y)\tilde{\mu}^*(du,dy),\ V_{g_N}\right)\right] = \lim_{i\to\infty} E\left[d\left(\int g_N(u,y)\tilde{\mu}_i(du,dy), V_{g_N}\right)\right] \geq \delta$$

Thus Corollary 4.4 is proved. □

Proof of Corollary 4.5. By (7.15) and (7.18), to prove (4.12), it is sufficient to prove that

$$\sup_{v\in V_{g_N}} d^E(v, V_{g_N}(S,y_0)) \leq \nu_{g_N}(S), \qquad \lim_{S\to\infty} \nu_{g_N}(S) = 0. \tag{7.21}$$

Assume it is not true. Then there exist a number $\delta > 0$ and sequences $\mu_i \in D$ and $S_i, i = 1,2,\ldots, (S_i \to \infty$ as $i \to \infty)$ such that

$$E\left[\left\|\int g_N(u,y)\mu_i(du,dy) - \int g_N(u,y)\mu(du,dy)\right\|\right] \geq \delta \qquad \forall \mu \in \mathcal{M}(S_i,y_0). \tag{7.22}$$

From Theorem 3.4(ii) (see (3.19)) it follows that there exists $\mu^{S_i} \in \mathcal{M}(S_i,y_0)$, such that

$$E[\rho(\mu_i,\mu^{S_i})] \leq 2\tilde{\nu}^{(C,\alpha)}(S_i), \qquad \lim_{i\to\infty} \tilde{\nu}^{(C,\alpha)}(S_i) = 0. \tag{7.23}$$

Without loss of generality, one may assume that $\mu_i \to \mu^*$ and $\mu^{S_i} \to \mu^{**}$ in law. Also, using Skorohod's theorem, one can verify (similar to the way it is done in the proof of Theorem 4.2) that there exist $\mathcal{P}(U \times R^m)$-valued random variables $\tilde{\mu}_i, \tilde{\tilde{\mu}}_i, \tilde{\mu}^*$ and $\tilde{\tilde{\mu}}^{**}$ defined on a common probability space such that they agree in law with μ_i, μ^{S_i}, μ^* and μ^{**} and such that, almost sure, $\tilde{\mu}_i \to \tilde{\mu}^*$, $\tilde{\tilde{\mu}}_i \to \tilde{\tilde{\mu}}^{**}$. Note that $E[\rho(\tilde{\mu}_i,\tilde{\tilde{\mu}}_i)] = E[\rho(\mu_i,\mu^{S_i})]$ and, hence, by (7.23),

$$E[\rho(\tilde{\mu}^*,\tilde{\tilde{\mu}}^{**})] \leq \lim_{i\to\infty} E[\rho(\tilde{\mu}^*,\tilde{\mu}_i)] + \limsup_{i\to\infty} E[\rho(\tilde{\mu}_i,\tilde{\tilde{\mu}}_i)] + \lim_{i\to\infty} E[\rho(\tilde{\tilde{\mu}}_i,\tilde{\tilde{\mu}}^{**})]$$

$$\leq \lim_{i\to\infty} 2\tilde{\nu}^{(C,\alpha)}(S_i) = 0 \qquad \Rightarrow \qquad E[\rho(\tilde{\mu}^*,\tilde{\tilde{\mu}}^{**})] = 0 \qquad \Rightarrow \qquad \tilde{\mu}^* = \tilde{\tilde{\mu}}^{**} \ \ a.s.$$

The latter implies (by virtue of Lemma 7.2) that

$$\lim_{i\to\infty}\int g_N(u,y)\tilde{\mu}_i(du,dy) = \lim_{i\to\infty}\int g_N(u,y)\tilde{\tilde{\mu}}_i(du,dy) = \int g_N(u,y)\tilde{\mu}^*(du,dy) \quad a.s.$$
$$\Rightarrow \quad 0 = \lim_{i\to\infty} E\left[\left\|\int g_N(u,y)\tilde{\mu}_i(du,dy) - \int g_N(u,y)\tilde{\tilde{\mu}}_i(du,dy)\right\|\right]$$
$$= \lim_{i\to\infty} E\left[\left\|\int g_N(u,y)\mu_i(du,dy) - \int g_N(u,y)\mu^{S_i}(du,dy)\right\|\right].$$

These equalities contradict (7.22) and, thus, prove the corollary. □

8. Proofs for section 5.

LEMMA 8.1. *Let the assumptions of Theorem* 5.1 *be satisfied. Then any admissible solution* $(y^\epsilon(\cdot), z^\epsilon(\cdot))$ *of the singularly perturbed CSDE* (5.1) *and* (5.2) *and any admissible solution* $z(\cdot)$ *of the averaged CSDE* (5.4) *satisfy the inequalities*

$$(8.1) \qquad E[||y^\epsilon(t)||^4] \le L \qquad E[||z^\epsilon(t)||^4] \le L \qquad E[||z(t)||^4] \le L \quad \forall t \in [0,T];$$

$$(8.2) \qquad E[||z^\epsilon(t) - z^\epsilon(\theta)||^2] \le L|t-\theta| \qquad E[||z(t)-z(\theta)||^2] \le L|t-\theta| \quad \forall t,\theta \in [0,T],$$

where L *is a positive constant.*

Proof. The proof follows a standard argument based on Lemma 4.12 in [40, p. 125] and an application of the Gronwall–Bellman lemma. □

Proof of Theorem 5.1(i). Let $u^\epsilon(t)$ be an admissible control and $(y^\epsilon(t), z^\epsilon(t))$ be the solution of the singularly perturbed CSDE (5.1) and (5.2) obtained with this control. Divide the interval $[0,T]$ by the points $t_l \overset{def}{=} l\Delta(\epsilon),\ l = 0,1,\dots,N_\epsilon$, where $\Delta(\epsilon)$ is a function of ϵ such that

$$(8.3) \qquad \lim_{\epsilon\to 0}\Delta(\epsilon) = 0, \qquad \lim_{\epsilon\to 0}\frac{\Delta(\epsilon)}{\epsilon} = \infty$$

and N_ϵ is the integer part of $\frac{T}{\Delta(\epsilon)}$. For $l = 1,\dots,N_\epsilon$, define a $\mathcal{P}(U\times\bar{R}^m)$-valued random variable $\bar{\mu}_l$ by

$$(8.4) \quad \int f_i(u,y)\bar{\mu}_l(du,dy) = \frac{1}{\Delta(\epsilon)}\int_{t_{l-1}}^{t_l} f_i(u^\epsilon(t), y^\epsilon(t))dt \ = \frac{1}{S_\epsilon}\int_0^{S_\epsilon} f_i(\bar{u}(\tau),\bar{y}(\tau))d\tau,$$

where $f_i(u,y),\ i = 1,2,\dots$, are as in (2.7) and

$$(\bar{u}(\tau),\bar{y}(\tau)) \overset{\text{def}}{=} (u(t_{l-1}+\epsilon\tau), y(t_{l-1}+\epsilon\tau)), \qquad S_\epsilon \overset{\text{def}}{=} \frac{\Delta(\epsilon)}{\epsilon}.$$

The equations in (8.4) imply that $\bar{\mu}_l$ is the occupational measure generated on the interval $[0,S_\epsilon]$ by the control $\bar{u}(\tau)$ and the corresponding solution $\bar{y}(\tau)$ of the associated system (2.4). Hence, by Theorem 4.2,

$$(8.5) \qquad E[\rho(\bar{\mu}_l, D)] \le \bar{\nu}^{(C,\alpha)}(S_\epsilon) \overset{\text{def}}{=} \nu(\epsilon), \qquad \lim_{\epsilon\to 0}\nu(\epsilon) = 0.$$

For any $\mu', \mu'' \in \mathcal{P}(U\times\bar{R}^m)$, let

$$\hat{\rho}(\mu',\mu'') \overset{\text{def}}{=} \rho(\mu',\mu'') + \left(\sum_{i=1}^{\infty} 2^{-2i}\left|\int f_i(u,y)\mu'(du,dy) - \int f_i(u,y)\mu''(du,dy)\right|^2\right)^{\frac{1}{2}}$$

(8.6)

It is easy to see that $\hat{\rho}(\cdot,\cdot)$ is a metric on $\mathcal{P}(U \times \bar{R}^m)$ and that

$$\rho(\mu', \mu'') \le \hat{\rho}(\mu', \mu'') \le 2\rho(\mu', \mu'') \qquad \forall \mu', \mu'' \in \mathcal{P}(U \times \bar{R}^m). \tag{8.7}$$

The advantage of using $\hat{\rho}(\cdot,\cdot)$ instead of $\rho(\cdot,\cdot)$ is that, for any μ, the solution of the problem $\min_{\mu' \in D} \hat{\rho}(\mu, \mu')$, called the projection of μ onto D, is unique (this being easily verifiable on the basis of the inequality $\hat{\rho}(\mu, (1-\lambda)\mu' + \lambda\mu'') < (1-\lambda)\hat{\rho}(\mu, \mu') + \lambda\hat{\rho}(\mu, \mu'')\ \ \forall \lambda \in (0, 1)$).

Let μ_l stand for the projection of $\bar{\mu}_l$ onto D. By (8.5) and (8.7),

$$E[\hat{\rho}(\bar{\mu}_l, \mu_l)] = E[\hat{\rho}(\bar{\mu}_l, D)] \le 2\nu(\epsilon) \tag{8.8}$$

Using (8.8) and slightly extending arguments in the proof of Corollary 4.5 (to take into account the dependence on z), one can verify that, for any $N > 0$,

$$E[||\tilde{g}(\bar{\mu}_l, z) - \tilde{g}(\mu_l, z)||] \le \nu_N(\epsilon) \quad \forall\ z : ||z|| \le N, \qquad \lim_{\epsilon \to 0} \nu_N(\epsilon) = 0. \tag{8.9}$$

Also, it can be verified that the estimates (8.1) and Assumption 2 (with $\alpha = 4$) imply that

$$E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||^4] \le L_1, \quad E[||\tilde{g}(\mu_l, z^\epsilon(t_l))||^4] \le L_1, \quad L_1\ = \text{const.} \tag{8.10}$$

as well as that

$$E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||\chi_N] \le \kappa(N), \quad E[||\tilde{g}(\mu_l, z^\epsilon(t_l))||\chi_N] \le \kappa(N), \quad \lim_{N \to \infty} \kappa(N) = 0, \tag{8.11}$$

where χ_N is the indicator function of the event: $||z^\epsilon(t_l)|| > N$ ($\bar{\chi}_N$ below will stand for the indicator function of $||z^\epsilon(t_l)|| \le N$). From (8.9) and (8.11) it follows that

$$E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z^\epsilon(t_l))||] \le E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z^\epsilon(t_l))||\bar{\chi}_N]$$

$$+ E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||\chi_N] + E[||\tilde{g}(\mu_l, z^\epsilon(t_l))||\chi_N]\ \le\ \nu_N(\epsilon) + 2\kappa(N),$$

which implies that there exists $\hat{\nu}(\epsilon)$, $\lim_{\epsilon \to 0} \hat{\nu}(\epsilon) = 0$, such that

$$E[||\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z^\epsilon(t_l))||] \le \hat{\nu}(\epsilon). \tag{8.12}$$

This estimate and (8.10) imply, in turn, that

$$E[||\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||^2]$$

$$\le \sqrt{E[||\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||]}\sqrt{E[||\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||^3]}$$

$$\le L_2\sqrt{\hat{\nu}(\epsilon)}, \qquad L_2 = \text{const.}$$

Thus, denoting $\bar{\nu}(\epsilon) \stackrel{\text{def}}{=} L_2\sqrt{\hat{\nu}(\epsilon)}$, one obtains

$$E[||\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\bar{\mu}_l, z^\epsilon(t_l))||^2] \le \bar{\nu}(\epsilon), \qquad \lim_{\epsilon \to 0} \bar{\nu}(\epsilon) = 0. \tag{8.13}$$

Now define the admissible control $\mu(t)$ of the averaged system as follows. On the intervals $[t_0, t_1)$ and $[t_{N_\epsilon}, T]$, take $\mu(t) = \mu$ (an arbitrary element of D). On any other interval $[t_l, t_{l+1}), l = 1, 2, \ldots, N_\epsilon - 1$, take $\mu(t) = \mu_l$. Let $z(t)$ be the solution of the averaged system (5.4) obtained with the control $\mu(t)$. By definition, it satisfies

$$z(t) = z_0 + \int_0^t \tilde{g}(\mu(t'), z(t'))dt' + \int_0^t \sigma(z(t'))dB_2(t').$$

Subtracting this from

$$z^\epsilon(t) = z_0 + \int_0^t g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t'))dt' + \int_0^t \sigma(z^\epsilon(t'))dB_2(t'),$$

one can obtain that

(8.14)

$$E[||z^\epsilon(t) - z(t)||^2] \leq K \left\{ E\left[\left\| \int_0^t g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t'))dt' - \int_0^t \tilde{g}(\mu(t'), z(t'))dt' \right\|^2 \right] + \int_0^t E[||z^\epsilon(t') - z(t')||]^2 dt' \right\},$$

where K is a positive constant. Let us evaluate the first term on the right-hand side of (8.14). Let k_t stand for the integer part of $\frac{t}{\Delta(\epsilon)} (k_t \Delta(\epsilon) \leq t \leq T)$. Then

$$\begin{aligned} &E\left[\left\| \int_0^t g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t'))dt' - \int_0^t \tilde{g}(\mu(t'), z(t'))dt' \right\|^2 \right] \\ &\leq K_1 \Bigg\{ E\Bigg[\Bigg\| \sum_{l=1}^{k_t} \int_{t_{l-1}}^{t_l} (g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t')) - g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t_l)))dt' \\ &+ \sum_{l=1}^{k_t} (\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z^\epsilon(t_l)))\Delta(\epsilon) + \sum_{l=1}^{k_t - 1} (\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z(t_l)))\Delta(\epsilon) \\ &+ \sum_{l=1}^{k_t - 1} \int_{t_l}^{t_{l+1}} (\tilde{g}(\mu(t'), z(t_l)) - \tilde{g}(\mu(t'), z(t')))dt' \Bigg\|^2 \Bigg] + \Delta(\epsilon) \Bigg\} \\ &\leq K_2 \Bigg\{ E\left[\left\| \sum_{l=1}^{k_t} \int_{t_{l-1}}^{t_l} (g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t')) - g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t_l)))dt' \right\|^2 \right] \\ &+ E\left[\left\| \sum_{l=1}^{k_t} (\tilde{g}(\bar{\mu}_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z^\epsilon(t_l))) \right\|^2 + \left\| \sum_{l=1}^{k_t - 1} (\tilde{g}(\mu_l, z^\epsilon(t_l)) - \tilde{g}(\mu_l, z(t_l))) \right\|^2 \right] \Delta^2(\epsilon) \\ &+ E\left[\left\| \sum_{l=1}^{k_t - 1} \int_{t_l}^{t_{l+1}} (\tilde{g}(\mu(t'), z(t_l)) - \tilde{g}(\mu(t'), z(t')))dt' \right\|^2 \right] + \Delta(\epsilon) \Bigg\}, \quad K_1, K_2 = \text{const.} \end{aligned}$$

(8.15)

Using Cauchy–Schwarz inequality (two times), one can obtain that

$$E\left[\left\| \sum_{l=1}^{k_t} \int_{t_{l-1}}^{t_l} (g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t')) - g(u^\epsilon(t'), y^\epsilon(t'), z^\epsilon(t_l)))dt' \right\|^2 \right]$$

$$\leq\ k_t \sum_{l=1}^{k_t} E\left[\left\|\int_{t_{l-1}}^{t_l} (g(u^\epsilon(t'),y^\epsilon(t'),z^\epsilon(t')) - g(u^\epsilon(t'),y^\epsilon(t'),z^\epsilon(t_l)))dt'\right\|^2\right]$$

$$\leq\ k_t\Delta(\epsilon) \sum_{l=1}^{k_t} \int_{t_{l-1}}^{t_l} E\left[\|\ g(u^\epsilon(t'),y^\epsilon(t'),z^\epsilon(t')) - g(u^\epsilon(t'),y^\epsilon(t'),z^\epsilon(t_l))\|^2\right] dt'$$

$$\leq K_3\Delta(\epsilon), \qquad K_3 = \text{const}, \tag{8.16}$$

where, to obtain the last inequality, it has been taken into account that $g(u,y,z)$ satisfies Lipschitz conditions in z and also that, by (8.2), $E[||z^\epsilon(t')-z^\epsilon(t_l)||^2] \leq L\Delta(\epsilon)\ \forall t' \in [t_{l-1},t_l]$. Similarly, using Cauchy–Schwarz inequality and the fact that $\tilde{g}(\mu,z)$ satisfies Lipschitz conditions in z as well as that $E[||z(t')-z(t_l)||^2] \leq L\Delta(\epsilon)\ \forall t' \in [t_l,t_{l+1}]$, one can obtain that

$$E\left[\left\|\sum_{l=1}^{k_t-1}\int_{t_l}^{t_{l+1}}(\tilde{g}(\mu(t'),z(t_l)) - \tilde{g}(\mu(t'),z(t')))dt'\right\|^2\right] \leq\ K_4\Delta(\epsilon),\ \ K_4 = \text{const}. \tag{8.17}$$

Also, by (8.13),

$$E\left[\left\|\sum_{l=1}^{k_t}(\tilde{g}(\bar{\mu}_l,z^\epsilon(t_l)) - \tilde{g}(\mu_l,z^\epsilon(t_l)))\right\|^2\right]\Delta^2(\epsilon)$$

$$\leq k_t \sum_{l=1}^{k_t} E[||\tilde{g}(\bar{\mu}_l,z^\epsilon(t_l)) - \tilde{g}(\mu_l,z^\epsilon(t_l))||^2]\Delta^2(\epsilon)\ \leq\ K_5\bar{\nu}(\epsilon), \quad K_5 = \text{const}, \tag{8.18}$$

and, by (8.1) and (8.2),

$$E\left[\left\|\sum_{l=1}^{k_t-1}(\tilde{g}(\mu_l,z^\epsilon(t_l)) - \tilde{g}(\mu_l,z(t_l)))\right\|^2\right]\Delta^2(\epsilon)$$

$$\leq k_t \sum_{l=1}^{k_t-1} E\left[\|\tilde{g}(\mu_l,z^\epsilon(t_l)) - \tilde{g}(\mu_l,z(t_l))\|^2\right]\Delta^2(\epsilon) \leq K_6 \sum_{l=1}^{k_t-1} E[||z^\epsilon(t_l)-z(t_l)||^2]\Delta(\epsilon)$$

$$\leq K_7\left(\int_0^t E[||z^\epsilon(t')-z(t')||^2]dt' + \Delta^{\frac{1}{2}}(\epsilon)\right), \qquad K_6,K_7 = \text{const}. \tag{8.19}$$

Substitution of (8.16)–(8.19) into (8.15) leads to

$$E\left[\left\|\int_0^t g(u^\epsilon(t'),y^\epsilon(t'),z^\epsilon(t'))dt' - \int_0^t \tilde{g}(\mu(t'),z(t'))dt'\right\|^2\right]$$

$$\leq\ K_8\left(\int_0^t E[||z^\epsilon(t')-z(t')||^2]dt' + \bar{\nu}(\epsilon) + \Delta^{\frac{1}{2}}(\epsilon)\right), \qquad K_8 = \text{const}. \tag{8.20}$$

The latter, in turn, being substituted into (8.14) implies (with the help of a Gronwall–Bellman lemma) the validity of (5.7) with $\tilde{\nu}(\epsilon) = K_9(\bar{\nu}(\epsilon) + \Delta^{\frac{1}{2}}(\epsilon)), K_9 = \text{const}$. The validity of (5.8) follows from the Lipschitz continuity of $G(z)$. ∎

Proof of Theorem 5.1(ii) *(Outline).* Let $S_\epsilon = \frac{\Delta(\epsilon)}{\epsilon}$ (as in the proof of Theorem 5.1(i)) and let

$$\nu(\epsilon) \stackrel{\text{def}}{=} \tilde{\nu}^{(C,\alpha)}(S_\epsilon), \tag{8.21}$$

where $\tilde{\nu}^{(C,\alpha)}(\cdot)$ is the function from the estimate (4.8). Note that $\lim_{\epsilon\to 0}\nu(\epsilon) = 0$. Let J^ϵ be the integer part of $\nu^{-\frac{1}{2}}(\epsilon)$, which implies, in particular, that

$$\lim_{\epsilon\to 0} J^\epsilon = \infty, \qquad \lim_{\epsilon\to 0}(\nu(\epsilon)J^\epsilon) = 0. \tag{8.22}$$

Using the fact that D is compact, one can show that, for any $\epsilon > 0$, there exists a finite subset $D^\epsilon \stackrel{\text{def}}{=} \{\Upsilon_1^\epsilon, \dots, \Upsilon_{J^\epsilon}^\epsilon\}$ of D such that $\rho_H(D^\epsilon, D) \le \delta(\epsilon)$, where $\delta(\epsilon)$ is some function tending to zero as ϵ tends to zero.

It can be verified (by standard applications of a Gronwall–Bellman lemma) that, given a solution $z'(t)$ of the averaged system (5.4) obtained with an arbitrary admissible control $\mu'(t)$, there exists a piecewise constant admissible control $\mu(t)$:

$$\mu(t) = \mu_l \in D^\epsilon \quad \forall t \in [t_{l-1}, t_l), \qquad l = 1, \dots, N_\epsilon \tag{8.23}$$

such that the solution $z(t)$ of the averaged system (5.4), obtained with the use of this control, satisfies the inequality

$$\max_{t\in[0,T]} E[||z'(t) - z(t)||^2] \le \kappa(\epsilon), \qquad \lim_{\epsilon\to 0}\kappa(\epsilon) = 0.$$

Let us show that corresponding to any solution $z(t)$ of the averaged system (5.4) obtained with a control (8.23), there exists an admissible control $u^\epsilon(t)$ the use of which in the singularly perturbed CSDE (5.1) and (5.2) leads to the solution $(y^\epsilon(t), z^\epsilon(t))$ satisfying (5.7).

Take $u^\epsilon(t) = u$ (an arbitrary element of U) on the intervals $[0, t_1)$ and $[t_{N_\epsilon}, T]$ and denote by $(y^\epsilon(t), z^\epsilon(t))$ the solution of (5.1) and (5.2) on the interval $[0, t_1]$ obtained with this control.

From Corollary 4.3 (see (4.8) and the notation (8.21)) it follows that there exist random variables $\bar{\Upsilon}_j^\epsilon \in \mathcal{M}(S_\epsilon, y^\epsilon(t_1))$ such that

$$\frac{E[\rho^2(\bar{\Upsilon}_j^\epsilon, \Upsilon_j^\epsilon)]}{2} \le E[\rho(\bar{\Upsilon}_j^\epsilon, \Upsilon_j^\epsilon)] \le \nu(\epsilon), \quad j = 1, \dots, J^\epsilon, \tag{8.24}$$

where the left inequality is obtained by taking into account that $\frac{\rho(\cdot,\cdot)}{2} \le 1$ (see (2.1)) and, hence, $\frac{\rho^2(\cdot,\cdot)}{2} \le \rho(\cdot,\cdot)$. Define $\bar{\mu}_1$ by

$$\bar{\mu}_1 \stackrel{\text{def}}{=} \sum_{j=1}^{J^\epsilon} \bar{\Upsilon}_j^\epsilon \chi(\mu_1 = \Upsilon_j^\epsilon), \tag{8.25}$$

where $\chi(A)$ is the indicator function of the "event A." By (8.24),

$$E[\rho(\bar{\mu}_1, \mu_1)] = \sum_{j=1}^{J^\epsilon} E[\rho(\bar{\Upsilon}_j^\epsilon, \Upsilon_j^\epsilon)\chi(\mu_1 = \Upsilon_j^\epsilon)]$$

$$\le \sum_{j=1}^{J^\epsilon} \sqrt{E[\rho^2(\bar{\Upsilon}_j^\epsilon, \Upsilon_j^\epsilon)]}\sqrt{E[\chi(\mu_1 = \Upsilon_j^\epsilon)]} \le \sqrt{2\nu(\epsilon)} \sum_{j=1}^{J^\epsilon} \sqrt{E[\chi(\mu_1 = \Upsilon_j^\epsilon)]}$$

$$(8.26) \qquad \le \sqrt{2\nu(\epsilon)}\sqrt{J^{\epsilon}}\sqrt{\sum_{j=1}^{J^{\epsilon}} E[\chi(\mu_1 = \Upsilon_j^{\epsilon})]} = \sqrt{2\nu(\epsilon)}\sqrt{J^{\epsilon}} \stackrel{\text{def}}{=} \nu^*(\epsilon).$$

Note that, as follows from (8.22), $\lim_{\epsilon\to 0} \nu^*(\epsilon) = 0$.

The fact that $\bar{\Upsilon}_j^{\epsilon}$ is an element of $\mathcal{M}(S_{\epsilon}, y^{\epsilon}(t_1))$ implies that there exists an admissible control $\bar{u}_j^{\epsilon}(\tau)$ and the corresponding solution $\bar{y}_j^{\epsilon}(\tau)$ of the associated system with $\bar{y}_j^{\epsilon}(0) = y^{\epsilon}(t_1)$ such that the occupational measure generated by this pair on the interval $[0, S_{\epsilon}]$ coincides with $\bar{\Upsilon}_j$. That is,

$$(8.27) \qquad \frac{1}{S_{\epsilon}}\int_0^{S_{\epsilon}} f_i(\bar{u}_j^{\epsilon}(\tau), \bar{y}_j^{\epsilon}(\tau))d\tau = \int f_i(u,y)\bar{\Upsilon}_j(du,dy) \qquad \forall i = 1, 2, \dots.$$

Now take

$$u^{\epsilon}(t) \stackrel{\text{def}}{=} \sum_{j=1}^{J^{\epsilon}} \bar{u}_j^{\epsilon}\left(\frac{t-t_1}{\epsilon}\right)\chi(\mu_1 = \Upsilon_j^{\epsilon}) \qquad \forall t \in [t_1, t_2)$$

and, using this control, extend the solution $(y^{\epsilon}(t), z^{\epsilon}(t))$ of the CSDE (5.1) and (5.2) to the interval $[t_1, t_2]$. By construction, $y^{\epsilon}(t) = \sum_{j=1}^{J^{\epsilon}} \bar{y}_j^{\epsilon}(\frac{t-t_1}{\epsilon})\chi(\mu_1 = \Upsilon_j^{\epsilon})$ and also (see (8.25) and (8.27))

$$\frac{1}{\Delta(\epsilon)}\int_{t_1}^{t_2} f_i(u^{\epsilon}(t), y^{\epsilon}(t))dt = \sum_{j=1}^{J^{\epsilon}}\left(\frac{1}{S_{\epsilon}}\int_0^{S_{\epsilon}} f_i(\bar{u}_j^{\epsilon}(\tau), \bar{y}_j^{\epsilon}(\tau))d\tau\right)\chi(\mu_1 = \Upsilon_j^{\epsilon})$$

$$= \sum_{j=1}^{J^{\epsilon}}\left(\int f_i(u,y)\bar{\Upsilon}_j^{\epsilon}(du,dy)\right)\chi(\mu_1 = \Upsilon_j^{\epsilon}) = \int f_i(u,y)\bar{\mu}_1(du,dy) \;\forall i = 1, 2, \dots.$$

Continuing in a similar fashion, one can define an admissible control $u^{\epsilon}(t)$ and the corresponding solution $(y^{\epsilon}(t), z^{\epsilon}(t))$ of the CSDE (5.1) and (5.2) such that, on any interval $[t_l, t_{l+1}), l = 1, \dots, N_{\epsilon} - 1$,

$$(8.28) \qquad \frac{1}{\Delta(\epsilon)}\int_{t_l}^{t_{l+1}} f_i(u^{\epsilon}(t), y^{\epsilon}(t))dt = \int f_i(u,y)\bar{\mu}_l(du,dy) \quad \forall i = 1, 2, \dots,$$

where $\bar{\mu}_l$ satisfy the inequalities

$$(8.29) \qquad E[\rho(\bar{\mu}_l, \mu_l)] \le \nu^*(\epsilon) \qquad \forall l = 1, \dots, N_{\epsilon} - 1,$$

with $\nu^*(\epsilon)$ being as in (8.26).

Using arguments similar to the proof of Corollary 4.5, one can verify that (8.29) implies the validity of (8.9) which, in turn, implies the validity of (8.12) and (8.13). The latter leads to the estimate similar to (8.20), which, being substituted into (8.14), leads to (5.7). Due to the Lipschitz continuity of $G(z)$, one can easily derive now that $\limsup_{\epsilon\to 0} G_{\epsilon}^* \le G_{av}^*$, which, along with (5.8), implies (5.9). ∎

Acknowledgment. The authors express their gratitude to Prof. H. Kushner for useful discussions at early stages of this work.

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 (2001/2002), pp. 1159–1188.

[3] O. Alvarez and M. Bardi, *Singular perturbations of nonlinear degenerate parabolic PDEs: A general convergence result*, Arch. Ration. Mech. Anal., 170 (2003), pp. 17–61.

[4] Z. Artstein, *Relaxation of singularly perturbed control systems*, Proceedings of CDC-2002, Control and Decision Conference, Las Vegas, 2002.

[5] Z. Artstein, *An occupational measure solution to a singularly perturbed optimal control problem*, Control Cybernet., 31 (2002), pp. 623–642.

[6] 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.

[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] K. E. Avrachenkov, J. Filar, and M. Haviv, *Singular perturbations of Markov chains and decision processes*, in Handbook of Markov Decision Processes. Methods and Applications, Kluwer Academic Publishers, Boston, 2002, pp. 113–150.

[10] 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.

[11] G. K. Basak, V. S. Borkar, and M. K. Ghosh, *Ergodic control of degenerate diffusions*, Stochastic Anal. Appl., 15 (1997), pp. 1–17.

[12] A. Bensoussan, *Perturbation Methods in Optimal Control*, John Wiley, Chichester, 1988.

[13] A. G. Bhatt and V. S. Borkar, *Occupation measures for controlled Markov processes: Characterization and optimality*, Ann. Probab., 24 (1996), pp. 1531–1562.

[14] V. S. Borkar, *Optimal control of diffusion processes*, Pitman Research Notes in Mathematics Series 203, Longman Scientific and Technical, Harlow, UK, 1989.

[15] V. S. Borkar, *Probability Theory: An Advanced Course*, Springer-Verlag, New York, 1995.

[16] V. S. Borkar, *Stability of annealing schemes and related processes*, Systems Control Lett., 41 (2000), pp. 325–331.

[17] F. Colonius and R. Fabri, *Controllability for systems with slowly varying parameters*, ESAIM Control Optim. Calc. Var., 9 (2003), pp. 207–216.

[18] T. D. 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] T. D. Donchev and A. L. Dontchev, *Singular perturbations in infinite-dimensional control systems*, SIAM J. Control Optim., 42 (2003), pp. 1795–1812.

[20] A. L. Dontchev, T. D. Donchev, and I. Slavov, *A Tikhonov-type theorem for singularly perturbed differential inclusions*, Nonlinear Anal., 26 (1996), pp. 1547–1554.

[21] J. A. Filar, V. Gaitsgory, and A. Haurie, *Control of singularly perturbed hybrid stochastic systems*, IEEE Trans. Automat. Control, 46 (2001), pp. 179–190.

[22] O. P. Filatov and M. M. Hapaev, *Averaging of Systems of Differential Inclusions*, Moscow University Publishing House, Moscow, 1998 (in Russian).

[23] G. B. Folland, *Real Analysis, Modern Techniques and their Applications*, John Wiley, New York, 1984.

[24] V. Gaitsgory, *Control of Systems with Slow and Fast Motions*, Nauka, Moscow, 1991 (in Russian).

[25] V. Gaitsgory, *Suboptimization of singularly perturbed control problem*, SIAM J. Control Optim., 30 (1992), pp. 1228–1249.

[26] V. Gaitsgory, *On a representation of the limit occupational measures set of a control system with applications to singularly perturbed control systems*, SIAM J. Control Optim., 43 (2004), pp. 325–340.

[27] 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.

[28] 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.

[29] V. Gaitsgory and S. Rossomakhine, *Linear programming approach to deterministic long run average problems of optimal control*, SIAM J. Control Optim., accepted.

[30] G. Grammel, *Averaging of singularly perturbed systems*, Nonlinear Anal., 28 (1997), pp. 1851–1865.

[31] G. Grammel, *On nonlinear control systems with multiple time scales*, J. Dynam. Control Systems, 10 (2004), pp. 11–28.

[32] Y. Kabanov and S. Pergamenshchikov, *On convergence of attainability sets for controlled two-scale stochastic linear systems*, SIAM J. Control Optim., 35 (1997), pp. 134–159.

[33] Y. Kabanov and S. Pergamenshchikov, *Two-Scale Stochastic Systems, Asymptotic Analysis and Control*, Springer-Verlag, Berlin, 2003.

[34] R. Z. Khasminskii, *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, Alphen aan den Rijn, Germantown, MD, 1980.

[35] R. Z. Khasminskii and G. Yin, *On averaging principles: An asymptotic expansion approach*, SIAM J. Math. Anal. 35, (2004), pp. 1534–1560.

[36] P. V. Kokotovic, *Applications of singular perturbation techniques to control problems*, SIAM Rev., 26 (1984), pp. 501–550.

[37] P. V. Kokotovic, H. K. Khalil, and J. O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*, Academic Press, London, 1986.

[38] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhauser, Boston, 1990.

[39] 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.

[40] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes* I: *General Theory*, Springer-Verlag, New York, 1977.

[41] S. D. Naidu, *Singular perturbations and time scales in control theory and applications: An overview*, Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms, 9 (2002), pp. 233–278.

[42] 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.

[43] A. A. Pervozvanskii and V. Gaitsgory, *Theory of Suboptimal Decisions*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

[44] V. A. Plotnikov, A. V. Plotnikov, and A. N. Vituk, *Differential Equations with Multivalued Right-Hand Sides: Asymptotic Methods*, AstroPrint, Odessa, Ukraine, 1999 (in Russian).

[45] M. Quincampoix and H. Zhang, *Singular perturbations in non-linear optimal control systems*, Differential Integral Equations, 8 (1995), pp. 931–944.

[46] M. Quincampoix and F. Watbled, *Averaging method for discontinuous Mayer's problem of singularly perturbed control systems*, Nonlinear Anal., 54 (2003), pp. 819–837.

[47] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series 5, Princeton University Press, Princeton, NJ, 1970.

[48] S. P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhauser, Boston, 1994

[49] R. H. Stockbridge, *Time-average control of a martingale problem: Existence of a stationary solution*, Ann. Probab., 18 (1990), pp. 190–205.

[50] V. Veliov, *A generalization of Tikhonov theorem for singularly perturbed differential inclusions*, J. Dynam. Control Systems, 3 (1997), pp. 291–319.

[51] A. Vigodner, *Limits of singularly perturbed control problems with statistical dynamics of fast motions*, SIAM J. Control Optim., 35 (1997), pp. 1–28.

[52] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.

[53] G. G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications. A Singular Perturbation Approach*, Springer-Verlag, New York, 1998.