

ON CONVERGENCE OF ATTAINABILITY SETS FOR CONTROLLED TWO-SCALE STOCHASTIC LINEAR SYSTEMS*

YURI KABANOV[†] AND SERGEI PERGAMENSHCHIKOV[‡]

Abstract. A limit of attainability sets is found for a linear two-scale stochastic system for the case when the diffusion coefficient of the fast variable is of order $\varepsilon^{1/2}$. The attainability set is defined as the set of distributions of attainable terminal values of solutions of stochastic differential equations. As a corollary we calculate a limit of the optimal value of the terminal cost in the stochastic Mayer problem.

Key words. controlled stochastic differential equations, two-scale system, singular perturbations, attainability sets, Mayer problem, Hausdorff metric

AMS subject classifications. 93E20, 93C73

PII. S0363012994269685

Introduction. In mathematical modeling of complex systems with processes having two essentially different “velocities,” fast variables are usually described by singularly perturbed differential equations, i.e., by equations having a small parameter ε on the left-hand side. In general, there is a hope that the reduced limiting model (when the parameter is equal to zero) is more simple and can be used as an approximation of the original one which may be rather complicated. This idea seems to be fruitful also in the set-up of controlled systems. However, here an additional difficulty arises since the optimal value of the cost function which depends smoothly on $\varepsilon \in]0, 1]$ may have a discontinuity at the most interesting point $\varepsilon = 0$.

To overcome this difficulty in the deterministic setting, an approach based on a study of the convergence of the attainability sets in the Hausdorff metric has been developed; see, e.g., recent work [10]. In the linear case it is possible to find a limit of the attainability sets in a rather explicit way which has been done by Dontchev and Veliou [8]; see also the book [7]. Their result is as follows.

Let us consider the controlled system

$$(0.1) \quad \dot{x}_t = A_1(t)x_t + A_2(t)y_t + B_1(t)u_t, \quad x_0 = 0,$$

$$(0.2) \quad \varepsilon \dot{y}_t = A_3(t)x_t + A_4(t)y_t + B_2(t)u_t, \quad y_0 = 0,$$

where ε is a small positive number; u is any measurable function with values in a convex compact subset of \mathbf{R}^d ; matrix-valued functions A_i , B_i are continuous; and the eigenvalues of $A_4(t)$ have strictly negative real parts.

Let $K_\varepsilon(t)$ be the attainability set of the system (0.1), (0.2), i.e., the set of all end points (x_T, y_T) corresponding to various admissible controls, and let $K_0^x(T)$ be the attainability set of the reduced system

$$\dot{x}_t = A_0(t)x_t + B_0(t)u_t, \quad x_0 = 0,$$

with the coefficients $A_0 := A_1 - A_2A_4^{-1}A_3$, $B_0 := B_1 - A_2A_4^{-1}B_2$.

*Received by the editors June 8, 1994; accepted for publication (in revised form) October 18, 1995.

<http://www.siam.org/journals/sicon/35-1/26968.html>

[†]Bilkent University, Bilkent, 06533, Ankara, Turkey, and Central Economics and Mathematics Institute, Krasikova str., 32, Moscow 117433, Russia. Current address: U.F.R. des Sciences, Laboratoire des Mathématiques, 16, Route de Gray, 25030 Besançon Cedex, France (kabanov@vega.univ-fcomte.fr).

[‡]Tomsk State University, Tomsk 634041, Russia.

Let us define the set $K_0(T) := \{(x, y) : x \in K_0^x(T), y \in R(T, x)\}$, where $R(T, x) := -A_4^{-1}(T)A_3(T)x + Y$,

$$Y := \int_0^\infty \exp\{A_4(T)s\}B_2(T)U ds = \left\{y : y = \int_0^\infty \exp\{A_4(T)s\}B_2(T)v_s ds, v_s \in V_U\right\}.$$

V_U is the set of all U -valued Borel functions. In other words, if we put $F(x, y) = (x, -A_4^{-1}(T)A_3(T)x + y)$, then $K_0(T)$ is the image of $K_0^x(T) \times Y$ under the mapping F .

THEOREM (see [8], [7]). *The sets $K_\varepsilon(T)$ tend to $K_0(T)$ in the Hausdorff metric as $\varepsilon \rightarrow 0$.*

Let us consider for the system (0.1), (0.2) the Mayer problem

$$g(x_T, y_T) \rightarrow \min,$$

where g is a continuous function. Then the optimal value for the perturbed problem is

$$J_\varepsilon^* = \min_{K_\varepsilon(T)} g(x, y).$$

From the above theorem it follows immediately that

$$\lim_{\varepsilon \rightarrow 0} J_\varepsilon^* = \min_{K_0(T)} g(x, y).$$

In the paper [13] the authors extended the theorem on the convergence of the attainability sets to stochastic differential equations of the form

$$(0.3) \quad dx_t = (A_1(t)x_t + A_2(t)y_t + B_1(t)u_t)dt + dw_t^x, \quad x_0 = 0,$$

$$(0.4) \quad \varepsilon dy_t = (A_3(t)x_t + A_4(t)y_t + B_2(t)u_t)dt + \sigma(\varepsilon)dw_t^y, \quad y_0 = 0,$$

where w^x, w^y are independent Wiener processes and $\sigma(\varepsilon) = O(\varepsilon^{1/2+\delta})$, $\delta > 0$. In the stochastic setting it is natural to define the attainability set as the set of distributions of all terminal random variables (x_T, y_T) when u runs through the set of admissible controls. There are several possible choices for the latter. It seems that the most adequate one is to consider all nonanticipating functions of the trajectories as admissible controls. This implies the need to understand the system (0.3), (0.4) in the weak sense; i.e., the Wiener processes are not given in advance and the solution is actually a probability measure $P^{\varepsilon, u}$ in the space of continuous functions $C[0, T]$. Such a solution can be constructed by the Girsanov theorem. In this case the attainability set $\mathcal{K}_\varepsilon(T)$ is a compact convex set in the space of probability measures equipped with the Prohorov metric. In [13] it was shown that $\mathcal{K}_\varepsilon(T) \rightarrow \mathcal{K}_0(T)$ in the Hausdorff metric, where $\mathcal{K}_0(T)$ is the set of probability measures μF^{-1} where $\mu = \mu(dx, dy)$ is such that $\mu(dx, \mathbf{R}^n)$ belongs to the attainable set $\mathcal{K}_0^x(T)$ of the reduced system and $\mu(\mathbf{R}^k, dy)$ belongs to the set $\mathbf{P}(Y)$ of probability measures on Y . The reduced system is given by

$$(0.5) \quad dx_t = (A_0(t)x_t + B_0(t)u_t)dt + dw_t^x, \quad x_0 = 0,$$

where, as in the deterministic case, the coefficients A_0 and B_0 can be obtained if we substitute in (0.3) the expression for y_t which is a formal solution of (0.4) with $\varepsilon = 0$.

Notice that the condition $\delta > 0$ provides a limiting degeneracy of the stochastic equation (0.4) (with a fixed control) to an algebraic one.

In the present paper we prove the convergence result for $\sigma(\varepsilon) = \varepsilon^{1/2}$. In this case $\mathcal{K}_0(T)$ is the set of all measures $\mu^{F^{-1}}$ such that $\mu(dx, \mathbf{R}^n) \in \mathcal{K}_0^x(T)$ and $\mu(x, dy)$ belong to the convex closure of the set of probability distributions of random variables

$$\xi_0 + \int_0^\infty \exp\{A_4(T)s\} B_2(T) v_s ds,$$

where ξ is the stationary Gaussian Markov process (called also Ornstein–Uhlenbeck) with the zero mean and covariance

$$K(s, t) := \Xi \exp\{A_4'(T)(t - s)\}, \quad s \leq t,$$

$$\Xi := \int_0^\infty \exp\{A_4(T)s\} \exp\{A_4'(T)s\} ds,$$

v is any measurable process with values in U such that for any t the random variable v_t is measurable with respect to the σ -algebra $\mathcal{F}_{\geq t}^\xi := \sigma\{\xi_s, s \geq t\}$, and prime denotes the matrix transpose. As a corollary of the theorem on convergence of the attainability sets we calculate a limit of the optimal value in the Mayer problem $Eg(x_T^{\varepsilon, u}, y_T^{\varepsilon, u}) \rightarrow \min$ when ε tends to zero.

In the last few years singularly perturbed controlled stochastic differential equations have been intensively studied by various methods, mainly based on the theory of weak convergence in the functional spaces or the Bellman–Hamilton–Jacobi equation; see monographs [3], [4], [20] and papers [2], [5] (and the collection [17] for early results). However, almost all studies concern models where the controlled fast variable does not affect the terminal cost. Harold Kushner wrote in his book [20, p. 64]:

It is hard to deal in any general way with the case where the fast system is also controlled. The main difficulty is due to the fact that the ‘stationary measures’ which are used to average out the fast variable depend on the control which is used in the fast system. This makes it hard to define the ‘averaged problem.’... Similar problems occur in the deterministic case, and it is commonly dealt with there by supposing that the choice of control for the fast system does not alter the steady state value of that system, for each value of the fast variable, i.e., that the fast system is asymptotically stable and the control chosen in a class such that the limit point of that fast system does not depend on the control *when x is fixed*. This assumption essentially ‘decouples’ the fast and slow system. The assumption seems reasonable and yields good results. Unfortunately, it does not seem possible to find a stochastic analog of this approach which works in any generality.

It worth noticing that the result presented here is nontrivial even for a system with only fast variables. In this case it is clear that the limit of the attainability sets shows to what extent optimal controls (acting on the drift of the process) can follow the change in the scale parameter near the point zero.

The structure of the paper is the following. In section 1 we give the formal description of the problem. Section 2 contains some preliminary explanations and the proof of the result for the simplest one-dimensional model with the fast variable only. The proof of Theorem 1.1 is given in sections 3 and 4. Section 5 is devoted to measure-theoretical aspects which may have some independent interest.

1. Formulations of the results. We consider here the linear stochastic controlled system given by

$$(1.1) \quad dx_t = (A_1(t)x_t + A_2(t)y_t + B_1(t)u_t)dt + dw_t^x, \quad x_0 = 0,$$

$$(1.2) \quad \varepsilon dy_t = (A_3(t)x_t + A_4(t)y_t + B_2(t)u_t)dt + \sqrt{\varepsilon}dw_t^y, \quad y_0 = 0,$$

where w^x and w^y are standard independent Wiener processes with values in \mathbf{R}^k and \mathbf{R}^n , $0 \leq t \leq T < \infty$, $\varepsilon \in]0, 1]$.

We shall understand (1.1), (1.2) as a symbolic notation for the stochastic differential equation in a weak sense when a Wiener process $W = (w^x, w^y)$ is not given in advance and u is a feedback control. Actually, in the following rigorous formulation we could avoid the above representation (which is, in fact, a bit ambiguous) altogether.

We consider as a phase space $\mathbf{R}^m = \mathbf{R}^k \times \mathbf{R}^n$. (\mathbf{R}^k corresponds to the slow and \mathbf{R}^n to the fast variables.) The phase space of control will be a compact convex set $U \subseteq \mathbf{R}^d$. In our matrix notations vectors are column vectors.

The path space of the system is the space $C[0, T]$ of continuous functions $W : [0, T] \rightarrow \mathbf{R}^m$. Let \mathcal{C}_T be the Borel σ -algebra on $C[0, T]$, $\mathcal{C}_t^o := \sigma\{W_s, s \leq t\}$, $\mathcal{C}_t := \mathcal{C}_{t+}^o$. Let \mathcal{P} be the predictable σ -algebra in $C[0, T] \times [0, T]$ corresponding to the filtration $\mathbf{C} = (\mathcal{C}_t)$.

The class of admissible controls \mathcal{U} is defined as the set of all predictable processes $u = (u_t)_{t \in [0, T]}$ with values in U .

Let $A_i = A_i(t)$, $B_i = B_i(t)$ be matrix-valued continuous functions of dimensions compatible with (1.1), (1.2); i.e., $A_1(t)$ is a $k \times k$ matrix, $A_4(t)$ is $n \times n$, etc.

We introduce the following notation:

$$(1.3) \quad f_\varepsilon(W, t, u) = \begin{pmatrix} A_1(t) & A_2(t) \\ \varepsilon^{-1}A_3(t) & \varepsilon^{-1/2}A_4(t) \end{pmatrix} W_t + \begin{pmatrix} B_1(t) \\ \varepsilon^{-1}B_2(t) \end{pmatrix} u_t,$$

$$(1.4) \quad D_\varepsilon := \begin{pmatrix} I_k & 0 \\ 0 & \varepsilon^{-1}I_n(t) \end{pmatrix},$$

where I_k, I_n are the identity matrices of corresponding dimensions.

Consider on $(C[0, T], \mathcal{C}_T)$ the probability measure P^ε such that with respect to P^ε the coordinate process W is the Wiener process with the correlation matrix $D_\varepsilon D_\varepsilon'$.

For any admissible control u we define the measure $P^{\varepsilon, u} := \rho_T^\varepsilon(u)P^\varepsilon$ with

$$(1.5) \quad \rho_T^\varepsilon(u) = \exp \left\{ \int_0^T f_\varepsilon(W, s, u_s)' dW_s - \frac{1}{2} \int_0^T |f_\varepsilon(W, s, u_s)' D_\varepsilon|^2 ds \right\}.$$

It is well known (see [1] or [16]) that $P^{\varepsilon, u}$ is a probability measure. By the Girsanov theorem the process

$$W_t - \int_0^t f_\varepsilon(W, s, u_s) ds$$

with respect to $P^{\varepsilon, u}$ is the Wiener process with the correlation matrix $D_\varepsilon D_\varepsilon'$. Thus, we can write that

$$dW_t = f_\varepsilon(W, t, u_t)dt + D_\varepsilon dB_t, \quad W_0 = 0,$$

where B is the standard Wiener process.

If we denote the first k components of W and B by x and w^x and the remaining n components by y and w^y , the above representation formally coincides with the system (1.1), (1.2) and the control u will be a nonanticipating functional of the phase trajectory. This explains the terminology where $P^{\varepsilon, u}$ is called a weak solution of (1.1), (1.2) and the model itself usually is referred to as the model with the feedback control.

Let $\mathcal{K}_\varepsilon := \{P^{\varepsilon, u} : u \in \mathcal{U}\}$, where $\varepsilon > 0$ is fixed. The set \mathcal{K}_ε is an analog of the “tube” of trajectories for deterministic systems. Correspondingly, the attainability set $\mathcal{K}_\varepsilon(T) := \{P^{\varepsilon, u} W_T^{-1} : u \in \mathcal{U}\}$ is the set of all probability measures on \mathbf{R}^m which are the images of elements of \mathcal{K}_ε under the mapping $W \mapsto W_T$. It was proved in [1] that \mathcal{K}_ε is a convex set, hence $\mathcal{K}_\varepsilon(T)$ is also convex. In [1] it was also shown that the set $\{\rho_T^\varepsilon(u) : u \in \mathcal{U}\}$ of the attainable densities is sequentially compact in the weak topology of $L^1(P^\varepsilon)$. It follows immediately that \mathcal{K}_ε and $\mathcal{K}_\varepsilon(T)$ are compact subsets of the corresponding spaces of probability measures $\mathbf{P}(C[0, T])$ and $\mathbf{P}(\mathbf{R}^m)$ equipped with the Prohorov metric.

To formulate the convergence result we need the following assumption.

(A) For all t the real parts of the eigenvalues of $A_4(t)$ have strictly negative real parts:

$$(1.6) \quad \operatorname{Re} \lambda(A_4(t)) \leq -2\kappa < 0.$$

Let $\mathcal{K}_0^x(T)$ be the attainability set of the stochastic differential equation

$$(1.7) \quad dx_t = (A_0(t)x_t + B_0(t)u_t)dt + dw_t^x, \quad x_0 = 0,$$

where $A_0 := A_1 - A_2 A_4^{-1} A_3$, $B_0 := B_1 - A_2 A_4^{-1} B_2$.

Let ξ be the (strong) solution of the following stochastic differential equation with constant coefficients on some filtered probability space $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t), P)$:

$$(1.8) \quad d\xi_t = A_4(T)\xi_t dt + db_t, \quad \xi_0 = \xi^o,$$

where b is a standard Wiener process in \mathbf{R}^n and ξ^o is an independent Gaussian random variable with the zero mean and covariance matrix

$$(1.9) \quad \Xi := \int_0^\infty \exp\{A_4(T)s\} \exp\{A_4'(T)s\} ds.$$

In other words, ξ is the stationary Gaussian Markov process with zero mean and covariance function

$$(1.10) \quad K(s, t) := E\xi_s x_t' = \Xi \exp\{A_4'(T)(t - s)\};$$

see, e.g., [16].

Let \mathcal{V}_U be the set of all U -valued processes $v = (v_t)_{t \geq 0}$ such that $v_{1/t}$ is a predictable process with respect to the filtration generated by the process $\xi_{1/t}$, $S_Y^o := \{\mathcal{L}(\xi_0 + I(v)) : v \in \mathcal{V}_U\}$, where

$$(1.11) \quad I(v) := \int_0^\infty \exp\{A_4(T)s\} B_2(T) v_s ds.$$

Here and in what follows we use the notation $\mathcal{L}(\eta) := P\eta^{-1}$ for the distribution of the random variable η . The set S_Y^o is compact in $\mathbf{P}(\mathbf{R}^n)$; see Lemma 5.5.

Put $S_Y := \overline{\operatorname{conv}} S_Y^o$, the convex closure of S_Y^o in $\mathbf{P}(\mathbf{R}^n)$.

Let S be the set of all probability measures $\mu = \mu(dx, dy)$ on $\mathbf{R}^m = \mathbf{R}^k \times \mathbf{R}^n$ such that

- (1) $\mu(x, dy) \in S_Y$;
- (2) $\mu(dx, \mathbf{R}^n) \in \mathcal{K}_0^x(T)$.

From the Proposition 5.2 it follows that S is compact in $\mathbf{P}(\mathbf{R}^m)$.

Define a linear mapping $F(x, y) := (x, -A_4^{-1}(T)A_3(T)x + y)$ of \mathbf{R}^m into itself. Put $\mathcal{K}_0(T) := \{\mu F^{-1} : \mu \in S\}$.

Our main result is the following theorem.

THEOREM 1.1. *The set $\cup_{\varepsilon \in]0, 1]} \mathcal{K}_\varepsilon(T)$ is compact, and as $\varepsilon \rightarrow 0$, $\mathcal{K}_\varepsilon(T)$ tend to $\mathcal{K}_0(T)$ in the Hausdorff metric in the space of compact subsets of $\mathbf{P}(\mathbf{R}^m)$.*

For the model (1.1), (1.2) we consider now the Mayer problem, which can be rigorously formulated as the problem to determine the minimal value of the functional

$$(1.12) \quad J_\varepsilon^* := \inf_{u \in \mathcal{U}} E^{\varepsilon, u} g(W_T) = \inf_{\mu \in \mathcal{K}_\varepsilon(T)} \int g(x, y) \mu(dx, dy),$$

where g is a function on \mathbf{R}^m which is integrable with respect to the measures μ from $\mathcal{K}_\varepsilon(T)$.

COROLLARY 1.1. *Assume that g is continuous and bounded. Then*

$$(1.13) \quad \lim_{\varepsilon \rightarrow 0} J_\varepsilon^* = \inf_{\mu \in \mathcal{K}_0(T)} \int g(x, y) \mu(dx, dy).$$

Remark 1.1. The definition of the set \mathcal{V}_U seems rather complicated. Essentially, \mathcal{V}_U contains measurable processes v such that for any t the random variable v_t is measurable with respect to the σ -algebra $\mathcal{F}_{\geq t}^\xi := \sigma\{\xi_s, s \geq t\}$. To avoid a discussion of the measurable structures related to a decreasing family of σ -algebras we prefer to consider the processes in reversed time.

Remark 1.2. There is an alternative description of the set S_Y . Let α be a random variable independent of ξ with values in some Polish space and with a nonatomic distribution. Define the set \mathcal{V}_U^α as the set of all U -valued processes $v = (v_t)_{t \geq 0}$ such that $v_{1/t}$ is a predictable process with respect to the filtration generated by the process $\xi_{1/t}$ and the random variable α . Then $S_Y = \{\mathcal{L}(\xi_0 + I(v)) : v \in \mathcal{V}_U^\alpha\}$; see section 5.

Remark 1.3. Evidently, Theorem 1.1 can be applied to the more general optimization problem $J^\varepsilon(u) = F(P^{\varepsilon, u}) \rightarrow \min$, where F is any continuous function on $\mathbf{P}(\mathbf{R}^m)$.

We also use in our proof another possible model based on a different (and more traditional) interpretation of the equations (1.1), (1.2). To describe this alternative approach we consider the standard Wiener measure P on $(C[0, T], \mathcal{C}_T)$. Let w^x be the notation for the first k coordinates of the function W and w^y be the notation for the remaining n coordinates. Then for any $u \in \mathcal{U}$ we can find the strong solution $X^{\varepsilon, u} = (x^{\varepsilon, u}, y^{\varepsilon, u})$ of (1.1), (1.2). This model is referred to as the model with the open loop controls (since in this case u is a nonanticipating functional of the “noise”).

Let $P_X^{\varepsilon, u} := P(X^{\varepsilon, u})^{-1}$ be the distribution in $C[0, T]$ of the process $X^{\varepsilon, u}$. Certainly, the measure $P_X^{\varepsilon, u}$ need not be equal to $P^{\varepsilon, u}$. Let us consider the sets $\tilde{\mathcal{K}}_\varepsilon := \{P_X^{\varepsilon, u} : u \in \mathcal{U}\} \subseteq \mathbf{P}(C[0, T])$ and $\tilde{\mathcal{K}}_\varepsilon(T) := \{P(X_T^{\varepsilon, u})^{-1} : u \in \mathcal{U}\} \subseteq \mathbf{P}(\mathbf{R}^m)$. We do not know whether the attainability set $\tilde{\mathcal{K}}_\varepsilon(T)$ coincides with the attainability set $\mathcal{K}_\varepsilon(T)$. However, in our paper [13] it has been shown that there are dense embeddings $\tilde{\mathcal{K}}_\varepsilon \subseteq \mathcal{K}_\varepsilon$ and $\tilde{\mathcal{K}}_\varepsilon(T) \subseteq \mathcal{K}_\varepsilon(T)$ in the sense of total variation convergence (thus, in the weak topology) and that the inclusion $\tilde{\mathcal{K}}_\varepsilon \subseteq \mathcal{K}_\varepsilon$ is strict even in the simplest cases.

This fact, certainly, does not exclude the coincidence of $\tilde{\mathcal{K}}_\varepsilon(T)$ and $\mathcal{K}_\varepsilon(T)$. Nevertheless, the result that there is a dense embedding $\tilde{\mathcal{K}}_\varepsilon(T) \subseteq \mathcal{K}_\varepsilon(T)$ is very helpful since it permits us to apply pathwise techniques similar to that of the deterministic theory.

2. Main ideas and the proof of Theorem 1.1 in the simplest case. We recall some basic facts concerning the Hausdorff metric and convergence of compact sets (for details see, e.g., [11]).

Let (X, d) be a metric space and let \mathbf{K}_X be the class of all its nonempty compact subsets. For $A, B \in \mathbf{K}_X$ put $l(A, B) := \sup_{z \in A} d(z, B)$. The Hausdorff distance between A and B is defined by the equality

$$d_H(A, B) := l(A, B) \vee l(B, A).$$

If $A_m \in \mathbf{K}_X$, $m \in \mathbf{Z}_+$, and all A_m are contained in some compact set, then $\lim d_H(A_m, A_0) = 0$ if and only if the following two much more tractable conditions are satisfied for any subsequences of indices (n) :

- (1) For any convergent sequence $z_n \in A_n$ its limit is a point in A_0 .
- (2) For any point $z \in A_0$ there exists a subsequence $z_{n_k} \in A_{n_k}$ converging to z .

Notice that if A_n are not subsets of some compact set, the above equivalence fails in general. For the subsets of the real line $A_n := [0, 1] \cup \{n\}$, conditions (1) and (2) are satisfied but A_n do not tend to A_0 in the Hausdorff metric.

The strategy of the proof of Theorem 1.1 is the following. In the first stage we show that for any $\mu_\varepsilon \in \mathcal{K}_\varepsilon(T)$, $\varepsilon \in]0, 1]$, there exists $\bar{\mu}_\varepsilon \in \mathcal{K}_0(T)$ such that $d(\bar{\mu}_\varepsilon, \mu_\varepsilon) \rightarrow 0$ (d here is the Prohorov metric). Since all $\mathcal{K}_\varepsilon(T)$ are compact this implies that $\cup_{\varepsilon > 0} \mathcal{K}_\varepsilon(T)$ is compact and all limit points of $\{\mu_\varepsilon\}$ belongs to $\mathcal{K}_0(T)$; i.e., (1) is fulfilled. Since $\tilde{\mathcal{K}}_\varepsilon(T)$ is dense in $\mathcal{K}_\varepsilon(T)$ it is sufficient to consider only the case when $\mu_\varepsilon \in \tilde{\mathcal{K}}_\varepsilon(T)$. Thus, we can argue with terminal random variables $(x_T^{\varepsilon, u}, y_T^{\varepsilon, u})$ with the distributions μ_ε and approximate them in probability (or in L^p) by random variables $(\bar{x}_T^{\varepsilon, u}, \bar{y}_T^{\varepsilon, u})$ with distributions from $\mathcal{K}_0(T)$.

In the second step of the proof we should find for a given measure $\mu \in \mathcal{K}_0(T)$ the sequence of measures μ_n which are elements of $\tilde{\mathcal{K}}_{\varepsilon_n}(T)$ converging to μ . Again we shall argue with suitably chosen random variables with distributions corresponding to the measures for which we are looking.

Since the proof for the general multidimensional two-scale system requires rather long arguments, we clarify main ideas on the example of a one-dimensional model with constant coefficients and containing only the fast variable.

Let us consider the controlled stochastic differential equation

$$(2.1) \quad \varepsilon dy_t^{\varepsilon, u} = (-\gamma y_t^{\varepsilon, u} + u_t)dt + \varepsilon^{1/2} dw_t^y, \quad y_0 = 0,$$

where u is a predictable process which takes values in $U = [0, 1]$. In this case the set $\mathcal{K}_0(T)$ is the convex closure of the set $\{\mathcal{L}(\xi_0 + I(v)), v \in \mathcal{V}_U\}$, where

$$I(v) := \int_0^\infty e^{-\gamma s} v_s ds,$$

ξ is an Ornstein–Uhlenbeck process on some probability space (Ω, \mathcal{F}, P) with correlation function $K(s, t) = (2\gamma)^{-1} e^{-\gamma|t-s|}$, and \mathcal{V}_U is the set of all U -valued processes v such that $v_{1/t}$ is a predictable process with respect to the filtration generated by the process $\xi_{1/t}$. For our purpose it is more convenient to use the alternative description of $\mathcal{K}_0(T)$ as the set $\{\mathcal{L}(\xi_0 + I(v)), v \in \mathcal{V}_U^\alpha\}$, where α is a random variable independent of ξ with values in a Polish space and nonatomic distribution and \mathcal{V}_U^α is the set of

all U -valued processes v such that $v_{1/t}$ is a predictable process with respect to the filtration generated by the process $\xi_{1/t}$ and the random variable α . We understand the equation (2.1) in the strong sense. Its solution can be represented in the following way:

$$(2.2) \quad y_t^{\varepsilon,u} = \varepsilon^{-1} \int_0^t e^{-\gamma(t-s)/\varepsilon} u_s ds + \eta_t^\varepsilon,$$

where

$$(2.3) \quad \eta_t^\varepsilon := \varepsilon^{-1/2} \int_0^t e^{-\gamma(t-s)/\varepsilon} dw_s^y.$$

Put $T_\varepsilon := T(1 - \varepsilon^{1/2})$. Let us consider on the interval $[T_\varepsilon, T]$ the Gaussian stationary process

$$\tilde{\xi}_t^\varepsilon := (2\gamma)^{-1/2} \exp\{-\gamma(t - T_\varepsilon)/\varepsilon\} \beta + \varepsilon^{-1/2} \int_{T_\varepsilon}^t e^{-\gamma(t-s)/\varepsilon} dw_s^y,$$

where β is a standard normal random variable independent of the Wiener process w^y (to define β we can extend our canonical coordinate probability space). The process ξ^ε is the solution of the linear equation

$$\varepsilon d\tilde{\xi}_t^\varepsilon = -\gamma \tilde{\xi}_t^\varepsilon dt + \varepsilon^{1/2} dw_t^y, \quad \tilde{\xi}_{T_\varepsilon}^\varepsilon = (2\gamma)^{-1/2} \beta.$$

Let us consider the Ornstein–Uhlenbeck process $\xi_t^\varepsilon = \tilde{\xi}_{T-\varepsilon t}^\varepsilon$, $t \in [0, T/\sqrt{\varepsilon}]$.

Evidently, $\eta_T^\varepsilon - \xi_0^\varepsilon = \eta_T^\varepsilon - \tilde{\xi}_T^\varepsilon \rightarrow 0$ in L^2 as $\varepsilon \rightarrow 0$.

For $u \in \mathcal{U}$ we define the process $v_s = v_s^\varepsilon := u_{T-\varepsilon s} I_{[0, T/\sqrt{\varepsilon}]}$.

Now we can write that

$$y_T^{\varepsilon,u} = \eta_T^\varepsilon + \int_0^{T/\sqrt{\varepsilon}} e^{-\gamma s} u_{T-\varepsilon s} ds + \int_{T/\sqrt{\varepsilon}}^{T/\varepsilon} e^{-\gamma s} u_{T-\varepsilon s} ds = \tilde{y}_T^{\varepsilon,u} + R^\varepsilon(u),$$

where $\tilde{y}_T^{\varepsilon,u} = \xi_0^\varepsilon + I(v)$,

$$R^\varepsilon(u) := \int_{T/\sqrt{\varepsilon}}^{T/\varepsilon} e^{-\gamma s} u_{T-\varepsilon s} ds + \eta_T^\varepsilon - \xi_0^\varepsilon.$$

Since $\sup_{u \in \mathcal{U}} |R^\varepsilon(u)| \rightarrow 0$ in probability, to accomplish the first step we need to check only that $\mathcal{L}(\xi_0^\varepsilon + I(v)) \in \mathcal{K}_0(T)$. Indeed, let us take for ξ the process ξ^ε defined above. For any $s \leq T/\sqrt{\varepsilon}$ the random variable v_s is measurable with respect to the σ -algebra $\mathcal{C}_{T-\varepsilon s}$. But

$$\begin{aligned} \mathcal{C}_{T-\varepsilon s} &= \sigma\{w_r, r \leq T_\varepsilon\} \vee \sigma\{w_r, T_\varepsilon \leq r \leq s\} \subseteq \sigma\{w_r, r \leq T_\varepsilon\} \vee \sigma\{\tilde{\xi}_r^\varepsilon, T_\varepsilon \leq r \leq s\} \\ &= \sigma\{w_r, r \leq T_\varepsilon\} \vee \sigma\{\xi_r^\varepsilon, s \leq r \leq T/\sqrt{\varepsilon}\}, \end{aligned}$$

and we see that $v \in \mathcal{V}_U^\alpha$ where the random variable α is defined as the projection mapping of $C[0, T]$ onto $C[0, T_\varepsilon]$. The above considerations show that the limit of any convergent sequence $\mu^n \in \mathcal{K}_{\varepsilon_n}(T)$ is an element of $\mathcal{K}_0(T)$.

Now we introduce the set $\mathcal{V}_U^{\alpha'}$ consisting of all processes

$$(2.4) \quad v_s = \sum_{i=1}^N \varphi_i I_{[s_i, s_{i+1}]}(s) + u^0 I_{[s_{N+1}, \infty)}(s),$$

where $0 = s_1 < \dots < s_{N+1}$, $u^0 \in U$, and the U -valued random variables φ_i have the form

$$(2.5) \quad \varphi_i = f_i(\alpha, \xi(r_1^i), \dots, \xi(r_{M_i}^i)), \quad s_{i+1} < r_j^i \leq s_N.$$

Let $\mathcal{K}'_0(T) := \{\mathcal{L}(\xi_0 + I(v)), v \in \mathcal{V}'_U\}$. It is easy to show that the set $\{I(v), v \in \mathcal{V}'_U\}$ is dense in $\{I(v), v \in \mathcal{V}_U\}$ in probability. Thus, $\mathcal{K}'(T)$ is dense in $\mathcal{K}_0(T)$ in $\mathbf{P}(\mathbf{R})$.

Let $\mu \in \mathcal{K}'(T)$. This means that μ is the distribution of a random variable $\chi := \xi_0 + I(v)$ where v is of the form (2.4). The result will be proved if we construct a random variable χ^ε and a control u^ε such that $\mathcal{L}(\chi^\varepsilon) = \mathcal{L}(\chi)$ and $\chi^\varepsilon - y_T^{u^\varepsilon, \varepsilon} \rightarrow 0$ in probability. To this aim it is enough to find on the coordinate probability space $(C[0, T], \mathcal{C}, P)$ a stationary Gaussian Markov process ξ^ε with correlation function $K(s, t)$, a standard normal random variable α^ε independent on ξ^ε , and an admissible control $u^\varepsilon \in \mathcal{U}$ such that $\xi_0^\varepsilon - \eta_T^\varepsilon \rightarrow 0$ in probability (η_T^ε is defined by (2.3)), and

$$\int_0^\infty e^{-\gamma s} v_s^\varepsilon ds - \varepsilon^{-1} \int_0^T e^{-\gamma(T-s)/\varepsilon} u_s^\varepsilon ds \rightarrow 0,$$

where v^ε is the process given by the formula (2.4) if we substitute ξ^ε , φ^ε , and α^ε for ξ , φ , and α . Indeed, in this case the random variable $\chi^\varepsilon := \xi_0^\varepsilon + I(v^\varepsilon)$ meets the required properties.

The process ξ^ε can be constructed in the following way. For sufficiently small ε let $T_\varepsilon^k := T(1 - k\varepsilon^{1/2})$, $k = 1, 2, 3$. Put

$$\begin{aligned} \alpha^\varepsilon &:= (w_{T_\varepsilon^2} - w_{T_\varepsilon^3}) / (T_\varepsilon^2 - T_\varepsilon^3)^{1/2}, \\ \beta^\varepsilon &:= (2\gamma)^{-1/2} (w_{T_\varepsilon^1} - w_{T_\varepsilon^2}) / (T_\varepsilon^1 - T_\varepsilon^2)^{1/2}, \\ \tilde{\xi}_t^\varepsilon &:= \exp\{(t - T_\varepsilon^1)/\varepsilon\} \beta^\varepsilon + \varepsilon^{-1/2} \int_{T_\varepsilon^1}^t e^{-\gamma(t-s)/\varepsilon} dw_s, \quad t \geq T_\varepsilon^1. \end{aligned}$$

Define the process ξ^ε on $[0, \varepsilon^{-1/2}T]$ by the equality $\xi_t^\varepsilon := \tilde{\xi}_{T-\varepsilon t}^\varepsilon$.

Evidently,

$$\xi_0^\varepsilon - \eta_T^\varepsilon = \exp\{(T - T_\varepsilon^1)/\varepsilon\} \beta^\varepsilon - \varepsilon^{-1/2} \int_0^{T_\varepsilon^1} e^{-\gamma(T-s)/\varepsilon} dw_s \rightarrow 0 \text{ in } L^2.$$

For sufficiently small ε we put

$$u^\varepsilon := u^0 I_{[0, t_{N+1}[} + \sum_{i=1}^{N+1} \varphi_i^\varepsilon I_{[t_{i+1}, t_i[},$$

where $t_i := T - \varepsilon s_i$, $i \leq N + 1$.

The random variables φ_i^ε are $\mathcal{C}_{t_{i+1}}$ -measurable. Thus, $u^\varepsilon \in \mathcal{U}$. It follows that

$$\begin{aligned} \int_0^\infty e^{-\gamma s} v_s^\varepsilon ds - \varepsilon^{-1} \int_0^T e^{-\gamma(T-s)/\varepsilon} u_s^\varepsilon ds &= \int_0^\infty e^{-\gamma s} v_s^\varepsilon ds - \int_0^{T/\varepsilon} e^{-\gamma s} u_{T-\varepsilon s}^\varepsilon ds \\ &= \int_{T/\varepsilon}^\infty e^{-\gamma s} v_s^\varepsilon ds \rightarrow 0. \end{aligned}$$

The proof of the result for this particular case is finished.

3. Proof of Theorem 1.1. Part 1. We use the notation $\|f\|_t := \sup_{s \leq t} |f_s|$ (omitting the subscript $t = T$) and denote by C different constants which do not depend on ε and u .

In the following statements the solution of (1.1), (1.2) (as well as that of (3.1)) is understood in the strong sense as given on the probability space $(C[0, T], \mathcal{C}_T, P)$.

PROPOSITION 3.1. *Let $(x_T^{\varepsilon,u}, y_T^{\varepsilon,u})$ be the solution of (1.1), (1.2) corresponding to some $u \in \mathcal{U}$, and let \bar{x}^u be the solution of the reduced equation*

$$(3.1) \quad d\bar{x}_t^u = (A_0(t)\bar{x}_t^u + B_0(t)u_t)dt + dw_t^x, \quad \bar{x}_0^u = 0.$$

Then for any $p \in [1, \infty[$

$$(3.2) \quad \sup_{\varepsilon} \sup_{u \in \mathcal{U}} E \|x^{\varepsilon,u}\|^p < \infty,$$

$$(3.3) \quad \lim_{\varepsilon \rightarrow 0} \sup_{u \in \mathcal{U}} E \|x^{\varepsilon,u} - \bar{x}^u\|^p = 0,$$

$$(3.4) \quad \sup_{\varepsilon} \sup_{u \in \mathcal{U}} \sup_{t \leq T} E |y_t^{\varepsilon,u}|^p < \infty.$$

Proof. Let us introduce for $\varepsilon^{-1}A_4(t)$ the fundamental matrix $\Psi^\varepsilon(t, s)$, which is the solution of the linear matrix equation

$$(3.5) \quad \frac{\partial \Psi^\varepsilon(t, s)}{\partial t} = \varepsilon^{-1}A_4(t)\Psi^\varepsilon(t, s), \quad \Psi^\varepsilon(s, s) = I_n.$$

Since A_4 is continuous and the eigenvalues satisfy (1.6), there exists a constant L such that

$$(3.6) \quad |\Psi^\varepsilon(t, s)| \leq Le^{-\kappa(t-s)/\varepsilon}$$

for all $s \leq t \leq T$ and $\varepsilon \in]0, 1]$; see, e.g., [18]. In particular, from the above bound it follows that for all $t \leq T$ and $\varepsilon \in]0, 1]$

$$(3.7) \quad \frac{1}{\varepsilon} \int_0^t |\Psi^\varepsilon(t, s)| ds \leq L/\kappa.$$

Using the fundamental matrix, the equation (1.2) can be solved with respect to $y = y^{\varepsilon,u}$ and we get the representation

$$(3.8) \quad y_t^{\varepsilon,u} = \frac{1}{\varepsilon} \int_0^t \Psi^\varepsilon(t, s)[A_3(s)x_s^{\varepsilon,u} + B_2(s)u_s] ds + \eta_t^\varepsilon,$$

where

$$(3.9) \quad \eta_t^\varepsilon := \frac{1}{\sqrt{\varepsilon}} \int_0^t \Psi^\varepsilon(t, s) dw_s^y.$$

The process η^ε is the solution of the linear stochastic equation

$$(3.10) \quad d\eta_t^\varepsilon = \varepsilon^{-1}A_4(t)\eta_t^\varepsilon dt + \varepsilon^{-1/2}dw_t^y, \quad \eta_0^\varepsilon = 0.$$

We shall use the following properties of η^ε following, e.g., from Theorem 3.1 in [14]: there exists a constant C_p such that

$$(3.11) \quad \sup_{t \geq 0} E |\eta_t^\varepsilon|^p \leq C_p$$

for any $p \in [1, \infty[$ and

$$(3.12) \quad E \|\eta^\varepsilon\|^p \leq C_p \varepsilon^{-1/4}$$

for any $p \in [4, \infty[$.

Substituting (3.8) in the equation (1.1) written in the integral form we come to the following representation for the slow variable:

$$(3.13) \quad \begin{aligned} x_t^{\varepsilon,u} &= \int_0^t [A_1(s)x_s^{\varepsilon,u} + B_1(s)u_s] ds \\ &+ \int_0^t \left\{ A_2(s) \frac{1}{\varepsilon} \int_0^s \Psi^\varepsilon(s,r) [A_3(r)x_r^{\varepsilon,u} + B_2(r)u_r] dr \right\} ds + \zeta_t^\varepsilon + w_t^x, \end{aligned}$$

where

$$(3.14) \quad \zeta_t^\varepsilon := \int_0^t A_2(s) \eta_s^\varepsilon ds.$$

LEMMA 3.1. *For any $p \in [1, \infty[$ there exists a constant c_p such that for all $\varepsilon \in]0, 1]$ it holds that*

$$(3.15) \quad E \|\zeta^\varepsilon\|^p \leq c_p,$$

$$(3.16) \quad \lim_{\varepsilon \rightarrow 0} E \|\zeta^\varepsilon\|^p = 0.$$

Proof. Since A_2 is bounded, (3.15) follows immediately from the Jensen inequality and (3.11). To prove (3.16) we consider the approximation of $D := A_2 A_4^{-1}$ by the step functions

$$D^N := \sum_{i=1}^N D_{t_i} I_{]t_{i-1}, t_i]},$$

where $t_i := iT/N$. Using (3.10) we have

$$\begin{aligned} \zeta_t^\varepsilon &= \int_0^t D_s^N A_4(s) \eta_s^\varepsilon ds + \int_0^t (D_s - D_s^N) A_4(s) \eta_s^\varepsilon ds \\ &= \varepsilon \sum_{i=1}^N D_{t_i} [\eta_{t_i \wedge t}^\varepsilon - \eta_{t_{i-1} \wedge t}^\varepsilon - \varepsilon^{1/2} (w_{t_i \wedge t}^y - w_{t_{i-1} \wedge t}^y)] + \int_0^t (D_s - D_s^N) A_4(s) \eta_s^\varepsilon ds. \end{aligned}$$

This implies the bound

$$(3.17) \quad \|\zeta^\varepsilon\| \leq 2\varepsilon^{1/2} (\varepsilon^{1/2} \|\eta^\varepsilon\| + \|w^y\|) + C\delta_N \int_0^T |\eta_s^\varepsilon| ds,$$

where $\delta_N := \|D - D^N\| \rightarrow 0$ as $N \rightarrow \infty$ due to continuity of α .

Notice that (3.12) implies that the family of random variables $\{\varepsilon^{1/2} \|\eta^\varepsilon\|, \varepsilon \in]0, 1]\}$ is bounded in L^p (for any finite p). It follows from (3.11) that the family of integrals on the right-hand side of (3.17) is also bounded in L^p . Thus,

$$\limsup_{\varepsilon \rightarrow 0} \|\zeta^\varepsilon\| \leq C\delta_N$$

and (3.16) holds.

From the representation (3.13) and bounds (3.6), (3.15) it is easy to deduce that

$$E \| x^{\varepsilon,u} \|_t^{2p} \leq C \left(1 + \int_0^t E \| x^{\varepsilon,u} \|_s^{2p} ds \right),$$

and the standard application of the Gronwall–Bellman lemma gives (3.2).

Put $\bar{\Delta}_t^{x,\varepsilon,u} := x_t^{\varepsilon,u} - \bar{x}_t^u$. The relations (3.1), (3.13) imply that

$$(3.18) \quad \bar{\Delta}_t^{x,\varepsilon,u} = \int_0^t A_0(s) \bar{\Delta}_s^{x,\varepsilon,u} ds + R_t^{\varepsilon,u},$$

where

$$(3.19) \quad \begin{aligned} R_t^{\varepsilon,u} := & \int_0^t A_2(s) \left[\frac{1}{\varepsilon} \int_0^s \Psi^\varepsilon(s,r) A_3(r) x_r^{\varepsilon,u} dr + A_4^{-1}(r) A_3(r) x_r^{\varepsilon,u} \right] ds \\ & + \int_0^t A_2(s) \left[\frac{1}{\varepsilon} \int_0^s \Psi^\varepsilon(s,r) B_2(r) u_r dr + A_4^{-1}(r) B_2(r) u_r \right] ds + \zeta_t^\varepsilon. \end{aligned}$$

It follows from (3.18) that

$$E \| \bar{\Delta}^{x,\varepsilon,u} \|_t^p \leq C \left(\int_0^t E \| \bar{\Delta}^{x,\varepsilon,u} \|_s^p ds + E \| R^{\varepsilon,u} \|_t^p \right),$$

and by the Gronwall–Bellman lemma we have

$$E \| \bar{\Delta}^{x,\varepsilon,u} \|_t^p \leq CE \| R^{\varepsilon,u} \|_t^p e^{CT}.$$

Thus, to prove (3.3) we need to show that

$$\limsup_{\varepsilon \rightarrow 0} E \| R^{\varepsilon,u} \|_t^p = 0.$$

But this relation follows from (3.2), (3.16) and the following statement (see [15, Lemma 3.1] or [13, Lemma 3.2]).

LEMMA 3.2. *For any $\varepsilon \in]0, 1]$, $\eta > 0$, and bounded measurable function h the following holds:*

$$(3.20) \quad \left\| \int_0^\cdot A_2(s) \left[\frac{1}{\varepsilon} \int_0^s \Psi^\varepsilon(s,r) h_r dr + A_2(s) A_4^{-1}(s) h_s \right] ds \right\| \leq \| h \| T(C_1 \eta + \varepsilon C_2(\eta)),$$

where $C_1, C_2(\eta)$ depend on A_2 and A_4 .

At last, the property (3.4) of uniform boundedness in L^p of values of the fast variables for the fixed time follows from the representation (3.8) and (3.2), (3.7), and (3.11).

PROPOSITION 3.2. *Let $(x^{\varepsilon,u}, y^{\varepsilon,u})$ be the solution of (1.1), (1.2) corresponding to some $u \in \mathcal{U}$, and let \bar{x}^u be the solution of the reduced equation (3.1). Let the random variable $\bar{y}_T^{\varepsilon,u}$ be defined by*

$$(3.21) \quad \bar{y}_T^{\varepsilon,u} := -A_4^{-1}(T) A_3(T) \bar{x}_T^u + \int_0^\infty \exp\{A_4(T)r\} B_2(T) v_r^\varepsilon dr + \tilde{\zeta}_T^\varepsilon,$$

where $v_r^\varepsilon := u_{T-r\varepsilon}I_{[0, T/\sqrt{\varepsilon}]}(r) + u^0I_{]T/\sqrt{\varepsilon}, \infty[}(r)$, u^0 is an arbitrary point in U ,

$$(3.22) \quad \tilde{\xi}_T^\varepsilon := \exp\{\varepsilon^{-1}A_4(T)(T - T^\varepsilon)\}\beta + \frac{1}{\sqrt{\varepsilon}} \int_{T^\varepsilon}^T \exp\{\varepsilon^{-1}A_4(T)(T - s)\}dw_s^y,$$

$T_\varepsilon := (1 - \sqrt{\varepsilon})T$, β is a Gaussian random variable with the zero mean and covariance Ξ , and the matrix Ξ is defined in (1.9).

Then for any $p \in [1, \infty[$

$$(3.23) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{u \in \mathcal{U}} E|y_T^{\varepsilon, u} - \tilde{y}_T^{\varepsilon, u}|^p = 0.$$

Proof. Let $\tilde{y}^{\varepsilon, u}$ be the solution of the stochastic differential equation

$$(3.24) \quad \varepsilon d\tilde{y}_t^{\varepsilon, u} = (A_3(T)\tilde{x}_t^u + A_4(T)\tilde{y}_t^{\varepsilon, u} + B_2(T)u_t)dt + \sqrt{\varepsilon}dw_t^y \quad \tilde{y}_0^{\varepsilon, u} = 0.$$

Put

$$\tilde{\Delta}_t^{y, \varepsilon, u} := y_t^{\varepsilon, u} - \tilde{y}_t^{\varepsilon, u}, \quad \tilde{x}_t^{\varepsilon, u} := x_t^{\varepsilon, u} - \tilde{x}_t^{\varepsilon, u},$$

$$\hat{A}_i(t) := A_i(t) - A_i(T), \quad \hat{B}_i(t) := B_i(t) - B_i(T).$$

The process $\tilde{\Delta}^{y, \varepsilon, u}$ is the solution of the ordinary differential equation

$$d\tilde{\Delta}_t^{y, \varepsilon, u} = (A_4(T)\tilde{\Delta}_t^{y, \varepsilon, u} + \varphi_t^{\varepsilon, u})dt, \quad \tilde{\Delta}_0^{y, \varepsilon, u} = 0,$$

where

$$\varphi_t^{\varepsilon, u} := \hat{A}_4(t)y_t^{\varepsilon, u} + \hat{A}_3(t)x_t^{\varepsilon, u} + A_3(T)\tilde{x}_t^{\varepsilon, u} + A_3(T)\tilde{\Delta}_t^{x, \varepsilon, u} + \hat{B}_2(t)u_t.$$

Thus,

$$(3.25) \quad \tilde{\Delta}_T^{y, \varepsilon, u} = \frac{1}{\varepsilon} \int_0^T \exp\{\varepsilon^{-1}A_4(T)(T - s)\}\varphi_s^{\varepsilon, u}ds.$$

By virtue of (1.6) for all $t \geq 0$ we have that

$$(3.26) \quad |\exp\{\varepsilon^{-1}A_4(T)t\}| \leq Ce^{-2\kappa t/\varepsilon}.$$

Taking into account (3.2), (3.4) and the boundedness of U , we get from (3.25) that the L^p -norm of $\tilde{\Delta}_T^{y, \varepsilon, u}$ is bounded by

$$(3.27) \quad C \frac{1}{\varepsilon} \int_0^T e^{-2\kappa(T-s)/\varepsilon} (|\hat{A}_4(s)| + |\hat{A}_3(s)| + f_s^\varepsilon + \bar{g}^\varepsilon + |\hat{B}_2(s)|)ds,$$

where

$$f_s^\varepsilon := \sup_{u \in \mathcal{U}} (E|x_s^{\varepsilon, u} - x_T^{\varepsilon, u}|^p)^{1/p}, \quad \bar{g}^\varepsilon := \sup_{u \in \mathcal{U}} (E|\tilde{\Delta}_T^{x, \varepsilon, u}|^p)^{1/p}.$$

Let \bar{f}_s be the function similar to f_s^ε but defined for \bar{x}^u . It follows from (3.3) that for any $\delta > 0$ we have $f_s^\varepsilon \leq \bar{f}_s + \delta$ for all sufficiently small ε . But it is clear from the equation (3.1) that $\lim_{s \rightarrow T} \bar{f}_s = 0$. Taking into account the above remarks we check easily that the expression (3.27) tends to zero as $\varepsilon \rightarrow 0$ and, hence,

$$(3.28) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{u \in \mathcal{U}} E|y_T^{\varepsilon, u} - \tilde{y}_T^{\varepsilon, u}|^p = 0.$$

Now we show that

$$(3.29) \quad \limsup_{\varepsilon \rightarrow 0} \sup_{u \in \mathcal{U}} E|\bar{y}_T^{\varepsilon,u} - \tilde{y}_T^{\varepsilon,u}|^p = 0.$$

Indeed,

$$\begin{aligned} \bar{y}_T^{\varepsilon,u} - \tilde{y}_T^{\varepsilon,u} &= \left(-A_4^{-1}(T) - \frac{1}{\varepsilon} \int_0^T \exp\{\varepsilon^{-1}A_4(T)(T-s)\}ds \right) A_3(T)\bar{x}_T^u \\ &+ \int_{T/\varepsilon}^\infty \exp\{A_4(T)r\}B_2(T)u_r^0 dr - \int_{T/\sqrt{\varepsilon}}^{T/\varepsilon} \exp\{A_4(T)r\}B_2(T)u_{T-\varepsilon r} dr \\ &+ \exp\{\varepsilon^{-1/2}A_4(T)T\}\beta - \frac{1}{\sqrt{\varepsilon}} \int_0^{T\varepsilon} \exp\{\varepsilon^{-1}A_4(T)(T-s)\}dw_s^y. \end{aligned}$$

Evidently, L^p -norms of all terms on the right-hand side of this identity tend to zero and the convergence of the first one is uniform in $u \in \mathcal{U}$ by virtue of (3.2) and (3.3). Thus, (3.29) holds. The relations (3.28), (3.29) imply (3.23).

Proposition 3.2 is proved.

Assume that sequence $\mathcal{L}(x_T^{\varepsilon_n, u_n}, y_T^{\varepsilon_n, u_n})$ converges in $\mathbf{P}(\mathbf{R}^m)$ to some μ . Choose in the representation (3.22) the random variable β independent of W . It follows from Propositions 3.1, 3.2 that the sequence $\mathcal{L}(\bar{x}_T^{u_n}, \bar{y}_T^{\varepsilon_n, u_n})$ converges to the same limit. Let us introduce the modified controls $\hat{u}_n = u_n I_{[0, T_{\varepsilon_n}]} + u^0 I_{]T_{\varepsilon_n}, T]}$, where u^0 is a fixed point from U . Since $\bar{x}_T^{u_n} - \bar{x}_T^{\hat{u}_n}$ tends to zero in probability, the sequence $\mathcal{L}(\bar{x}_T^{\hat{u}_n}, \bar{y}_T^{\varepsilon_n, u_n})$ converges to μ and we need to check only that $\mathcal{L}(\bar{x}_T^{\hat{u}_n}, \bar{y}_T^{\varepsilon_n, u_n}) \in \mathcal{K}_0(T)$. To show this notice that $\bar{x}_T^{\hat{u}_n}$ is a function of the natural projection

$$i^{\varepsilon_n} : \{w_t^x, w_t^y, t \in [0, T]\} \mapsto (\{w_t^x, t \in [0, T]\}, \{w_t^y, t \in [0, T_{\varepsilon_n}]\}).$$

As in section 2 it can be shown that the regular conditional distribution of the random variable $\xi_0^{\varepsilon_n} + I(v^{\varepsilon_n})$ for a fixed value i^{ε_n} belongs to S . Since S is a convex closed set and $\bar{x}_T^{\hat{u}_n}$ is a measurable function on i^{ε_n} , it follows from Lemma 5.6 that the regular conditional distribution of $\xi_0^{\varepsilon_n} + I(v^{\varepsilon_n})$ for a fixed value $\bar{x}_T^{\hat{u}_n}$ also belongs to S , implying the result.

4. Proof of Theorem 1.1. Part 2. Now we must show that for any measure $\mu F^{-1} \in \mathcal{K}_0(T)$ there exists a sequence $\mu_n \in \mathcal{K}_{\varepsilon_n}(T)$ which converges to μF^{-1} in $\mathbf{P}(\mathbf{R}^n)$. It is sufficient to find such a sequence for an arbitrary μF^{-1} from the set $\tilde{\mathcal{K}}_0(T)$ which is dense in $\mathcal{K}_0(T)$ in the total variation topology. The latter property holds since the attainability set $\tilde{\mathcal{K}}_0^x$ corresponding to the strong solutions of (2.1) is dense in \mathcal{K}_0^x in the total variation topology. Thus, there are dense embeddings $\tilde{\mathcal{K}}_0 \subseteq \mathcal{K}_0$ and $\tilde{\mathcal{K}}_0(T) \subseteq \mathcal{K}_0(T)$.

Let us fix $\delta > 0$ and a measure $\mu = m(x, dy)\nu(dx)$ such that $\mu F^{-1} \in \mathcal{K}_0(T)$. By definition $\nu = \mathcal{L}(\bar{x}_T^u)$, where \bar{x}^u is a solution of the reduced equation (2.1) corresponding to some admissible control u . Let $\nu_h := \mathcal{L}(\bar{x}_{T-h}^u)$, $\mu_h(dx, dy) := m(x, dy)\nu_h(dx)$, $h \in [0, T]$. Then there exists $h_0 > 0$ such that

$$(4.1) \quad d(\mu F^{-1}, \mu_h F^{-1}) \leq \delta$$

for all $h \in]0, h_0]$.

To prove (4.1) we use the following.

LEMMA 4.1. *Let \bar{x}^u be the solution of (3.1). Then*

$$(4.2) \quad \limsup_{s \rightarrow 0} \sup_{u \in \mathcal{U}} \text{Var}(\mathcal{L}(\bar{x}_{T-s}^u) - \mathcal{L}(\bar{x}_T^u)) = 0.$$

Proof. For any $u \in \mathcal{U}$ let $u^r := uI_{[0, T-r]} + u^0I_{[T-r, T]}$, where u^0 is an arbitrary point in \mathcal{U} . It follows from the bound for the total variation distance in terms of the Hellinger process h_t (see [12, Theorems 2.2 and 5.1]) that

$$(4.3) \quad \text{Var}(\mathcal{L}(\bar{x}^u) - \mathcal{L}(\bar{x}^{u^r})) \leq Cr^{1/2}.$$

(Notice that in the considered situation the Hellinger process for the pair $(\mathcal{L}(\bar{x}^u), \mathcal{L}(\bar{x}^{u^r}))$ has the form

$$h_t = \int_0^t I_{[r, T]}(\tau) |B_0(\tau)(\hat{u}_\tau - u^0)|^2 d\tau,$$

where \hat{u}_s takes values in U .)

Fix $\gamma > 0$ and $r > 0$ such that $Cr^{1/2} \leq \gamma$. For any $s \in [0, r]$ we have

$$\mathcal{L}(\bar{x}_{T-s}^{u^r}) = \mathcal{L}(\bar{x}_{T-r}^u) * \mathcal{N}(a_s, K_s),$$

where $*$ denotes the convolution, $\mathcal{N}(a_s, K_s)$ is the nondegenerate Gaussian distribution with the mean

$$a_s := \int_{T-r}^{T-s} B_0(\tau) u^0 d\tau$$

and covariance

$$K_s := \int_{T-r}^{T-s} \Phi_0(T-s, \tau) \Phi_0'(T-s, \tau) d\tau,$$

and $\Phi_0(T-s, \tau)$ is the fundamental matrix corresponding to $A_0(t)$. In particular,

$$\mathcal{L}(\bar{x}_T^{u^r}) = \mathcal{L}(\bar{x}_{T-r}^u) * \mathcal{N}(a_0, K_0).$$

The well-known inequality

$$\text{Var}(F * G - F * \tilde{G}) \leq \text{Var}(G - \tilde{G})$$

implies that

$$\text{Var}(\mathcal{L}(\bar{x}_{T-s}^{u^r}) - \mathcal{L}(\bar{x}_T^{u^r})) \leq \text{Var}(\mathcal{N}(a_s, K_s) - \mathcal{N}(a_0, K_0)),$$

where the right-hand side tends to zero as $s \rightarrow 0$.

Thus, for sufficiently small s we have

$$(4.4) \quad \sup_{u \in \mathcal{U}} \text{Var}(\mathcal{L}(\bar{x}_{T-s}^{u^r}) - \mathcal{L}(\bar{x}_T^{u^r})) \leq \gamma.$$

It follows from (4.3) and (4.4) that

$$\sup_{u \in \mathcal{U}} \text{Var}(\mathcal{L}(\bar{x}_{T-s}^u) - \mathcal{L}(\bar{x}_T^u)) \leq 3\gamma$$

and the lemma is proved.

Since

$$\text{Var}(\mu F^{-1} - \mu_h F^{-1}) = \text{Var}(\mu - \mu_h) = \text{Var}(\nu - \nu_h) \rightarrow 0$$

by virtue of the above lemma, the relation (4.1) holds.

Furthermore, there exists $h_1 > 0$

$$(4.5) \quad \sup_{\varepsilon} \sup_{z \in \mathcal{U}_h(u)} d(\mathcal{L}(x_{T-h}^{\varepsilon,z}, y_T^{\varepsilon,z}), \mathcal{L}(x_T^{\varepsilon,z}, y_T^{\varepsilon,z})) \leq \delta,$$

where $\mathcal{U}_h(u)$ is the set consisting of all $z \in \mathcal{U}$ such that

$$(4.6) \quad zI_{[0, T-h]} = uI_{[0, T-h]}.$$

The relation (4.5) is an evident corollary of Proposition 3.1 and the following.

LEMMA 4.2. *Let $(\xi_{\iota, h}^{(i)})$, $\iota \in I(h)$, $h \in [0, T]$, $i = 1, 2$, be two families of random variables with values in \mathbf{R}^m such that*

$$\begin{aligned} \sup_h \sup_{\iota \in I(h)} E|\xi_{\iota, h}^{(i)}|^p &< \infty, \quad i = 1, 2, \\ \lim_{h \rightarrow 0} \sup_{\iota \in I(h)} E|\xi_{\iota, h}^{(1)} - \xi_{\iota, h}^{(2)}|^p &= 0 \end{aligned}$$

for some $p > 0$. Then for any bounded continuous function f on \mathbf{R}^m

$$\lim_{h \rightarrow 0} \sup_{\iota \in I(h)} |Ef(\xi_{\iota, h}^{(1)}) - f(\xi_{\iota, h}^{(2)})| = 0.$$

The proof of Lemma 4.2 is easy and is omitted.

Lemma 4.2 implies also the existence of $h_2 > 0$ such that

$$(4.7) \quad \sup_{\iota} d(\mathcal{L}(\bar{x}_{T-h}^u, -A_4(T)A_3(T)\bar{x}_{T-h}^u + \eta_{\iota}), \mathcal{L}(\bar{x}_{T-h}^u, -A_4(T)A_3(T)\bar{x}_{T-h}^u + \eta_{\iota})) \leq \delta,$$

where the family (η_{ι}) consists of all random variables with distribution from S_Y .

Let us consider some $h \leq h_0 \wedge h_1 \wedge h_2$. The desired result will be proved if we find for any sufficiently small ε an admissible control $z = z^{\varepsilon}$ satisfying (4.6) such that

$$(4.8) \quad d(\mathcal{L}(x_{T-h}^{\varepsilon, z}, y_T^{\varepsilon, z}), \mu_h F^{-1}) \leq 2\delta.$$

Indeed, it follows from (4.1), (4.5), and (4.8) that

$$d(\mathcal{L}(x_T^{\varepsilon, z}, y_T^{\varepsilon, z}), \mu_h F^{-1}) \leq 4\delta,$$

and this means that any point in $\mathcal{K}_0(T)$ can be approximated by points from $\mathcal{K}_{\varepsilon}(T)$.

Let (Ω, \mathcal{F}, P) be a probability space with a countably generated σ -algebra. Assume that on this space we have independent random elements ζ , α , ξ , where ζ has the distribution ν_h , i.e., the same distribution as \bar{x}_{T-h}^u ; α has the standard normal distribution; ξ is a stationary Gaussian Markov process with zero mean and covariance function given by (1.8), (1.9). Let us consider the set \mathcal{V}_U^{α} of all U -valued processes which are predictable with respect to the filtration generated by $\xi_{1/t}$ and α (we denote by \mathcal{P} the corresponding predictable σ -algebra in $\Omega \times \mathbf{R}_+$).

LEMMA 4.3. *There is a function $v : \Omega \times \mathbf{R}_+ \times \mathbf{R}^m \rightarrow U$ which is measurable with respect to $\mathcal{P} \otimes \mathcal{B}(\mathbf{R}^m)$ such that $v(\cdot, x) \in \mathcal{V}^{\alpha}$ for all $x \in \mathbf{R}^m$ and $\mathcal{L}(\xi_0 + I(v(\cdot, x)))$ is equal to $\mu(x, dy)$ for ν_h almost all $x \in \mathbf{R}^m$.*

Proof. Evidently, $v \mapsto \mathcal{L}(\xi_0 + I(v))$ is a continuous, hence measurable, mapping from the space $\mathbf{V} := L^1(\Omega \times \mathbf{R}_+, \mathcal{P}, \rho)^d$ into $\mathbf{P}(\mathbf{R}^n)$, where $\rho(d\omega, dt) = e^{-2\kappa t} P(d\omega) dt$. Thus, the multivalued mapping

$$\Gamma : x \mapsto \{v \in \mathbf{V} : v(\omega, t) \in U \text{ } \rho \text{ a.e., } \mathcal{L}(\xi_0 + I(v)) = \mu(x, \cdot)\}$$

has a measurable graph. Hence, it admits a measurable selector $x \mapsto V(x)$. Notice that $V(x)$ as an element of \mathbf{V} is a class of ρ -equivalent functions. To choose from $V(x)$ a representative in a measurable way we proceed as follows. Let (v^i) be a sequence of elements from \mathcal{V}_U^α which is dense in $\mathcal{V}_U^\alpha \cap \mathbf{V}$, $j(x, l) := \min\{i : \|v(x) - v^i\| \leq 1/l\}$. Then $v^{j(l)} = v^{j(x,l)}(\omega, t)$ is a $\mathcal{P} \otimes \mathcal{B}(\mathbf{R}^m)$ -measurable function with values in U . The sequence $v^{j(x,l)}$ converges to $V(x)$ in \mathbf{V} . Since U is bounded, the sequence $v^{j(l)}$ converges to V in $L^1(\Omega \times \mathbf{R}_+ \times \mathbf{R}^m, \mathcal{P} \otimes \mathcal{B}(\mathbf{R}^m), \rho \times \nu_h)^d$. Hence, there exists a subsequence which converges $\rho \times \nu_h$ a.e. to some $\mathcal{P} \otimes \mathcal{B}(\mathbf{R}^m)$ -measurable function $v = v(\omega, t, x)$. For ν_h almost all x we have the inclusion $v(\cdot, x) \in V(x)$ implying that $\mathcal{L}(\xi_0 + I(v(\cdot, x))) = \mu(x, dy)$ for such x .

It follows from the above lemma that the measure μ_h is the distribution of the random variable $(\zeta, \xi_0 + I(v(\cdot, \zeta)))$, i.e.,

$$(4.9) \quad \mu_h = \mathcal{L}(\zeta, \xi_0 + I(v(\cdot, \zeta))).$$

Generalizing the arguments of section 2 we introduce a set $\mathcal{V}_U^{(\alpha, \zeta)}$ consisting of all functions

$$(4.10) \quad v(s, x) = \sum_{i=1}^N \varphi_i(x) I_{]s_i, s_{i+1}]}(s) + u^0 I_{]s_{N+1}, \infty[}(s),$$

where $0 = s_1 < \dots < s_{N+1}$, $u^0 \in U$, and $\varphi_i(x)$ have the form

$$(4.11) \quad \varphi_i(x) = f_i(\alpha, \xi(r_1^i), \dots, \xi(r_{M_i}^i), x), \quad s_{i+1} < r_j^i \leq s_N,$$

and the functions f_i are measurable with respect to their arguments and take values in U .

Assume that the representation (4.9) holds with $v \in \mathcal{V}_U^{(\alpha, \zeta)}$. There is a freedom in the choice of ζ , α , and ξ which we use in the following constructions.

Put $T_\varepsilon^k := T(1 - k\varepsilon^{1/2})$, $k = 1, 2, 3$, $\zeta := \bar{x}_{T-h}^u$.

Define

$$\alpha^\varepsilon := (w_{T_\varepsilon^2}^{y,1} - w_{T_\varepsilon^3}^{y,1}) / (T_\varepsilon^2 - T_\varepsilon^3)^{1/2},$$

where $w^{y,1}$ is the first component of the vector process w^y ,

$$\beta^\varepsilon := \Xi^{1/2} (w_{T_\varepsilon^1}^y - w_{T_\varepsilon^2}^y) / (T_\varepsilon^1 - T_\varepsilon^2)^{1/2}.$$

Let us consider on $[T_\varepsilon^1, T]$ the linear stochastic differential equation

$$\varepsilon d\tilde{\xi}_t^\varepsilon = A_4(T)\tilde{\xi}_t^\varepsilon dt + \varepsilon^{1/2} dw_t^y, \quad \tilde{\xi}_{T_\varepsilon^1}^\varepsilon = \beta^\varepsilon.$$

Put $\xi_t^\varepsilon := \tilde{\xi}_{T-\varepsilon t}^\varepsilon$, $t \in [0, \varepsilon^{-1/2}T]$. For sufficiently small ε we define the admissible control

$$z^\varepsilon := u I_{]0, t_{N+1}[} + \sum_{i=1}^{N+1} \varphi_i^\varepsilon(\bar{x}_{T-h}^u) I_{]t_{i+1}, t_i[},$$

where $t_i := T - \varepsilon s_i$, $i \leq N + 1$, and φ_i^ε is constructed in accordance with (4.11).

It follows from Propositions 3.1 and 3.2 that

$$(x_{T-h}^{\varepsilon, z^\varepsilon}, y_{T-h}^{\varepsilon, z^\varepsilon}) - (\bar{x}_{T-h}^u, -A_4(T)A_3(T)\bar{x}_T^u + \xi_0^\varepsilon + I(v(\cdot, \bar{x}_{T-h}^u))) \rightarrow 0$$

in probability as $\varepsilon \rightarrow 0$. Thus,

$$(4.12) \quad d(\mathcal{L}(x_{T-h}^{\varepsilon, z^\varepsilon}, y_{T-h}^{\varepsilon, z^\varepsilon}), \mathcal{L}(\bar{x}_{T-h}^u, -A_4(T)A_3(T)\bar{x}_T^u + \xi_0^\varepsilon + I(v(\cdot, \bar{x}_{T-h}^u)))) \leq \delta$$

for all sufficiently small ε . Taking into account (4.7) we get from here the desired inequality (4.8).

Part 2 of Theorem 1.1 is proved now for the case when μ_h is given by (4.9) with $v \in \mathcal{V}_U^{(\alpha, \zeta)}$. Since the set $\{I(v) : v \in \mathcal{V}_U^{(\alpha, \zeta)}\}$ is dense in probability in the set $\{I(v) : v \in \mathcal{V}_U^{\alpha, \zeta}\}$, the result holds for the general case as well.

5. On a compactness of some subsets in the space of probability measures.

5.1. Notations and preliminaries. Let X be a Polish space with the Borel σ -algebra \mathcal{X} and $\mathbf{P}(X)$ be a space of all probability measures on X with the topology of weak convergence. It is well known that $\mathbf{P}(X)$ equipped by the Prohorov metric is again a Polish space. The relative compactness of a subset $A \subseteq \mathbf{P}(X)$ is equivalent to its tightness. The last means that for any $\varepsilon > 0$ there exists a compact set $K \subseteq X$ such that $m(K) \geq 1 - \varepsilon$ for all $m \in A$.

We shall use the notation $m(f) = \int_X f(x)m(dx)$. We denote by $\mathcal{L}(\xi)$ the distribution of a random variable ξ .

Let (X, \mathcal{X}) and (Y, \mathcal{Y}) be two Polish spaces. We denote by $\mathcal{M}(X, Y)$ the set of stochastic kernels from (X, \mathcal{X}) to (Y, \mathcal{Y}) that is mappings $\mu : X \times \mathcal{Y} \rightarrow ([0, 1], \mathcal{B}[0, 1])$ such that $x \mapsto \mu(x, \Gamma)$ is \mathcal{X} -measurable for any $\Gamma \in \mathcal{Y}$ and $\mu(x, \cdot) \in \mathbf{P}(Y)$ for any $x \in X$.

It is easy to check that the mapping $\mu : X \times \mathcal{Y} \rightarrow ([0, 1], \mathcal{B}[0, 1])$ is in $\mathcal{M}(X, Y)$ if and only if one of the following equivalent conditions is satisfied:

(1) The mapping $x \mapsto \mu(x, \cdot)$ is \mathcal{X} -measurable (i.e., $\mu(x, \cdot)$ is a $\mathbf{P}(Y)$ -valued random variable).

(2) For any $f \in C_b(Y)$ (the set of all bounded continuous functions on Y) the mapping $x \mapsto \mu(x, f)$ is \mathcal{X} -measurable (i.e., $\mu(x, f)$ is a real-valued random variable).

THE SKOROHOD REPRESENTATION THEOREM. *Let Y be a Polish space and $m_n \in \mathbf{P}(Y)$ be a sequence converging in $\mathbf{P}(Y)$ to some m . Then on the probability space $([0, 1], \mathcal{B}[0, 1], dx)$ there exist Y -valued random variables ξ_n and ξ such that $\mathcal{L}(\xi_n) = m_n$, $\mathcal{L}(\xi) = m$, and $\xi_n \rightarrow \xi$ pointwise.*

THE MEASURABLE ISOMORPHISM THEOREM. *Let (X, \mathcal{X}) be an uncountable Polish space. Then there is a one-to-one mapping $i : X \rightarrow [0, 1]$ such that $i(\Gamma) \in \mathcal{B}[0, 1]$ for any $\Gamma \in \mathcal{X}$ and $i^{-1}(A) \in \mathcal{X}$ for any $A \in \mathcal{B}[0, 1]$.*

Another useful result is that any Polish space X is homeomorphic to a G_δ -subset of the Hilbert cube $[0, 1]^{\mathbb{N}}$. For further information see, e.g., [6], [9].

5.2. For $\mu \in \mathcal{M}(X, Y)$, $m \in \mathbf{P}(X)$, and $\Gamma \in \mathcal{Y}$, the integral $\int_X \mu(x, \Gamma)m(dx)$ defines a probability measure on (Y, \mathcal{Y}) which we shall denote by $\int_X \mu(x, \cdot)m(dx)$.

LEMMA 5.1. *Let (X, \mathcal{X}) be a Polish space with nonatomic measure ν on it, let S be a compact set in $\mathbf{P}(Y)$, and let \mathcal{M} be the set consisting of all stochastic kernels μ from (X, \mathcal{X}) to (Y, \mathcal{Y}) such that $\mu(x, \cdot) \in S$ for all $x \in X$. Then the set*

$$K = \left\{ m \in \mathbf{P}(Y) : m(\cdot) = \int_X \mu(x, \cdot)\nu(dx), \mu \in \mathcal{M} \right\}$$

is a convex compact subset in $\mathbf{P}(Y)$ coinciding with $\overline{\text{conv}S}$.

Proof. By virtue of the measurable isomorphism theorem we can consider only the case when $(X, \mathcal{X}) = ([0, 1], \mathcal{B}[0, 1])$. Assume at first that $\nu(dx) = dx$, i.e., ν is the Lebesgue measure. Convexity of \mathcal{M} is clear: if measures $m_i(\cdot) = \int_X \mu_i(x, \cdot) dx$, $i = 1, 2$, belong to K , $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$, then the measure $\alpha m_1(\cdot) + \beta m_2(\cdot) = \int_X \mu(x, \cdot) dx$ with

$$\mu(x, \cdot) = I_{[0, \alpha]}(x) \mu_1(\alpha^{-1}x, \cdot) + I_{]1-\beta, 1]}(x) m_2(\beta^{-1}(x-1+\beta), \cdot)$$

also belonging to K . The tightness of K follows easily from the tightness of S . To prove that K is closed, let us consider the sequence $m_n(\cdot) = \int \mu_n(x, \cdot) dx \in K$ converging to some $m(\cdot)$ in $\mathbf{P}(Y)$. Notice that elements of \mathcal{M} are random variables with values in the compact subset S of a Polish space. Thus, the set of distributions of these random variables $\{\mathcal{L}(\mu) : \mu \in \mathcal{M}\}$ is relatively compact in $\mathbf{P}(\mathbf{P}(Y))$. Taking, if necessary, a subsequence we can assume that $\mathcal{L}(\mu_n)$ tend to some \mathcal{L} in $\mathbf{P}(\mathbf{P}(Y))$. By the Skorohod representation theorem on the probability space $([0, 1], \mathcal{B}[0, 1], dx)$ there exist S -valued random variables $\tilde{\mu}_n$ and $\tilde{\mu}$ such that $\tilde{\mu}_n(x, \cdot) \rightarrow \tilde{\mu}(x, \cdot)$ for all x when $n \rightarrow \infty$ and $\mathcal{L}(\tilde{\mu}) = m$, $\mathcal{L}(\tilde{\mu}_n) = \mathcal{L}(\mu_n)$ for all n .

The last equality means that for any $f \in C_b(Y)$ the distribution of the random variable $\tilde{\mu}_n(f)$ coincides with the distribution of $\mu_n(f)$. It follows that for any $f \in C_b(Y)$

$$m(f) = \lim_{n \rightarrow \infty} m_n(f) = \lim_{n \rightarrow \infty} \int \mu_n(x, f) dx = \lim_{n \rightarrow \infty} \int \tilde{\mu}_n(x, f) dx = \int \tilde{\mu}(x, f) dx.$$

Thus, $m(\cdot) = \int \tilde{\mu}(x, \cdot) dx \in K$.

The general case when ν is any nonatomic measure on $[0, 1], \mathcal{B}[0, 1]$ is easily reduced to the considered one by the quantile transformation. Indeed, let $F(t) := \nu([0, t])$, $C(t) := \inf\{s : F(s) > t\}$. Then we have the identities

$$\int \mu(x, \cdot) dx = \int \mu(F(x), \cdot) \nu(dx), \quad \int \mu(x, \cdot) \nu(dx) = \int \mu(C(x), \cdot) dx$$

which show that K does not depend on the measure ν .

Evidently, $S \subseteq K$. Hence, $\overline{\text{conv}} S \subseteq K$. Let $m_0(\cdot) = \int \mu(t, \cdot) dt$ be a point in K which does not belong to $\overline{\text{conv}} S$. By the separation theorem a convex compact set and a point outside it can be strictly separated by a continuous linear functional. This means that there exists $f \in C_b(Y)$ such that $\inf_{m \in \overline{\text{conv}} S} m(f) < m_0(f)$. It follows that $\int \mu(t, f) dt < m_0(f)$ in contradiction with the assumption that $m_0 \in K$.

Remark 5.1. If ν has atoms, then we can assert only that K is a subset of $\overline{\text{conv}} S$, even when S is compact.

5.3. Convergence of measure-valued martingales.

PROPOSITION 5.1. *Let (Ω, \mathcal{F}, P) be a probability space with an increasing family of σ -algebras (\mathcal{F}_n) such that $\mathcal{F} = \sigma\{\mathcal{F}_n, n \in \mathbf{N}\}$. Let $\mu_n(\omega, \cdot)$ be a stochastic kernel from (Ω, \mathcal{F}_n) to (Y, \mathcal{Y}) such that for any $f \in C_b(Y)$ the sequence $(\mu_n(f), \mathcal{F}_n)$ is a martingale. Assume that for almost all ω the sequence $\mu_n(\omega, \cdot)$ is tight. Then for almost all ω there exists a limit $\mu(\cdot)$ of $\mu_n(\omega, \cdot)$ in $\mathbf{P}(Y)$ and $E(\mu(f) | \mathcal{F}_n) = \mu_n(f)$ for all $f \in C_b(Y)$ and $n \in \mathbf{N}$.*

Proof. To clarify ideas we start from the case when $Y = \mathbf{R}$. Let $M_n(\omega, y) = \mu_n(\omega,]-\infty, y])$ be the distribution function of $\mu_n(\omega, \cdot)$. Evidently, $(M_n(y), \mathcal{F}_n)$ is a bounded martingale for all $y \in \mathbf{R}$ and by the Doob theorem it converges almost surely (a.s.) to $M^0(y)$. There is a set Ω_1 with $P(\Omega_1) = 1$ such that for all $\omega \in \Omega_1$

and all rationals r we have convergence of $M_n(\omega, r)$ to $M^0(\omega, r)$. Put $M(\omega, y) = \inf\{M^0(\omega, r) : r \in \mathbf{Q}, r > y\}$ for $\omega \in \Omega_1$. Let $M(\omega, \cdot)$ be equal to any distribution function outside Ω_1 . The assumption on tightness implies that $M(\omega, \cdot)$ is a probability distribution function and for any $\omega \in \Omega_1$ we have that $M_n(\omega, y)$ tends to $M(\omega, y)$ at any point y where the function $M(\omega, \cdot)$ is continuous.

As any Polish space is homeomorphic to a G_δ -subset of $H = [0, 1]^{\mathbf{N}}$ we can assume in general case that Y is the intersection of open subsets G_n in H . The closure \bar{Y} of Y is a compact subset of H . Thus, $C_b(\bar{Y})$ is separable. Let A be a countable dense subset of $C_b(\bar{Y})$ closed under finite sums and multiplication by rationals. For any $f \in A$ the sequence $\mu_n(\omega, f)$ converges to some $\mu_f(\omega)$ for all ω from a set Ω_f with $P(\Omega_f) = 1$. It is possible to find a set Ω_1 with $P(\Omega_1) = 1$ such that for all $\omega \in \Omega_1$, $f, g \in A$, and rational a and b

$$\mu_{af+bg}(\omega) = a\mu_f(\omega) + b\mu_g(\omega).$$

Evidently,

$$|\mu_f(\omega) - \mu_g(\omega)| \leq \|f - g\|, \omega \in \Omega_1,$$

where $\|\cdot\|$ is a uniform norm in $C_b(\bar{Y})$, and the function $f \mapsto \mu_f(\omega)$ can be extended uniquely to the continuous positive linear functional on $C_b(\bar{Y})$ which by the Riesz theorem has the form $\mu_f(\omega) = \mu(\omega, f)$ for some measure $\mu(\omega, \cdot)$ on \bar{Y} . For $\omega \in \Omega_1$ we put $\mu(\omega, \cdot)$ equal to any fixed probability measure on Y . We show that μ is the kernel we are seeking. Notice that $\mu(\omega, Y) = 1$. Fix $\omega \in \Omega_1$. By the assumption there exists a subsequence $\mu_{n'}(\omega, \cdot)$ which converges in $\mathbf{P}(Y)$ to a measure $\mu'(\omega, \cdot)$ on Y . We can extend $\mu_{n'}(\omega, \cdot)$ and $\mu'(\omega, \cdot)$ to \bar{Y} in a trivial way. Then for $f \in A$ we have

$$\begin{aligned} \int_{\bar{Y}} f(y)\mu'(\omega, dy) &= \int_Y f(y)\mu'(\omega, dy) = \lim_{n \rightarrow \infty} \int_Y f(y)\mu_{n'}(\omega, dy) \\ &= \lim_{n \rightarrow \infty} \int_{\bar{Y}} f(y)\mu_{n'}(\omega, dy) = \int_{\bar{Y}} f(y)\mu(\omega, dy). \end{aligned}$$

It follows that the probability measures $\mu'(\omega, \cdot)$ and $\mu(\omega, \cdot)$ coincide, and, since any convergent subsequence has the same limit, the whole sequence $\mu_n(\omega, \cdot)$ converges in $\mathbf{P}(Y)$ to $\mu_n(\omega, \cdot)$.

The result is proved.

5.4. Let X and Y be Polish spaces. Any measure $m \in \mathbf{P}(X \times Y)$ can be disintegrated, that is, can be represented as $m(dx, dy) = \mu(x, dy)\nu(dx)$, where ν is the image of m under the projection mapping $X \times Y$ onto X and μ is an element of $\mathcal{M}(X, Y)$ (regular conditional probability) defined ν a.s. uniquely.

LEMMA 5.2. *Let S_Y be a convex compact subset in $\mathbf{P}(Y)$, and let S be the set of all $m \in \mathbf{P}([0, 1] \times Y)$ such that $m(dx, dy) = \mu(x, dy)dx$ with $\mu(x, \cdot) \in S_Y$ for all $t \in [0, 1]$. Then S is a convex compact set.*

Proof. The problem is to prove that S is closed. Let us consider for any $\Delta = [a, b] \subseteq [0, 1]$, $b > a$, the set

$$K_\Delta = \left\{ m \in \mathbf{P}(Y) : m(\cdot) = \frac{1}{b-a} \int_\Delta \mu(x, \cdot) dx, \mu(x, \cdot) \in S_Y \text{ for all } x \in \Delta \right\},$$

which is, by Lemma 5.1, a convex compact set in $\mathbf{P}(Y)$. Let L be the set of all $m \in \mathbf{P}([0, 1] \times Y)$ such that the image of m under the projection mapping $X \times Y$

onto X is the Lebesgue measure (this means that $m(dx, dy) = \mu(x, dy)dx$ without any restriction on μ). Evidently, L is a closed convex set in $\mathbf{P}([0, 1] \times Y)$.

Define the continuous affine mapping $f_\Delta : L \rightarrow \mathbf{P}(Y)$ by the formula $f_\Delta : m \mapsto m_\Delta$ where $m_\Delta(\Gamma) = m(\Delta \times \Gamma)/(b - a)$. The result will be proved if we show that $S = \cap_\Delta f_\Delta^{-1}(K_\Delta)$. The inclusion $S \subseteq \cap_\Delta f_\Delta^{-1}(K_\Delta)$ is evident. To prove the opposite inclusion let us consider the measure m from L which belongs to $\cap_\Delta f_\Delta^{-1}(K_\Delta)$. Let us define the dyadic σ -algebras $\mathcal{F}_l = \sigma\{\Delta_{k,l}, k = 1, \dots, 2^l\}$, where $\Delta_{0,l} = [0, 2^{-l}]$, $\Delta_{k,l} = [(k - 1)2^{-l}, k2^{-l}]$, $k \geq 1$. Using Lemma 5.1 it is easy to show that for any l there exists a stochastic kernel μ_l such that $\mu_l(x, \cdot) \in S_Y$ for all $x \in [0, 1]$ and

$$m(A \times \cdot) = \int_A \mu_l(x, \cdot) dx$$

for all $A \in \mathcal{F}_l$. Put

$$m_l(t, \cdot) = \sum_{k=1}^{2^l} I_{\Delta_{k,l}}(t) m_{l,k}(\cdot)$$

where

$$m_{l,k}(\cdot) = 2^l \int_{\Delta_l} \mu_l(x, \cdot) dx \in S$$

according to Lemma 5.1. By Proposition 5.1 on convergence of measure-valued martingales, the sequence $\mu_l(x, \cdot)$ tends to $\mu(x, \cdot)$ in $\mathbf{P}(Y)$ for almost all x and

$$\int_A \mu_l(x, \cdot) dx = \int_A \mu(x, \cdot) dx$$

for all $A \in \mathcal{F}_l$. Thus, we find a stochastic kernel μ such that $\mu(x, \cdot) \in S_Y$ for all $x \in [0, 1]$ and $m(A \times \Gamma) = \int_A \mu(x, \Gamma) dx$ for all $A \in \mathcal{B}_l$, $l \in \mathbf{N}$, and $\Gamma \in \mathcal{Y}$. It follows that $m(dx, dy) = \mu(x, dy)dt$. Hence, $m \in S$ and the lemma is proved.

5.5.

LEMMA 5.3. *Let (X, \mathcal{X}) be any uncountable Polish space with a probability measure ν on it. Then there exists an increasing family of σ -algebras (\mathcal{X}_l) , $l \in \mathbf{N}$, such that*

- (1) \mathcal{X}_l is generated by a finite partition of X to the sets $A_{k,l}$, $k = 1, \dots, r_l$;
- (2) $\mathcal{X} = \sigma\{\mathcal{X}_l, l \in \mathbf{N}\}$;
- (3) $\nu(\partial A_{k,l}) = 0$ for any k and l (∂A denotes the boundary of A).

Proof. Since a Polish space is homeomorphic to G_δ -subsets of $H = [0, 1]^{\mathbf{N}}$, we can assume without loss of generality that X is a Borel subset of H . Moreover, it is sufficient to construct the family (\mathcal{X}_l) for the space H (then the σ -algebras $\mathcal{X}_l \cap X = \{A \cap X, X \in \mathcal{X}_l\}$ will have the desired properties for X). Let $\varepsilon \in [0, 1/2]$. Let us define the partitions of the interval $[0, 1]$ by points $a_{k2^{-l}}^\varepsilon$, $k = 0, \dots, 2^l$, in the following recurrent way. Let $a_0^\varepsilon = 0$, $a_1^\varepsilon = 1$, $a_{2^{-l}}^\varepsilon = 2^{-1} + \varepsilon$. Starting from the l th partition we define for k even the point $a_{k2^{-l-1}}^\varepsilon = (a_{k2^{-l}}^\varepsilon + a_{(k+1)2^{-l}}^\varepsilon)/2$; i.e., we construct the ordinary dyadic partitions on both intervals $[0, 2^{-1} + \varepsilon]$ and $]2^{-1} + \varepsilon, 1]$.

Evidently, diameters of the partitions tend to zero as $l \rightarrow \infty$.

Put

$$\Delta_{1,l}^\varepsilon = [0, a_{2^{-l}}^\varepsilon], \Delta_{k,l}^\varepsilon =]a_{(k-1)2^{-l}}^\varepsilon, a_{k2^{-l}}^\varepsilon], k = 1, \dots, 2^l,$$

$$\Gamma^\varepsilon = \{a_{k2^{-l}}^\varepsilon, k = 1, \dots, 2^l, l \in \mathbf{N}\}.$$

Let $\Delta_{k_1, \dots, k_l, l}^\varepsilon = \{x : x_1 \in \Delta_{k_1, l}^\varepsilon, \dots, x_l \in \Delta_{k_l, l}^\varepsilon\}$, $\mathcal{X}_l^\varepsilon = \sigma\{\Delta_{k_1, \dots, k_l, l}^\varepsilon, k_i \leq 2^l\}$. Notice that the set N_d of superscripts $\varepsilon \in [0, 1/2[$ such that Γ^ε are disjoint is uncountable (this follows from the observation that $\Gamma^\varepsilon \cap \Gamma^\eta = \emptyset$ if $\mathbf{Q}\varepsilon + \mathbf{Q} \neq \mathbf{Q}\eta + \mathbf{Q}$ and there are uncountably many different sets $\mathbf{Q}\varepsilon + \mathbf{Q}$). Let's consider the countable subset N_p of N_d containing all superscripts ε such that at least one of the probabilities $\nu(x : x_k \in \Gamma^\varepsilon)$, $k \in \mathbf{N}$, is positive. Thus, $N_d \setminus N_p$ is uncountable. It is clear that for any $\varepsilon \in N_d \setminus N_p$ the sequence of σ -algebras $\mathcal{X}_l^\varepsilon$ has the needed properties.

5.6. The following assertion is a generalization of Lemma 5.2.

PROPOSITION 5.2. *Let S_X be a compact subset in $\mathbf{P}(X)$, and let S_Y be a convex compact subset in $\mathbf{P}(Y)$. Assume that all elements of S_X are nonatomic. Let S be the set of all $m \in \mathbf{P}(X \times Y)$ such that $m(dx, dy) = \mu(x, dy)\nu(dx)$ with $\mu(x, \cdot) \in S_Y$ for all x and $\nu(\cdot) \in S_X$. Then S is a compact set.*

Proof. Since the relative compactness is evident, we need to show only that S is closed. Let us consider the sequence $m_n \in S$ with $m_n(dx, dy) = \mu_n(x, dy)\nu_n(dx)$ which tends in $\mathbf{P}(X \times Y)$ to $m(dx, dy) = \mu(x, dy)\nu(dx)$. As ν_n tends to ν in $\mathbf{P}(X)$ and S_X is a compact, $\nu \in S$.

To prove that $m \in S$ for all x , we construct a sequence of stochastic kernels $\tilde{\mu}_l$ such that $\tilde{\mu}_l(x, \cdot) \in S_Y$ for any x , $\tilde{\mu}_l(x, \cdot)$ converges ν -a.s. to some $\tilde{\mu}(x, \cdot)$, and $\tilde{\mu}(x, dy)\nu(dx) = \mu(x, dy)\nu(dx)$.

Let us consider the σ -algebras $\mathcal{X}_l = \sigma\{A_{k,l}, k = 1, \dots, r_l\}$, $l \in \mathbf{N}$, defined in Lemma 5.3. Since $\nu(\partial A_{k,l}) = 0$, the sequence of measures $m_n(A_{k,l} \times \cdot)$ converges in $\mathbf{P}(Y)$ to the measure $m(A_{k,l} \times \cdot)$ for any set $A_{k,l}$. From Lemma 5.1 it follows that for any $l \in \mathbf{N}$ there exists a stochastic kernel μ_l such that $\mu_l(t, \cdot) \in S_Y$ for all $t \in [0, 1]$ and

$$m(A \times \cdot) = \int_A \mu_l(x, \cdot)\nu(dx)$$

for all $A \in \mathcal{X}_l$. Let

$$\tilde{\mu}_l(x, \cdot) = \sum_{k=1}^{2^l} I_{A_{k,l}}(x)m_{l,k}(\cdot),$$

where

$$m_{l,k}(\cdot) = \frac{1}{\nu(A_{k,l})} \int_{A_{k,l}} \mu_l(x, \cdot)\nu(dx) \in S_Y$$

according to Lemma 5.1 (if $\nu(A_{k,l}) = 0$ we can put $m_{l,k}(\cdot)$ to be equal to any point of S_Y). By Proposition 5.1 on the convergence of measure-valued martingales the sequence $\tilde{\mu}_l(x, \cdot)$ tends to $\tilde{\mu}(x, \cdot)$ in $\mathbf{P}(Y)$ for almost all x and

$$\int_A \tilde{\mu}_l(x, \cdot)\nu(dx) = \int_A \tilde{\mu}(x, \cdot)\nu(dx)$$

for all $A \in \mathcal{X}_l$. Thus, we found a stochastic kernel μ such that $\tilde{\mu}(x, \cdot) \in S_Y$ for all $x \in [0, 1]$ and $m(A \times \Gamma) = \int_A \tilde{\mu}(x, \Gamma)\nu(dx)$ for all $A \in \mathcal{X}_l$, $l \in \mathbf{N}$, and $\Gamma \in \mathcal{Y}$. It follows that $m(dx, dy) = \tilde{\mu}(x, dy)\nu(dx)$. Hence, $m \in S$.

Remark 5.2. Walter Schachermayer suggested the following simpler proof of the above result without the assumption that measures from S_X are nonatomic. At first, notice that $S_Y = \cup_{j=1}^n \Gamma_j$, where $\Gamma_j := \{\mu : \mu(f_j) \leq \beta_j\}$, $f_j \in C_b(Y)$, $\beta_j \in \mathbf{R}$.

Indeed, from the Hahn–Banach theorem it follows that S_Y is an intersection of sets of this type. Their complements form an open covering of the open set $\mathbf{P}(Y) \setminus S_Y$. Since a Polish space is Lindelöf it contains a countable covering $\bar{\Gamma}_j$, $j \in \mathbf{N}$. Assume now that for the limiting measure $m(dx, dy) = \mu(x, dy)\nu(dx)$ there exists a set of positive ν -measure where $\mu(x, \cdot) \notin S_Y$. The above representation for S_Y implies that there exists a set $B = \{x : \mu(x, f) > \beta\}$ with $\nu(B) > 0$. Let $g_k \in C_b(X)$ be a sequence converging in $L^1(\nu)$ to I_B . Since $\mu_n(x, \cdot) \in S_Y$ we have that $\mu_n(x, f) \leq \beta$. Thus,

$$\begin{aligned} \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int \int g_k(x) f(y) m_n(dx, dy) &= \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int g_k(x) \mu_n(x, f) \nu_n(dx) \\ &\leq \lim_{k \rightarrow \infty} \beta \int g_k(x) \nu(dx) = \beta \nu(B). \end{aligned}$$

From the other side,

$$\begin{aligned} \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int \int g_k(x) f(y) m_n(dx, dy) &= \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int \int g_k(x) f(y) m(dx, dy) \\ &= \lim_{k \rightarrow \infty} \int g_k(x) \mu(dx, f) \nu(dx) = \int_B \mu(dx, f) \nu(dx) > \beta \nu(B), \end{aligned}$$

and we get a contradiction to the assumption that $\mu(x, \cdot)$ does not belong to S_Y ν -a.s.

5.7. Now we consider the following problem.

Let (Ω, \mathcal{F}, P) be a probability space, \mathcal{P} be a σ -algebra in the product $\Omega \times \mathbf{R}_+$ such that $\mathcal{P} \subseteq \mathcal{F} \otimes \mathcal{B}(\mathbf{R}_+)$, Γ is a measurable set-valued mapping from $(\mathbf{R}_+, \mathcal{B}(\mathbf{R}_+))$ to \mathbf{R}^q . Measurability means that the graph $\text{Gr } \Gamma = \{(t, x) : x \in \Gamma(t)\}$ is a $\mathcal{B}(\mathbf{R}_+) \otimes \mathcal{B}(\mathbf{R}^q)$ -measurable set. We shall assume that $\Gamma(t)$ are closed sets and there exists a function $r \in L^1(\mathbf{R}, dt)$ such that $|\Gamma(t)| \leq r_t$ for all t . Let \mathcal{V} be a set of all \mathcal{P} -measurable functions f on $\Omega \times \mathbf{R}_+$ such that $f(\omega, t) \in \Gamma(t)$. Define the set K in $\mathbf{P}(\mathbf{R}^q)$ as

$$K := \left\{ \mathcal{L}(\phi) : \phi = \int_0^\infty f(t) dt, f \in \mathcal{V} \right\}.$$

The question is if K is a compact set. We give here only a partial answer to this question imposing some specific assumption on the structure of the σ -algebra \mathcal{P} .

Let $w = (w_t)$ be a d -dimensional Wiener process on (Ω, \mathcal{F}, P) , $\mathcal{F}_t^{o,w} = \sigma\{w_s, s \leq t\}$, $\mathcal{F}_t^w = \mathcal{F}_{t+}^{o,w} \vee \mathcal{N}$, where \mathcal{N} is a family of all sets from \mathcal{F} of zero probability. In other words, $\mathbf{F}^w = (\mathcal{F}_t^w)$ is the minimal filtration generated by the Wiener process and satisfying the usual assumptions.

LEMMA 5.4. *Assume that \mathcal{P} is the predictable σ -algebra generated by \mathbf{F}^w and $\Gamma(t)$ is a convex set for all t . Then K is a compact set.*

Proof. Since random variables ϕ are bounded by some constant, K is relatively compact and it remains to show that K is closed.

Let us consider the sequence $f^n \in \mathcal{V}$ such that the corresponding sequence of distribution $\mathcal{L}(\phi^n)$ converges in $\mathbf{P}(\mathbf{R}^q)$. Define the random processes

$$\phi_t^n = \int_0^t f^n(\omega, s) ds.$$

Using the criteria of relative compactness in $\mathbf{P}(C^{q+d}(\mathbf{R}_+))$ (the space $C^{q+d}(\mathbf{R}_+)$ is equipped with the metric $\sum_j 2^{-j} \|x\|_j (1 + \|x\|_j)^{-1}$), we can assume without loss

of generality that the sequence $\mathcal{L}((\phi^n, w))$ converges to some \mathcal{L} in $\mathbf{P}(C^{q+d}(\mathbf{R}_+))$. The Skorohod theorem asserts that on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ (actually, on the standard unit interval) there are processes $(\tilde{\phi}^n, \tilde{w}^n)$, $n \in \mathbf{N}$, and $(\tilde{\phi}, \tilde{w})$ such that $\mathcal{L}(\tilde{\phi}^n, \tilde{w}^n) = \mathcal{L}(\phi^n, w)$, $\mathcal{L}(\tilde{\phi}, \tilde{w}) = \mathcal{L}$, and $(\tilde{\phi}^n, \tilde{w}^n)$ converges to $(\tilde{\phi}, \tilde{w})$ in $C^{q+d}(\mathbf{R}_+)$ pointwise.

It is easy to show that the following properties hold:

(1) The process $\tilde{\phi}^n$ is adapted with respect to $(\tilde{\mathcal{F}}_t^n)$, where $\tilde{\mathcal{F}}_t^n := \sigma\{\tilde{w}_s^n, s \leq t\}$ and

$$(5.1) \quad \tilde{\phi}_t^n(\tilde{\omega}) = \int_0^t \tilde{f}^n(\tilde{\omega}, s) ds$$

with $\tilde{\mathcal{P}}^n$ -measurable \tilde{f}^n such that $\tilde{f}^n(\tilde{\omega}, s) \in \Gamma(s)$ for $(\tilde{\omega}, s)$ (where $\tilde{\mathcal{P}}^n$ is the predictable σ -algebra generated by $(\tilde{\mathcal{F}}_t^n)$).

(2) The process $\tilde{\phi}$ is adapted with respect to $(\tilde{\mathcal{F}}_t)$, where $\tilde{\mathcal{F}}_t := \sigma\{\tilde{w}_s, s \leq t\}$ and

$$(5.2) \quad \tilde{\phi}_t(\tilde{\omega}) = \int_0^t \tilde{f}(\tilde{\omega}, s) ds$$

with $\tilde{\mathcal{P}}$ -measurable \tilde{f} such that $\tilde{f}(\tilde{\omega}, s) \in \Gamma(s)$ for $(\tilde{\omega}, s)$ (where $\tilde{\mathcal{P}}$ is the predictable σ -algebra generated by the minimal filtration with the usual assumptions for \tilde{w}).

Let us prove that $\tilde{\phi}^n$ is adapted with respect to $(\tilde{\mathcal{F}}_t^n)$. Fix $t \in \mathbf{R}_+$ and define the Wiener process $\tilde{w}_s^n = \tilde{w}_{s+t}^n - \tilde{w}_t^n$, $s \in \mathbf{R}_+$, which is independent of $\tilde{\mathcal{F}}_t^n$. It is sufficient to show that $\tilde{E}(\tilde{\phi}_t^n | \tilde{\mathcal{F}}_t^n) = \tilde{\phi}_t^n$ (\tilde{P} -a.s.) or, equivalently, that

$$\tilde{E}\tilde{E}(\tilde{\phi}_t^n | \tilde{\mathcal{F}}_t^n)h(\tilde{w}^n)g(\hat{w}^n) = \tilde{E}\tilde{\phi}_t^n h(\tilde{w}^n)g(\hat{w}^n)$$

for any bounded continuous functions $h : C^d[0, t] \rightarrow \mathbf{R}$ and $g : C^d(\mathbf{R}_+) \rightarrow \mathbf{R}$ (the argument of h , in fact, is the restriction of \tilde{w}^n to $[0, t]$). Since $h(\tilde{w}^n)$ is $\tilde{\mathcal{F}}_t^n$ -measurable, it follows from properties of the conditional expectations that the above equality holds if and only if

$$(5.3) \quad \tilde{E}\tilde{\phi}_t^n h(\tilde{w}^n)\tilde{E}g(\hat{w}^n) = \tilde{E}\tilde{\phi}_t^n h(\tilde{w}^n)g(\hat{w}^n).$$

But $\mathcal{L}(\tilde{\phi}^n, \tilde{w}^n) = \mathcal{L}(\phi^n, w)$, and the last identity is equivalent to the following one:

$$E\phi_t^n h(w)Eg(w) = E\phi_t^n h(w)g(w'),$$

where $w'_s = w_{s+t} - w_t$, $s \in \mathbf{R}_+$, which holds because ϕ^n is adapted with respect to (\mathcal{F}_t^n) .

Taking a limit in (5.3) we get that

$$\tilde{E}\tilde{\phi}_t h(\tilde{w})\tilde{E}g(\hat{w}) = \tilde{E}\tilde{\phi}_t h(\tilde{w})g(\hat{w}),$$

where $\hat{w}_s = \tilde{w}_{s+t} - \tilde{w}_t$, $s \in \mathbf{R}_+$. As above, this means that $\tilde{\phi}_t = \tilde{E}(\tilde{\phi}_t | \tilde{\mathcal{F}}_t)$; i.e., $\tilde{\phi}$ is adapted with respect to $(\tilde{\mathcal{F}}_t)$. The representation (5.1) follows from the definition of ϕ^n and coincidence of $\mathcal{L}(\tilde{\phi}^n, \tilde{w}^n)$ and $\mathcal{L}(\phi^n, w)$. To obtain the representation (5.2) we notice that by the Komlós theorem [19] for the bounded sequence \tilde{f}^n , there exists a subsequence (n_j) such that $(\tilde{f}^{n_1} + \dots + \tilde{f}^{n_k})/k$ converge to some function \tilde{f}^0 for almost all $(\tilde{\omega}, t)$. It follows that

$$(5.4) \quad \tilde{\phi}_t(\tilde{\omega}) = \int_0^t \tilde{f}^0(\tilde{\omega}, s) ds.$$

The convexity assumption implies that $\tilde{f}^0(\tilde{\omega}, s) \in \Gamma(s)$ for almost all $(\tilde{\omega}, s)$, and we can assume without loss of generality that $\tilde{f}^0(\tilde{\omega}, s) \in \Gamma(s)$ for all $(\tilde{\omega}, s)$. This means that the trajectories of $\tilde{\phi}$ are absolutely continuous functions. Let

$$\tilde{f}'(\tilde{\omega}, s) = \limsup_{m \rightarrow \infty} \sum_{i=2}^m I_{\Delta_i}(s) 2^m (\tilde{\phi}_{t_{i-1}}(\tilde{\omega}) - \tilde{\phi}_{t_{i-2}}(\tilde{\omega})),$$

where $t_i = i2^{-m}$, $\Delta_i = t_i - t_{i-1}$. Clearly, $\tilde{f}'(\tilde{\omega}, s)$ is a $\tilde{\mathcal{P}}$ -measurable function, and for all $\tilde{\omega}$ and almost all s it coincides with $\tilde{f}^0(\tilde{\omega}, s) \in \Gamma(s)$. Thus, the following function gives the representation (5.2) with the required properties:

$$\tilde{f}(\tilde{\omega}, s) = \tilde{f}'(\tilde{\omega}, s)I_A + x(s)I_{\bar{A}},$$

where $A = \{(\tilde{\omega}, s) : \tilde{f}'(\tilde{\omega}, s) \in \Gamma(s)\}$, $x(s)$ is any Borel function such that $x(s) \in \Gamma(s)$.

Properties (1) and (2) imply the result. Indeed, it follows from (2) and Lemma 2.1 in [13] that there exists a predictable function $a(x, s) : C^d(\mathbf{R}_+) \times \mathbf{R}_+ \rightarrow \mathbf{R}^q$ such that $\tilde{f}(\tilde{\omega}, s) = a(\tilde{w}(\tilde{\omega}), s)$. Evidently, we can modify $a(x, s)$ in such a way that $a(x, s) \in \Gamma(s)$ for all (x, s) . Let us define on the original probability space (Ω, \mathcal{F}, P) the process

$$\phi_t(\omega) = \int_0^t f(\omega, s) ds$$

with $f(\omega, s) = a(w(\omega), s)$. Since $f \in \mathcal{V}$ and $\mathcal{L}(\phi) = \mathcal{L}(\tilde{\phi}) = \mathcal{L}$ it follows that the limit of $\mathcal{L}(\phi^n)$ belongs to K and the lemma is proved.

5.8. Now we apply the previous result to our specific setting.

LEMMA 5.5. *The set $S_Y^0 := \{\mathcal{L}(\xi_0 + I(v)) : v \in \mathcal{V}_U\}$ is compact in $\mathbf{P}(\mathbf{R}^n)$.*

Proof. Reversing the time and taking into account the notations of the previous subsection we can reduce the problem to the question of whether the set

$$K := \left\{ \mathcal{L}(\phi) : \phi = \int_0^\infty f(t) dt, f \in \mathcal{V} \right\}$$

is compact. Here $\Gamma(t) = -s^{-2} \exp\{A_4(T)/s\} B_2(T)U$ and the σ -algebra \mathcal{P} is generated by the time reverse of the Ornstein–Uhlenbeck process $\xi_{1/t}$, or, equivalently, by the process $\eta_t := t\xi_{1/t}$. The process η (as well as ξ) is defined in the present context only up to the distribution. For example, we can take as η the process defined by the stochastic differential equations

$$(5.5) \quad d\eta_t = t^{-2}(tI - A)\eta_t dt + dw_t, \quad \eta_0 = 0,$$

where I is the unit matrix and w is the Wiener process. This representation can be deduced from the differential equation for the Ornstein–Uhlenbeck process by the Ito formula. But from equation (5.5) it follows that $\mathcal{F}_t^{\eta, w} = \sigma\{\eta_s, s \leq t\}$ and the needed result is a corollary of Lemma 5.4.

5.9. Let η_i be random variables with values in Polish spaces (X_i, \mathcal{X}_i) , $i = 1, 2, 3$, let ν_i be the distribution of η_i , and let $\mu_{ij}(x_j, dx_i)$ be the regular conditional distribution of η_i given η_j .

LEMMA 5.6. *Let $\eta_3 = f(\eta_2)$ for some measurable function $f : X_2 \rightarrow X_3$, and let S_1 be a compact convex set in $\mathbf{P}(X_1)$. Assume that $\mu_{12}(x_2, dx_1) \in S_1$ for all x_2 . Then $\mu_{13}(x_3, dx_1) \in S_1$ for ν_3 -almost all x_3 .*

Proof. The assertion follows from the relation

$$\mu_{13}(x_3, dx_1) = \int_{X_2} \mu_{12}(x_2, dx_1) \mu_{23}(x_3, dx_2) \quad (\nu_3\text{-a.e.})$$

and Remark 5.1.

Acknowledgment. The authors express their thanks to Walter Schachermayer for helpful discussions of functional-analytic aspects of the problem.

REFERENCES

- [1] V. E. BENEŠ, *Existence of optimal control laws*, SIAM J. Control, 9 (1971), pp. 446–475.
- [2] A. BENSOUSSAN, *On some singular perturbation problems arising in stochastic control*, Stochastic Anal. Appl., 2 (1984), pp. 13–53.
- [3] A. BENSOUSSAN, *Perturbation Methods in Optimal Control*, J. Wiley/Gauthier Villars, New York, 1988.
- [4] A. BENSOUSSAN, *Optimal Control of Partially Observed Systems*, Cambridge University Press, Cambridge, UK, 1992.
- [5] A. BENSOUSSAN AND G. L. BLANKENSHIP, *Singular perturbations in stochastic control*, in Singular Perturbations and Asymptotic Analysis in Control Systems, Lecture Notes in Control and Inform. Sci. 90, P. Kokotović, A. Bensoussan, and G. Blankenship, eds. Springer-Verlag, Berlin, 1987.
- [6] C. DELLACHERIE AND P.-A. MEYER, *Probabilities and Potentials*, North-Holland, Amsterdam, 1978.
- [7] A. L. DONTCHEV, *Perturbations, Approximations and Sensitivity Analysis of Optimal Control Systems*, Lecture Notes in Control and Inform. Sci. 52, Springer-Verlag, Berlin, 1983.
- [8] A. L. DONTCHEV AND V. M. VELIOV, *Singular perturbation in Mayer's optimization problem for linear systems*, SIAM J. Control Optim., 21 (1983), pp. 566–581.
- [9] S. N. ETHIER AND T. G. KURTZ, *Markov Processes: Characterization and Convergence*, J. Wiley, New York, 1986.
- [10] V. GAITSGORY, *Suboptimization of singularly perturbed control systems*, SIAM J. Control Optim., 30 (1992), pp. 1228–1249.
- [11] W. HILDENBRANDT, *Core and Equilibria of a Large Economy*, Princeton University Press, Princeton, NJ, 1974.
- [12] YU. M. KABANOV, R. SH. LIPTSER, AND A. N. SHIRYAEV, *On the variation distance for the probability measures defined on a filtered space*, Probability Theory Related Fields, 71 (1986), pp. 19–35.
- [13] YU. M. KABANOV AND S. M. PERGAMENSHCHIKOV, *Optimal control of singularly perturbed linear stochastic systems*, Stochastics Stochastics Rep., 36 (1991), pp. 109–135.
- [14] YU. M. KABANOV, S. M. PERGAMENSHCHIKOV, AND J. M. STOYANOV, *Asymptotic expansions for singularly perturbed stochastic differential equations*, in New Trends in Probability and Statistics, Vol. 1., V. V. Sazonov and T. L. Shervashidze, eds., Coronet Books, Philadelphia, PA, 1991, pp. 413–435.
- [15] YU. M. KABANOV AND W. RUNGALDIER, *On control of two-scale stochastic systems with linear dynamics in the fast variables*, Math. Control Systems Signals, to appear.
- [16] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.
- [17] P. V. KOKOTOVIĆ AND H. K. KHALIL, eds., *Singular Perturbations in Systems and Control*, IEEE Press, New York, 1986.
- [18] P. V. KOKOTOVIĆ, H. K. KHALIL, AND J. O'REILLY, *Singular Perturbation Methods in Control: Analysis and Design*, Academic Press, New York, 1986.
- [19] J. KOMLOŠ, *A generalization of a problem of Steinhaus*, Acta Math. Sci. (Hung.), 18 (1967), pp. 217–229.
- [20] H. J. KUSHNER, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser, Boston, 1990.