Dynamic Systems and Applications 16 (2007) 49-72

DIFFERENTIAL INCLUSIONS WITH MEAN DERIVATIVES

SVETLANA V. AZARINA AND YURI E. GLIKLIKH

Department of Algebra and Topological Methods of Analysis, Mathematics Faculty, Voronezh State University, Universitetskaya pl. 1, 394006 Voronezh, Russia

azarinas@mail.ru yeg@math.vsu.ru

ABSTRACT. We introduce and investigate a new sort of stochastic differential inclusions given in terms of mean derivatives of a stochastic process, introduced by E. Nelson for the needs of the so called stochastic mechanics. This class of stochastic inclusions is ideologically the closest one to ordinary differential inclusions. We consider three types of inclusions: with forward mean derivatives, with backward mean derivatives and with current velocities (symmetric mean derivatives). These types have different properties and physical meaning. Some existence of solutions results are proved.

AMS (MOS) Subject Classification. 34A60 60H10 60H99.

Key words: Mean derivatives, differential inclusions, stochastic processes

INTRODUCTION

The notion of mean derivatives was introduced by Edward Nelson (see [14, 15, 16]) for the needs of stochastic mechanics (a version of quantum mechanics). The equation of motion in this theory (called the Newton-Nelson equation) was the first example of equations in mean derivatives. Later it turned out that the equations in mean derivatives arose also in the description of motion of viscous incompressible fluid (see, e.g., [5, 6, 9, 10]), in the description of Navier-Stokes vortices (see, e.g., [11]), etc. In [7, 8] (see also [10]) investigation of the equations in mean derivatives as a special class of stochastic differential equations was started.

In all above-mentioned cases the solutions of the equations were supposed to be Ito diffusion type processes (or even Markov diffusion processes) whose diffusion summand was given a priory since the classical Nelson's mean derivatives yield, roughly speaking, only the drift term (forward, backward, etc.) of a stochastic process. In this paper we present a two-fold generalization of the theory. First, giving a slight modification of a certain Nelson's idea, we introduce a new type of mean derivative that is responsible for diffusion term. And second, we investigate the differential inclusions with mean derivatives, i.e., equations with set-valued right-hand sides.

Received August 30, 2006

The stochastic differential inclusions with mean derivatives is ideologically the closest class to ordinary differential inclusions and so they arise in applications in more direct way than usual stochastic differential inclusions. Here we introduce and investigate the first order inclusions with forward mean derivatives, with backward mean derivatives and with current velocities (symmetric mean derivatives) and prove some existence of solutions theorems for them. As usual, the inclusions with forward mean derivatives play the basic role for the entire theory. Probably the inclusions with current velocities are the most interesting from the physical point of view since the current velocity is considered for stochastic processes as an analog of ordinary velocity of a deterministic trajectory.

The structure of our paper is as follows. In Section 1 we describe necessary preliminary facts from the theory of mean derivatives and from stochastic differential equations.

Section 2 is devoted to differential inclusions with forward mean derivatives. This type of inclusions looks the most natural for describing stochastic processes with control and in other cases analogous to those where ordinary differential inclusions arise. The results of this section also give the basis for the next sections.

Equations and inclusions with backward mean derivatives appear in applications as well. Say, the above-mentioned equation, arising in description of viscous incompressible fluids, is an equation with backward mean derivatives. In Section 2 we present the simplest results on inclusions with backward derivatives.

In Section 3 we consider inclusions with current velocities, probably the most interesting from the physical point of view, but essentially more complicated for investigation than the other types of inclusions with mean derivatives.

Some remarks on notations. In this paper we deal with equations and inclusions in the linear space $\mathbb{R}^n$, for which we always use coordinate presentation of vectors and linear operators. Vectors in $\mathbb{R}^n$ are considered as columns. If X is such a vector, the transposed row vector is denoted by X^*. Linear operators from $\mathbb{R}^n$ to $\mathbb{R}^n$ are represented as $n \times n$ matrices, the symbol $*$ means transposition of a matrix (pass to the matrix of conjugate operator). The space of $n \times n$ matrices is denoted by $L(\mathbb{R}^n, \mathbb{R}^n)$.

By $\mathrm{S}(n)$ we denote the linear space of symmetric $n \times n$ matrices that is a subspace in $L(\mathbb{R}^n, \mathbb{R}^n)$. The symbol $\mathrm{S}_+(n)$ denotes the set of positive definite symmetric $n \times n$ matrices that is a convex open set in $\mathrm{S}(n)$. Its closure, i.e., the set of positive semi-definite symmetric $n \times n$ matrices, is denoted by $\bar{\mathrm{S}}_+(n)$.

Everywhere below for a set B in $\mathbb{R}^n$ or in $L(\mathbb{R}^n, \mathbb{R}^n)$ we use the norm introduced by usual formula $\|B\| = \sup_{y \in B} \|y\|$.

For the sake of simplicity we consider equations, their solutions and other objects on a finite time interval $t \in [0, T]$.

Everywhere in the paper we use Einstein's summation convention with respect to the coinciding upper and lower indices.

1. MEAN DERIVATIVES

In this section we briefly describe preliminary facts about mean derivatives. See details in [15, 16, 5, 6, 10].

Consider a stochastic process $\xi(t)$ in $\mathbb{R}^n$, $t \in [0, T]$, given on a certain probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and such that $\xi(t)$ is an L_1 random element for all t. It is known that such a process determines 3 families of σ-subalgebras of the σ-algebra $\mathcal{F}$:

(i) "the past" $\mathcal{P}_t^\xi$ generated by preimages of Borel sets from $\mathbb{R}^n$ under all mappings $\xi(s) : \Omega \to \mathbb{R}^n$ for $0 \le s \le t$;

(ii) "the future" $\mathcal{F}_t^\xi$ generated by preimages of Borel sets from $\mathbb{R}^n$ under all mappings $\xi(s) : \Omega \to \mathbb{R}^n$ for $t \le s \le T$;

(iii) "the present" ("now") $\mathcal{N}_t^\xi$ generated by preimages of Borel sets from $\mathbb{R}^n$ under the mapping $\xi(t) : \Omega \to \mathbb{R}^n$.

All the above families we suppose to be complete, i.e., containing all sets of probability zero.

For the sake of convenience we denote by E_t^ξ the conditional expectation $E(\cdot|\mathcal{N}_t^\xi)$ with respect to the "present" $\mathcal{N}_t^\xi$ for $\xi(t)$.

Following [14, 15, 16], introduce the following notions of forward and backward mean derivatives.

Definition 1.1. (i) The forward mean derivative $D\xi(t)$ of $\xi(t)$ at the time instant t is an L_1 random element of the form

$$D\xi(t) = \lim_{\triangle t \to +0} E_t^\xi \Big(\frac{\xi(t + \triangle t) - \xi(t)}{\triangle t}\Big), \tag{1.1}$$

where the limit is supposed to exist in $L_1(\Omega, \mathcal{F}, \mathsf{P})$ and $\triangle t \to +0$ means that $\triangle t$ tends to 0 and $\triangle t > 0$.

(ii) The backward mean derivative $D_*\xi(t)$ of $\xi(t)$ at t is the L_1-random element

$$D_*\xi(t) = \lim_{\Delta t \to +0} E_t^\xi \Big(\frac{\xi(t) - \xi(t - \Delta t)}{\Delta t}\Big) \tag{1.2}$$

where (as well as in (i)) the limit is assumed to exist in $L^1(\Omega, \mathcal{F}, \mathsf{P})$ and $\Delta t \to +0$ means that $\Delta t \to 0$ and $\Delta t > 0$.

Remark 1.2. If $\xi(t)$ is a Markov process then evidently E_t^ξ can be replaced by $E(\cdot|\mathcal{P}_t^\xi)$ in (1.1) and by $E(\cdot|\mathcal{F}_t^\xi)$ in (1.2). In initial Nelson's works there were two versions of definition of mean derivatives: as in our Definition 1.1 and with conditional expectations with respect to "past" and "future" as above that coincide for Markov processes. We shall not suppose $\xi(t)$ to be a Markov process and give the definition with conditional expectation with respect to "present" taking into account the physical principle of locality: the derivative should be determined by the present state of the system, not by its past or future.

We also shall use the following generalizations of the notions of forward and backward mean derivatives (see, e.g., [10]):

Definition 1.3. The forward mean derivative $D^{\xi}\eta(t)$ of $\eta(t)$ with respect to $\xi(t)$ at the time instant t is an L_1 random element of the form

$$D^{\xi}\eta(t) = \lim_{\triangle t \to +0} E^{\xi}_{t}\Big(\frac{\eta(t+\triangle t)-\eta(t)}{\triangle t}\Big), \tag{1.3}$$

and backward derivative of $D^{\xi}_{*}\eta(t)$ of $\eta(t)$ with respect to $\xi(t)$ by the formula

$$D^{\xi}_{*}\eta(t) = \lim_{\Delta t \to +0} E^{\xi}_{t}\Big(\frac{\eta(t)-\eta(t-\Delta t)}{\Delta t}\Big) \tag{1.4}$$

where the limits are supposed to exist in $L_1(\Omega, \mathcal{F}, \mathsf{P})$ and $\triangle t \to +0$ means that $\triangle t$ tends to 0 and $\triangle t > 0$.

Introduce the differential operator D_2 that differentiates an L_1 random process $\xi(t)$, $t \in [0,T]$ according to the rule

$$D_2\xi(t) = \lim_{\triangle t \to +0} E^{\xi}_{t}\Big(\frac{(\xi(t+\triangle t)-\xi(t))(\xi(t+\triangle t)-\xi(t))^{*}}{\triangle t}\Big), \tag{1.5}$$

where $(\xi(t+\triangle t)-\xi(t))$ is considered as a column vector (vector in $\mathbb{R}^n$), $(\xi(t+\triangle t)-\xi(t))^{*}$ is a row vector (transposed, or conjugate vector) and the limit is supposed to exists in $L_1(\Omega, \mathcal{F}, \mathsf{P})$. We emphasize that the matrix product of a column on the left and a row on the right is a matrix so that $D_2\xi(t)$ is a symmetric semi-positive definite matrix function on $[0,T]\times\mathbb{R}^n$. We call D_2 the quadratic mean derivative.

Remark 1.4. From the properties of conditional expectation (see, e.g., [17]) it follows that there exist Borel mappings $a(t,x)$, $a_*(t,x)$ and $\alpha(t,x)$ from $R\times\mathbb{R}^n$ to $\mathbb{R}^n$ and to $\bar{\mathrm{S}}_+$, respectively, such that $D\xi(t) = a(t,\xi(t))$, $D_*\xi(t) = a_*(t,\xi(t))$ and $D_2\xi(t) = \alpha(t,\xi(t))$. Following [17] we call $a(t,x)$, $a_*(t,x)$ and $\alpha(t,x)$ the regressions.

Ito process is a process $\xi(t)$ of the form

$$\xi(t) = \xi_0 + \int_0^t a(s)ds + \int_0^t A(s)dw(s),$$

where $a(t)$ is a process in $\mathbb{R}^n$ whose sample paths a.s. have bounded variation; $A(t)$ is a process in $L(\mathbb{R}^n,\mathbb{R}^n)$ such that for any element $A^j_i(t)$ of matrix $A(t)$ the condition

$\mathsf{P}(\omega|\int_0^T (A_i^j)^2 dt < \infty) = 1$ holds; $w(t)$ is a Wiener process in $\mathbb{R}^n$; the first integral is Lebesgue integral, the second one is Itô integral and all integrals are well-posed.

Recall that for an Itô process the column vector $a(t)$ is called drift and $\alpha(t) = A(t)A^*(t) \in \bar{\mathrm{S}}_+(n)$, where $A^*(t)$ is the transposed matrix, is called the diffusion coefficient. Notice that indeed AA^* is a square symmetric positive semi-definite $n \times n$ matrix, i.e., a matrix from $\bar{\mathrm{S}}_+(n)$.

Definition 1.5. An Itô process $\xi(t)$ is called a process of diffusion type if $a(t)$ and $A(t)$ are not anticipating with respect to $\mathcal{P}_t^\xi$ and the Wiener process $w(t)$ is adapted to $\mathcal{P}_t^\xi$. If $a(t) = a(t, \xi(t))$ and $A(t) = A(t, \xi(t))$, where $a(t, x)$ and $A(t, x)$ are Borel measurable mappings from $[0, T] \times \mathbb{R}^n$ to $\mathbb{R}^n$ and to $L(\mathbb{R}^n, \mathbb{R}^n)$, respectively, the Itô process is called a diffusion process.

Diffusion type processes are solutions of the so called diffusion type equations that are described as follows. Denote by $C^0([0, T], \mathbb{R}^n)$ the Banach space of continuous maps (curves) from the interval $[0, T] \subset \mathbb{R}$ to $\mathbb{R}^n$. Consider the mappings

$$a : [0, T] \times C^0([0, T], \mathbb{R}^n) \to \mathbb{R}^n,$$

$$A : [0, T] \times C^0([0, T], \mathbb{R}^n) \to L(\mathbb{R}^n, \mathbb{R}^n).$$

Let $a(t, x(\cdot))$ and $A(t, x(\cdot))$ be continuous jointly in all variables and let for all $t \in [0, T]$ the mappings $a(t, \cdot)$ and $A(t, \cdot)$ be measurable with respect to the σ-algebra, generated by cylinder sets with bases on $[0, t]$.

Definition 1.6. The Itô type equation

$$\xi(t) = \int_0^t a(t, \xi(\cdot)) d\tau + \int_0^t A(t, \xi(\cdot)) dw(\tau) \tag{1.6}$$

is called a diffusion type stochastic differential equation.

Let us turn back to mean derivatives. Taking into account the properties of conditional expectation and the fact that $\mathcal{N}_t^\xi$ is a σ-subalgebra in $\mathcal{P}_t^\xi$, it is clear (see, e.g., [10]) that for any martingale $\eta(t)$ with respect to $\mathcal{P}_t^\xi$ the equality $D^\xi \eta(t) = 0$ holds. Since for a diffusion type process the integral $\int_0^t A(s) dw(s)$ is a martingale with respect to $\mathcal{P}_t^\xi$, the following statement takes place (see, e.g., [5, 6, 10]):

Theorem 1.7. *For an Itô diffusion type process $\xi(t)$ the mean derivative $D\xi(t)$ exists and equals $E_t^\xi(a(t))$. In particular, if $\xi(t)$ a diffusion process, $D\xi(t) = a(t, \xi(t))$.*

Theorem 1.8. *Let $\xi(t)$ be a diffusion type process. Then $D_2\xi(t) = E_t^\xi[\alpha(t)]$ where $\alpha(t)$ is the diffusion coefficient. In particular, if $\xi(t)$ is a diffusion process, $D_2\xi(t) = \alpha(t, \xi(t))$ where α is the diffusion coefficient.*

Proof. Since $\xi(t+\triangle t)-\xi(t)=\int_t^{t+\triangle t}a(s)ds+\int_t^{t+\triangle t}A(s)dw(s)$, taking into account the properties of Lebesgue and Itô integrals one can see that $(\xi(t+\triangle t)-\xi(t))(\xi(t+\triangle t)-\xi(t))^*$ is approximated by $a(t)a^*(t)(\Delta t)^2+(a(t)\Delta t)(A(t)\Delta w(t))^*+(A(t)\Delta w(t))(a(t)\Delta t)^*+A(t)A^*(t)\Delta t$. Applying formula (1.5) we obtain the assertion of Theorem since $AA^*=\alpha$ (see above). □

Definition 1.9. The derivative $D_S=\frac{1}{2}(D+D_*)$ is called the symmetric mean derivative. The derivative $D_A=\frac{1}{2}(D-D_*)$ is called the antisymmetric mean derivative.

Consider the vectors $v^\xi(t,x)=\frac{1}{2}(a(t,x)+a_*(t,x))$ and $u^\xi(t,x)=\frac{1}{2}(a(t,x)-a_*(t,x))$.

Definition 1.10. $v^\xi(t)=v^\xi(t,\xi(t))=D_S\xi(t)$ is called the current velocity of the process $\xi(t)$; $u^\xi(t)=u^\xi(t,\xi(t))=D_A\xi(t)$ is called the osmotic velocity of the process $\xi(t)$.

The physical meaning of v^ξ and u^ξ is as follows (cf. [15, 16]). Let $\xi(t)$ describe the motion of a physical process, say the motion of a particle (we are sure that all physical motions are random with very small dispersion so that it usually looks natural to omit randomness from consideration). Then the current velocity v^ξ is what we usually consider as ordinary physical velocity while the osmotic velocity u^ξ shows how fast the particle "diffuses" into the enveloping continuum, i.e., how fast the "randomness" is changing. This interpretation has the following mathematical motivation.

Consider an autonomous smooth field of non-degenerate linear operators $A(x):\mathbb{R}^n\to\mathbb{R}^n$, $x\in\mathbb{R}^n$. Suppose that $\xi(t)$ is a diffusion type process whose diffusion integrand is $A(\xi(t))$. Then its diffusion coefficient $A(x)A^*(x)$ is a smooth field of symmetric positive definite matrices $\alpha(x)=(\alpha^{ij}(x))$. Since all those matrices are non-degenerate, the field of inverse matrices (α_{ij}) exists and is smooth and at any x the matrix $(\alpha_{ij})(x)$ is symmetric and positive definite. Thus it defines a new Riemannian metric $\alpha(\cdot,\cdot)=\alpha_{ij}dx^idx^j$ on $\mathbb{R}^n$. Consider the Riemannian volume form of this Riemannian metric $\Lambda_\alpha=\sqrt{\det(\alpha_{ij})}dx^1\wedge dx^2\wedge\cdots\wedge dx^n$ (see, e.g., [19]).

Denote by $\rho^\xi(t,x)$ the probability density of $\xi(t)$ with respect to the volume form $dt\wedge\Lambda_\alpha=\sqrt{\det(\alpha_{ij})}dt\wedge dx^1\wedge dx^2\wedge\cdots\wedge dx^n$ on $[0,T]\times\mathbb{R}^n$, i.e., for any continuous bounded function $f:[0,T]\times\mathbb{R}^n\to R$ the relation

$$\int_0^T E(f(t,\xi(t)))dt=\int_0^T(\int_\Omega f(t,\xi(t))d\mathsf{P})dt=\int_{[0,T]\times\mathbb{R}^n} f(t,x)\rho^\xi(t,x)dt\wedge\Lambda_\alpha$$

holds. Then (see [16])

$$u^\xi(t,x)=\frac{1}{2}Grad\log\rho^\xi(t,x)=Grad\log\sqrt{\rho^\xi(t,x)},\tag{1.7}$$

where $Grad$ denotes the gradient with respect to the Riemannian metric $\alpha(\cdot,\cdot)$, and for $v^\xi(t,x)$ and $\rho^\xi(t,x)$ the so called equation of continuity

$$\frac{\partial \rho^\xi(t,x)}{\partial t} = -Div(v^\xi(t,x)\rho^\xi(t,x)) \tag{1.8}$$

holds, where Div denotes divergence with respect to the Riemannian metric $\alpha(\cdot,\cdot)$.

2. DIFFERENTIAL INCLUSIONS WITH FORWARD MEAN DERIVATIVES

Let Borel measurable mappings $a(t,x)$ and $\alpha(t,x)$ from $[0,T]\times\mathbb{R}^n$ to $\mathbb{R}^n$ and to $\bar{\mathrm{S}}_+(n)$, respectively, be given. We call the system of the form

$$\begin{cases} D\xi(t) = a(t,\xi(t)), \\ D_2\xi(t) = \alpha(t,\xi(t)), \end{cases} \tag{2.1}$$

a first order differential equation with forward mean derivatives.

Taking into account Remark 1.4, Theorem 1.7 and Theorem 1.8, one can easily see that the problem of finding a diffusion type process that P-a.s. satisfies (2.1), is well-posed. It is clear that the first equation of (2.1) determines the drift and the second one determines the diffusion coefficient of the process.

Let $\mathbf{a}(t,x)$ and $\boldsymbol{\alpha}(t,x)$ be set-valued mappings from $[0,T]\times\mathbb{R}^n$ to $\mathbb{R}^n$ and to $\bar{\mathrm{S}}_+(n)$, respectively. The system of the form

$$\begin{cases} D\xi(t) \in \mathbf{a}(t,\xi(t)), \\ D_2\xi(t) \in \boldsymbol{\alpha}(t,\xi(t)). \end{cases} \tag{2.2}$$

will be called a first order differential inclusion in forward mean derivatives.

Definition 2.1. We say that (2.2) has a weak solution on $[0,T]$ with initial condition $\xi(0)=x_0$, if there exist a probability space $(\Omega,\mathcal{F},\mathsf{P})$ and a process $\xi(t)$ given on $(\Omega,\mathcal{F},\mathsf{P})$ and taking values in $\mathbb{R}^n$ such that P-a.s. and for almost all t (2.2) is satisfied.

Analogous definitions are also valid for inclusions with backward derivatives and with current velocities below.

In this section we shall mainly look for weak solutions in the class of diffusion type processes.

Further on we shall need the following technical statement.

Lemma 2.2. *Let $\alpha(t,x)$ be a jointly continuous (measurable, smooth) mapping from $[0,T]\times\mathbb{R}^n$ to $\mathrm{S}_+(n)$. Then there exists a jointly continuous (measurable, smooth, respectively) mapping $A(t,x)$ from $[0,T]\times\mathbb{R}^n$ to $L(\mathbb{R}^n,\mathbb{R}^n)$ such that for all $t\in R$, $x\in\mathbb{R}^n$ the equality $A(t,x)A^*(t,x)=\alpha(t,x)$ holds.*

Proof. Since the symmetric matrices $\alpha(t,x)$ are positive definite, all diagonal minors of $\alpha(t,x)$ are positive and, in particular, are not equal to zero. Then for $\alpha(t,x)$ the Gauss decomposition is valid (see Theorem II.9.3 [23]): $\alpha = \zeta\delta z$, where ζ is a lower-triangle matrix with units on the diagonal, z is an upper-triangle matrix with units on the diagonal and δ is a diagonal matrix. In addition, the elements of matrices ζ, δ and z are rationally expressed via the elements of α, hence if the matrices $\alpha(t,x)$ are continuous (measurable, smooth) jointly in t,x, the matrices ζ, δ and z are also continuous (measurable, smooth, respectively) jointly in variables t,x. From the fact that α are symmetric matrices one can easily derive that $z=\zeta^*$ (i.e., z equals the transposed ζ). One also can easily see that the elements of diagonal matrix δ are positive. Thus the diagonal matrix $\sqrt{\delta}$ is well-posed: its diagonal contains the square roots of the corresponding diagonal elements of δ. Consider the matrix $A(t,x)=\zeta\sqrt{\delta}$. By the construction $A(t,x)$ is jointly continuous (measurable, smooth, respectively) in t,x and $A(t,x)A^*(t,x)=\zeta(t,x)\delta(t,x)z(t,x)=\alpha(t,x)$. □

Theorem 2.3. *Suppose that* $\mathbf{a}(t,x)$ *is a uniformly bounded, Borel measurable set-valued mapping from* $[0,T]\times\mathbb{R}^n$ *to* $\mathbb{R}^n$ *with closed images.*

Let $\boldsymbol{\alpha}(t,x)$ *be a uniformly bounded, Borel measurable set-valued mapping from* $[0,T]\times\mathbb{R}^n$ *to* $\mathrm{S}_+(n)$ *with closed images and let there exist* $\varepsilon_0>0$ *such that for all* t,x *the* ε_0*-neighbourhood of* $\boldsymbol{\alpha}(t,x)$ *in* $\mathrm{S}(n)$ *does not intersect the set* $\mathrm{S}_0(n)$ *of symmetric degenerate* $n\times n$ *matrices.*

Then for any initial condition $\xi(0)=\xi_0\in\mathbb{R}^n$ *inclusion* (2.2) *has a weak solution that is well-posed on the entire interval* $t\in[0,T]$.

Proof. As $\mathbf{a}(t,x)$ and $\boldsymbol{\alpha}(t,x)$ are Borel measurable, they have Borel measurable selectors $a(t,x)$ and $\alpha(t,x)$, and from the hypothesis of Theorem it follows that those selectors are uniformly bounded. By Lemma 2.2 there exists Borel measurable $A(t,x)$ such that $\alpha(t,x)=A(t,x)A^*(t,x)$. By the construction $A(t,x)$ is uniformly bounded and uniformly separated from $\mathrm{S}_0(n)$. Then the equation

$$\xi(t)=\xi_0+\int_0^t a(s,\xi(s))ds+\int_0^t A(s,\xi(s))dw(s)$$

satisfies the hypothesis of theorem II.6.1 [13] and so it has a weak solution that is evidently a weak solution of (2.2). □

For the next two existence results we need the following technical statement:

Lemma 2.4. *For a solution of the diffusion type stochastic differential equation*

$$\xi(t)=\xi_0+\int_0^t a(s,\xi(\cdot))ds+\int_0^t A(s,\xi(\cdot))dw(s)$$

in $\mathbb{R}^n$, $t \in [0,T]$, *whose coefficients satisfy the estimates*

$$\|a(t,x(\cdot))\| < K(1+\|x(\cdot)\|), \tag{2.3}$$

$$\|A(t,x(\cdot))\| < K(1+\|x(\cdot)\|) \tag{2.4}$$

for some $K>0$, *for any integer* $p>1$ *there exists a constant* $C_p>0$, *depending only on* K *and* T, *such that the inequality* $E(\sup_{t\le T}\|\xi(t)\|^p) < C_p$ *holds.*

The proof of Lemma 2.4 can be found in [4] (see Lemma III.2.1 and the remark after it).

For considering upper semicontinuous mean forward differential inclusions we need to recall the following

Definition 2.5. Let X and Y be metric spaces. For a given $\varepsilon>0$ a continuous single-valued mapping $f_\varepsilon : X \to Y$ is called an ε-approximation of the set-valued mapping $F : X \to Y$, if the graph of f, as a set in $X \times Y$, belongs in ε-neighbourhood of the graph of F.

It is known (see, e.g., [3]), that for upper semicontinuous set-valued mappings with convex closed images in normed linear spaces the ε-approximations exist for each $\varepsilon>0$.

Theorem 2.6. *Let* $\mathbf{a}(t,x)$ *be an upper semicontinuous set-valued mapping with closed convex images from* $[0,T]\times\mathbb{R}^n$ *to* $\mathbb{R}^n$ *and let it satisfy the estimate*

$$\|\mathbf{a}(t,x)\|^2 < K(1+\|x\|^2) \tag{2.5}$$

for some $K>0$.

Let $\boldsymbol{\alpha}(t,x)$ *be an upper semicontinuous set-valued mapping with closed convex images from* $[0,T]\times\mathbb{R}^n$ *to* $\bar{\mathrm{S}}_+(n)$ *such that for each* $\alpha(t,x)\in\boldsymbol{\alpha}(t,x)$ *the estimate*

$$\|tr\alpha(t,x)\| < K(1+\|x\|^2) \tag{2.6}$$

takes place for some $K>0$.

Then for any initial condition $\xi(0)=\xi_0\in\mathbb{R}^n$ *inclusion* (2.2) *has a weak solution that is well-posed on the entire interval* $t\in[0,T]$.

Proof. As the norm in $\mathrm{S}(n)$ we take the restriction to $\mathrm{S}(n)$ of Euclidean norm (i.e., the square root from the sum of squares of all elements of a matrix) in the space $L(\mathbb{R}^n,\mathbb{R}^n)$ isomorphic to $\mathbb{R}^{n^2}$. Since all norms in the finite-dimensional space $\mathrm{S}(n)$ are equivalent to each other, for this norm (2.6) is valid as well, perhaps with another constant, for which we keep the notation K.

Since $\mathbf{a}(t,x)$ is an upper semicontinuous set-valued mapping with closed convex images, for any there $\varepsilon>0$ there exists its ε-approximation (see above).

Choose a sequence $\varepsilon_i \to 0$ such that $\varepsilon_i > 0$ for all $i \in N$. Denote by $a_i(t,x)$ continuous ε_i-approximations of $\boldsymbol{a}(t,x)$ in $\mathbb{R}^n$. It is clear that all $a_i(t,x)$ satisfy (2.5) with a certain constant that is bigger than K from the condition of Theorem. Nevertheless we keep the notation K for this constant. Since clearly $1+\|x\|^2 \le (1+\|x\|)^2$, for $a_i(t,x)$ estimate (2.3) is valid as well.

As well as $\mathbf{a}(t,x)$, $\boldsymbol{\alpha}(t,x)$ has in $\bar{\mathrm{S}}_+(n)$ an ε-approximation for any $\varepsilon > 0$ since $\boldsymbol{\alpha}(t,x)$ is an upper semicontinuous set-valued mapping with closed convex images. For the sequence ε_i (see above) consider $\frac{\varepsilon_i}{2}$-approximations $\bar{\alpha}_i(t,x)$ of $\boldsymbol{\alpha}(t,x)$. Introduce $\alpha_i(t,x) = \bar{\alpha}_i(t,x) + \frac{\varepsilon_i}{4} I$ where I is the unit matrix. Immediately from the construction it follows that $\alpha_i(t,x)$ for any i is a continuous ε_i-approximation of $\boldsymbol{\alpha}(t,x)$ and that at any (t,x) it belongs to $\mathrm{S}_+(n)$, i.e., it is strictly positive definite. Besides, $\alpha_i(t,x)$ satisfy (2.6) where the constant $K > 0$ is bigger than the constant from the hypothesis of Theorem but nevertheless we keep the notation K for it.

By Lemma 2.2 there exist continuous $A_i(t,x)$ such that $\alpha_i(t,x) = A_i(t,x)A_i^*(t,x)$. Directly from the definition of trace we obtain that $tr\alpha_i(t,x)$ is equal to the sum of squares of all elements of $A_i(t,x)$, i.e., it is the square of the Euclidean norm in $L(\mathbb{R}^n,\mathbb{R}^n)$. Hence form (2.6) and from the obvious inequality $1+\|x\|^2 \le (1+\|x\|)^2$ it follows that $A_i(t,x)$ satisfies (2.4).

Thus the stochastic differential equation

$$\xi(t) = \xi_0 + \int_0^t a_i(s,\xi(s))ds + \int_0^t A_i(s,\xi(s))dw(s), \tag{2.7}$$

satisfies the hypothesis of Theorem III.2.4 [4] and so there exists its weak solution that is well-posed on the entire interval $[0,T]$. Denote by $\xi_i(t)$ that solution.

Consider the Banach space $\Omega = C^0([0,T],\mathbb{R}^n)$ with usual norm $\|x(\cdot)\|_{C^0} = \sup\limits_{t\in[0,T]} \|x(t)\|$. Via $\mathcal{F}$ we denote the σ-algebra on it, generated by cylinder sets. By $\mathcal{P}_t$ we denote the σ-subalgebra of $\mathcal{F}$, generated by cylinder sets with bases on $[0,t]$.

On the measure space $([0,T],\mathcal{B})$, where $\mathcal{B}$ is Borel σ-algebra, by λ_1 we denote the Lebesgue measure.

The process $\xi_i(t)$ determines a measure μ_i on $(\Omega,\mathcal{F})$. On the probability space $(\Omega,\mathcal{F},\mu_i)$ the process $\xi_i(t)$ is the coordinate one, i.e., $\xi_i(t,x(\cdot)) = x(t)$, $x(\cdot)\in\Omega$.

Since $a_i(t,x)$ satisfies (2.3) and $A_i(t,x)$ satisfies (2.4) (see above), equations (2.7) satisfy the hypothesis of Lemma 2.4 for all i and so the estimate

$$E(\sup_{t\le 1}\|\xi_i(t)\|^2) \le C_2. \tag{2.8}$$

is valid for all ξ_i. In addition by corollary III.2 [4] the set of measures $\{\mu_i\}$ is weakly compact, i.e., it is possible to select a subsequence weakly convergent to a certain measure μ. Denote by $\xi(t)$ the coordinate process on the probability space $(\Omega,\mathcal{F},\mu)$.

Let us use the following fact (see [20, 18]): since $\Omega = C^0([0,T],\mathbb{R}^n)$ is a separable metric space and the measures μ_i weakly converge to μ, there exists a certain probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{\mathsf{P}})$ and random elements $\bar{\xi}_i : \bar{\Omega} \to \Omega$ and $\bar{\xi} : \bar{\Omega} \to \Omega$ such that the measures on $(\Omega,\mathcal{F})$, generated by them, coincide with μ_i and μ, respectively, and $\bar{\xi}_i$ converge to $\bar{\xi}$ $\bar{\mathsf{P}}$-almost surely. Denote elementary events from $\bar{\Omega}$ by $\bar{\omega}$.

As $\|a_i(t,\bar{\xi}_i(t))\|^2 \le K(1+\|\bar{\xi}_i(t)\|^2)$ by (2.5), then, taking into account (2.8), one can easily see that

$$\int\limits_{\bar{\Omega}\times[0,T]} \|a_i(t,\bar{\xi}_i(t))\|^2 d\bar{\mathsf{P}} \times d\lambda_1 \le K_1 \tag{2.9}$$

and so $a_i(t,\bar{\xi}_i(t))$ are uniformly bounded with respect to norm in the Hilbert space $L_2([0,T]\times\bar{\Omega},\mathbb{R}^n)$. Hence this set is weakly relatively compact in $L_2([0,T]\times\bar{\Omega},\mathbb{R}^n)$. Thus it is possible to select a subsequence that weakly in $L_2([0,T]\times\bar{\Omega},\mathbb{R}^n)$ converges to a certain $\bar{a} : [0,T]\times\bar{\Omega} \to \mathbb{R}^n$. Notice (see [17]) that there exists a measurable mapping $\mathsf{a} : [0,T]\times\Omega \to \mathbb{R}^n$ such that $\bar{a}(t,\bar{\omega}) = \mathsf{a}(t,\bar{\xi}(\bar{\omega}))$. By general properties of weak convergence in L_p-spaces (see [12]) this means that for any set $B \in \bar{\mathcal{F}}\times\mathcal{B}$ the convergence $\lim\limits_{i\to\infty}\int_B a_i(t,\bar{\xi}_i) d\bar{\mathsf{P}}\times d\lambda_1 = \int_B \mathsf{a}(t,\bar{\xi}) d\bar{\mathsf{P}}\times d\lambda_1$ takes place.

Denote by $a(t,x(\cdot))$ the conditional expectation $E(\mathsf{a}(t,x(\cdot)) \mid \mathcal{P}_t)$ on the probability space $(\Omega,\mathcal{F},\mu)$. Thus, $a(t,x(\cdot))$ is measurable with respect to $\mathcal{P}_t$ and for any set $Q \in \mathcal{P}_t$ the equality

$$\int_Q a(t,x(\cdot)) d\mu = \int_{\bar{\xi}(\bar{\omega})\in Q} \bar{a}(t,\bar{\omega}) d\bar{\mathsf{P}}$$

holds.

From weak convergence of $a_i(t,\bar{\xi}_i)$ to $\mathsf{a}(t,\bar{\xi})$ in $L_2([0,T]\times\bar{\Omega},\mathbb{R}^n)$ it is easy to derive that for any continuous bounded real function f on Ω at any $t\in[0,T]$ and for any $\Delta t > 0$ the convergence $\lim\limits_{i\to\infty}\int\limits_{\bar{\Omega}} f(\bar{\xi})[\int\limits_t^{t+\Delta t}(a_i(s,\bar{\xi}_i) - \mathsf{a}(s,\bar{\xi}))ds] d\bar{\mathsf{P}} = 0$ takes place. Hence for any continuous bounded real function f_t on Ω, that is measurable with respect to $\mathcal{P}_t$, we get

$$\lim_{i\to\infty}\int\limits_{\bar{\Omega}} f_t(\bar{\xi})[\int\limits_t^{t+\Delta t}(a_i(s,\bar{\xi}_i) - a(s,\bar{\xi}))ds] d\bar{\mathsf{P}} = 0. \tag{2.10}$$

Choose $\delta > 0$. By Egorov theorem there exists a set $\mathsf{K}_\delta \subset \bar{\Omega}$ such that $\bar{\mathsf{P}}\mathsf{K}_\delta > 1-\delta$ and on this set $\bar{\xi}_i$ converge to $\bar{\xi}$ uniformly. Let h_t be a uniformly continuous bounded real function on Ω that is measurable with respect to $\mathcal{P}_t$. Then on K_δ the functions $h_t(\bar{\xi}_i)$ uniformly converge to $h_t(\bar{\xi})$. From this it is easy to see that

$$\lim_{i\to\infty}\int_{\mathsf{K}_\delta}[(h_t(\bar{\xi}) - h_t(\bar{\xi}_i))(\int\limits_t^{t+\Delta t} a_i(t,\bar{\xi}_i)ds)] d\bar{\mathsf{P}} = 0. \tag{2.11}$$

The random elements $[h_t(\bar\xi)-h_t(\bar\xi_i)]a_i(t,\bar\xi_i)$ are uniformly integrable. This follows from the facts that $h_t(\bar\xi)-h_t(\bar\xi_i)$ are bounded, that

$$\|a_i(t,\xi_i(t))\| < K(1+\|\xi_i(t)\|),$$

that by Lemma 2.4 $\sup\limits_i \int\limits_{\bar\Omega} \|\bar\xi_i\|^2_{C^0} d\bar{\mathsf P} < C$ and that

$$\int\limits_{\|\bar\xi_i\|>c} \|\bar\xi_i\|_{C^0} d\mu_i < \frac{1}{c} \int\limits_{\|\bar\xi_i\|>c} \|\bar\xi_i\|^2_{C^0} d\bar{\mathsf P}$$

(see [2]). Thus $\| \int_{\bar\Omega\setminus\mathsf K_\delta}[(h_t(\bar\xi)-h_t(\bar\xi_i))(\int_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)]d\bar{\mathsf P}\|$ becomes smaller than any positive number when $\delta\to 0$. Together with (2.11) this means that

$$\lim_{i\to\infty}\int_{\bar\Omega}[(h_t(\bar\xi)-h_t(\bar\xi_i))(\int_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)]d\bar{\mathsf P}=0. \tag{2.12}$$

Replace f_t in (2.10) with h_t from (2.12). Then, taking into account (2.10) and (2.12), we obtain

$$\lim_{i\to\infty}(\int\limits_{\bar\Omega} h_t(\bar\xi)(\int\limits_t^{t+\Delta t} a(s,\bar\xi)ds)d\bar{\mathsf P} - \int\limits_{\bar\Omega} h_t(\bar\xi_i)(\int\limits_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)d\bar{\mathsf P}) =$$

$$= \lim_{i\to\infty}(\int\limits_{\bar\Omega} h_t(\bar\xi)(\int\limits_t^{t+\Delta t} a(s,\bar\xi)ds)d\bar{\mathsf P} - \int\limits_{\bar\Omega} h_t(\bar\xi)(\int\limits_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)d\bar{\mathsf P}+$$

$$+\int\limits_{\bar\Omega} h_t(\bar\xi)(\int\limits_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)d\bar{\mathsf P} - \int\limits_{\bar\Omega} g(\bar\xi_i)(\int\limits_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)d\bar{\mathsf P}) =$$

$$\lim_{i\to\infty}(\int_{\bar\Omega} h_t(\bar\xi)[\int\limits_t^{t+\Delta t}(a(s,\bar\xi)-a_i(s,\bar\xi_i))ds]d\bar{\mathsf P}+$$

$$+\int_{\bar\Omega}[h_t(\bar\xi)-h_t(\bar\xi_i)](\int\limits_t^{t+\Delta t} a_i(s,\bar\xi_i)ds)d\bar{\mathsf P})=0. \tag{2.13}$$

Notice that the random elements $h_t(\bar\xi_i)\bar\xi_i$ are uniformly integrable. Indeed, h_t is bounded, i.e., $|h_t(\bar\xi_i)|<\Xi$ for all i where $\Xi>0$ is a certain constant, by Lemma 2.4 $\sup\limits_i \|\bar\xi_i\|^2_{C^0} < C_2$ and

$$\int_{|h_t(\bar\xi_i)|\|\bar\xi_i\|_{C^0}>c} |h_t(\bar\xi_i)|\|\bar\xi_i\|_{C^0} d\mathsf P < \frac{1}{c}\int_{|h_t(\bar\xi_i)|\|\bar\xi_i\|_{C^0}>c} |h_t(\bar\xi_i)|\|\bar\xi_i\|^2_{C^0} d\mathsf P < \frac{\Xi C_2}{c}.$$

Since $h_t(\bar\xi_i)\bar\xi_i$ converge to $h_t(\bar\xi)\bar\xi$ $\bar{\mathsf P}$-a.s., from this it follows that

$$\lim_{i\to\infty}\int\limits_{\bar\Omega} h_t(\bar\xi_i)\bar\xi_i d\bar{\mathsf P} = \int\limits_{\bar\Omega} h_t(\bar\xi)\bar\xi d\bar{\mathsf P}. \tag{2.14}$$

From (2.13) and (2.14) it follows that on $(\Omega, \mathcal{F})$ for any uniformly continuous bounded real function $h_t : \Omega \to R$, that is measurable with respect to $\mathcal{P}_t$, the relation

$$\lim_{i\to\infty} \int_\Omega [(x(t+\Delta t) - x(t)) - \int_t^{t+\Delta t} a_i(s, x(s))ds] h_t(x(\cdot))d\mu_i =$$

$$= \int_\Omega [(x(t+\Delta t) - x(t)) - \int_t^{t+\Delta t} a(s, x(\cdot))ds] h_t(x(\cdot))d\mu$$

takes place. Since, by construction, the process $\xi_i(t) - \int_0^t a_i(s, \xi_i(s))ds$ is a martingale with respect to $\mathcal{P}_t$ for any i, then

$$\int_\Omega [(x(t+\Delta t) - x(t)) - \int_t^{t+\Delta t} a_i(s, x(s))ds] h_t(x(\cdot))d\mu_i = 0$$

for all i. Hence

$$\int_\Omega [(x(t+\Delta t) - x(t)) - \int_t^{t+\Delta t} a(s, x(s))ds] h_t(x(\cdot))d\mu = 0.$$

From this evidently follows

Lemma 2.7. *The process* $\xi(t) - \int_0^t a(s, \xi(s))ds$ *is a martingale with respect to* $\mathcal{P}_t$.

Now let us turn to α_i and A_i.

Since $\|A_i(t, x(t))\|^2 \le K(1+\|x(t)\|^2)$, on the probability space $(\bar\Omega, \bar{\mathcal{F}}, \bar{\mathsf{P}})$ we obtain

$$\int_{[0,T]\times\bar\Omega} \|A_i(t, \bar\xi_i)\|^2 d\bar{\mathsf{P}} \times d\lambda_1 \le K_2 \tag{2.15}$$

and so $A_i(t, \bar\xi_i)$ are uniformly bounded by the norm in Hilbert space $L_2([0,T] \times \bar\Omega, L(\mathbb{R}^n, \mathbb{R}^n))$, i.e., the set of $A_i(t, \bar\xi_i)$ is weakly relatively compact in $L_2([0,T] \times \bar\Omega, L(\mathbb{R}^n, \mathbb{R}^n))$.

Notice that from the equality $\alpha_i(t,x) = A_i(t,x)A_i^*(t,x)$ it follows that the elements of the matrix $\alpha_i(t,x)$ are sums of products of elements of $A_i(t,x)$. Then from the fact that $tr\alpha_i(t,x)$ is equal to the sum of squares of all elements of $A_i(t,x)$, from (2.6) and from Lemma 2.4 one can easily derive that for all i the estimate

$$\int_{[0,T]\times\bar\Omega} \|\alpha_i(t, \bar\xi_i)\|^2 d\bar{\mathsf{P}} \times d\lambda_1 \le K_3 \tag{2.16}$$

holds, i.e., the set of $\alpha_i(t, \bar\xi_i)$ for all i is uniformly bounded by the norm in the space $L_2([0,T] \times \bar\Omega, \mathrm{S}(n))$ and so this set is weakly relatively compact.

Thus we can select a sequence of indices i so that $A_i(t,\bar{\xi}_i)$ weakly in $L_2([0,T]\times\bar{\Omega}, L(\mathbb{R}^n,\mathbb{R}^n))$ converge to a certain $\bar{A}:[0,T]\times\bar{\Omega}\to L(\mathbb{R}^n,\mathbb{R}^n)$ and also $\alpha_i(t,\bar{\xi}_i)$ weakly in $L_2([0,T]\times\bar{\Omega},\mathrm{S}(n))$ converge to a certain $\bar{\alpha}:[0,T]\times\bar{\Omega}\to\mathrm{S}(n)$. By the construction $\alpha(t,\bar{\omega})=A(t,\bar{\omega})A^*(t,\bar{\omega})$.

From the properties of conditional expectations it follows that there exist the measurable mappings $\mathsf{A}:[0,T]\times\Omega\to L(\mathbb{R}^n,\mathbb{R}^n)$ such that $\bar{A}(t,\bar{\omega})=\mathsf{A}(t,\bar{\xi}(\bar{\omega}))$, and $\aleph:[0,T]\times\Omega\to\mathrm{S}(n)$ such that $\bar{\alpha}(t,\bar{\omega})=\aleph(t,\bar{\xi}(\bar{\omega}))$. Consider the following conditional expectations on probability space $(\Omega,\mathcal{F},\mu)$: $A(t,x(\cdot))=E(\mathsf{A}(t,x(\cdot))\mid\mathcal{P}_t)$ and $\alpha(t,x(\cdot))=E(\aleph(t,x(\cdot))\mid\mathcal{P}_t)$. By construction, at each $t\in[0,T]$ they are measurable with respect to $\mathcal{P}^\xi_t$.

By an elementary modification of arguments, used above for $a_i(t,x(t))$, one can show that for any uniformly continuous, bounded, real function $h_t:\Omega\to R$, that is measurable with respect to $\mathcal{P}_t$, the convergence

$$\lim_{i\to\infty}\int\limits_{\Omega}[(x(t+\Delta t)-x(t))(x(t+\Delta t)-x(t))^*$$

$$-\int\limits_t^{t+\Delta t}A_i(s,x(s))A_i^*(s,x(s))ds]h_t(x(\cdot)d\mu_i=$$

$$=\int\limits_{\Omega}[(x(t+\Delta t)-x(t))(x(t+\Delta t)-x(t))^*$$

$$-\int\limits_t^{t+\Delta t}A(s,x(\cdot))A^*(s,x(\cdot))ds]h_t(x(\cdot)d\mu$$

holds. Also for each i

$$\int\limits_{\Omega}[(x(t+\Delta t)-x(t))(x(t+\Delta t)-x(t))^*$$

$$-\int\limits_t^{t+\Delta t}A_i(s,x(s))A_i^*(s,x(s))ds]h_t(x(\cdot)d\mu_i=0$$

and, consequently,

$$\int\limits_{\Omega}[(x(t+\Delta t)-x(t))(x(t+\Delta t)-x(t))^*$$

$$-\int\limits_t^{t+\Delta t}A(s,x(\cdot))A^*(s,x(\cdot))ds]h_t(x(\cdot)d\mu=0.$$

By Lemma 2.7 for the coordinate process $\xi(t)$ on the probability space $(\Omega,\mathcal{F},\mu)$ the process $\xi(t)-\int_t^{t+\Delta t}a(s,\xi(\cdot))ds$ is a martingale with respect to $\mathcal{P}_t$. From this, by methods of [4], one can derive from the above arguments that $\xi(t)$ satisfies the equality

$$\xi(t) = \xi_0 + \int_0^t a(s, \xi(\cdot))ds + \int_0^t A(s, \xi(\cdot))dw(s),$$

where $w(t)$ is a certain Wiener process on $(\Omega, \mathcal{F}, \mu)$, adapted to $\mathcal{P}_t$. By the construction of $\xi(t)$, for each t the σ-algebra $\mathcal{P}_t$ is the "past" for $\xi(t)$. Hence $\xi(t)$ is a diffusion type process. Then by Theorem 1.7 $D\xi = E_t^{\xi}(a(t, \xi(\cdot)))$ and by Theorem 1.8 $D_2\xi(t) = E_t^{\xi}(A(t, \xi(\cdot))A^*(t, \xi(\cdot))) = E_t^{\xi}(\alpha(t, \xi(\cdot)))$.

It remains to show that μ-a.s. $E_t^{\xi}(a(t, \xi(\cdot))) \in \mathbf{a}(t, \xi(t))$ and that $E_t^{\xi}(\alpha(t, \xi(\cdot)) \in \boldsymbol{\alpha}(t, \xi(t))$. By Mazur's lemma (see, e.g., [22]), for a weakly convergent sequence $a_i(t, \bar{\xi}_i(t))$ to $\bar{a} : [0, T] \times \bar{\Omega} \to \mathbb{R}^n$ in $L_2([0, T] \times \bar{\Omega}, \mathbb{R}^n)$ there exists a sequence of finite convex combinations of its elements that converges in the same space strongly (in norm). The convex combinations have the form

$$\tilde{a}_k(t, \bar{\xi}_i) = \sum_{i=j(k)}^{n(k)} \beta_i a_i(t, \bar{\xi}_i)$$

where $\beta_i \geq 0$, $i = j(k), \ldots, n(k)$ and $\sum_{i=j(k)}^{n(k)} \beta_i = 1$. As $\mathbf{a}$ has convex images, $\tilde{a}_i$ are ε-approximations with $\varepsilon \to 0$. Since the convergence is strong in $L_2([0, T] \times \bar{\Omega}, \mathbb{R}^n)$, the limit $a(t, \bar{\xi})$ belongs to $\mathbf{a}$ $\bar{\mathsf{P}}$-a.s.

For α the proof is completely analogous. □

Theorem 2.8. *Suppose that* $\boldsymbol{\alpha}(t, x)$ *takes values in the space* $\bar{\mathrm{S}}_+(n)$ *of positive semi-definite symmetric matrices, has closed convex images, it is lower semicontinuous and for each* $\alpha \in \boldsymbol{\alpha}(t, x)$ *the following estimate*

$$\|tr\alpha(t, x)\| < K(1 + \|x\|)^2 \tag{2.17}$$

holds for some $K > 0$. *Let also* $\mathbf{a}(t, x)$ *be Borel measurable set-valued mapping and satisfy the estimate*

$$\|\mathbf{a}(t, x)\| < K(1 + \|x\|) \tag{2.18}$$

for some $K > 0$. *Then for any initial condition* $\xi(0) = \xi_0$ *there exists a weak solution of* (2.2) *that is well-posed on the entire interval* $t \in [0, T]$.

Proof. From Michael's theorem it follows that under the conditions of Theorem 2.8 the set-valued mapping $\boldsymbol{\alpha}(t, x)$ has a single-valued continuous selector $\alpha(t, x)$. Obviously $\alpha(t, x)$ belongs to $\bar{\mathrm{S}}_+(n)$ at any t, x. The Borel measurable set-valued mapping $\mathbf{a}(t, x)$ has a Borel measurable single-valued selector $a(t, x)$.

Take a sequence of positive $\varepsilon_i \to 0$. Introduce $\alpha_i(t, x) = \alpha(t, x) + \varepsilon_i I$ where I is the unit $n \times n$ matrix. Obviously α_i are strictly positive definite and continuous. Then by Lemma 2.2 there exists a continuous $A_i(t, x)$ such that $A_i(t, x)A_i^*(t, x) = \alpha_i(t, x)$.

Recall that $tr\alpha_i(t,x)$ is equal to the sum of squares of all elements of $A_i(t,x)$, i.e., it is the square of the Euclidean norm in $L(\mathbb{R}^n,\mathbb{R}^n)$. Since in the finite-dimensional linear space $\mathrm{S}(n)$ all norms are equivalent, from (2.17) it immediately follows that $\|A(t,x)\| < K(1+\|x\|)$ for some $K>0$. As $\alpha_i(t,x)$ is positive definite, the matrix $A_i(t,x)$ is not degenerate at all t,x. Since $a(t,x)$ is measurable and satisfies (2.17), under the above-mentioned properties of $A_i(t,x)$ by Theorem III.3.3 [4] there exists a weak solution of the stochastic differential equation

$$\xi_i(t) = \xi_0 + \int_0^t a(s,\xi_i(s))ds + \int_0^t A_i(s,\xi_i(s))dw(s), \tag{2.19}$$

well-posed on the entire interval $t\in[0,T]$. Denote by $\xi_i(t)$ this solution of (2.19). It determines a measure μ_i on $(\Omega,\mathcal{F})$ where $(\Omega,\mathcal{F})$ was introduced in the Proof of Theorem 2.6.

The rest of the proof is quite analogous to that of Theorem 2.6. All equations (2.19) satisfy the hypothesis of Lemma 2.4. The set of measures $\{\mu_i\}$ is weakly compact so that there exists a subsequence that weakly converges to a certain measure μ. Denote by $\xi(t)$ the coordinate process on the probability space $(\Omega,\mathcal{F},\mu)$. Construct $A(t,x(\cdot))$ in complete analogy with that in Theorem 2.6, i.e., as a weak limit in the corresponding space L_2 of the bounded (and so weakly compact) set A_i. The process $\xi(t)$ satisfies the equality $\xi(t)=\xi_0+\int_0^t a(s,\xi(s))ds+\int_0^t A(s,\xi(\cdot))dw$ where $w(t)$ is a certain Wiener process. Since by construction α_i converge to α uniformly, one can easily show that
$E^{\xi}_t(AA^*) = \alpha$. Taking into account Theorems 1.7 and 1.8, this means that $\xi(t)$ is a weak solution of (2.2) that we are looking for. □

3. DIFFERENTIAL INCLUSIONS WITH BACKWARD MEAN DERIVATIVES

Equations and inclusions with backward mean derivatives arise in description of some special stochastic processes of mathematical physics. Say (see, e.g., [9, 10]) a second order equation in backward mean derivatives of the group of Sobolev diffeomorphisms is derived that describes a process whose expectation is a flow of viscous incompressible fluid. It should be pointed out that such equations and inclusions are much more complicated for investigation than those with forward mean derivatives. Nevertheless there exists a simple method of using inverse time direction for solutions of equations and inclusions with forward mean derivatives, that allows one to obtain some results for the case of backward mean derivatives. In this section we illustrate this method on some examples.

The system

$$\begin{cases} D_*\xi(t) = a(t, \xi(t)) \\ D_2\xi(t) = \alpha(t, \xi(t)) \end{cases} \tag{3.1}$$

is called a first order differential equation with backward mean derivatives.

Notice that we do not introduce the notion of backward analog of operator D_2 since, applying the properties of Itô integral, one can easily prove that for a diffusion type process $\xi(t)$ the result of application of that analog coincides with $D_2\xi(t)$ (for the case of diffusion processes this follows from the results of [15, 16]).

Let $\mathbf{a}(t, x)$ and $\boldsymbol{\alpha}(t, x)$ be set-valued mappings from $[0, T] \times \mathbb{R}^n$ to $\mathbb{R}^n$ and to $\bar{\mathrm{S}}_+(n)$, respectively. The system of the form

$$\begin{cases} D_*\xi(t) \in \mathbf{a}(t, \xi(t)), \\ D_2\xi(t) \in \boldsymbol{\alpha}(t, \xi(t)). \end{cases} \tag{3.2}$$

is called a first order differential inclusion in backward mean derivatives.

Consider a weak solution $\eta(t)$, given on $t \in [0, T]$, with initial condition $\eta(0) = 0$ of the following differential inclusion with forward mean derivatives

$$\begin{cases} D\eta(t) \in -\mathbf{a}(1 - t, \eta(t)), \\ D_2\eta(t) \in \boldsymbol{\alpha}(1 - t, \eta(t)). \end{cases} \tag{3.3}$$

Theorem 3.1. *The process $\xi(t) = \xi_0 - \eta(T) + \eta(T - t)$ is a weak solution of* (3.2) *with initial condition $\xi(0) = \xi_0$ where $\eta(t)$ is a solution of* (3.3) *with initial condition $\eta(0) = 0$.*

Indeed, $D_*\xi(t) = -D\eta(T - t) \in \boldsymbol{a}(t, \eta(T - t)) = \boldsymbol{a}(t, \xi(t))$. For $D_2\xi(t)$ the arguments are analogous.

4. DIFFERENTIAL INCLUSIONS WITH CURRENT VELOCITIES

As it is mentioned in Section 1, the meaning of current velocities is analogous to that of ordinary velocity for a non-random process. Thus the case of equations and inclusions with current velocities is probably the most natural from the physical point of view.

The system

$$\begin{cases} D_S\xi(t) = v(t, \xi(t)) \\ D_2\xi(t) = \alpha(t, \xi(t)) \end{cases} \tag{4.1}$$

is called a first order differential equation with current velocities.

Theorem 4.1. *Let $v : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ be smooth and $\alpha : \mathbb{R}^n \to \mathrm{S}_+(n)$ be smooth and autonomous (so, it determines the Riemannian metric $\alpha(\cdot,\cdot)$ on $\mathbb{R}^n$, introduced in Section* 1*). Let also they satisfy the estimates*

$$\|v(t,x)\| < K(1 + \|x\|), \tag{4.2}$$

$$tr\ \alpha(x) < K(1 + \|x\|^2) \tag{4.3}$$

for some $K > 0$. Let ξ_0 be a random element with values in $\mathbb{R}^n$ whose probability density ρ_0 with respect to the the volume form Λ_α of $\alpha(\cdot,\cdot)$ on $\mathbb{R}^n$ (see Section 1*), is smooth and nowhere equal to zero. Then for the initial condition $\xi(0) = \xi_0$ equation* (4.1) *has a weak solution that is well posed on the entire interval $t \in [0, T]$.*

Proof. Since $v(t, x)$ is smooth and estimate (4.2) is fulfilled, its flow g_t is well posed on the entire interval $t \in [0, T]$. By $g_t(x)$ we denote the orbit of the flow (i.e., the solution of equation $x'(t) = v(t, x)$) with the initial condition $g_0(x) = x$. Since $v(t, x)$ is smooth, its flow is also smooth.

Continuity equation (1.8) obviously can be transformed into the form

$$\frac{\partial \rho}{\partial t} = -\alpha(v, Grad\ \rho) - \rho\ Div\ v. \tag{4.4}$$

Suppose that $\rho(t, x)$ nowhere in $[0, T] \times \mathbb{R}^n$ equals zero. Then we can divide (4.4) by ρ so that it is transformed into the equation

$$\frac{\partial p}{\partial t} = -\alpha(v, Grad\ p) - Div\ v \tag{4.5}$$

where $p = \log \rho$. Introduce $p_0 = \log \rho_0$.

Show that the solution of (4.5) with initial condition $p(0) = p_0$ is described by the formula $p(t, x) = p_0(g_{-t}(x)) - \int_0^t (Div\ v)(s, g_s(g_{-t}(x))\ ds$. Introduce the product $[0, T] \times \mathbb{R}^n$ and consider the function p_0 as given on the level surface $(0, \mathbb{R}^n)$. Consider the vector field $(1, v(t, x))$ on $[0, T] \times \mathbb{R}^n$. The orbits of its flow $\hat{g}_t$, starting at the points of $(0, \mathbb{R}^n)$, have the form $\hat{g}_t(0, x) = (t, g_t(x))$ and the flow is smooth as well as g_t. Also introduce on $[0, T] \times \mathbb{R}^n$ the Riemannian metric $\hat{\alpha}(\cdot,\cdot)$ by the formula $\hat{\alpha}((X_1, Y_1), (X_2, Y_2)) = X_1 X_2 + \alpha(Y_1, Y_2)$. Notice that for any (t, x) the point $\hat{g}_{-t}(t, x)$ belongs to $(0, \mathbb{R}^n)$ where the function p_0 is given. Thus on the one hand $(1, v)p(t, x)$, the derivative of $p(t, x)$ in the direction of $(1, v)$, by construction equals $-Div\ v(t, x)$. And on the other hand one can easily calculate that $(1, v)p(t, x) = \frac{\partial}{\partial t} p(t, x) + \alpha(v(t, x), Grad\ p(t, x))$. Thus (4.5) is satisfied.

Notice that $\rho = e^p$ is indeed nowhere zero and so our arguments are well-posed.

Since $\rho(t, x)$ is well-posed for all $t \in [0, T]$, it determines a process $\xi(t)$ with this probability density and so with initial density ρ_0. By the construction $D_S \xi(t) = v(t, \xi(t))$.

Introduce $u = \frac{1}{2} Grad\ p = Grad \log \sqrt{\rho}$ and $a(t, x) = v(t, x) + u(t, x)$.

From Lemma 2.2 and from the hypothesis of Theorem it follows that there exists smooth $A(t,x)$ such that $A(t,x)A^*(t,x) = \alpha(t,x)$ and the relation $\|A(t,x)\| < K(1+\|x\|)$ holds. Then $\xi(t)$ satisfies the stochastic differential equation

$$\xi(t) = \xi_0 + \int_0^t a(s,\xi(s))ds + \int_0^t A(s,\xi(s))dw(s) \tag{4.6}$$

and so by Theorem 1.8 $D_2\xi(t) = \alpha(\xi(t))$. □

Lemma 4.2. *Let $\alpha(x)$, $\rho(t,x)$ and Λ_α be the same as in Theorem 4.1. Let also the vector field v from Theorem 4.1 be autonomous. Then the flow $\hat{g}_t$ of vector field $(1, v(x))$ on $[0,T]\times\mathbb{R}^n$ preserves the volume form $\rho(t,x)dt\wedge\Lambda_\alpha$ (i.e., $\hat{g}_t^*(\rho(t,x)dt\wedge\Lambda_\alpha) = \rho_0(x)dt\wedge\Lambda_\alpha$ where $\hat{g}_t^*$ is the pull back) and so for any measurable set $Q\subset\mathbb{R}^n$ and for any $t\in[0,T]$*

$$\int_Q \rho_0(x)\Lambda_\alpha = \int_{g_t(Q)} \rho(t,x)\Lambda_\alpha.$$

Proof. It is enough to show that $L_{(1,v)}(\rho(t,x)dt\wedge\Lambda_\alpha) = 0$ where $L_{(1,v)}$ is the Lie derivative along $(1,v)$. Obviously

$$L_{(1,v)}(\rho(t,x)dt\wedge\Lambda_\alpha) = (L_{(1,v)}\rho(t,x))dt\wedge\Lambda_\alpha + \rho(t,x)(L_{(1,v)}dt\wedge\Lambda_\alpha).$$

For a function the Lie derivative coincides with the derivative in direction of vector field, hence $L_{(1,v)}\rho(t,x) = \frac{\partial\rho}{\partial t} + \alpha(v, Grad\,\rho)$ (see the proof of Theorem 4.1) and so $(L_{(1,v)}\rho(t,x))dt\wedge\Lambda_\alpha = (\frac{\partial\rho}{\partial t} + \alpha(v, Grad\,\rho))dt\wedge\Lambda_\alpha$. Since neither the form Λ_α nor the vector filed $v(x)$ depend on t, $L_{(1,v)}dt\wedge\Lambda_\alpha = dt\wedge(L_v\Lambda_\alpha) = Div\,v\,(\,dt\wedge\Lambda_\alpha)$ as the Lie derivative along v of the volume form Λ_α equals $(Div\,v)\Lambda_\alpha$ (see, e.g., [19]). Taking into account (4.4), we obtain $L_{(1,v)}(\rho(t,x)dt\wedge\Lambda_\alpha) = 0$. □

Let $\mathbf{v}(t,x)$ and $\boldsymbol{\alpha}(t,x)$ be set-valued mappings from $[0,T]\times\mathbb{R}^n$ to $\mathbb{R}^n$ and to $\bar{\mathrm{S}}_+(n)$, respectively. The system of the form

$$\begin{cases} D_S\xi(t)\in\mathbf{v}(t,\xi(t)),\\ D_2\xi(t)\in\boldsymbol{\alpha}(t,\xi(t)).\end{cases} \tag{4.7}$$

is called a first order differential inclusion with current velocities.

To avoid some technical difficulties we investigate the first order differential inclusion with current velocities on flat n-dimensional torus $\mathbb{T}^n$, i.e. the Riemannian metric on $\mathbb{T}^n$ is inherited from $\mathbb{R}^n$ after factorization with respect to the integer lattice. As usual this problem can be considered as the one in $\mathbb{R}^n$ with periodical data. The key fact that makes the use of $\mathbb{T}^n$ more simple for investigation is that $\mathbb{T}^n$ is a compact manifold and so all our smooth objects become bounded.

It should be also pointed out that the theory of mean derivatives and so the corresponding equations or inclusions on general Riemannian manifolds are naturally connected with the theory of Itô equations in Belopolskaya-Daletskii form (see, e.g., [1, 5, 6, 10]) that involves serious geometric structures (we suppose to consider this theory in another publication). But this is not the case for flat torus since locally it is the same as $\mathbb{R}^n$.

Notice also that the tangent bundle $T\mathbb{T}^n$ to $\mathbb{T}^n$ is trivial: $T\mathbb{T}^n = \mathbb{T}\times\mathbb{R}^n$, and so all vector fields on $\mathbb{T}^n$ can be considered as mappings from $\mathbb{T}^n$ to $\mathbb{R}^n$ as well as symmetric semi-positive definite $(2,0)$-tensor fields (like α above) can be considered as mappings from $\mathbb{T}^n$ to $\bar{\mathrm{S}}_+(n)$.

Here we shall proof existence of solutions of (4.7) only in the simplest situation where a single-valued $\boldsymbol{\alpha}$ takes the form $\boldsymbol{\alpha} = \sigma^2 I$ for I being the unit matrix, $\sigma > 0$ being a constant, and a set-valued $\mathbf{v}$ satisfies a rather strong hypothesis:

Theorem 4.3. *Let $\mathbf{v}(x)$ be a uniformly bounded autonomous set-valued mapping from $\mathbb{T}^n$ to $\mathbb{R}^n$ with closed convex images. Suppose that there exists a sequence of positive numbers $\varepsilon_i \to 0$ such that for any ε_i the mapping $\mathbf{v}(x)$ has a smooth ε_i-approximation $v_i(x)$ and all those approximations have uniformly bounded first partial derivatives $\frac{\partial v_i}{\partial x_j}$.*

Let ξ_0 be a random element with values in $\mathbb{T}^n$ whose probability density ρ_0 with respect to the Euclidean volume form $dt\wedge dx^1\wedge\cdots\wedge dx^n$ on $\mathbb{T}^n$ is smooth and nowhere equal to zero. Then for the initial condition $\xi(0)=\xi_0$ inclusion

$$\begin{cases} D_S\xi(t) \in \mathbf{v}(\xi(t)) \\ D_2\xi(t) = \sigma^2 I \end{cases} \tag{4.8}$$

has a weak solution that is well-posed on the entire interval $t\in[0,T]$.

Proof. Notice that the Riemannian metric on $\mathbb{T}^n$, generated by the field of matrices $\sigma^2 I$, is $\frac{1}{\sigma^2}(\cdot,\cdot)$ where $(\cdot,\cdot)$ is the Euclidean metric inherited from $\mathbb{R}^n$. In particular this means that for any function f we get $Grad\, f = \sigma^2 grad\, f$ where $grad\, f$ is the ordinary gradient corresponding to the Euclidean metric.

Denote by $\xi_i(t)$ the solution of equation

$$\begin{cases} D_S\xi_i(t) = v_i(\xi_i(t)) \\ D_2\xi_i(t) = \sigma^2 I \end{cases}$$

with initial condition $\xi(0)=\xi_0$ that exists by Theorem 4.1. By $\rho_i(t,x)$ we denote the corresponding density. Recall that $\xi_i(t)$ satisfies the Itô equation

$$\xi_i(t) = \xi_0 + \int_0^t a_i(s,\xi_i(s))ds + \sigma w(t) \tag{4.9}$$

where $a_i(t,x) = v_i(x) + Grad\, p_i(t,x)$ for $p_i(t,x) = \log\sqrt{\rho_i(t,x)}$.

Consider $\Omega = C^0([0,T],\mathbb{T}^n)$ and the σ-algebra $\mathcal{F}$ on it generated by cylinder sets. Denote by μ_i the measure on $(\Omega,\mathcal{F})$ generated by $\xi_i(t)$.

By Lemma 4.2 $\rho_i(t,x)dt\wedge dx^1\wedge\cdots\wedge dx^n = \hat{g}_t^{(i)*}\rho_0(t,x)dt\wedge dx_1\wedge\cdots\wedge dx^n$ where $\hat{g}_t^{(i)}$ is the flow of vector field $(1,v_i)$ on $[0,1]\times\mathbb{T}^n$. Hence $\frac{\partial\rho_i}{\partial x_j}$ equals $(T\hat{g}_{-t}\frac{\partial}{\partial x^j})\rho_0$, the derivative of ρ_0 in the direction of vector field $(T\hat{g}_{-t}\frac{\partial}{\partial x^j})$ where $T\hat{g}_{-t}$ is the tangent mapping to $\hat{g}_{-t}$. Since all partial derivatives of all v_i are uniformly bounded, all $T\hat{g}_{-t}$ are also uniformly bounded and we obtain that the derivatives $\frac{\partial\rho_i}{\partial x_j}$ are uniformly bounded for all $i=1,\dots\infty$ and all $j=1,2,\dots,n$ as well as all $\frac{\partial p_i}{\partial x^j}$. Thus all vector fields $Grad\, p_i$ for all i are uniformly bounded. Since all $v_i(x)$ are evidently also bounded, this means that all $a_i(t,x)$ form equations (4.9) are uniformly bounded.

Hence corollary III.2 [4] is valid for equations (4.9), i.e., the set of measures $\{\mu_i\}$ is relatively weakly compact and so we can choose a subsequence in $\{\mu_i\}$ that weakly converges to a certain measure μ. For the sake of simplicity let $\{\mu_i\}$ itself be that subsequence. Denote by $\xi(t)$ the coordinate process on probability space $(\Omega,\mathcal{F},\mu)$.

As well as in the proof of Theorem 2.6 we can introduce a probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{\mathsf{P}})$ and random elements $\bar{\xi}_i:\bar{\Omega}\to\Omega$ and $\bar{\xi}:\bar{\Omega}\to\Omega$ such that the measures on $(\Omega,\mathcal{F})$, generated by them, coincide with μ_i and μ, respectively, and $\bar{\xi}_i$ converge to $\bar{\xi}$ $\bar{\mathsf{P}}$-almost surely. Recall that in this case $\bar{\mathsf{P}}$-a.s. convergence means that for $\bar{\mathsf{P}}$-almost all $\bar{\omega}$ the curves $\bar{\xi}_i(\bar{\omega})$ uniformly on $[0,T]$ converge to $\bar{\xi}(\bar{\omega})$. Evidently, if h is a uniformly continuous function on Ω, then for $\bar{\mathsf{P}}$-almost all $\bar{\omega}$ we obtain that $h(\bar{\xi}_i(\bar{\omega}))\bar{\xi}_i(\bar{\omega})(s)$ uniformly in $s\in[0,T]$ converge to $h(\bar{\xi}(\bar{\omega}))\bar{\xi}(\bar{\omega})(s)$.

Notice that the sets $\{a_i(t,\bar{\xi}_i(t))\}$ and $\{v_i(\bar{\xi}_i(t))\}$ are weakly compact in $L_2([0,T]\times\bar{\Omega},\mathbb{R}^n)$ and so we can take weakly convergent subsequences from them. As usual we suppose that the sequences $\{a_i(t,\bar{\xi}_i(t))\}$ and $\{v_i(\bar{\xi}_i(t))\}$ are weakly convergent. In complete analogy with the proof of Theorem 2.6 introduce $\bar{a}:[0,T]\times\bar{\Omega}\to\mathbb{R}^n$ as a weak limit of $a_i(t,\bar{\xi}_i(t))$ and $\bar{v}:[0,T]\times\bar{\Omega}\to\mathbb{R}^n$ as a weak limit of $v_i(\bar{\xi}_i(t))$. Also denote by $\mathsf{a}:[0,T]\times\Omega\to\mathbb{R}^n$ and by $\mathsf{v}:[0,T]\times\Omega\to\mathbb{R}^n$ measurable mappings, existing by [17], such that $\bar{a}(t,\bar{\omega})=\mathsf{a}(t,\bar{\xi}(\bar{\omega}))$ and $\bar{v}(t,\bar{\omega})=\mathsf{v}(t,\bar{\xi}(\bar{\omega}))$. As well as in Theorem 2.6 $\xi(t)$ satisfies the equation

$$\xi(t)=\xi_0+\int_0^t a(s,\xi(s))ds+\sigma w(t).$$

where $a(t,x(\cdot)) = E(\mathsf{a}(t,x(\cdot))\mid\mathcal{P}_t)$ on the probability space $(\Omega,\mathcal{F},\mu)$, $\mathcal{P}_t$ is the σ-algebra generated by cylinder sets with bases over $[0,t]$ (see the proof of Theorem 2.6 for details). From this by Theorem 1.8 it follows that $D_2\xi(t)=\sigma^2 I$.

Denote by $\mathcal{N}_t$ the σ-algebra generated by cylinder sets with bases over the point t and by $v(x)=E(\mathsf{v}(t,x(\cdot))\mid\bar{\xi}(t)=x)$ denote the regression for the conditional expectation $E(\mathsf{v}(t,x(\cdot))\mid\mathcal{N}_t)$ on the probability space $(\Omega,\mathcal{F},\mu)$. Notice that by construction $v(x)$ is a measurable vector field on $\mathbb{R}^n$ such that $v(\bar{\xi}(t))=\mathsf{v}(t,\bar{\xi})$.

Again analogously to the proof of Theorem 2.6 we can derive from weak convergence of $v_i(\bar\xi_i(t))$ to $\mathsf{v}(t,\bar\xi)$ in $L_2([0,t]\times\bar\Omega,\mathbb{R}^n)$ that for any $s\in[0,T]$ and for any continuous bounded real function f_t on Ω, that is measurable with respect to $\mathcal{N}_t$, at any $s\in[0,T]$ we get

$$\lim_{i\to\infty}\int\limits_{\bar\Omega} f_t(\bar\xi)(v_i(\bar\xi_i(s))-v(\bar\xi(s)))d\bar{\mathsf{P}}=0 \tag{4.10}$$

and that for a uniformly continuous bounded real function h_t on Ω, measurable with respect to $\mathcal{N}_t$, we can derive from $\bar{\mathsf{P}}$-a.s. uniform in $s\in[0,T]$ convergence of $h_t(\bar\xi_i)\bar\xi_i$ to $h_t(\bar\xi)\bar\xi$ (see above) that for any $s\in[0,T]$

$$\lim_{i\to\infty}\int\limits_{\bar\Omega} h_t(\bar\xi_i)\bar\xi_i(s)d\bar{\mathsf{P}}=\int\limits_{\bar\Omega} h_t(\bar\xi)\bar\xi(s)d\bar{\mathsf{P}}. \tag{4.11}$$

By definition of D_S we get for any i

$$\lim_{\Delta t\to+0}\frac{1}{2\Delta t}\int\limits_{\bar\Omega} h_t(\bar\xi_i)(\bar\xi_i(t+\Delta t)-\bar\xi_i(t-\Delta t))d\bar{\mathsf{P}}=\int\limits_{\bar\Omega} h_t(\bar\xi_i)v_i(\bar\xi_i(t))d\bar{\mathsf{P}}.$$

To show that $D_S\xi(t)=v(\xi(t))$ it is enough to prove that

$$\lim_{\Delta t\to+0}\frac{1}{2\Delta t}\int\limits_{\bar\Omega} h_t(\bar\xi)(\bar\xi(t+\Delta t)-\bar\xi(t-\Delta t))d\bar{\mathsf{P}}=\int\limits_{\bar\Omega} h_t(\bar\xi)v(\bar\xi(t))d\bar{\mathsf{P}}.$$

From (4.11) it follows that for given $\varepsilon>0$ there exists N such that for all $i>N$

$$\|\int\limits_{\bar\Omega} h_t(\bar\xi_i)(\bar\xi_i(t+\Delta t)-\bar\xi_i(t-\Delta t))d\bar{\mathsf{P}}-\int\limits_{\bar\Omega} h_t(\bar\xi)(\bar\xi(t+\Delta t)-\bar\xi(t-\Delta t))d\bar{\mathsf{P}}\|<\varepsilon$$

for any Δt. Thus for Δt small enough

$$\|\frac{1}{2}\int\limits_{\bar\Omega} h_t(\bar\xi)(\bar\xi(t+\Delta t)-\bar\xi(t-\Delta t))d\bar{\mathsf{P}}-\int\limits_{\bar\Omega} h_t(\bar\xi_i)v_i(\bar\xi_i(t))d\bar{\mathsf{P}}\Delta t\|<\varepsilon.$$

Taking into account (4.10) and $\bar{\mathsf{P}}$-a.s. uniform convergence of $h_t(\bar\xi_i)$ to $h_t(\bar\xi)$ one can easily show that

$$\int\limits_{\bar\Omega} h_t(\bar\xi_i)v_i(\bar\xi_i(t))d\bar{\mathsf{P}}-\int\limits_{\bar\Omega} h_t(\bar\xi)v(\bar\xi(t))d\bar{\mathsf{P}}=$$

$$\int\limits_{\bar\Omega}(h_t(\bar\xi_i)-h_t(\bar\xi))v_i(\bar\xi_i(t))d\bar{\mathsf{P}}+\int\limits_{\bar\Omega} h_t(\bar\xi)(v_i(\bar\xi_i(t))-v(\bar\xi(t))d\bar{\mathsf{P}}$$

becomes less than any ε for i large enough. This completes the proof that $D_S\xi(t)=v(\xi(t))$.

The fact that $v(t,\xi(\cdot))\in\mathbf{v}(\xi(t))$ μ-a.s. is proved analogously to that in the Proof of Theorem 2.6 by application of Mazur's lemma. □

REFERENCES

[1] Belopolskaya Ya.I., Dalecky Yu.L. Stochastic processes and differential geometry.- Dordrecht: Kluwer Academic Publishers, 1989.

[2] Billingsley P. Convergence of Probability Measures.- New York et al.: Wiley, 1969.

[3] Borisovich Yu.G., Gelman B.D., Myshkis A.D., Obukhovskii V.V. Introduction to the theory of multi-valued mappings and differential inclusions.- Mscow: KomKniga (URSS), 2005. (in Russian)

[4] Gihman I.I., Skorohod A.V. Theory of Stochastic Processes. Vol.3.- New York: Springer-Verlag, 1979.

[5] Gliklikh Yu.E. Ordinary and Stochastic Differential Geometry as a Tool for Mathematical Physics.- Dordrecht: Kluwer, 1996.

[6] Gliklikh Yu.E. Global Analysis in Mathematical Physics. Geometric and Stochastic Methods.- N.Y.: Springer-Verlag, 1997.

[7] Gliklikh Yu.E. Stochastic equations in mean derivatives and their applications. I. Transections of Russian Acacdemy of Natural Sciences, Series MMMIC.- 1997.- Vol. 1, No. 4.- P. 26-52. (in Russian)

[8] Gliklikh Yu.E. Stochastic equations in mean derivatives and their applications. II. Transections of Russian Acacdemy of Natural Sciences, Series MMMIC.- 2000.- Vol. 4, No. 4.- P. 17-36. (in Russian)

[9] Gliklikh Yu.E. Deterministic viscous hydrodynamics via stochastic processes on groups of diffeomorphisms. Probabilistic Metohds in Fluids (I.M. Davis et al., eds).- Singapore: World Scientific, 2003.- P. 179-190

[10] Gliklikh Yu.E. Stochastic and global analysis in problems of mathematical physics.- Moscow: KomKniga (URSS), 2005. (in Russian)

[11] He X. A probabilistic method for Navier-Stokes vorticies. J. Appl. Probab., 2001.- P. 1059-1066.

[12] Kantorovich L.V., Akilov G.P. Functional analysis.- Oxford: Pergamon Press, 1982.

[13] Krylov, N.V.: Controlled diffusion processes.- New York et al.: Springer-Verlag, 1980.

[14] Nelson E. Derivation of the Schrödinger equation from Newtonian mechanics. Phys. Reviews, 1966.- Vol. 150, No. 4.- P. 1079-1085

[15] Nelson E. Dynamical theory of Brownian motion.- Princeton: Princeton University Press, 1967.- 142 p.

[16] Nelson E. Quantum fluctuations.- Princeton: Princeton University Press, 1985.- 147 p.

[17] Parthasarathy K.R. Introduction to Probability and Measure.- New York: Springer-Verlag, 1978.

[18] Pollard D. Convergence of stochastic processes.- Berlin etc.: Springer-Verlag, 1984.- 215 p.

[19] Schutz B.F.: Geometrical methods of mathematical physics.- Cambridge: Cambridge University Press, 1982.

[20] Shiryaev A.N.: Probability.- New York et al.: Springer-Verlag, 1984.

[21] Trenogin V.A. Functional analysis.- Moscow: Fizmatlit, 1980. (in Russian)

[22] Yosida Y. Functional Analysis.- Berlin et. al.: Springer-Verlag, 1965.

[23] Zhelobenko D.P. Compact Lie groups and their representations.- Providence, RI: Amer. Math. Soc., 1973.