

A TRANSFORMATION OF THE PHASE SPACE OF A DIFFUSION PROCESS THAT REMOVES THE DRIFT

To cite this article: A K Zvonkin 1974 *Math. USSR Sb.* **22** 129

View the [article online](#) for updates and enhancements.

Related content

- [ON THE DIRICHLET PROBLEM FOR BELLMAN'S EQUATION IN A PLANE DOMAIN](#)
M V Safonov
- [SOME NEW RESULTS IN THE THEORY OF CONTROLLED DIFFUSION PROCESSES](#)
N V Krylov
- [FUNCTIONALS OF RANDOM PROCESSES AND INFINITE-DIMENSIONAL OSCILLATORY INTEGRALS CONNECTED WITH THEM](#)
V I Bogachëv

Recent citations

- [Alexander Yur'evich Veretennikov *et al*](#)
- [A Zvonkin's transformation for stochastic differential equations with singular drift and applications](#)
Shao-Qin Zhang and Chengui Yuan
- [Existence, uniqueness and ergodic properties for time-homogeneous Itô-SDEs with locally integrable drifts and Sobolev diffusion coefficients](#)
Haesung Lee and Gerald Trutnau

A TRANSFORMATION OF THE PHASE SPACE OF A DIFFUSION PROCESS THAT REMOVES THE DRIFT

UDC 519.2

A. K. ZVONKIN

Abstract. In this paper we construct a one-to-one (and quasi-isometric) transformation of a phase space that allows us to pass from a diffusion process with nonzero drift coefficient to a process without drift. Using this transformation we construct strong solutions of stochastic differential equations with a "bad" drift coefficient and give other applications.

Bibliography: 21 items.

§1. Introduction

Let x_t be a (Markov) diffusion process satisfying an Itô stochastic differential equation:

$$\begin{aligned} dx_t &= b(t, x_t)dt + \sigma(t, x_t)d\omega_t, \\ x_0 &= \xi. \end{aligned} \tag{1}$$

Here x_t and ξ are random variables with values in \mathbb{R}^n , $t \in [0, T]$, ω_t is the standard n -dimensional Wiener process, and for all t and x , $b(t, x)$ is an n -dimensional drift vector, and $\sigma(t, x)$ is a diffusion matrix of size $n \times n$.

In this paper we construct a mapping $u(t, x): [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that the process $y_t = u(t, x_t)$ has drift coefficient zero, and this mapping $u(t, x)$ possesses the following properties:

a) For each $t \in [0, T]$ the mapping $u(t, x) = u_t(x): \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a one-to-one mapping on all of \mathbb{R}^n ; moreover, it is a quasi-isometry, i.e.

$$M \geq \frac{|u(t, x_2) - u(t, x_1)|}{|x_2 - x_1|} \geq m > 0.$$

b) Itô's formula is applicable to the functions $u(t, x)$ and $v(t, x)$ (here for fixed $t \in [0, T]$ the mapping $v_t(y) = v(t, y)$ is an inverse to the mapping $u_t(x)$).

Obviously, $x_t = v(t, y_t)$. The mappings $u(t, x)$ and $v(t, y)$ allow us to reduce

many problems about processes satisfying equation (1) to the same problems about processes that already do not have drift (for such processes certain problems are solved more easily).

The decisive feature in this paper is an estimate *from below* of the derivatives of a parabolic differential equation. The problem of obtaining such estimates deserves more detailed independent consideration. Here we are concerned with it to the minimal extent necessary for our purposes (cf. Theorem 2).

In §3 we construct the mappings $u(t, x)$ and $v(t, y)$ and study their properties (Theorem 2), and also give the first applications: a very important estimate due to N. V. Krylov is generalized to the case of functions depending on t and x , and using this we prove that Itô's formula applies not only to functions of class $C^{1,2}$, but also to functions of class $W_p^{1,2}$ with sufficiently large p (Theorem 3).

In §4, using the mappings u and v , we prove new existence and uniqueness theorems for solutions of equations of the form (1) (Theorem 4).

In §5 Theorem 4 is used to prove the existence of an optimal strategy for a one-dimensional controllable diffusion process.

The author expresses deep thanks to A. N. Širjaev and N. V. Krylov for their constant attention to this work, valuable discussion and useful comments.

§2. Preliminary remarks

1. *Notation.* The action takes place in n -dimensional Euclidean space \mathbf{R}^n , equipped with an affine structure, i.e. to each point $x \in \mathbf{R}^n$ there is associated an n -dimensional linear space L_x^n "with origin at the point x " (and with the natural operation of identification between all $L_x^n, x \in \mathbf{R}^n$). A scalar product generated by the metric of \mathbf{R}^n is fixed in L_x^n . A system of Cartesian coordinates is fixed in \mathbf{R}^n .

1) $x \in \mathbf{R}^n, t \in [0, T]$.

$b(t, x)$ is a vector field on \mathbf{R}^n , i.e. for all $(t, x) \in [0, T] \times \mathbf{R}^n$ the vector

$$b(t, x) = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in L_x^n.$$

For each $(t, x) \in [0, T] \times \mathbf{R}^n, \sigma(t, x)$ is the matrix

$$\begin{pmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \dots & \dots & \dots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{pmatrix}$$

of the linear map $\sigma: L_x^n \rightarrow L_x^n$.

The process

$$\omega_t = \begin{pmatrix} \omega_t^{(1)} \\ \vdots \\ \omega_t^{(n)} \end{pmatrix} \in \mathbf{R}^n$$

is the standard n -dimensional Wiener process.

Equation (1) is understood in an integral sense: with probability 1 for all $t \in [0, T]$ simultaneously

$$x_t = \xi + \int_0^t b(s, x_s) ds + \int_0^t \sigma(s, x_s) d\omega_s, \tag{2}$$

where the second integral on the right is Itô's stochastic integral (cf. [1]).

2) For a mapping $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ we have (cf., for example, [2]):

The derivative $f'(x)$ is a linear map $f'(x): L_x^n \rightarrow L_y^m$ (here $y = f(x) \in \mathbf{R}^m$) with matrix $(\partial f_i / \partial x_j)$, consisting of m rows and n columns (here x_1, \dots, x_n are the coordinates of the point x in \mathbf{R}^n and f_1, \dots, f_m are functions $f_k: \mathbf{R}^n \rightarrow \mathbf{R}$; namely, $f_k(x) = y_k$ if $f(x) = y$, where the y_k are the coordinates of the point y in \mathbf{R}^m).

The second derivative $f''(x)$ is a bilinear mapping $f''(x): L_x^n \times L_x^n \rightarrow L_y^m$.

The partial derivatives of the mappings are defined analogously.

3) We introduce a symmetric matrix $a(t, x) = \sigma \cdot \sigma^*$ (here σ^* is the transpose matrix of the matrix σ); it will play the role of the matrix of the bilinear form on L_x^n , $a: L_x^n \times L_x^n \rightarrow \mathbf{R}$. (1)

Let e_1, \dots, e_n be an arbitrary basis of L_x^n , and d_1, \dots, d_n the basis biorthogonal to it (i.e. $(e_i, d_j) = \delta_{ij}$). (2)

The differential operator $\mathcal{L}^{(x_t)}$ corresponding to the process x_t is defined in the following way: for $u: [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}^m$

$$\mathcal{L}^{(x_t)} u(t, x) = u'_t + u'_x \cdot b(t, x) + \frac{1}{2} \sum_{i,j=1}^n u''_{xx}(e_i, e_j) \cdot a(t, x)(d_i, d_j). \tag{3}$$

4) As the norm of matrices it is convenient for us to choose $|(a_{ij})| = \max_{i,j} |a_{ij}|$; in correspondence with this

$$|f'(x)| = \max_{i,j} \left| \frac{\partial f_i}{\partial x_j}(x) \right|,$$

$$|f''(x)| = \max_{i,j,k} \left| \frac{\partial^2 f_k}{\partial x_i \partial x_j}(x) \right|.$$

For a domain $D \subseteq \mathbf{R}^n$ the space $W_p^{1,2}([0, T] \times D)$, $p > 1$, is the completion of the family of C^∞ functions with compact support with respect to the norm

$$\|u(t, x)\|_{W_p^{1,2}} = \|u(t, x)\|_{L_p} + \|u'_t(t, x)\|_{L_p} + \|u'_x(t, x)\|_{L_p} + \|u''_{xx}(t, x)\|_{L_p},$$

where

$$\|f(t, x)\|_{L_p} = \left(\int_{[0,T] \times D} |f(t, x)|^p dt dx \right)^{1/p}.$$

(1) It would be legal to assume a to be a bilinear form on the dual space $(L_x^n)^*$ to L_x^n ; but we identify L_x^n and $(L_x^n)^*$ by fixing a scalar product in L_x^n .

(2) We note that if the basis e_1, \dots, e_n is orthonormal, then $e_x = d_x$.

(3) The sum on the right side of (3) does not depend on the choice of basis (for example, see [3]).

Analogously we define the space $W_p^2(D)$ (the term with the derivative in t is omitted and the integration is taken over D). See [4] and [5] for more details about the properties of the spaces W .

2. The main facts about equation (1) and the associated parabolic partial differential equations that we shall need are collected in the following theorem.

Theorem 1. *Suppose that the coefficients $b(t, x)$ and $\sigma(t, x)$ in equation (1) satisfy the following conditions:*

A. *Continuity of the diffusion coefficient: $\sigma(t, x)$ is continuous with respect to the pair of arguments.*

B. *Uniform ellipticity of the operator $\mathcal{L}^{(x_t)}$: there exists a number $\mu > 0$ such that for any vector $e \in L_x^n$*

$$a(t, x)(e, e) \geq \mu^2 |e|^2.$$

C. *Boundedness of the coefficients: there is a number C such that*

$$|b(t, x)| + |\sigma(t, x)| \leq C.$$

Then the following assertions hold:

1) *There exist a probability space $\{\Omega, \mathcal{F}, P\}$ and a pair of processes $\{(x_t, w_t), \mathcal{F}_t\}$ on it such that $\{w_t, \mathcal{F}_t\}$ is a Wiener process and relation (2) holds.*

2) *If the processes (x_t', w_t') and (x_t'', w_t'') satisfy relation (2), then all the finite-dimensional distributions of the processes x_t' and x_t'' coincide.*

3) *If we consider the solutions of equation (1) on segments $[t_0, T] \subseteq [0, T]$ for distinct initial values $x_{t_0} = x \in \mathbf{R}^n$, then this family of solutions forms an (inhomogeneous) strong Markov process. The mathematical expectation with respect to the measure corresponding to this process will be denoted $\mathbf{E}_{(t_0, x)}$.*

Consider the equation

$$\begin{aligned} \mathcal{L}^{(x_t)} u(t, x) &= 0, \\ u(T, x) &= \varphi(x), \end{aligned} \tag{4}$$

where $u: [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}^m$ and $\varphi: \mathbf{R}^n \rightarrow \mathbf{R}^m$, where the function $\varphi(x)$ grows to infinity slower than $e^{k|x|^2}$ for any $k > 0$ (i.e. $\varphi(x)$ is a "slowly growing" function), and $\varphi \in W_p^2(D)$ for any bounded domain $D \subseteq \mathbf{R}^n$ with $p > (n+2)/2$.

4) *In the class of slowly growing functions there exists a solution to equation (4) that is unique and $u \in W_p^{1,2}([0, T] \times D)$ for any bounded domain $D \subset \mathbf{R}^n$.*

5) *This solution has the probability representation*

$$u(t, x) = \mathbf{E}_{(t, x)} \varphi(x_T).$$

Proof. Assertions 1), 2) and 3) proved in [6] (Corollary 3.2 and Theorem 6.2). For assertion 4), see [5], Theorem IV.9.1 and §14 of Chapter IV there. Assertion 5) is proved in §11 of [6] (Corollary 11.1, Remark 2).

§3. A transformation that removes the drift

Theorem 2. Let $b(t, x)$ and $\sigma(t, x)$ satisfy conditions A, B, C, and let the function $u: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a solution of the equation

$$\begin{aligned} \mathcal{L}^{(x_t)} u(t, x) &= 0, \\ u(T, x) &= x. \end{aligned} \tag{5}$$

Then there is a $T > 0$ ⁽⁴⁾ satisfying the following conditions:

1) For any fixed $t \in [0, T]$ the function $u(t, x) = u_t(x)$ is a one-to-one mapping $u_t: \mathbb{R}^n \rightarrow \mathbb{R}^n$ onto all of \mathbb{R}^n . The inverse mapping to $u_t(x)$ is denoted by $v_t(y) = v(t, y)$. The inverse of the mapping $\tilde{u}: [0, T] \times \mathbb{R}^n \rightarrow [0, T] \times \mathbb{R}^n$, $\tilde{u}(t, x) = (t, u(t, x))$, will be $\tilde{v}(t, y) = (t, v(t, y))$.

2) The mappings $u(t, x)$ and $v(t, y)$ have bounded derivatives $u'_x(t, x)$ and $v'_y(t, y)$; in particular, they are quasi-isometries: there is a constant $m > 0$ such that

$$m \leq \frac{|u(t, x_2) - u(t, x_1)|}{|x_2 - x_1|} = \frac{|y_2 - y_1|}{|v(t, y_2) - v(t, y_1)|} \leq \frac{1}{m};$$

the mappings \tilde{u} and \tilde{v} take bounded sets into bounded sets.

3) The functions $u(t, x)$ and $v(t, y)$ belong to the classes $W_p^{1,2}([0, T] \times D)$ for any bounded set $D \subset \mathbb{R}^n$ and all $p > 1$. In particular, the derivatives $u'_x(t, x)$ and $v'_y(t, y)$ satisfy a Hölder condition in t and x with arbitrary exponent $\alpha < 1$.

Proof. I. We denote by V_x^r the ball of radius r in \mathbb{R}^n with center at the point x and we take $r = \max\{T, 1\}$. We shall show that the derivatives $\partial u_k / \partial x_i$, $i, k = 1, \dots, n$, satisfy a Hölder condition in $[0, T] \times V_x^r$, where the Hölder exponent can be chosen as close to 1 as desired, and the Hölder constant is the same for all cylinders $[0, T] \times V_x^r$, $x \in \mathbb{R}^n$.

In fact, the function $u_k(t, x)$ is a solution of the equation

$$\begin{aligned} \mathcal{L}^{(x_t)} u_k(t, x) &= 0, \\ u_k(T, x) &= x_k. \end{aligned} \tag{5k}$$

Therefore, by assertion 4) of Theorem 1, $u_k \in W_p^{1,2}([0, T] \times V_x^r)$ with arbitrary $p > (n + 2)/2$ (since the function $\phi_k(x) = x_k$ belongs to $W_p^2(V_x^r)$ with arbitrary $p > (n + 2)/2$). By the second assertion of Theorem II.3.3 of [5]⁽⁵⁾ if $p > n + 2$, then any derivative $\partial u_k / \partial x_i$ satisfies a Hölder condition in x and t in $[0, T] \times V_x^r$, where the Hölder exponent α can be chosen arbitrarily in the limits $0 < \alpha < 1 - (n + 2)/p$ (since p can be taken as large as desired, then α is as close to 1 as desired), and the Hölder constant is estimated by the norm of the function $u_k(t, x)$ in the space $W_p^{1,2}([0, T] \times V_x^r)$. But this norm, by the remark at the end of §10 of Chapter IV of [5], in addition to the

⁽⁴⁾ By more complicated arguments we can show that all the results of Theorem 2 are true for arbitrary finite T . However, for our purposes it suffices to obtain them for small T .

⁽⁵⁾ The possibility of using this lemma was indicated to the author by N. V. Krylov.

constants μ and C from conditions B and C of Theorem 1, depends only on the norm of the function $\phi_k(x) = x_k$ in the space $W_p^2(V_x^{r+1})$ and on $\max_{[0, T] \times V_x^{r+1}} |u_k(t, x)|$.

In view of the probabilistic representation of the solution of (5k) (see assertion 5) of Theorem 1) the last quantity is easily estimated by $\max_{V_x^{r+1} + CT} |x_k|$ (C is the constant of condition C).

Finally, because of the linearity of the operator $\mathcal{Q}^{(x,t)}$ the estimate of the Hölder constant must be preserved if we add an arbitrary constant to the function $\phi_k(x) = x_k$ or, what is the same, if we translate the balls V_x^{r+1} and V_x^{r+1+CT} to an arbitrary place in the space \mathbb{R}^n . Thus we have shown that the Hölder constant can be chosen to be the same for all the domains $[0, T] \times V_x^r, x \in \mathbb{R}^n$.

II. From the assertion of part I it immediately follows that all the derivatives $\partial u_k / \partial x_i, i, k = 1, \dots, n$, satisfy a Hölder condition in t that is uniform in x . But since

$$\left. \frac{\partial u_k}{\partial x_i} \right|_{t=T} \equiv \frac{\partial \varphi_k}{\partial x_i} = \begin{cases} 0, & i \neq k, \\ 1, & i = k, \end{cases}$$

they all are furthermore bounded in absolute value by a common constant (and hence the norm of the matrix of u'_x is bounded: $|u'_x(t, x)| \leq K$).

III. If two bounded functions satisfy a Hölder condition with exponents α and $\beta, \alpha \leq \beta$, then their product (and obviously, their sum) satisfies a Hölder condition with exponent α .

In fact, let

$$|f(z_2) - f(z_1)| \leq K_1 |z_2 - z_1|^\alpha, \quad |f(z)| \leq M_1,$$

$$|g(z_2) - g(z_1)| \leq K_2 |z_2 - z_1|^\beta, \quad |g(z)| \leq M_2,$$

then

$$\begin{aligned} & |f(z_2)g(z_2) - f(z_1)g(z_1)| = |f(z_2)g(z_2) - f(z_2)g(z_1) + f(z_2)g(z_1) - f(z_1)g(z_1)| \\ & \leq |f(z_2)g(z_2) - f(z_2)g(z_1)| + |f(z_2)g(z_1) - f(z_1)g(z_1)| \leq M_1 |g(z_2) - g(z_1)| \\ & \quad + M_2 |f(z_2) - f(z_1)| \leq M_1 K_2 |z_2 - z_1|^\beta + M_2 K_1 |z_2 - z_1|^\alpha \leq \text{Const} \cdot |z_2 - z_1|^\alpha. \end{aligned}$$

Corollary. $\det u'_x(t, x)$ satisfies a Hölder condition in t uniformly in x .

Analyzing the proofs of parts I and II, it is easy to see that the estimate of the Hölder constant for $\det u'_x(t, x)$ depends only on T and on the constants μ and C of conditions B and C of Theorem 1, where this estimate does not grow with a decrease in T . Hence, if T is sufficiently close to 0, then $\det u'_x(t, x)$ (uniformly in x) is close to $\det u'_x(T, x) = \det \phi'(x) \equiv 1$, i.e.

$$\det u'_x(t, x) \geq k > 0, \tag{6}$$

In what follows, unless the contrary is stated, we shall assume that T has been chosen to be so small that inequality (6) holds for it.

IV. Let $\lambda_{\min}(t, x)$ be the smallest (in absolute value) eigenvalue of the matrix $u'_x(t, x)$. From the inequalities $|u'_x| \leq K$ and $\det u'_x \geq k$ it follows that $|\lambda_{\min}(t, x)| \geq c > 0$, from which we get

$$|u(t, x_2) - u(t, x_1)| \geq c |x_2 - x_1|. \tag{7}$$

Condition (7), as shown in [7] (Corollary to Theorem II), is sufficient for the *global invertibility* of the mapping $u_t(x)$. We note that for dimension $n > 1$ this fact is not trivial. In particular, the single inequality (6) is in general insufficient for the global invertibility of $u_t(x)$ (although it is sufficient for local invertibility).

From (7) we also get the boundedness of all the derivatives $\partial v_k / \partial y_j$ by the constant $1/c$.

The mapping \tilde{u} is obviously also invertible. Its derivative at the point (t, x) is an $(n + 1) \times (n + 1)$ matrix:

$$\tilde{u}'(t, x) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ u'_t & \dots & \dots & u'_x \end{pmatrix}.$$

Since $u'_t \in L_p$ (with any $p > 1$), it follows that \tilde{u} satisfies a Hölder condition (with any $\alpha < 1$) and hence takes bounded subsets of $[0, T] \times \mathbf{R}^n$ into bounded sets. The Jacobian of $\tilde{u}(t, x)$ coincides with the Jacobian of $u_t(x)$.

V. In order to complete the proof of the theorem it remains to show that for any bounded domain $E \subset \mathbf{R}^n$ the function $v(t, y)$ belongs to $W_p^{1,2}([0, T] \times E)$ for all $p > 1$.

First suppose $u \in C^\infty$; then (see, for example, [2]) the following assertions hold:

1) v'_t is an n -vector.

$$v'_t(t, y) = -(u'_x)^{-1}(t, v(t, y)) \cdot u'_t(t, v(t, y)) = -(u'_x)^{-1} \cdot u'_t.$$

2) v'_y is a linear mapping $v'_y: L_y^n \rightarrow L_x^n$, inverse to the mapping $u'_x: L_x^n \rightarrow L_y^n$:

$$v'_y(t, y) = (u'_x)^{-1}(t, v(t, y)).$$

3) The second derivative v''_{yy} is a bilinear mapping $v''_{yy}: L_y^n \times L_y^n \rightarrow L_x^n$, which is given by the formula

$$v''_{yy}(l_1, l_2) = -(u'_x)^{-1} \cdot u''_{xx}((u'_x)^{-1} \cdot l_1, (u'_x)^{-1} \cdot l_2),$$

where $l_1, l_2 \in L_y^n$ and $u''_{xx}: L_x^n \times L_x^n \rightarrow L_y^n$.

Now suppose $u \in W_p^{1,2}([0, T] \times D)$, where the set $[0, T] \times D$ contains the inverse image of the set $[0, T] \times E$ under the map \tilde{u} . Consider the functions $u_k \in C^\infty$,

$\|u_k - u\|_{W_p^{1,2}} \rightarrow 0$ as $k \rightarrow \infty$. Without loss of generality we may assume that for all

functions u_k the inequalities (6), (7) and $|(u_k)'_x| \leq K$ hold. We shall show that the derivatives $(v_k)'_t, (v_k)'_y$ and $(v_k)''_{yy}$ converge in L_p to the expressions 1), 2) and 3) respectively.

a) $v_k \rightarrow v$ uniformly in each bounded domain. In fact, put $v_k(t, y) = x_k$ and $v(t, y) = x$; then

$$u_k(t, x_k) = y = u(t, x);$$

but, beginning with some k ,

$$|u_k(t, x_k) - u(t, x_k)| < \varepsilon$$

(since if $p > (n+2)/2$, then $u_k \rightarrow u$ uniformly in each bounded domain; [4], §8.1), and hence

$$|u(t, x) - u(t, x_k)| \leq \varepsilon;$$

but in view of (7)

$$|u(t, x) - u(t, x_k)| \geq c|x - x_k|,$$

from which $|x - x_k| < \varepsilon/c$, as was required.

b) We shall show that if $f_k(t, x) \rightarrow f(t, x)$ in L_p , then the functions $\hat{f}_k(t, y) = f(t, v(t, y))$. In fact,

$$\begin{aligned} & \int |f_k(t, v_k(t, y)) - f(t, v(t, y))|^p dt dy \\ & \leq 2^p \int |f_k(t, v_k(t, y)) - f(t, v_k(t, y))|^p dt dy \\ & + 2^p \int |f(t, v_k(t, y)) - f(t, v(t, y))|^p dt dy. \end{aligned}$$

By the change of variables $v_k(t, y) = x$ the first integral transforms to

$$\int |f_k(t, x) - f(t, x)|^p \det (u_k)_x' dt dx \leq K \int |f_k(t, x) - f(t, x)|^p dt dx \rightarrow 0,$$

and the second integral tends to 0 since v_k converges uniformly to v .

c) Our assertion now follows from the fact that $u_k \rightarrow u$ in $W_p^{1,2}$ (i.e. the derivatives of u_k converge in L_p to the corresponding derivatives of u) and all the elements of the matrices $(u_x')^{-1}$ and $((u_k)_x')^{-1}$ are bounded.

Theorem 2 is proved.

Put $y_t = u(t, x_t)$. Then obviously $x_t = v(t, y_t)$. If we could apply Itô's formula, we would obtain

$$dy_t = [\mathcal{L}^{(x_t)} u(t, x_t)] dt + [u_x(t, x_t) \cdot \sigma(t, x_t)] dw_t, \quad (8)$$

and since $\mathcal{L}^{(x_t)} u(t, x) \equiv 0$ in view of (5), then $dy_t = s(t, y_t) dw_t$, where

$$s(t, y) = u_x'(t, v(t, y)) \cdot \sigma(t, v(t, y)). \quad (9)$$

But the function $u(t, x)$ does not have two continuous derivatives in x and one in t , as is required by the usual Itô formula. Nevertheless, we shall show that in fact Itô's formula is applicable to a wider class of functions than $C^{1,2}$ (and, in particular, to $u(t, x)$ and $v(t, y)$).

Theorem 3. *Suppose the coefficients of equation (1) satisfy conditions A, B, C. Then the following assertions are true:*

a) *If $f(t, x) \in L_p([0, T] \times D)$, $D \subset \mathbf{R}^n$ is a bounded domain and $p > n + 2$, then*

$$\mathbf{E} \int_0^{T \wedge \tau} |f(t, x_t)| dt \leq N \|f\|_{L_p}, \tag{10}$$

where τ is the moment of the first passage of the process x_t from the domain D , $T \wedge \tau = \min\{T, \tau\}$, and the constant N depends only on T, μ and C .

b) If the function $u(t, x) \in W_p^{1,2}([0, T] \times D)$, $p > n + 2$, then for $y_t = u(t, x_t)$ relation (8) holds.

Estimate (10) plays a very important role in the theory of stochastic differential equations. For functions $f(x)$ not depending on t (and for a wider class of processes x_t), this estimate is obtained in [8] by means of the theory of convex polygons. For functions $f(t, x)$ depending on t , but for processes x_t with zero drift coefficient, it is obtained in [6]. Estimate (10) plays the main role in the proof of Itô's formula: our argument is completely analogous to the proof of Itô's formula for functions not depending on t and belonging to W^2 (Theorem 4 of [9]).

Proof of Theorem 3. I. The matrix $s(t, y)$ defined in (9) satisfies conditions A, B, C. In fact:

1) $s(t, y)$ is bounded and continuous, since $\sigma(t, x)$ (by conditions A and C) and $u'_x(t, x)$ (Theorem 2) are bounded and continuous, and $v(t, y)$ is continuous.

$$2) \quad A = s \cdot s^* = (u'_x \cdot \sigma) \cdot (u'_x \cdot \sigma)^* = u'_x \cdot \sigma \cdot \sigma^* \cdot (u'_x)^* = u'_x \cdot a \cdot (u'_x)^*,$$

the uniform ellipticity of the operator A follows from the uniform ellipticity of the matrix a (Condition B) and the uniform nondegeneracy and boundedness of the matrix u'_x (Theorem 2).

II. Consider a pair of processes $\{(y_t, w_t), \mathcal{F}_t\}$ satisfying the condition

$$\begin{aligned} dy_t &= s(t, y_t) dw_t, \\ y_0 &= \zeta. \end{aligned} \tag{11}$$

Such a pair of processes exists (cf. Theorem 1, assertions 1)–3)). Let $v(t, y) \in W_p^{1,2}([0, T] \times E)$, $p > n + 2$. We shall prove an equality that is equivalent to Itô's formula, in an integral formulation: if δ is the first exit time of y_t from E , then

$$v(T \wedge \delta, y_{T \wedge \delta}) - v(0, \zeta) = \int_0^{T \wedge \delta} \mathcal{L}^{(y_t)} v(t, y_t) dt + \int_0^{T \wedge \delta} v'_y(t, y_t) \cdot s(t, y_t) dw_t. \tag{12}$$

1) In Lemma 5.2 of [6] for processes y_t with zero drift coefficient the following estimate is proved (the same as in (10)): if $f(t, y) \in L_p$ and $p \geq n + 2$, then

$$\mathbf{E} \int_0^{T \wedge \delta} |f(t, y_t)| dt \leq N \cdot \|f\|_{L_p}. \tag{13}$$

2) Applying estimate (13) to the first integral on the right side of (12), we find that the integral is defined and

$$\mathbf{E} \int_0^{T \wedge \delta} |\mathcal{L}^{(y_t)} v(t, y_t)| dt \leq N \cdot \|\mathcal{L}^{(y_t)} v\|_{L_p} \leq N_1 \cdot \|v\|_{W_p^{1,2}}. \tag{14}$$

3) Applying estimate (13) to the second integral on the right side of (12), we will get that the integral is defined and

$$\begin{aligned} \mathbf{E} \left| \int_0^{T \wedge \delta} v'_y(t, y_t) \cdot s(t, y_t) d\omega_t \right|^2 &\leq N_2 \cdot \mathbf{E} \int_0^{T \wedge \delta} |v'_y(t, y_t)|^2 dt \\ &\leq N_3 \cdot \|v'_y\|_{L^p}^2 \leq N_4 \cdot \|v\|_{W^{1,2}_p}^2. \end{aligned} \tag{15}$$

4) If the function $v(t, y)$ is sufficiently smooth, then (12) coincides with the usual Itô formula. Now suppose that $v \in W^{1,2}_p$. We choose a sequence of smooth functions $v_k(t, y)$ such that $\|v_k - v\|_{W^{1,2}_p} \rightarrow 0$ as $k \rightarrow \infty$. Then

a) by Sobolev's theorem ([4], §8.1) $v_k \rightarrow v$ for $p > (n + 2)/2$ uniformly in $[0, T] \times E$, which guarantees the (almost sure) convergence of the left side of (12);

b) the integrals of v_k in the right side of (12) converge to the corresponding integrals of v in view of the estimates (14) and (15).

III. We take $\zeta = u(0, \xi)$, and the function $v(t, y)$ of Theorem 2; then the Itô formula can be applied to $v(t, y_t)$; we shall show that the process $x_t = v(t, y_t)$ satisfies equation (1). Actually, according to (3) and (12),

$$\begin{aligned} dx_t = &\left[v'_t(t, y_t) + \frac{1}{2} \sum_{i,j=1}^n v''_{yy}(t, y_t)(l_i, l_j) \cdot A(t, y_t)(k_i, k_j) \right] dt \\ &+ v'_y(t, y_t) \cdot s(t, y_t) d\omega_t, \end{aligned}$$

where l_1, \dots, l_n is an arbitrary basis of L^n_y and k_1, \dots, k_n a basis biorthogonal to it.

1) We compute the diffusion coefficient: $v'_y \cdot s = v'_y \cdot u'_x \cdot \sigma = \sigma(t, x)$, since $v'_y = (u'_x)^{-1}$.

2) We compute the drift coefficient (cf. part V of the proof of Theorem 2):

a) $v'_t = -(u'_x)^{-1} \cdot u'_t$.

b) $v''_{yy}(l_i, l_j) = -(u'_x)^{-1} \cdot u''_{xx}((u'_x)^{-1} \cdot l_i, (u'_x)^{-1} \cdot l_j) = -(u'_x)^{-1} \cdot u''_{xx}(e_i, e_j)$, where e_1, \dots, e_n is a basis in L^n_x which is the inverse image of the basis l_1, \dots, l_n under the mapping u'_x .

c) The matrix of the bilinear form $A = s \cdot s^* = u'_x \cdot \sigma \cdot \sigma^* \cdot (u'_x)^*$; the value of this form on the pair of vectors $k_i, k_j \in L^n_y$ coincides with the value of the form $a = \sigma \cdot \sigma^*$ on the pair of vectors $d_i = (u'_x)^* \cdot k_i, d_j = (u'_x)^* \cdot k_j$ (cf. [3]); the vectors d_1, \dots, d_n obviously form a basis of L^n_x biorthogonal to the basis e_1, \dots, e_n (since $(e_i, d_j) = (e_i, (u'_x)^* \cdot k_j) = (u'_x \cdot e_i, k_j) = (l_i, k_j) = \delta_{ij}$).

d) Hence the coefficient of dt is equal to

$$-(u'_x)^{-1} \left[u'_t + \frac{1}{2} \sum_{i,j=1}^n u''_{xx}(e_i, e_j) \cdot a(d_i, d_j) \right]. \tag{16}$$

But $\mathcal{L}^{(x_t)} u = 0$ by the definition of the function $u(t, x)$, i.e.

$$u'_t + \frac{1}{2} \sum_{i,j=1}^n u''_{xx}(e_i, \theta_j) \cdot a(d_i, d_j) = -u'_x \cdot b;$$

and hence expression (16) is equal to $(u'_x)^{-1} \cdot u'_x \cdot b = b(t, x)$.

IV. Now we shall prove (10). We have

$$\begin{aligned} \mathbf{E} \int_0^{T \wedge \tau} |f(t, x_t)| dt &= \mathbf{E} \int_0^{T \wedge \tau} |f(t, v(t, y_t))| dt = \mathbf{E} \int_0^{T \wedge \tau} |g(t, y_t)| dt \\ &\leq \mathbf{E} \int_0^{T \wedge \delta} |g(t, y_t)| dt \leq N_1 \cdot \|g\|_{L_p([0, T] \times E)}, \end{aligned} \tag{17}$$

where $[0, T] \times E$ is a cylinder contained in the set Q , which is the image of the set $[0, T] \times D$ under the mapping \tilde{u} , the function

$$g(t, y) = \begin{cases} f(t, v(t, y)), & (t, y) \in Q, \\ 0, & (t, y) \in [0, T] \times E \setminus Q, \end{cases}$$

and δ is the first exit time of the process y_t from the domain E ($\delta \geq \tau$ almost surely). The last inequality in (17) follows from (13). Here we also use the fact that the process x_t satisfying equation (1) is unique in the sense of measure (cf. assertion 2) of Theorem 1)), and therefore, in calculating the expectation, as x_t we can take $v(t, y_t)$. Furthermore,

$$\begin{aligned} \|g\|_{L_p}^p &= \int_{[0, T] \times E} |g(t, y)|^p dt dy = \int_Q |g(t, y)|^p dt dy = \int_Q |f(t, v(t, y))|^p dt dy \\ &= \int_{[0, T] \times D} |f(t, x)|^p \det u'_x(t, x) dt dx \leq K \int_{[0, T] \times D} |f(t, x)|^p dt dx = K \|f\|_{L_p([0, T] \times D)}^p \end{aligned}$$

and hence $\|g\|_{L_p} \leq \text{const} \cdot \|f\|_{L_p}$, which in conjunction with (17) proves (10).

V. In order to prove the Itô formula (8), it is sufficient to repeat the arguments of part II verbatim, using estimate (10) each time instead of (13).

Theorem 3 is proved.

Remark. The results of Theorem 3 extend trivially from small T to arbitrary finite T .

§4. Strong and weak solutions of stochastic differential equations.

Existence and uniqueness

1. The classical existence and uniqueness theorem for a solution of equation (1) (cf. [10]) requires that the coefficients b and σ satisfy (uniformly in t) a Lipschitz condition in x . Here the solution x_t is constructed as a measurable functional of the process w_t . The later development of the theory, however, required an essential extension of the class of admissible coefficients. Many extremely important results in this direction have been obtained, for example, in [11], [9] and [6]. However in all these papers a solution of equation (1) is not understood in the sense of [10]: it is constructed on another probability space, with another Wiener process \tilde{w}_t , etc.; briefly, a measure is constructed which corresponds to the coefficients b and σ , or a weak solution (for a complete formulation see below). The uniqueness as a rule is not

understood in the sense of the coincidence of the trajectories of the two solutions, but in the sense of the coincidence of the corresponding measures (an exception is presented by [12], where trajectorial uniqueness is studied). The well-known example of Tanaka given below shows that the existence of a weak solution does not mean the existence of a strong solution, and weak uniqueness does not imply strong uniqueness.

For studying many questions it is entirely sufficient to obtain a measure as a solution of equation (1): to this relate questions connected with the study of the corresponding parabolic and elliptic equations, or limit theorems of various sorts. Meanwhile other results (e.g. comparison theorems like Lemma 4 of §16 of [13]) simply lose all meaning for weak solutions.

In Theorem 4 we shall prove the existence of a strong solution and its uniqueness in the trajectory sense under substantially weaker (as compared with [10] or [12]) assumptions on the drift coefficient. Here we impose a nondegeneracy condition on the diffusion coefficient (which was not done in [10] or [12]).

2. Definitions.

1) *Strong solution.* We start with a probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$, a Wiener process $\{w_t, \mathcal{F}_t^w\}$ ⁽⁶⁾ on it and measurable functions $b_k: [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}$, $k = 1, \dots, n$, and $\sigma_{ij}: [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}$, $i, j = 1, \dots, n$. We must construct a process x_t that for each t is measurable with respect to the σ -algebra $\overline{\mathcal{B}}(\mathbf{R}^n) \times \mathcal{F}_t^w$, generated by the random variables ξ and w_s , $s \leq t$ (here $\overline{\mathcal{B}}(\mathbf{R}^n)$ is the σ -algebra of the absolutely measurable subsets of \mathbf{R}^n) and such that equality (2) holds with probability 1. Such a process $\{x_t, \overline{\mathcal{B}}(\mathbf{R}^n) \times \mathcal{F}_t^w\}$ is called a *strong solution* of equation (1).

2) *Weak solution.* We start with the functions $b_k(t, x)$ and $\sigma_{ij}(t, x)$. We must construct a probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$ and a pair of processes $\{(x_t, w_t), G_t\}$ on it such that $\{w_t, G_t\}$ is a Wiener process and relation (2) holds. The pair of processes $\{(x_t, w_t), G_t\}$ is called a *weak solution* of equation (1).

Here the process x_t is not required to be measurable with respect to the σ -algebra $\overline{\mathcal{B}}(\mathbf{R}^n) \times \mathcal{F}_t^w$, generated separately by the random variables ξ and w_s , $s \leq t$. In other words, x_t is not a measurable functional on the Wiener trajectory. And therefore from the existence of a solution for one Wiener process w_t we do not obtain the existence of an analogous solution for another Wiener process \tilde{w}_t .

3) *Uniqueness.* We shall say that a solution (weak or strong) of equation (1) is *unique in the weak sense* or *unique in the sense of measure*, if all the finite-dimensional distributions coincide for any two processes $x_t^{(1)}$ and $x_t^{(2)}$.

We shall say that a solution (weak or strong) of equation (1) is *unique in the strong sense* or *unique with respect to trajectories* if:

- a) For strong solutions: if $x_t^{(1)}$ and $x_t^{(2)}$ are two solutions, then

$$\mathbf{P} \left\{ \sup_{0 \leq t \leq T} |x_t^{(2)} - x_t^{(1)}| > 0 \right\} = 0; \quad (18)$$

⁽⁶⁾ \mathcal{F}_t^w is the σ -algebra generated by the events $\{\omega: w_s(\omega) \in \Gamma, s \leq t\}$; we shall assume that it is complete with respect to \mathbf{P} .

b) for weak solutions: if $(x_t^{(1)}, w_t^{(1)})$ and $(x_t^{(2)}, w_t^{(2)})$ are two solutions, then from the equality $w_t^{(1)} \equiv w_t^{(2)}$ for all $t \in [0, T]$ a.s. (naturally it is also assumed that $\{\Omega^{(1)}, \mathcal{F}^{(1)}, \mathbf{P}^{(1)}\} = \{\Omega^{(2)}, \mathcal{F}^{(2)}, \mathbf{P}^{(2)}\}$) equality (18) follows.

The main results about weak solutions of equation (1) are collected in Theorem 1: for the existence of a weak solution and its uniqueness in the sense of measure conditions A, B, C of Theorem 1 are sufficient.

3. **Example.** The following example was first considered by Tanaka. It was later studied by many authors. We consider the equation

$$\begin{aligned} dx_t &= \sigma(x_t) dw_t, \\ x_0 &= 0, \end{aligned} \tag{19}$$

where

$$\sigma(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0, \end{cases}$$

(note that $\sigma^2(x) \equiv 1$).

a) *Weak uniqueness.* The martingale $x_t = \int_0^t \sigma(x_s) dw_s$, according to §2.5 of [1], is a Wiener process relative to the new time $\tau_t = \int_0^t \sigma^2(x_s) ds = t$, which in the given case coincides with the old time. Hence any solution x_t has the same (namely, the Wiener) finite-dimensional distributions.

b) *Existence of a weak solution.* As x_t we take an arbitrary Wiener process and construct $\tilde{w}_t = \int_0^t \sigma(x_s) dx_s$. The process \tilde{w}_t is also a Wiener process (we can make the same change of time as in part a)), where $d\tilde{w}_t = \sigma(x_t) dx_t$, from which we conclude that

$$\sigma(x_t) d\tilde{w}_t = \sigma(x_t) \sigma(x_t) dx_t = dx_t.$$

Hence (x_t, \tilde{w}_t) is a (weak) solution of equation (19).

c) *Absence of strong uniqueness.* Along with the solution (x_t, \tilde{w}_t) equation (19) also has the solution $(-x_t, \tilde{w}_t)$.

d) *Absence of a strong solution.* No solution of equation (19) is strong, i.e. x_t is automatically *not measurable* with respect to the σ -algebra \mathcal{F}_t^w . In order to see this, we consider the random variable θ_t , the local time described by the process x_t at 0. Since x_t is a Wiener process, θ_t is represented as (cf. [1], §3.8)

$$\nu_t = x_t^+ - \int_0^t \chi_{[0, \infty)}(x_s) dx_s.$$

In view of the obvious symmetry we have

$$\theta_t = x_t^- + \int_0^t \chi_{(-\infty, 0]}(x_s) dx_s.$$

By adding these two equalities we obtain

$$2\theta_t = |x_t| - \int_0^t \operatorname{sgn} x_s dx_s = |x_t| - \int_0^t \operatorname{sgn} x_s \cdot \sigma(x_s) d\omega_s = |x_t| - \omega_t,$$

from which, by the definition of local time,

$$\omega_t = |x_t| - 2\theta_t = |x_t| - \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \chi_{[0, \varepsilon]}(|x_s|) ds,$$

i.e. ω_t is measurably expressed by $|x_t|$, and $\mathcal{F}_t^w \subseteq \mathcal{F}_t^{|x|}$. Assuming that the process x_t is \mathcal{F}_t^w -measurable, we obtain the absurd consequence $\mathcal{F}_t^x \subseteq \mathcal{F}_t^{|x|}$.

It would be interesting to construct an analogous example with continuous coefficients $\sigma(x)$.

4. **Theorem 4.** *Let the coefficients $b(t, x)$ and $\sigma(t, x)$ satisfy conditions A, B, C of Theorem 1 and suppose that one of the following two conditions holds:*

1) *The diffusion satisfies a Lipschitz conditions, and the drift, a Dini condition:*

$$|\sigma(t, x_2) - \sigma(t, x_1)| \leq K |x_2 - x_1|,$$

$$\int_0^\varepsilon \frac{\rho(r)}{r} dr < \infty,$$

where $\rho(r)$ is the modulus of continuity of the function $b(t, x)$.

2) x is one-dimensional, $b(t, x)$ is bounded and measurable, and

$$|\sigma(t, x_2) - \sigma(t, x_1)| \leq K |x_2 - x_1|^\alpha, \quad \alpha \geq \frac{1}{2}.$$

Then equation (1) has a strong solution that is unique with respect to trajectories.

Proof. We shall first prove the theorem for small t (namely, for $t \in [0, T]$, where T comes from Theorem 2).

1) If the coefficients of equation (5) satisfy a Dini condition, then the second derivative u''_{xx} is bounded (cf. [14]), and hence $u'_x(t, x)$ satisfies a Lipschitz condition in x . Since $\sigma(t, x)$ is Lipschitz in x and $v(t, y)$ is Lipschitz in y , the function $s(t, y)$ defined by (9) is Lipschitz in y .

Hence, according to Itô's theorem [10], equation (11) has a strong solution y_t that is unique with respect to trajectories. In this case the process $x_t = v(t, y_t)$, according to part III of the proof of Theorem 3, is a solution of equation (1). The measurability of x_t with respect to the σ -algebra $\overline{\mathcal{B}}(\mathbb{R}^n) \times \mathcal{F}_t^w$ is obvious in view of the measurability of y_t and the continuity of $v(t, y)$.

Finally, if equation (1) has two distinct (strong) solutions $x_t^{(1)}$ and $x_t^{(2)}$, then equation (11) also has two (strong) solutions $y_t^{(1)} = u(t, x_t^{(1)})$ and $y_t^{(2)} = u(t, x_t^{(2)})$ (they are different in view of the single-valuedness of the mapping $u(t, x)$), which contradicts the theorem of Itô.

2) In the one-dimensional case (for measurable $b(t, x)$), $u'_x(t, x)$ satisfies a Hölder condition with arbitrary exponent $\alpha < 1$ (cf. Theorem 2), and $\sigma(t, x)$ satisfies one with $\alpha \geq \frac{1}{2}$; hence both $\hat{s}(t, x) = u'_x(t, x) \cdot \sigma(t, x)$ and $s(t, y) = \hat{s}(t, v(t, y))$ are Hölder with exponent $\alpha \geq \frac{1}{2}$.

x_{t_k} , obtained as a result of solving the preceding equation on $[t_{k-1}, t_k]$.

Each equation (22- κ), $\kappa = 0, \dots, m-1$, has a strong solution that is unique with respect to trajectories. Hence the process x_t given by (22) exists and is unique. Its measurability with respect to $\mathcal{B}(\mathbb{R}^n) \times \mathcal{F}_t^w$ is obvious. To prove the theorem it remains to note the following facts:

a) The process x_t satisfies equation (1) on the whole interval $[0, T]$. In fact, for example, for $t \in [t_1, t_2]$

$$x_t = x_{t_1} + \int_{t_1}^t b(s, x_s) ds + \int_{t_1}^t \sigma(s, x_s) d\omega_s;$$

but

$$x_{t_1} = \xi + \int_0^{t_1} b(s, x_s) ds + \int_0^{t_1} \sigma(s, x_s) d\omega_s,$$

and, substituting the expression for x_{t_1} in the first equality, we obtain

$$x_t = \xi + \int_0^t b(s, x_s) ds + \int_0^t \sigma(s, x_s) d\omega_s,$$

the procedure for the remaining intervals $[t_k, t_{k+1}]$ is analogous;

b) Any solution x_t of equation (1) also satisfies equation (22).

Theorem 4 is proved.

5. We make several remarks about possible generalizations of Theorem 4.

1) If the drift coefficient satisfies a Dini condition, then the Lipschitz condition on the diffusion coefficient can be replaced by any condition that guarantees the strong uniqueness for a process without drift (cf. Part II of [12]). In fact, all these conditions in [12] are formulated in terms of the modulus of continuity, and the modulus of continuity of $s(t, y)$ varies in comparison with $\sigma(t, x)$ by no more than a constant factor (since u'_x is Lipschitz).

2) If the drift coefficient does not satisfy a Dini condition (for example, if it is discontinuous), then, as the example in [15] shows, u''_{xx} can be unbounded in a neighborhood of some point. By changing this example a little, perhaps even by setting $\sigma(t, x) \equiv E$ (the identity matrix), it is possible to obtain for $s(t, y)$ a modulus of continuity which already leads to nonuniqueness of the solution; cf. the example in Part II of [12]. Thus, so far we have not succeeded in extending the result of Theorem 4 to the case of a measurable coefficient $b(t, x)$ (in dimensions $n \geq 2$).

However, the nonuniqueness example in [12] uses the degeneracy of the diffusion coefficient in an essential way. If, using the uniform nondegeneracy of $s(t, y)$, we were able to extend the strong uniqueness theorem to an arbitrary Hölder class (or even to functions satisfying all the Hölder conditions with $\alpha < 1$, i.e. just as u'_x), the result of Theorem 4 would automatically remain true for any measurable (and bounded) drift coefficient.

3) *On unbounded coefficients.* Conditions B and C of Theorem 1 can be weakened, by requiring the following:

a) Conditions B and C hold on $[0, T] \times D$ for any bounded domain $D \subseteq \mathbb{R}^n$ (the constants C and μ can depend on the domain).

b) On the coefficients $b(t, x)$ and $\sigma(t, x)$ we impose any conditions guaranteeing the existence (or extendability) of a weak solution (for example, the linearity of the growth of the coefficients to infinity; for other possible conditions see [16] and [1]).

In fact, let $\hat{b}(t, x)$ and $\hat{\sigma}(t, x)$ coincide with $b(t, x)$ and $\sigma(t, x)$ on $[0, T] \times D$, and let them be defined on the complement so that conditions B and C hold everywhere. Then the solution \hat{x}_t of equation (1) with coefficients \hat{b} and $\hat{\sigma}$ exists and is unique in the strong sense and (almost surely) coincides with the solution x_t of (1) with coefficients b and σ up to the moment of time τ_D of passage of the process x_t from the domain D . But the question of whether $\tau_D \rightarrow \infty$ (a.s.), if the domain D , extended, fills up the whole space, is in no way connected with the question of whether x_t is a strong or weak solution. The answer to this question depends only on the measure generated by the process x_t .

§5. The existence of optimal strategies

1. In the study of controllable diffusion processes proofs of existence theorems for optimal strategies meet with difficulties connected with the fact that already in the simplest examples optimal strategies lead to discontinuous drift coefficients and it is unclear whether a process with such a drift exists. Therefore for a rigorous account it is necessary either to use weak solutions and to actually replace the original control problem along individual trajectories by another—an extremal problem on the choice of best measure (for example, see [17])—or to impose on the problem strong enough restrictions so that the optimal strategy starts in a Lipschitz class (for example, see [18], [19]).

Using assertion 2) of Theorem 4, we shall prove the existence of an optimal strategy for a one-dimensional controllable diffusion process (the controlling parameter is only included in the drift coefficient, not in the diffusion coefficient).

2. **Lemma.** *Consider the equation*

$$\begin{aligned} \mathcal{L}^{(x_t)} u(t, x) &= -f(t, x), \\ u(T, x) &= \varphi(x), \end{aligned} \tag{23}$$

where $u: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$, and where the function $f(t, x)$ is bounded and measurable, and $\phi(x)$ is bounded and twice differentiable. We make the following assertions:

1) Equation (23) has a unique solution, which for any $p > 1$ and any bounded domain $D \subset \mathbb{R}^n$ belongs to the class $W_p^{1,2}([0, T] \times D)$.

2) This solution has the probabilistic representation

$$u(t, x) = \mathbf{E}_{(t,x)} \left\{ \int_0^T f(s, x_s) ds + \varphi(x_T) \right\}.$$

Proof. 1) Assertion 1 is proved in [5], Chapter IV, §14.

2) We apply Itô's formula to $u(t, x_t)$; we obtain

$$u(T, x_T) - u(t, x_t) = \int_t^T \mathcal{L}^{(x_t)} u(s, x_s) ds + \text{martingale} = - \int_t^T f(s, x_s) ds + \text{martingale};$$

taking the expectation $\mathbf{E}(t, x)$ of both sides and replacing $u(T, x_T)$ by $\phi(x_T)$, we obtain

$$\mathbf{E}(t, x) \phi(x_T) - u(t, x) = - \mathbf{E}(t, x) \int_0^T f(s, x_s) ds,$$

as was required.

Remark. The lemma admits many natural generalizations: solutions of equation (23) with increasing functions f and ϕ have a probabilistic representation (under suitable restrictions on their growth), with a function $f \in L_p([0, T] \times D)$ for sufficiently large p , and also solutions of problems in bounded domains.

3. 1) A family of (Markov) diffusion processes

$$\begin{aligned} dx_t^d &= b(t, x_t^d, d) dt + \sigma(t, x_t^d) d\omega_t, \\ x_t^d &= x, \end{aligned}$$

is given on $[0, T] \times \mathbf{R}$, which satisfies the following conditions:

a) $d \in D \subset \mathbf{R}^n$; the set D is compact; it is called the *control set*.

b) For each d the coefficients $b(t, x, d) = b_d(t, x)$ and $\sigma(t, x)$ satisfy conditions A, B, C of Theorem 1; the constants in conditions B and C do not depend on d .

c) The function $b(t, x, d)$ is measurable with respect to all the variables and continuous in d .

d) $|\sigma(t, x_2) - \sigma(t, x_1)| \leq |x_2 - x_1|^\alpha$, $\alpha \geq 1/2$.

2) A *strategy* on the interval $[t, T]$ is a mapping $\delta: [t, T] \times \Omega \rightarrow D$ which is $B[t, T] \times \mathcal{F}$ -measurable in all the variables and \mathcal{F}_t^s -measurable in ω for each $s \in [t, T]$.

We substitute the random variable $\delta(t, \omega)$ in the coefficient $b(t, x, d)$; if the equation (1) with the new coefficient $b(t, \omega) = b(t, x_t(\omega), \delta(t, \omega))$ has a solution, then the expectation with respect to the measure thus obtained will be denoted $\mathbf{E}_{(t, x)}^\delta$.

A *problem* consists of the determination of a strategy δ^* which would give

$$\mathbf{E}_{(t, x)}^{\delta^*} \left\{ \int_t^T f(s, x_s, \delta_s^*) ds + \phi(x_T) \right\} = \sup_{\delta} \mathbf{E}_{(t, x)}^{\delta} \left\{ \int_t^T f(s, x_s, \delta_s) ds + \phi(x_T) \right\} = U(t, x),$$

where

e) The function $\phi(x)$ is bounded and twice differentiable; the function $f(t, x, d)$ is bounded, measurable with respect to all the variables and continuous in d .

3) We write the Bellman equation:

$$\frac{\partial u}{\partial t} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 u}{\partial x^2} + \sup_{d \in D} \left[b(t, x, d) \frac{\partial u}{\partial x} + f(t, x, d) \right] = 0, \tag{24}$$

$$u(T, x) = \varphi(x).$$

a) Equation (24) is quasilinear and hence it has a unique solution possessing sufficient smoothness (cf. [5], Chapter V, §6).

b) The function $b(t, x, d)du/dx + f(t, x, d)$ (where u is a solution of equation (24)) for each fixed (t, x) is a continuous function on the compact set D ; hence it admits a maximum at some point $d^* = d^*(t, x)$; according to Theorem IX of Appendix III of [20] the function $d^*(t, x)$ can be taken to be measurable.

c) Hence the function $u(t, x)$ also satisfies the equation

$$\frac{\partial u}{\partial t} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 u}{\partial x^2} + b(t, x, d^*(t, x)) \frac{\partial u}{\partial x} + f(t, x, d^*(t, x)) = 0,$$

$$u(T, x) = \varphi(x),$$

and, according to the lemma,

$$u(t, x) = \mathbf{E}_{(t,x)}^{d^*} \left\{ \int_t^T f(s, x_s, d^*(s, x_s)) ds + \varphi(x_T) \right\},$$

where x_t is a solution of the equation

$$\begin{aligned} dx_t &= b(t, x_t, d^*(t, x_t)) dt + \sigma(t, x_t) d\omega_t, \\ x_t &= x, \end{aligned}$$

which exists according to assertion 2) of Theorem 4. By the same token, $u(t, x)$ is the payoff for the strategy $\delta(t, \omega) = d^*(t, x_t(\omega))$. Therefore $u(t, x) \leq U(t, x)$.

d) We take an arbitrary $d \in D$ and the function

$$u^d(t, x) = \mathbf{E}_{(t,x)}^d \left\{ \int_t^T f(s, x_s^d, d) ds + \varphi(x_T) \right\}.$$

Then $\mathcal{L}^d u^d(t, x) + f(t, x, d) = 0$ (by the lemma) and $\mathcal{L}^d u(t, x) + f(t, x, d) \leq 0$ (by (24)), whence $\mathcal{L}^d(u(t, x) - u^d(t, x)) \leq 0$. Hence for each fixed $d \in D$ the function $u(t, x) - u^d(t, x)$ is excessive for the process x_t^d . Therefore, by Theorem 4 of [21], $u(t, x) \geq U(t, x)$.

Thus we have proved the following result.

Theorem 5. 1) *The cost $U(t, x)$ is the (unique) solution of the Bellman equation (24).*

2) *The strategy $\delta^*(t, \omega) = d^*(t, x_t)$, where $d^*(t, x)$ is the function constructed in part 3b above, is optimal.*

BIBLIOGRAPHY

1. H. P. McKean, Jr., *Stochastic integrals*, Probability and Math. Statist., no. 5, Academic Press, New York, 1969. MR 40 #947.
2. L. Schwartz, *Analyse mathématique*. I, Hermann, Paris, 1967. MR 37 #2558a.
3. I. M. Gel'fand, *Lectures on linear algebra*, 4th ed., "Nauka", Moscow, 1971; English transl. of 1st ed., Interscience Tracts in Pure and Appl. Math., no. 9, Interscience, New York, 1961. MR 23 #A152.
4. S. L. Sobolev, *Applications of functional analysis in mathematical physics*, Izdat. Leningrad. Gos. Univ., Leningrad, 1950; English transl., Transl. Math. Monographs, vol. 7, Amer. Math. Soc., Providence, R. I., 1963. MR 14, 565; 29 #2624.
5. O. A. Ladyženskaja, V. A. Solonnikov and N. N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, "Nauka", Moscow, 1967; English transl., Transl. Math. Monographs, vol. 23, Amer. Math. Soc., Providence, R. I., 1968. MR 39 #3159a,b.
6. D. V. Stroock and S. R. S. Varadhan, *Diffusion processes with continuous coefficients*. I, II, Comm. Pure Appl. Math. 22 (1969), 345–400, 479–530. MR 40 #6641; #8130.
7. F. John, *On quasi-isometric mappings*. I, Comm. Pure Appl. Math. 21 (1968), 77–110. MR 36 #5716.
8. N. V. Krylov, *A certain estimate from the theory of stochastic integrals*, Teor. Verojatnost. i. Primenen. 16 (1971), 446–457 = Theor. Probability Appl. 16 (1971), 438–448. MR 45 #7841.
9. ———, *On the stochastic integral of $It\hat{o}$* , Teor. Verojatnost. i. Primenen. 14 (1969), 340–348 = Theor. Probability Appl. 14 (1969), 330–336. MR 42 #5350.
10. K. Itô, *On stochastic differential equations*, Mem. Amer. Math. Soc. No. 4 (1951). MR 12, 724.
11. A. V. Skorohod, *On the existence and uniqueness of solutions of stochastic differential equations*, Sibirsk. Mat. Ž. 2 (1961), 129–137; English transl., Selected Transl. Math. Statist. and Probability, vol. 5, Amer. Math. Soc., Providence, R. I., 1965, pp. 191–200. MR 24 #A2435.
12. T. Yamada and S. Watanabe, *On the uniqueness of solutions of stochastic differential equations*. I, II, J. Math. Kyoto Univ. 11 (1971), 155–167, 553–563. MR 43 #4150; 44 #6071.
13. I. I. Gihman and A. V. Skorohod, *Stochastic differential equations*, "Naukova Dumka", Kiev, 1968; English transl., Ergebnisse der Math. und ihrer Grenzgebiete, Band 72, Springer-Verlag, Berlin and New York, 1972. MR 41 #7777.
14. M. D. Ivanovič, *On the nature of continuity of solutions of linear parabolic equations of the second order*, Vestnik Moskov. Univ. Ser. I Mat. Meh. 21 (1966), no. 4, 31–41. (Russian) MR 34 #4672.
15. S. N. Kružkov, *Estimates for the highest derivatives of solutions of elliptic and parabolic equations with continuous coefficients*, Mat. Zametki 2 (1967), 549–560 = Math. Notes 2 (1967), 824–830. MR 36 #4123.
16. R. Z. Has'minskiĭ, *Stability of systems of differential equations under random perturbations of their parameters*, "Nauka", Moscow, 1969; English transl., Transl. Math. Monographs, Amer. Math. Soc., Providence, R. I. (to appear). MR 41 #3925.
17. V. E. Beneš, *Existence of optimal stochastic control laws*, SIAM J. Control 9 (1971), 446–472. MR 45 #9771.
18. W. H. Fleming and M. Nisio, *On the existence of optimal stochastic controls*, J. Math. Mech. 15 (1966), 777–794. MR 33 #7170.
19. W. M. Wonham, *On the separation theorem of stochastic control*, SIAM J. Control 6 (1968), 312–326. MR 38 #5509.

20. M. A. Naimark, *Normed rings*, 2nd rev. ed., "Nauka", Moscow, 1968; English transl, *Normed algebras*, Wolters-Noordhoff, Groningen, 1972.
21. A. K. Zvonkin, *On sequentially controlled Markov processes*, *Mat. Sb.* 86 (228) (1971), 611–621 = *Math. USSR Sb.* 15 (1971), 607–618. MR 45 #3113.

Translated by J. S. JOEL