"ayv_sdes_autumn2021_manual lectures tmp".tex

Stochastic differential equations

A.Yu. Veretennikov*

24 декабря 2021 г.

Содержание

*The course for mexmat MSU 508 & 509 groups, autumn 2021

Аннотация

Note *exercises* and *examples* in the manual; some of them may be asked in the exam. Some additional exercises will be collected in a separate file[1]. By the blue colour additional phrases, examples, sections, etc. are shown which are new in comparison to the earlier versions of this file. If the colour is magenta, it means corrections in comparison to the earlier versions. This colour may be used just to highlight some thought.

1 Lecture 1, 03.09.2021

These lecture notes are provisional and will be gradually updated, corrected, etc.

1.0 Introduction

The ultimate goal of this mini-course it to introduce you to the theory of stochastic differential equations (SDEs) of the form

$$dX_t = b(X_t)dt + dW_t, \quad X_0 = x \in \mathbb{R}, \tag{1}$$

and more general, where W_t is a **Wiener process**[2]. Consider, as a first example, the case of $b \in C_b$ ($b \in C_b^1$). Denote

$$Y_t := X_t - W_t.$$

Then it turns out that the equation on Y_t may be solved on each elementary event (aka outcome) ω as an ordinary differential equation. This will be commented on the whiteboard in the lecture. However, for more general SDEs this simple consideration does not work and this is why you will be listening these lectures during the first part of this Summer school on 17-21.08.2021. In particular, in the lectures it will be shown that the equation (1) has a unique "strong" solution (the term will be explained in what follows) for, at least, any bounded Borel measurable function b. Equations like (1) are important not only as such, i.e., as "randomly perturbed ODEs", but also in their connections to PDEs.

[1] Mostly without solutions.

[2] Introduced shortly in the sequel; at the moment you should know that it is a continuous but not differentiable random process.

1.1 A bit of measure theory

The state space is $\mathbb{R}^d$, $d \geq 1$, but for the first reading think that $d = 1$. All functions are Borel measurable. A short reminder about what the reader/participant of the course should remember about the measure theory and integration can be found in the Problems file. It is important that you read it.

Theorem 1. *Any sigma-finite measure on $\mathbb{R}^d$ is* regular*, that is, its value at any Borel measurable set A can be approximated by its values on open (O) and compact (C) approximations $C \subset A \subset O$.*

Theorem 2. *Any sigma-finite measure on $\mathbb{R}^d$ is uniquely determined by its integration with only continuous bounded functions*[3].

1.2 Wiener process

Definition 1. *A standard (one-dimensional) Wiener process (WP) is a stochastic process $(W_t,\, t \geq 0)$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following properties:*

1. $W_0 = 0$*;* W_t *is continuous for each* $\omega \in \Omega$.
2. $W_t - W_s \sim \mathcal{N}(0, |t-s|), \quad \forall\, t, s \geq 0.$
3. *For any n and any set of disjoint intervals $(s_1, t_1), \dots, (s_n, t_n)$ with $s_i \leq t_i$ the differences (increments) $(W_{t_i} - W_{s_i})$ are independent.*

Existence of a WP may be derived in different ways. One is to use Kolmogorov's theorem about distributions *(compulsory to revise!);* one more will be suggested briefly in the lecture.

In the sequel a little extension of this definition will be convenient.

Definition 2. *The process $(W_t,\, t \geq 0)$ on some probability space $(\Omega, (\mathcal{F}_t), \mathcal{F}, \mathbb{P})$* **with a filtration** *$(\mathcal{F}_t)$ is $(\mathcal{F}_t)$-Wiener iff it is a WP itself, and for each t the random variable W_t is $\mathcal{F}_t$-adapted, and for each $s > t$ the increment $W_s - W_t$ is $\mathcal{F}_t$-independent.*

[3] Most likely the class already knows both results; it is just for refreshing; if not, note that the first theorem is needed here just as a step for establishing the second one; in turn, the second one will be used in the proof of the Markov property of SDE solutions.

The notion of filtration means just a family of nested sigma - algebras (i.e., $\mathcal{F}_t \subset \mathcal{F}_s$, $\forall t < s$); all of them are $\mathbb{P}$-completed, and usually right-continuous, that is, $\mathcal{F}_{t+} := \bigcap_{\epsilon>0}(\mathcal{F}_{t+\epsilon}) = \mathcal{F}_t$.

Definition 3. *A d-dimensional Wiener process is a vector-valued process $W_t = (W_t^1, \ldots, W_t^d)$ where all components are independents standard WP.*

For the first reading always assume $d = 1$.

1.3 WP's quadratic variation QV

Definition 4. *The quadratic variation of a WP is the function*

$$t \mapsto (P) \lim_{\lambda=\max \Delta t_i \to 0} \sum_{t_i \le t} (W_{t_i} - W_{t_{i-1}})^2 =: QV(W)_t, \quad t > 0,$$

if this limit exists.

Existence a.s. follows from two properties of the prelimiting sums:

$$\mathbb{E} \sum_{t_i \le t} (W_{t_i} - W_{t_{i-1}})^2 = t,$$

and

$$\mathrm{Var}\left(\sum_{t_i \le t} (W_{t_i} - W_{t_{i-1}})^2\right) = \sum_{t_i \le t} \mathrm{Var}(W_{t_i} - W_{t_{i-1}})^2 \le 3 \sum (\Delta t_i)^2 \to 0, \quad \lambda \to 0.$$

Both properties are suggested as *exercises.* The existence of the QV(W) is the basis for the construction of a stochastic integral with respect to a WP. Note that for any C^1 function f, its QV(f)=0.

This result has some important interpretation formulated below as a lemma.

Lemma 1.

$$(W_{t+\delta} - W_t)^2 \approx \delta, \quad \delta \to 0,$$

in the following strict sense:

$$\mathbb{E} \left|(W_{t+\delta} - W_t)^2 - \delta\right|^2 = o(\delta), \quad \delta \to 0.$$

Indeed, repeating the calculus above, we can say $(W_{t+\delta} - W_t)^2 - \delta$ is a r.v. with a mean value zero and a variance

$$\mathbb{E}\left|(W_{t+\delta} - W_t)^2 - \delta\right|^2 = 3\delta^2 - 2\delta \times \delta + \delta^2 = 2\delta^2 = o(\delta).$$

In some old papers a notation was accepted (nowadays it is not in use) $dW_t = \sqrt{dt}$, with the convention that this $\sqrt{dt}$ may have a positive or negative sign, assigned randomly. Intuitively, it can be expressed by the approximate formula

$$(W_t - W_s)^2 \approx |t - s|, \quad \text{or} \quad (dW_t)^2 = dt.$$

1.4 Martingales: definitions

Definition 5. *The $(\mathcal{F}_t)$-adapted stochastic process (X_t) is called a martingale with respect to the given filtration iff for each $t < s$*

$$X_t \stackrel{a.s.}{=} \mathbb{E}(X_s|\mathcal{F}_t);$$

it is called a sub-martingale with respect to the given filtration iff for each $t < s$

$$X_t \stackrel{a.s.}{\leq} \mathbb{E}(X_s|\mathcal{F}_t);$$

it is called a super-martingale with respect to the given filtration iff for each $t < s$

$$X_t \stackrel{a.s.}{\geq} \mathbb{E}(X_s|\mathcal{F}_t).$$

Example 1. *A Wiener process $(W_t, \mathcal{F}_t)$ is an $(\mathcal{F}_t)$-martingale.*

Example 2. *Let X be a random variable with $\mathrm{E}|X| < \infty$, and let $(\mathcal{F}_n)$ be any filtration (a set of increasing sigma-fields). Let*

$$X_n := \mathrm{E}(X|\mathcal{F}_n).$$

Show that $(X_n, \mathcal{F}_n)$ is a martingale.

Definition 6. *The random variable τ is called* stopping time *with respect to the filtration $(\mathcal{F}_t)$ iff the event $(\omega : \tau \leq t)$ is $\mathcal{F}_t$-measurable for any t.*

Exercise 1. *Show that in discrete time τ is a stopping time iff*[4] *$(\tau = n) \in \mathcal{F}_n$ for each n.*
Let τ be a stopping time. Show that $\tau \wedge t, \tau \vee t, \tau + 1$ are also stopping times for each $t \geq 0$. Think of an example where $\tau - 1$ is not a stopping time.

[4] "Iff" = "if and only if".

1.5 Stochastic integral

The first construction of the Ito stochastic integral (SI) is based on approximations via step functions and *Ito's isometry.*

1. Of course, we start with simple ("ladder") random functions,

$$f_t = \sum_j f_{t_j}(\omega)(t_j \le t < t_{j+1}),$$

with a finitely many nonrandom values t_j & $\mathrm{E} f_{t_i}^2 < \infty$, & f_{t_i} is $\mathcal{F}_{t_i}$-measurable, on each finite interval, in which case the SI (stochastic integral) of f with respect to W_t is defined by the formula

$$I_T(f) := \sum_j f_{t_j}(W_{t_{j+1}\wedge T} - W_{t_j \wedge T}) =: \int_0^T f_s dW_s,$$

(where T is any fixed nonrandom value, including $T = \infty$; in the latter case we require $\mathrm{E}\int_0^\infty f_s^2 ds < \infty$) and

$$I_t(f) := \sum_j f_{t_j}(W_{t_{j+1}\wedge t} - W_{t_j \wedge t}).$$

In other words, any ladder function is constant on some nonrandom interval, and the value of this function is a random variable, the same on any such interval. **We will assume that for each value of t, the r.v. f_t is $\mathcal{F}_t$- adapted (measurable),** and will call it an adapted ladder function. In the sequel we will also assume

$$\mathbb{E} f_t^2 < \infty, \quad \forall\, t \ge 0;$$

the set of all such ladder functions satisfying the latter condition is denoted by $H_0[0, T]$.

2. Let us establish three *main properties* of an SI for (adapted) ladder functions from $H_0[0, T]$.

(a)

$$\mathbb{E} \int_0^t f_s \, dW_s = 0.$$

(b)

$$\mathbb{E}(\int_0^t f_s \, dW_s)^2 = \int_0^t \mathrm{E}\, f_s^2 \, ds.$$

(c)
$$\int_0^t f_s\, dW_s \qquad \text{is a martingale (short notation \textbf{``mart''}).}$$

(d) (additional)
For any $f \in H_0[0,T]$ the SI $I_t(f)$ is a.s. **continuous** in t.

(e) (additional)
SI for processes from $H_0[0,T]$ are linear: for any constants a, b
$$I_T(af^1 + bf_2) = aI_T(af^1) + bI_T(f_2).$$

The proofs are very simple and are all left to the independent study as **exercises**.

The second property is called **Ito's isometry**[5]. It plays – along with the quadratic variation of the WP – the main role in the general definition of an SI "driven by the WP".

Definition 7. *Let $(W_t, \mathcal{F}_t)$ be a WP, and let the process (f_t) satisfy the following conditions:*
$$f_t \in \mathcal{F}_t$$
(this will be the short notation for the property that f_t is $\mathcal{F}_t$-measurable), and
$$\int_0^T \mathbb{E} f_s^2\, ds < \infty.$$
Such a class of processes will be denoted by $H_2[0,T]$.

Note that $T = +\infty$ is allowed.

Lemma 2 (J.L.Doob)**.** *Let $f \in H_2[0,T]$. Then there exists a sequence of ladder functions $(f_t^n) \in H_2[0,T]$ such that*
$$\int_0^T \mathbb{E}|f_t^n - f_t|^2 dt \to 0. \tag{2}$$
Moreover, f^n may be chosen by the following algorithm (yet, with unknown $a \in [0,1]$ and with the unknown subsequence $(n_k), k = 1, 2, \ldots$):
$$f_t^k = f_{\kappa_{n_k}(t+a)-a}, \quad \textit{with a.e. } a \in [0,1]. \tag{3}$$
(Here $\kappa_n(t) = 2^{-n}[2^n t]$.)

[5] Later you will also see its multiidimensional version.

For the proof[6] of this lemma see [Krylov, Theorem 2.8.2 (English edition)], or [Krylov, Theorem I.10.2 or II.6.8 (Russian edition)].

Proof. **1.** There is an integration theory lemma: for any (nonrandom) function $g \in L_2[0,T]$

$$\int_0^T |g(s+v) - g(s)|^2 ds \to 0, \quad v \to 0.$$

(Outside $[0,T]$ we assume that $f = 0$.) Prove it yourself!

2. Now consider the limit (we are to show that it equals zero)

$$\mathrm{E} \underbrace{\int_0^T |f_{s+v} - f_s|^2 ds}_{} \stackrel{?}{\to} 0, \quad v \to 0. \tag{4}$$

The claim (that the integral tends to zero) is correct and follows from the Lebesgue dominated convergence theorem due to the domination bound

$$\int_0^T |f_{s+v} - f_s|^2 ds \le 2\int_0^T |f_{s+v}|^2 ds + 2\int_0^T |f_s|^2 ds \le 4\int_0^T |f_s|^2 ds.$$

and because of the integration lemma mentioned in the step 1 above. Indeed, the "underbraced"integral under the sign of expectation above is a random variable (depending also on v), which tends to zero as $v \to 0$, and does not exceed the value $4\int_0^T |f_s|^2 ds$ which has a finite expectation.

3. We conclude that

$$\int_0^1 da \, \underbrace{\mathrm{E} \int_0^T |f_{\kappa_n(s+a)-a} - f_s|^2 ds}_{=:h_n(a)} \to 0, \quad n \to \infty.$$

Indeed, denoting $\kappa_n(s+a) - a = s + v$ with $v = \kappa_n(s+a) - a - s$ we easily see that since $\kappa_n(s+a) \to s+a$, then $v \to 0$. Moreover, $|v| = |\kappa_n(s+a) - (s+a)| \le 2^{-n} \to 0$ uniformly with respect to a. Therefore, the "underbraced"expression here tends to zero uniformly by the step 2, and as a consequence, the integral over a also goes to zero.

[6] It is shown below with some little changes.

4. So we have $\int_0^1 h_n(a)\,da \to 0$ and $0 \le h_n(a) \le 4\mathrm{E}\int_0^T |f_s|^2 ds$ which last expression is a finite nonrandom value. By the well-known integration lemma (the version of Borel – Cantelli lemma in probability[7]) there exists a subsequence $n_k \to \infty$ such that

$$h_{n_k}(a) \to 0, \quad k \to \infty, \quad \text{for almost every } a \in [0,1]$$

In other words,

$$\mathrm{E}\int_0^T |f_{\kappa_{n_k}(s+a)-a} - f_s|^2 ds \to 0, \quad k \to \infty, \quad \text{for almost every } a \in [0,1],$$

as required. The lemma is proved.

Definition 8. *If (2) holds, then the stochastic integral (SI, aka Ito's integral) $I_T(f)$ may be defined as a square limit of $I_T(f^n) = \int_0^T f_s^n dW_s$, that is,*

$$\int_0^T f_s\,dW_s := l.i.m._{n\to\infty} \int_0^T f_s^n\,dW_s.$$

Remark 1. *For continuous $f \in H_2[0,T]$ it is true that*

$$\int_0^T f_s\,dW_s := l.i.m._{\lambda\to 0} \sum_i f_{s_i}\left(W_{s_{i+1}\wedge T} - W_{s_i\wedge T}\right),$$

where $\lambda = \max_i(s_{i+1}-s_i)$. More than that, we can define the integral Ito from 0 to t for any $0 \le t \le T$ as follows,

$$\int_0^t f_s\,dW_s := l.i.m._{\lambda\to 0} \sum_i f_{s_i}\left(W_{s_{i+1}\wedge t} - W_{s_i\wedge t}\right),$$

however, the problem whether this definition may correctly define $I_t(f)$ for all $t \ge 0$ at the moment remains open (until the section 1.7), because when we use approximations, we apply SI for all ω at once up to almost surely, and this almost surely, in principle, may depend on t.

Of course, it should be also verified that this square limit (both in the definition and in the remark) exists and that it is defined correctly, i.e., that it does not depend on the choice of the approximating functions (f^n), or of the sequence of partitions (s_i); it can be regarded as an additional *exercise*. Note that we have f_{s_i}, not $f_{s_{i+1}}$ here and this is rather important.

[7] It suffices to choose n_k so that $\sum_k \int_0^1 h_{n_k}(a)da < \infty$, for example, so that $\int_0^1 h_{n_k}(a)da < 1/k^2$.

1.6 Martingale inequalities

Firstly, two major results will be shown for continuous square integrable martingales $M_t := \int_0^t f_s dW_s$ with $f \in H_0[0,T]$ (the notation for simple functions). They use analogous *discrete time* martingale inequalities; so, we start with discrete time.

Definition 9. *The random variable τ is called a stopping time for the filtration $\mathcal{F} = (\mathcal{F}_t)$ iff the set $(\tau \le n)$ is $\mathcal{F}_n$-measurable for each n.*

Lemma 3. *For a martingale M_n, $n = 0, 1, \ldots$ with $\mathbb{E}M_n^2 < \infty$, the process M_n^2 is a sub-martingale. More generally, for any convex function g such that $\mathbb{E}g(M_n) < \infty$, the process $g(M_n)$ is a sub-martingale.*

Proof follows from the properties of convex functions and conditional expectations: there exists a (countable) set of pairs $a_i, b_i \in \mathbb{R}$ such that

$$g(x) = \sup_i (a_i x + b_i), \quad \& \quad g(x) \ge (a_i x + b_i)\ \forall i.$$

(It is useful to draw a picture.) Then for any $n < m$ we have

$$\mathbb{E}(g(M_m)|\mathcal{F}_n) \ge \sup_i \mathbb{E}(a_i M_m + b_i)|\mathcal{F}_n) = \sup_i (a_i \mathbb{E}(M_m|\mathcal{F}_n) + b_i)$$

$$= \sup_i (a_i M_n + b_i) = g(M_n) \quad (a.s.)$$

Remark 2. *If g is not only convex, but also increasing (non-strictly), then the claim of the lemma remains valid for non-negative submartingales. Note in this respect that the modulus of any square-integrable stochastic integral is a non-negative submartingale (namely, because it is a convex function of a square-integrable martingale).*

Theorem 3 (J.L.Doob)**.** *Let M_n, $n = 0, 1, \ldots$ be a sub-martingale wrt a filtration $(\mathcal{F}_n)$, and let τ be a stopping time. Then the process $M_{n\wedge\tau}$, $n = 0, 1, \ldots$ is a sub-martingale with respect to the same filtration.*

Hint: For the proof, use the decompositions

$$M_{n\wedge\tau} = M_0 1(\tau = 0) + \ldots + M_n 1(\tau = n) + M_n 1(\tau > n),$$

and

$$1(\tau \le n) M_{n\wedge\tau} = M_0 1(\tau = 0) + \ldots + M_n 1(\tau = n),$$

where $1(A) = 1_A$ is the indicator function of the event or set A, i.e., it is a random variable which takes value 1 for $\omega \in A$ and zero otherwise.

Theorem 4 (Doob's "optional stopping theorem" (without proof)). *Let M_n, $n = 0, 1, \ldots$ be a sub-martingale wrt a filtration $(\mathcal{F}_n)$, and let $\tau_1 \le \tau_2 \le \ldots$ be a sequence of stopping times. Then the process $(M_{\tau_n}, \mathcal{F}_{\tau_n})\, n = 0, 1, \ldots$ is a sub-martingale.*

Theorem 5 (Doob–Kolmogorov). *For a martingale (and sub-martingale) M_n, $n = 0, 1, \ldots$ for any $N > 0$, $c > 0$,*

$$\mathbb{P}(\max_{n\le N} M_n \ge c) \le \frac{1}{c} \mathrm{E} M_N 1(\max_{n\le N} M_n \ge c) \le \frac{1}{c}\mathrm{E}(M_N)_+. \tag{5}$$

Proof. Let $\tau := \inf(n \ge 0 : M_n \ge c)$. Then

$$\mathbb{P}(\max_{n\le N} M_n \ge c) = \mathbb{P}(M_\tau 1(\tau \le N) \ge c)$$

By the Bienaymé – Chebyshev – Markov's (BCM for what follows) inequality,

$$\mathbb{P}(M_\tau 1(\tau \le N) \ge c) \le \frac{1}{c}\mathrm{E} M_\tau 1(\tau \le N).$$

Further,

$$\mathrm{E} M_\tau 1(\tau \le N) = \mathrm{E}\left(M_0 1(\tau = 0) + \ldots + M_n 1(\tau = N)\right) 1(\tau \le N).$$

By the martingale definition, for any $k \le n$

$$\mathrm{E}\left(M_k 1(\tau = k)\right) 1(\tau \le N) = \mathrm{E}\left(M_N 1(\tau = k)\right) 1(\tau \le N),$$

since $(\tau = k) \cap (\tau \le N) \in \mathcal{F}_N$. So,

$$\mathrm{E}\left(M_0 1(\tau = 0) + \ldots + M_n 1(\tau = n)\right) 1(\tau \le n)$$

$$= \mathrm{E}\left(M_N 1(\tau = 0) + \ldots + M_N 1(\tau = n)\right) 1(\tau \le n) = \mathrm{E} M_N 1(\tau \le n).$$

Hence, we get

$$\mathbb{P}(\max_{n\le N} M_n \ge c) \le \frac{1}{c}\mathrm{E} M_N 1(\tau \le N).$$

Then, we easily obtain

$$\frac{1}{c}\mathrm{E} M_N 1(\tau \le N) \le \frac{1}{c}\mathrm{E}(M_N)_+ 1(\max_{n\le N} M_n \ge c) \le \frac{1}{c}\mathrm{E}(M_N)_+,$$

as required. The theorem is proved.

Theorem 6 (Doob)**.** *For a martingale (and sub-martingale) M_n, $n = 0, 1, \ldots$ with $\mathbb{E}M_n^2 < \infty$ for each n, for any $N > 0$,*

$$\mathbb{E} \max_{n \le N} (M_n)^2 \le 4M_N^2. \tag{6}$$

For the proof, see, e.g., [Krylov (English edition), theorem 3.4.15].

Proof. Firstly, let us notice that in (6) both parts are finite by the assumption. Secondly,

$$\mathrm{E} \max_{n \le N} (M_n)^2 = 2 \int_0^\infty c \, \mathrm{E} 1(\max_{n \le N} (M_n) \ge c) dc$$

This equation holds true even without E, by integration over c in the r.h.s.: for any finite $X \ge 0$

$$X^2 = 2 \int_0^\infty c \, 1(X \ge c) \, dc.$$

Then we get

$$\mathbb{E} \max_{n \le N} (M_n)^2 = 2 \int_0^\infty c \, \mathrm{E} 1(\max_{n \le N} (M_n) \ge c) dc \stackrel{(Thm\, 5)}{\le} 2\mathrm{E} \int_0^\infty c \, \frac{1}{c} M_N 1(\max_{n \le N} M_n \ge c) \, dc$$

$$= 2\mathrm{E} M_N \max_{n \le N} M_n \stackrel{Cauchy}{\le} 2(\mathrm{E} M_N^2)^{1/2} (\mathrm{E} \max_{n \le N} M_n^2)^{1/2}.$$

In what follows Cauchy – Buniakowskii – Schwarz inequality will be denoted as CBS. Here $\mathrm{E} \max_{n \le N} M_n^2 < \infty$. In the case of $\mathrm{E} \max_{n \le N} M_n^2 = 0$, we automatically have (6), and, in fact, the r.h.s. also equals zero. If $\mathrm{E} \max_{n \le N} M_n^2 > 0$, it follows that

$$(\mathbb{E} \max_{n \le N} (M_n)^2)^{1/2} \le 2(\mathrm{E} M_N^2)^{1/2},$$

which implies (6) again after squaring both sides here. The theorem is proved.

Doob – Kolmogorov's inequality for SI: for any $c > 0$

$$\mathbb{P}(\sup_{0 \le t \le T} \int_0^t f_s dW_s \ge c)$$

$$\le \frac{1}{c} \mathrm{E}(\int_0^T f_s dW_s) 1(\sup_{0 \le t \le T} \int_0^t f_s dW_s \ge c) \le \frac{1}{c} \mathrm{E}(\int_0^T f_s dW_s)^+.$$

The idea of the proof: The SI is a continuous martingale. Its supremum over any interval can be regarded as a supremum over just all rational points from this interval. In turn, the latter may be approximated by the maximum over the first n values from the (any) numerated sequence of these rational numbers, which maximum satisfies the discrete version of Doob–Kolmogorov's inequality.

Doob's inequality for SI (partial case $p = 2$)

$$\mathrm{E} \sup_{0\le t\le T} (\int_0^t f_s dW_s)^2 \le 4 \int_0^T \mathrm{E} f_s^2 ds.$$

Note that for continuous martingales sup *can be evaluated over just a countable set of times (e.g., rational). This allows to extend martingale inequalities form discrete time to continuous.*

Ito's inequality ([Krylov, Theorem 2.7.8] (Russian edition)) For[8] any $c, N > 0$ and $f \in H_0$

$$\mathbb{P}(\sup_{0\le t\le T} | \int_0^t f_s dW_s| \ge c)$$

$$\le P(\int_0^T f_s^2 ds \ge N) + \frac{1}{c^2} E(N \wedge \int_0^T f_s^2 ds).$$

1.7 Applications to stochastic integrals

Doob–Kolmogorov's inequality (or, Doob's) plays a crucial role in the construction of SI because it allows to show that there is modification defined for each t, and this modification is continuous in t a.s., as desired.

Let $f \in H_2[0, T]$, and let (f^n) be some approximating sequence from $H_0[0, T]$ satisfying

$$\|f^n - f\|_{H_2}^2 < 2^{-n}.$$

[8] For SI over martingales known in French literature as *Lenglart's inequality*.

Then by Doob-Kolmogorov's inequality we estimate for $m > n$

$$\mathbb{P}(\sup_{0\le t\le T} |\int_0^t f_s^n dW_s - \int_0^t f_s^m dW_s| \ge \frac{1}{n^2}) \le 4n^4 \mathrm{E}|\int_0^T f_s^n dW_s - \int_0^T f_s dW_s|^2$$

$$= 4n^4 \mathrm{E}\int_0^T |f_s^n - f_s^m\,(\mp f_s)|^2 ds \le 8n^4 \mathrm{E}\int_0^T (|f_s^n - f_s|^2 + |f_s^m - f_s|^2) ds \le \frac{8n^4}{2^n}.$$

Therefore, the sum of probabilities

$$\sum_{n=1}^{\infty} \mathbb{P}(\sup_{0\le t\le T} |\int_0^t f_s^n dW_s - \int_0^t f_s^{n+1} dW_s| \ge \frac{1}{n^2}) < \infty$$

converges. Hence, by Borel – Cantelli lemma for $\mathbb{P}$-almost all ω (aka with probability one)

$$\sup_{0\le t\le T} |\int_0^t f_s^n dW_s - \int_0^t f_s^{n+1} dW_s| < \frac{1}{n^2},$$

starting with some finite $n_0 = n_0(\omega)$. Thus, the sequence of *continuous* functions $I_t(f^n) := \int_0^t f_s^n dW_s$ is a Cauchy sequence in $C[0,T]$ with probability one. So, we can choose a subsequence (n') over which $I_t(f^{n'})$ converges in $C[0,T]$, and this limit is continuous in t. This limit is by definition a stochastic integral (SI) $I_T(f) = \int_0^T f_s dW_s$.

On the other hand, *for each* t the sequence of SI $(I_t(f^n) = \int_0^T 1(s \le t) f_s^n dW_s)$ converges in $H_2[0,T]$ to the SI $(I_t(f) = \int_0^T 1(s \le t) f_s dW_s)$. We conclude that this limit $I_t(f)$ is a.s. continuous.

NB: Alternatively, Doob's inequality in a combination with Bienaymé – Chebyshev – Markov's inequality may be applied for the same goal.

Exercise 2. For your independent consideration.*: Give a full proof of the latter statement: once we know that the sequence (or subsequence) $(I_t(f^n)$ converges in $H_2[0,T]$ and at the same time it converges for almost all ω in $C[0,T]$, then it is almost surely the same limit.*

Exercise 3. *Show that all five properties of SI – three main ones and two additional – remain valid for $I_t(f)$ with $f \in H_2$. (Hint: Use CBS inequality.)*

1.8 Stochastic integrals under the condition $\mathbb{P}(\int_0^T f_s^2 ds < \infty) = 1$

Under this condition, which is, clearly, much weaker than $f \in H_2[0,T]$, the SI can be also defined using *stopping times.* Let us introduce the sequence of stopping times

$$\tau_n := \inf(t \geq 0 : \int_0^t f_s^2 ds \geq n).$$

Then

$$f_s 1(s < \tau_n) \in H_2[0,T],$$

so, the SI $I_T(f_s 1(s < \tau_n)) = \int_0^T f_s 1(s < \tau_n) dW_s = \int_0^{T\wedge\tau_n} f_s dW_s$ is well-defined. The condition $\mathbb{P}(\int_0^T f_s^2 ds < \infty) = 1$ implies that $\tau_n \uparrow \infty$, $n \to \infty$ a.s. Hence, the value $T \wedge \tau_n$ stabilises at T for n large enough. This allows us to define

$$\int_0^T f_s dW_s := \lim_{n\to\infty} \int_0^{T\wedge\tau_n} f_s dW_s,$$

and we know that the limit exists a.s. The main properties of the SI in this case may be lost. However, it is a very useful extension.

Exercise[9]*: Verify the correctness of this definition in the sense that if some other "localising sequence of stopping times" is chosen, then the limit will be a.s. the same. It is left as a non-compulsory exercise.*

1.9 Ito's formulae

Definition 10. *The process X_t possesses a stochastic differential*

$$dX_t = b_t dt + \sigma_t dW_t$$

with $b, \sigma \in H_2[0,T]$, iff for each t

$$\mathbb{P}\left(X_t - X_0 = \int_0^t b_s ds + \int_0^t \sigma_s dW_s\right) = 1.$$

Here is the idea of Ito's calculus.

[9]For your independent consideration.

Rule 1 (Ito's formal calculus)**.** *While multiplying formally two stochastic differentials, use the rules:*

$$(dW_t)^2 = dt,\ dt\, dW_t = 0, (dt)^2 = 0.$$

NB *(repetition): In some old papers on SDEs (for example, by A.V.Skorokhod, and others) instead of dW_t the expression $\sqrt{dt}$ was used, with the convention that the sign of this $\sqrt{dt}$ may be random, plus or minus.* It corresponds to the equality $\boxed{(dW_t)^2 = dt}$.

Exercise 4. *Show "by hands"(i.e., without Ito's formula) that*

$$\int_0^t W_s dW_s = \frac{W_t^2}{2} - \frac{t}{2}.$$

Hint: use Riemann sums

$$\sum W_{t_i}(W_{t_{i+1}} - W_{t_i})$$

and compare them to the quadratic variation

$$l.i.m. \sum (W_{t_{i+1}} - W_{t_i})(W_{t_{i+1}} - W_{t_i});$$

for evaluating $\sum W_{t_{i+1}}(W_{t_{i+1}} - W_{t_i})$ use Abel's transformation[10] *(aka discrete "summation by parts")*

$$\sum_{n=1}^{N} a_n(b_n - b_{n-1}) = a_N b_N - a_0 b_0 - \sum_{n=1}^{N} b_{n-1}(a_n - a_{n-1}).$$

Exercise 5. *Let f_s, $s \geq 0$ be a non-random function $f \in C^1$. Show that the SI $I_t(f)$ in this case may be defined by the "integration by parts" formula*[11]

$$\int_0^t f_s dW_s = f_t W_t - \int_0^t W_s df_s. \tag{7}$$

Hint: Firstly, note that this formula is valid for any $f \in H_0$, in which case the integral $\int_0^t W_s df_s$ is understood as a sum $\sum_{s \leq t} W_s(f_s - f_{s-})$, and (7) can be regarded as Abel's transformation (aka "summation by parts"); then use approximations and Ito's isometry.

[10] It is a good time to revise it.

[11] Well before Ito, this was suggested by N. Wiener.

Exercise 6. *If f_s, $s \geq 0$ be a non-random function $f \in C^1$ and the SI $I_t(f)$ is defined for it by the formula (7), show for this integral the three main properties of SI: expectation zero, Ito's isometry, and the martingale property.*

Theorem 7 (Ito's formula for product)**.** *Let*

$$dX_t = b_t dt + \sigma_t dW_t,$$

and

$$dY_t = c_t dt + e_t dW_t.$$

Then

$$d(X_t Y_t) = b_t Y_t dt + c_t X_t dt + (\sigma_t e_t) dt + e_t X_t dW_t + \sigma_t Y_t dW_t$$

$$= (b_t Y_t + c_t X_t + \sigma_t e_t)\, dt + (e_t X_t + \sigma_t Y_t)\, dW_t$$

The proof may be performed first for integrands from H_0, reduced just to single summands of the form aW_t and bt, which result is then extended to the general case by approximations and by Ito's isometry.

Proof (sketch for simple (ladder) piecewise-constant functions). Consider just one small interval where all coefficients b, σ, c, e are (random) constants. In this case

$$X_t = x + bt + \sigma W_t, \quad Y_t = y + ct + eW_t,$$

and we are to show that the following two expressions are equal:

$$I_1 = (x + bt + \sigma W_t)(y + ct + eW_t)$$

and

$$I_2 = xy + \int_0^t (bY_s + cX_s + \sigma e)\, ds + \int_0^t (eX_s + \sigma Y_s)\, dW_s.$$

For I_1 we rewrite as

$$\begin{aligned} I_1 &= \sigma e W_t^2 + ((x + bt)e + (y + ct)\sigma)\, W_t + (x + bt)(y + ct). \\ &= xy + \sigma e W_t^2 + (be + c\sigma)tW_t + (xe + y\sigma)W_t + (cx + by)t + bct^2. \end{aligned} \tag{8}$$

For I_2 we have for Ito's stochastic integral part,

$$\int_0^t (eX_s + \sigma Y_s)\, dW_s = \int_0^t (ex + ebs + e\sigma W_s + \sigma y + \sigma cs + \sigma e W_s)\, dW_s$$

$$= \int_0^t (ex + ebs + \sigma y + \sigma cs + 2e\sigma W_s)\, dW_s$$

$$= (ex + \sigma y)W_t + (eb + \sigma c)\int_0^t s dW_s + 2e\sigma \underbrace{\int_0^t W_s dW_s}_{=W_t^2/2 - t/2}$$

$$= (ex + \sigma y)W_t + (eb + \sigma c)\int_0^t s dW_s + e\sigma W_t^2 - e\sigma t,$$

and for the Lebesgue's integral part

$$\int_0^t (bY_s + cX_s + \sigma e)\, ds = \sigma e t + byt + cxt + \int_0^t b(cs + eW_s)ds + \int_0^t c(bs + \sigma W_s)ds$$

$$= (\sigma e + by + cx)t + bct^2 + (c\sigma + be)\int_0^t W_s ds.$$

Note that according to the exercise 5, the random variable (r.v. in the sequel) tW_t possesses the following representation:

$$tW_t = \int_0^t s dW_s + \int_0^t W_s ds. \tag{9}$$

Combining all terms after the rearrangements, we get

$$I_2 = xy + (ex+\sigma y)W_t + (eb+\sigma c)\int_0^t s dW_s + e\sigma W_t^2 - e\sigma t$$

$$+(\sigma e + by + cx)t + bct^2 + (c\sigma + be)\int_0^t W_s ds$$

$$= xy + (ex+\sigma y)W_t + (by+cx)t + bct^2 + e\sigma W_t^2 + (c\sigma+be)\left(\int_0^t W_s ds + \int_0^t s dW_s\right)$$

$$\overset{(9)}{=} xy + (ex+\sigma y)W_t + (by+cx)t + bct^2 + e\sigma W_t^2 + (c\sigma + be)tW_t.$$

By inspection, these are exactly the terms in the expression for I_1 in (8). Hence, $I_1 = I_2$, as required.

NB: The case of several partitions may be, in principle, considered in a similar way. Then the general case follows by approximations of the integrands by simple functions from H_0. See [Krylov, Lemma 2.10.2 (Russian edition)].

Theorem 8 (Ito's formula ($d=1$))**.** *Let*

$$dX_t = b_t dt + \sigma_t dW_t,$$

$g(x) \in C_b^2(\mathbb{R}), f(t,x) \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R})$. *Then*

$$\begin{aligned} dg(X_t) &= g'(X_t)dX_t + \frac{1}{2}g''(X_t)(dX_t)^2 \\ &= \left(\frac{1}{2}g''(X_t)\sigma_t^2 + g'(X_t)b_t\right)dt + g'(X_t)\sigma_t dW_t, \end{aligned} \tag{10}$$

and[12]

$$df(t, X_t) = f_t(t, X_t)dt + f_x(t, X_t)dX_t + \frac{1}{2}f_{xx}(X_t)(dX_t)^2$$

$$= \left(f_t(t, X_t) + \frac{\sigma_t^2}{2}f_{xx}(t, X_t) + b_t f_x(t, X_t)\right)dt + f_x(t, X_t)\sigma_t dW_t.$$

NB: Ito's formulae hold for stochastic integrals under the relaxed condition $\mathbb{P}(\int_0^T f_s^2 ds < \infty) = 1$, *too.*

In the case of $X_t = W_t$ Ito's formulae simplify a little. The intuitive sketch of the proof in this case may be performed in the following way. We say that

$$dg(W_t) = g'(W_t)dW_t + (1/2)g''(W_t)dt,$$

loosely speaking, if

$$\Delta g(W_t) \approx g'(W_t)\Delta W_t + (1/2)g''(W_t)\Delta t.$$

Hence, let us consider $\Delta g(W_t)$. By Taylor's expansion,

$$\Delta g(W_t) \equiv g(W_{t+\delta}) - g(W_t)$$

$$\approx g'(W_t)\Delta W_t + (1/2)g''(W_t)(\Delta W_t)^2 + o((\Delta W_t)^2).$$

Hence,

$$g(W_t) - g(W_0) \approx \sum_i g'(W_{s_i})\Delta W_{s_i} + (1/2)\sum_i g''(W_{s_i})(\Delta W_{s_i})^2 + o((\Delta W_{s_i})^2).$$

Since

$$(\Delta W_{s_i})^2 \approx \Delta s_i,$$

(see the Rule 1), then we get the desired assertion concerning function g. Indeed,

$$g(W_t) - g(W_0) = \sum_i g'(W_{s_i})\Delta W_{s_i} + (1/2)\sum_i g''(W_{s_i})(\Delta W_{s_i})^2$$

$$\approx \sum_i g'(W_{s_i})\Delta W_{s_i} + (1/2)\sum_i g''(W_{s_i})(\Delta s_i).$$

[12] The standard PDE notation will be used frequently, $f_x = \partial f/\partial x$, $f_t = \partial f/\partial t$, etc.

Here the second integral sum is an approximation to the Riemann integral $\int_0^t (1/2)g''(W_s)\,ds$ while the first one approximates the stochastic integral $\int_0^t g'(W_s)\,dW_s$.

Exercise 7. *Show using Ito's formula for product that if Ito's formula (10) in dimension 1 is valid for $g(x) = x^n$, then it is also valid for $g(x) = x^{n+1}$.*

Ito's formula for the function $f(t,x)$ and for W_t can be commented similarly. An original rigorous proof may be based on Ito's formula for products.

Both versions of Ito's formula can be justified for many unbounded functions such as x^2, $\exp(x)$, tx^3, $\exp(-t+x)$, etc.[13] *However, remember that for some unbounded functions the integral form of Ito's formula may require some localisation, i.e., not the form with $\int_0^t \ldots$*

Example 3. *1. By the Itô formula,*

$$d(W_t^2/2) = W_t\,dW_t + (1/2)\,(dW_t)^2 = W_t\,dW_t + (1/2)\,dt.$$

2. So,

$$d(W_t^2/2 - t/2) = W_t\,dW_t + (1/2)\,dt - (1/2)\,dt = W_t\,dW_t.$$

3. Hence[14]*,*

$$\int_0^t W_s\,dW_s = W_t^2/2 - t/2.$$

Ito's formula multidimensional

Rule 2 (Ito's formal calculus, $d > 1$). *While multiplying formally two stochastic differentials, use the rules:*

$$dW_t^i dW_t^j = \delta_{ij}dt,\ dt\,dW_t^i = 0, (dt)^2 = 0.$$

Here δ_{ij} is a Kronecker symbol.

[13] We will use this result for such functions without special notice.

[14] In fact, we already know this result from earlier exercises.

Definition 11. *The* **multidimensional stochastic integral**

$$\int_0^T \sigma_t dW_t = I_T(\sigma)$$

for a d-dimensional WP $W_t = (W_t^1, \ldots, W_t^d)$ and a matrix-valued process $\sigma \in H_2[0,T]$ componentwise is defined as

$$(I_T(\sigma))^j = \sum_{k=1}^d \int_0^T \sigma_t^{jk} dW_t^k.$$

All main properties of this SI are preserved (also componentwise); **Ito's isometry** reads in the multidimensional case as

$$\mathrm{E}\left(I_T(\sigma)I_T(\sigma)^*\right) \equiv \mathrm{E}\left(\int_0^T \sigma_t dW_t\right)\left(\int_0^T \sigma_t dW_t\right)^* = \int_0^T \mathrm{E}\sigma_t\sigma_t^* dt \tag{11}$$

In particular, for two independent WP W^1 and W^2,

$$\mathrm{E}\int_0^T f_t dW_t^1 \int_0^T g_t dW_t^2 = 0. \tag{12}$$

Exercise 8. *Justify the property (11).*
Hint: Firstly do it for ladder inregrands from H_0 with one step each, etc.

Exercise 9. *Justify the multidimensional property (12).*
Hint: For the jk-element of the martix in the left side here write

$$\mathrm{E}\sum_{\ell,m}\int_0^T \sigma_t^{j\ell} dW_t^\ell \int_0^T \sigma_t^{km} dW_t^m = \sum_{\ell,m}\mathrm{E}\int_0^T \sigma_t^{j\ell} dW_t^\ell \int_0^T \sigma_t^{km} dW_t^m$$

and now use the result of the Exercise 8.

Theorem 9 (Ito's formula for WP ($d \geq 1$))**.** *Let function $g : \mathbb{R}^d \to \mathbb{R}$ be of the class C^2. Then*

$$dg(W_t) = \nabla g(W_t) dW_t + \frac{1}{2}\Delta g(W_t) dt,$$

where $\nabla g(W_t) dW_t = \sum_i \frac{\partial g(W_t)}{\partial x^i} dW_t^i$ (a scalar product), $\Delta g(W_t) = \sum_i \frac{\partial^2 g(W_t)}{(\partial x^i)^2}$. Also,

$$df(t, W_t) = f_t(t, W_t) dt + \nabla_x f(t, W_t) dW_t + \frac{1}{2}\Delta f(t, W_t) dt, \tag{13}$$

for any $f(t,x) \in C^{1,2}$.

NB: The Laplace operator here is a consequence of the Ito calculus rule $dW_t^i dW_t^j = \delta_{ij} dt$. *A more general Ito formula for a generic d-dimensional Ito process (the process with a stochastic differential) can be written similarly; it looks awful, but you will easily get it right if you use Ito calculus formal rules. The important message is that it will involve some second order differential operator, elliptic or parabolic.*

Remark 3. *It is important to remember that any equality for stochastic differentials may be rewritten by definition in the integral form, e.g., for (13)* ($\nabla_x = \frac{\partial}{\partial x}, \Delta = \frac{\partial^2}{\partial x^2}$ *for* $d = 1$*):*

$$f(t, W_t) - f(0,0) = \int_0^t \left(f_s(s, W_s) + \frac{1}{2}\Delta f(s, W_s) \right) ds + \int_0^t \nabla_x f(s, W_s) dW_s. \quad (14)$$

As a corollary, if the function $f(s, x)$ *satisfies the* **heat equation** $f_s(s, x) + \frac{1}{2}\Delta f(s, x) = 0$ *for each* s, x, *then*

$$f(t, W_t) - f(0,0) = \int_0^t \nabla_x f(s, W_s) dW_s, \quad (15)$$

and taking expectations, we (hope to (explain why?!)) get

$$f(0,0) = \mathrm{E} f(t, W_t). \quad (16)$$

Similarly, we may obtain the probability representation for $f(0, x)$, $\forall\, x$*:*

$$f(0, x) = \mathrm{E} f(t, x + W_t). \quad (17)$$

NB: *This is not "just"a one more formula, but a* probabilistic representation of the solution of heat equation. *More general PDEs also possess similar representations based on more general SDEs instead of a Wiener process. This will be one of the important themes in this course.*

Multidimensional Ito's formula, general coordinate form

Let X be a d-dimensional Ito process (i.e., possessing a stochastic differential),

$$dX_t = b_t dt + \sigma_t dW_t.$$

Theorem 10 (general Ito formula in coordinates and in the vector form)**.** *Let* $g(x) = (g^1(x), \dots g^\ell(x)) \in C^2$, *and* $f(t,x) = (f^1(t,x), \dots f^\ell(t,x)) \in C^{1,2}$. *Then (summation over the repeated indices, notation* $\boxed{a(x) = \sigma\sigma^*(x)}$*)*

$$dg^i(X_t) = \frac{\partial}{\partial x^j} g^i(X_t)\sigma_t^{jk} dW_t^k + \left(\frac{1}{2}\frac{\partial^2}{\partial x^j \partial x^k} g^i(X_t) a_t^{jk} + \frac{\partial}{\partial x^j} g^i(X_t) b_t^j\right) dt,$$

or[15]*,*

$$dg(X_t) = \nabla_x g(X_t)\sigma_t dW_t + \left(\frac{1}{2} Tr(a_t g_{xx})(X_t) + \nabla_x g(X_t) b_t\right) dt;$$

here $\nabla_x g\, b = \langle \nabla_x g, b\rangle$ *(the inner product); the expression* $\nabla_x g(X_t)\sigma_t dW_t$ *is also understood in a similar way, as product of three matrices of corresponding dimensions*[16]*; the Ito "correction term" reads*

$$(g_{x^k x^n}\sigma^{km}\sigma^{nm}) = (g_{x^k x^n} a^{kn}) \overset{check!}{=} Tr(a g_{xx});$$

for the function f *we have,*

$$df^i(t, X_t) = \frac{\partial}{\partial x^j} f^i, (t X_t)\sigma^{jk}(t, X_t) dW_t^k$$

$$+ \left(\frac{\partial}{\partial t} f_t^i(t, X_t) + \frac{1}{2}\frac{\partial^2}{\partial x^j \partial x^k} f^i(t, X_t) a_t^{jk} + \frac{\partial}{\partial x^j} f^i(t, X_t) b_t^j\right) dt,$$

or[17]*,*

$$df(t, X_t) = \nabla_x f(t, X_t)\sigma_t dW_t + \left(f_t(t, X_t) + \frac{1}{2} Tr(a_t f_{xx})(t, X_t) + \nabla_x f(t, X_t) b_t\right) dt.$$

Список литературы

[1] Н.В.Крылов, Введение в теорию случайных процессов, части 1, 2, МГУ, 1986-1987; N.V. Krylov, Introduction to the Theory of Random Processes, AMS, 2002.

[2] А.В.Булинский, А.Н.Ширяев, Теория случайных процессов. Физматлит, 2005.

[3] MAGIC: Stochastic Processes, 2011/12, http://maths-magic.ac.uk/course.php?id=205

[4] А.Д.Вентцель, Курс теории случайных процессов, М., Наука, 1975 (2е изд. 1996).

[15] For the first reading assume that $\ell = 1$
[16] In fact, the dimension d_1 of W may differ from d, then σ should be a non-square matrix $d \times d_1$
[17] For the first reading assume that $\ell = 1$

2 Lecture 2, 10.09.2021

2.1 Stochastic differential equations

Recall that the main object of the course is, actually, a stochastic differential equation (SDE) in $\mathbb{R}^d$ (in some cases in $\mathbb{R}^1$ only)

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \ t \geq 0, \qquad X_0 = x_0, \tag{18}$$

or, equivalently in the integral form,

$$X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{19}$$

Here $(W_t, \mathcal{F}_t)$ is a standard d-dimensional Wiener process, b and σ are vector and matrix Borel functions of corresponding dimensions d and $d \times d$. However, if you see this setting for the first time, consider $d = 1$. The initial value x_0 may be non-random, or random but $\mathcal{F}_0$-measurable. In this lecture/section we will study Ito's theorems on existence and uniqueness, and in addition the continuity property of solutions with respect to the initial conditions (i.c.), which is (one of) the key step(s) towards the important **Markov property**.

2.2 Strong solutions: Kiyoshy Ito's theorems

So, we will discuss solutions of Ito SDEs

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \ t \geq 0, \quad X_0 = x_0. \tag{20}$$

Definition 12. *Solution $X_t, t \geq 0$ of the equation (20) is called* **strong** *iff for each t the random variable X_t is $\mathcal{F}_t^W$-measurable. In all other cases solution is called* **weak**[18]. *When we are talking of a solution without specifying strong or weak, usually we mean weak solution.*

Theorem 11 (Uniqueness by Ito)**.** *Assume that there exists $C > 0$ such that for any t, x, x',*

$$|b(t,x) - b(t,x')| + \|\sigma(t,x) - \sigma(t,x')\| \leq C|x - x'|,$$
$$\text{and}$$
$$|b(t,x)| + \|\sigma(t,x)\| \leq C(1 + |x|).$$

[18]In case if we do not know whether or not solution is strong, it is also called weak; eventually, it may happen that it is strong.

Then[19] *there is no more than one solution of the equation (180) on any probability space with any Wiener process.*

Proof. Suppose there are two solutions X_t and Y_t *on the same probability space and with the same WP*. We have,

$$|X_t - Y_t|^2 \le 2|\int_0^t (b(s,X_s) - b(s,Y_s))ds|^2 \\ +2|\int_0^t (\sigma(s,X_s) - \sigma(s,Y_s))dW_s|^2.$$

(We used that $(a+b)^2 \le 2a^2 + 2b^2$.) Hence,

$$\begin{aligned} \mathrm{E}|X_t - Y_t|^2 &\le 2\mathrm{E}|\int_0^t |b(s,X_s) - b(s,Y_s)|ds|^2 \\ +2\mathrm{E}\int_0^t Tr(\sigma(s,X_s) - \sigma(s,Y_s))(\sigma(s,X_s) - \sigma(s,Y_s))^* ds \\ &\le C\mathrm{E}|\int_0^t |X_s - Y_s|ds|^2 \\ +C\mathrm{E}\int_0^t Tr(\sigma(s,X_s) - \sigma(s,Y_s))(\sigma(s,X_s) - \sigma(s,Y_s))^* ds \\ &\le (C + Ct)\int_0^t \mathrm{E}|X_s - Y_s|^2 ds. \end{aligned}$$

Since $\sup_{s\le t}\mathrm{E}|X_s - Y_s|^2 \le 2(\sup_{s\le t}\mathrm{E}(|X_s|^2 + |Y_s|^2)) < \infty$, now Gronwall's (Grönwall's) inequality

$$0 \le \varphi(t) \le C_1 + C_2\int_0^t \varphi(s)ds \;\forall\, t \quad \mapsto \quad \varphi(t) \le C_1\exp(C_2 t)\;\forall\, t.$$

with $C_1 = \varphi(0) = \mathrm{E}|X_0 - Y_0|^2$ implies $\mathrm{E}|X_t - Y_t|^2 = 0$, and $\mathbb{P}(X_t = Y_t, \forall t \ge 0) = 1$, as required.

Theorem 12 (Existence by Ito)**.** *Assume that there exists* $c > 0$ *such that for any* t, x, x',

$$|b(t,x) - b(t,x')| + \|\sigma(t,x) - \sigma(t,x')\| \le C|x - x'|,$$

[19]For the first reading assume b and σ bounded along with Lipschitz condition with respect to x.

and

$$|b(t,x)| + \|\sigma(t,x)\| \le C(1+|x|).$$

Then there exists a (strong) solution of the equation (180).

Proof. The proof will use successive approximations. Let $X_t^0 := x_0$, $t \ge 0$. Given $X_t^n, t \ge 0$, let

$$X_t^{n+1} := x_0 + \int_0^t b(s, X_s^n)ds + \int_0^t \sigma(s, X_s^n)dW_s.$$

Let us consider the difference $X_t^{n+1} - X_t^n$. We have, similarly to the calculus in the uniqueness theorem, Again using $(a+b)^2 \le 2a^2 + 2b^2$, we get

$$|X_t^{n+1} - X_t^n|^2 \le 2|\int_0^t (b(s, X_s^n) - b(s, X_s^{n-1}))ds|^2$$

$$+2|\int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1}))dW_s|^2.$$

So, with some $C_t = C + Ct$,

$$\mathrm{E}|X_t^{n+1} - X_t^n|^2 \le 2\mathrm{E}|\int_0^t (b(s, X_s^n) - b(s, X_s^{n-1}))ds|^2$$

$$+2\mathrm{E}|\int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1}))dW_s|^2$$

$$\le C_t \mathrm{E}\int_0^t |X_s^n - X_s^{n-1}|^2 ds.$$

Moreover, by using Doob's inequality for stochastic integrals we also have (with different constants $C_T > 0$ on different lines, but not depending on n)

$$\mathrm{E}\sup_{t\le T}|X_t^{n+1}-X_t^n|^2 \le 2\mathrm{E}\sup_{t\le T}|\int_0^t (b(s,X_s^n)-b(s,X_s^{n-1}))ds|^2$$

$$+2\mathrm{E}\sup_{t\le T}|\int_0^t (\sigma(s,X_s^n)-\sigma(s,X_s^{n-1}))dW_s|^2$$

$$\le C_T \mathrm{E}\int_0^T |X_s^n - X_s^{n-1}|^2 ds \stackrel{(<\infty!)}{\le} C_T\int_0^T \mathrm{E}\sup_{t\le s}|X_t^n - X_t^{n-1}|^2 ds.$$

Denoting $a_T^n := \mathrm{E}\sup_{t\le T}|X_t^{n+1}-X_t^n|^2$, by induction we get,

$$0 \le a_T^n \le \frac{C_T^n T^n a_T^0}{n!}.$$

Clearly, the series $\sum_n a_T^n$ converges. Let us estimate the probability

$$\mathbb{P}(\sup_{t\le T}|X_t^{n+1}-X_t^n| \ge 2^{-n}) \le 4^n a_T^n \le \frac{(4C_T T)^n a_T^0}{n!}.$$

So, by the Borel – Cantelli lemma, with probability one for all n starting with some $n_0(\omega)$,

$$\sup_{t\le T}|X_t^{n+1}-X_t^n| < 2^{-n}.$$

This implies that the following object is well-defined in $L_2(\Omega\times[0,T], P\times\Lambda)$ and simultaneously in $C([0,T]$ a.s. in ω:

$$X_t := X_t^0 + \sum_{n\ge 0}(X_t^{n+1}-X_t^n).$$

Moreover, there is a uniform in $t\in[0,T]$ convergence

$$X_t^{n+1} = X_t^0 + \sum_{k=0}^n (X_t^{k+1}-X_t^k) \rightrightarrows X_t, \quad n\to\infty.$$

We are now able to pass to the limit in the equation

$$X_t^{n+1} := x_0 + \int_0^t b(s,X_s^n)ds + \int_0^t \sigma(s,X_s^n)dW_s,$$

which, of course, results in the limiting version of it,

$$X_t := x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s,$$

as required.

To justify this limiting equation, we firstly note that the process X_t is $(\mathcal{F}_t)$-adapted, along with all X^n. Now, in the right hand side we clearly have $X_t^{n+1} \to X_t$ a.s. (and even uniformly in $t \le T$).

If the drift b is, actually, bounded, then we immediately obtain convergence a.s.

$$\int_0^t b(s, X_s^n)ds \to \int_0^t b(s, X_s)ds, \quad n \to \infty,$$

from the Lebesgue's bounded convergence theorem. Under the more relaxed linear growth condition on b the same result for this Lebesgue integral follows from the a priori bounds (uniform in n, by induction, and for X_t, too)

$$\sup_{t \le T} \mathrm{E}|X_t^n|^2 < \infty.$$

This exercise is left as a homework to the readers.

Finally, similarly, but now with the help of the Ito isometry we obtain convergence

$$\int_0^t \sigma(s, X_s^n)dW_s - \int_0^t \sigma(s, X_s)dW_s \to 0, \quad n \to \infty.$$

straightforwardly in the case of bounded σ. Indeed,

$$\begin{aligned} &\mathrm{E}|\int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s))dW_s|^2 \\ &= \mathrm{E}\int_0^t \|\sigma(s, X_s^n) - \sigma(s, X_s))\|^2 ds \to 0, \quad n \to \infty. \end{aligned}$$

For unbounded σ the same convergence of stochastic terms follows from similar a priori bounds for the fourth moment,

$$\sup_{t \le T} \mathrm{E}|X_t^n|^4 < \infty$$

(also left as a homework). The existence theorem is proved. QED

Let $x \in \mathbb{R}^d$, and let there exist $K > 0$ such that for all x, x' the following *linear growth condition* holds,

$$|b(t,x)|^2 + \|\sigma(t,x)\|^2 \le K(1+|x|^2). \tag{21}$$

Suppose X_t is a solution (weak or strong) of the equation

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \quad X_0 = x,$$

on the whole half-line $t \ge 0$, where the i.c. x is non-random.

Exercise 10 (a priori bound for the 2nd moment)**.** *Under the assumptions above, there exists $C > 0$ such that*

$$\mathrm{E} \sup_{0\le t\le T} |X_t|^2 \le C(1+T+x^2)\exp(CT). \tag{22}$$

(See, e.g., [Gikhman & Skorokhod, SDEs 1968].)

Assume

$$\begin{aligned} |b(t,x) - b(t,x')| + \|\sigma(t,x) - \sigma(t,x')\| \le K|x-x'|, \\ |b(t,x)| + \|\sigma(t,x)\| \le K(1+|x|). \end{aligned} \tag{23}$$

Let with random $\xi_n, \xi_0 \in \mathcal{F}_0$

$$X_t = \xi_0 + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s,$$

$$X_t^n = \xi_n + \int_0^t b(s,X_s^n)ds + \int_0^t \sigma(s,X_s^n)dW_s,$$

Theorem 13 (version of K.6.9.3: let $\mathrm{E}|\xi_0|^2 + \sup_n \mathrm{E}|\xi_n|^2 < \infty$)**.** *Assume also (23) and* $\mathrm{E}|\xi_n - \xi_0|^2 \to 0$*. Then* $\forall T > 0$, $\forall c > 0$,

$$\mathbb{P}(\sup_{0\le t\le T} |X_t^n - X_t|^2 > c) \stackrel{\exists C}{\le} Cc^{-1}\mathrm{E}|\xi_n - \xi_0|^2 \to 0, \; n \to \infty. \tag{24}$$

Proof. We subtract,

$$\begin{aligned} X_t - X_t^n = \xi_0 - \xi_n &+ \int_0^t (b(s,X_s) - b(s,X_s^n))ds \\ &+ \int_0^t (\sigma(s,X_s) - \sigma(s,X_s^n))dW_s. \end{aligned}$$

Using Doob's inequality and a priori 2nd moment bounds (22) both for X^n and X, we get

$$\mathrm{E}\sup_{0\le s\le t}|X_s - X_s^n|^2 \le 3t\mathrm{E}\int_0^t |b(s,X_s)-b(s,X_s^n)|^2 ds$$
$$+3\mathrm{E}|\xi_0-\xi_n|^2 + 12\,E\int_0^t \|\sigma(s,X_s)-\sigma(s,X_s^n)\|^2 ds.$$

Hence, from the Lipschitz condition it follows,

$$\mathrm{E}\sup_{0\le s\le t}|X_s - X_s^n|^2 \le 3\mathrm{E}|\xi_0-\xi_n|^2 + C(1+t)\mathrm{E}\int_0^t |X_s - X_s^n|^2 ds,$$

and now due to Gronwall's inequality we obtain $\forall\, t>0$,

$$\mathrm{E}\sup_{0\le s\le t}|X_s - X_s^n|^2 \le 3\mathrm{E}|\xi_0-\xi_n|^2 \exp(C(1+t)t).$$

Bienaymé–Chebyshev–Markov's inequality now implies (24).

Let us state *without proof* the exact version of the Theorem in Krylov's textbook.

Theorem 14 (K.6.9.3)**.** *Assume (23) and let*

$$\xi_n - \xi_0 \to 0, \;\; n\to\infty, \quad a.s.$$

Then $\forall T>0$, $\forall c>0$,

$$\mathbb{P}(\sup_{0\le t\le T}|X_t^n - X_t|^2 > c) \to 0, \; n\to\infty. \tag{25}$$

2.3 Markov property for SDE solutions

Definition 13. *The process $X_t, t\ge 0$ is called Markov iff for any finite sequence of non-random times $0\le t_1 < \ldots < t_{n+1}$ the equality holds a.s. $\mathrm{E}(g(X_{t_{n+1}})|X_{t_n},\ldots X_{t_1}) = \mathrm{E}(g(X_{t_{n+1}})|X_{t_n})$, $\forall g\in B$. The process adapted to the filtration $(\mathcal{F}_t)$ is called Markov* with respect to this filtration *iff for any $s\le t$, the equality holds a.s.* $\mathrm{E}(g(X_t)|\mathcal{F}_s) = \mathrm{E}(g(X_t)|X_s)$.

Definition 14. *The process X is strong Markov iff for any a.s. finite stopping times $\tau_1\le\tau_2$, $\mathrm{E}(g(X_{\tau_2})|\mathcal{F}^X_{\tau_1}) = \mathrm{E}(g(X_{\tau_2})|X_{\tau_1})$, and strong Markov with respect to $(\mathcal{F}_t)$ iff almost surely $\mathrm{E}(g(X_{\tau_2})|\mathcal{F}_{\tau_1}) = \mathrm{E}(g(X_{\tau_2})|X_{\tau_1})$ for any Borel bounded g.*

For any $t_n \geq 0$ (there will be $0 \leq t_1 \leq \ldots \leq t_n$) an SDE may be considered

$$\tilde{X}_t = \xi_{t_n} + \int_{t_n}^{t} b(s, \tilde{X}_s)ds + \int_{t_n}^{t} \sigma(s, \tilde{X}_s)dW_s, \quad t \geq t_n, \tag{26}$$

where the r.v. ξ_{t_n} is $\mathcal{F}^W_{t_n}$-adapted. Now, "clearly"solution $\tilde{X}$ after time t_n is determined by ξ_{t_n} and by the increments of W after time t_0, which increments do not depend on $\tilde{X}_{t_n}$. Hence, if we have a solution X of the equation (180) from 0 to t_n, and define $\xi_{t_n} = X_{t_n}$ (X being the unique solution of (180)), then this (strong) solution does not depend on the past trajectory X before t_n *given* X_{t_n}. Due to the uniqueness, solutions $\tilde{X}_t$ and X_t after t_n coincide a.s.

Let us fix y and consider the SDE (181) with a non-random i.c.

$$Y_t = y + \int_{t_n}^{t} b(s, Y_s)ds + \int_{t_n}^{t} \sigma(s, Y_s)dW_s, \quad t \geq t_n. \tag{27}$$

The theory of such SDEs under our conditions is similar to the one in the case $t_n = 0$. The solution is denoted by Y^y_t.

Lemma 4 (K.6.11.2)**.** *Y^y_t is measurable wrt $\mathcal{F}^{\Delta W}_{[t_n,t]} \equiv \sigma(W_s - W_{t_n}, t_n \leq s \leq t)$.*

Follows from the change of time $t - t_n = t'$.

Lemma 5 (K.6.11.3)**.** *The σ-fields $\mathcal{F}^{\Delta W}_{[t_n,t]}$ and $\mathcal{F}^W_{t_n}$ are independent.*

For the Proof of the Lemma [K.6.11.3] it suffices to show the equality

$$\mathbb{P}(B; W_{t_n+s_1} - W_{t_n} \in \Gamma_1, \ldots W_{t_n+s_k} - W_{t_n+s_{k-1}} \in \Gamma_1)$$

$$= \mathbb{P}(B)\mathbb{P}(W_{t_n+s_1} - W_{t_n} \in \Gamma_1, \ldots W_{t_n+s_k} - W_{t_n+s_{k-1}} \in \Gamma_1)$$

for any $B \in \mathcal{F}^W_{t_n}$, $\forall k$, $\forall \Gamma_i \in \mathcal{B}(R^d)$, and $\forall\, 0 \leq s_1 < \ldots < s_k$. This equality is a consequence of the definition of WP, that is, of the independence of the increments

$$W_{t_n+s_1} - W_{t_n}, \ldots W_{t_n+s_k} - W_{t_n+s_{k-1}}$$

of the sigma-algebra $\mathcal{F}^W_{t_n}$. The Lemma follows.

Corollary 1 (K.6.11.4)**.** *For any $t \geq t_n$ and any $x \in R^d$, the r.v. Y^y_t and sigma - algebra $\mathcal{F}^W_{t_n}$ are independent.*

Follows from the previous two Lemmata.

Lemma 6 (K.6.11.5)**.** *Let* $\xi^m = 2^{-m}[2^m X_{t_n}]$*, where* X_t *is a solution of the equation (180). Then*

$$Y_t^{\xi^m} \xrightarrow{\mathbb{P}} X_t^{t_n, X_{t_n}} = X_t, \quad m \to \infty, \;\; \forall\, t \geq t_n.$$

Proof. On the set $(\omega \in \Omega : \xi^m = y)$ we have $Y_t^{\xi^m} = Y_t^y$ (by definition of $Y_t^{\xi^m}$), where Y_t^y solves (31) with the i.c. y. The union of all (no more than countably many) such events is the whole Ω. This means that the process $Y_t^{\xi^m}$ satisfies the equation (31) **with** y **replaced by** ξ^m:

$$Y_t^{\xi^m} = \xi^m + \int_{t_n}^t b(s, Y_s^{\xi^m})ds + \int_{t_n}^t \sigma(s, Y_s^{\xi^m})dW_s, \quad t \geq t_n.$$

Also, X_t for $t \geq t_n$ satisfies (31) **with** y **replaced by** X_{t_n}:

$$X_t = X_{t_n} + \int_{t_n}^t b(s, X_s)ds + \int_{t_n}^t \sigma(s, X_s)dW_s, \quad t \geq t_n.$$

Indeed,

$$X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad t \geq t_n.$$

and

$$X_{t_n} = x_0 + \int_0^{t_n} b(s, X_s)ds + \int_0^{t_n} \sigma(s, X_s)dW_s.$$

Subtract.

This follows from the equation on X_t if we subtract $X_t - X_{t_n}$. In other words, X_t on $t \geq t_n$ serves as a solution of the equation for Y_t^{ξ}:

$$Y_t^{\xi} = \xi + \int_{t_n}^t b(s, Y_s^{\xi})ds + \int_{t_n}^t \sigma(s, Y_s^{\xi})dW_s, \quad t \geq t_n,$$

with $\xi \in \mathcal{F}_{t_n}^W$. Compare these two equations: if $\xi = X_{t_n}$ then they coincide. However, the solution of this equation is unique (by Ito's theorem for the equations on $t \geq t_n$). Hence, we have,

$$X_t^{t_n, X_{t_n}} = X_t, \quad t \geq t_n.$$

They both coincide with $Y_t^{X_{t_n}}$ by definition of Y_t^{ξ}. The claim of the lemma follows from the continuity theorem as $\xi^m \rightrightarrows X_{t_n}$.

Theorem 15 (K.6.11.6)**.** *The solution of equation (180) is a Markov process.*

Proof. Let us take $t \ge t_n$ and any bounded *continuous* function $f(x) \ge 0$. We will now show

$$\mathrm{E}(f(X_t)|\mathcal{F}^W_{t_n}) = \mathrm{E}(f(X_t)|X_{t_n}), \quad \text{a.s.}$$

This will not be the "full Markov property but some restricted version of it, just for $f \in C_b$; the full Markov property would be this equality a.s. for any Borel bounded function f. We just start with a continuous f, which will be later extended to all Borel bounded ones. Denote $\Phi(y) := \mathrm{E}f(Y^y_t)$. This function is *continuous.* Let us now take any $B \in \mathcal{F}^W_{t_n}$. By Corollary K.6.11.4 and Lemma K.6.11.5, we have a sequence of equalities

$$\begin{aligned}
\mathrm{E}1(B)f(X_t) &\overset{Le\,K.6.11.5}{=} \lim_{m\to\infty} \mathrm{E}1(B)f(Y^{\xi^m}_t) \\
&= \lim_{m\to\infty} \sum_{r=-\infty}^{\infty} \mathrm{E}1(B)f(Y^{r/2^m}_t)1(\xi^m = r/2^m) \\
&\overset{Cor\,K.6.11.4}{=} \lim_{m\to\infty} \sum_{r=-\infty}^{\infty} \mathrm{E}1(B,\xi^m = r/2^m)E(f(Y^{r/2^m}_t)|\mathcal{F}^W_{t_n}) \\
&= \lim_{m\to\infty} \sum_{r=-\infty}^{\infty} \mathrm{E}1(B)1(\xi^m = r/2^m)\Phi(r/2^m) \\
&= \lim_{m\to\infty} \mathrm{E}1(B)\Phi(\xi^m) = \mathrm{E}1(B)\Phi(X_{t_n}).
\end{aligned}$$

$$\mathrm{E}1(B)f(X_t) = \mathrm{E}1(B)\Phi(X_{t_n}).$$

By definition of conditional expectations this implies the equality

$$\mathrm{E}(f(X_t)|\mathcal{F}^W_{t_n}) = \Phi(X_{t_n}),$$

which, in turn, implies

$$\mathrm{E}(f(X_t)|X_{t_n}) = \mathrm{E}(\mathrm{E}(f(X_t)|\mathcal{F}^W_{t_n})|X_{t_n}) = \Phi(X_{t_n}),$$

and

$$\mathrm{E}(f(X_t)|X_{t_n}, \dots X_{t_1}) = \mathrm{E}(\mathrm{E}(f(X_t)|\mathcal{F}^W_{t_n})|X_{t_n}, \dots X_{t_1}) = \Phi(X_{t_n}).$$

So, $\forall f \in C_b$

$$\mathrm{E}(f(X_t)|X_{t_n}) = \mathrm{E}(f(X_t)|X_{t_n}, \dots X_{t_1}) \quad (a.s.) \tag{28}$$

It remains to extend this "restricted Markov property"on all $f \in B(R^d)$ Let us fix any $B \in \mathcal{F}^W_{t_n}$ and consider two measures,

$$\mu(\Gamma) = \mathrm{E}1(B)\mathrm{E}(1(X_t \in \Gamma)|X_{t_n}, ...X_{t_1}), \quad \nu(\Gamma) = \mathrm{E}1(B)\mathrm{E}(1(X_t \in \Gamma)|X_{t_n}).$$

We will integrate bounded function f over each measure μ and ν. For any step function f we have,

$$\int f(x)\mu(dx) == \mathrm{E}1(B)\mathrm{E}(f(X_t)|X_{t_n},...X_{t_1}), \qquad \int f(x)\nu(dx) = \mathrm{E}1(B)\mathrm{E}(f(X_t)|X_{t_n}).$$

So, it is also true for any $f \in C_b$, as such f can be *uniformly* approximated by a sequence of step functions. Thus, (28) implies

$$\int f(x)\mu(dx) = \int f(x)\nu(dx), \ \forall f \in C_b.$$

So, since any regular measure is uniquely determined by integrals with all bounded continuous functions, we obtain

$$\mu \overset{K.1.2.4}{=} \nu \Longrightarrow \mathrm{E}(f(X_t)|X_{t_n}) = \mathrm{E}(f(X_t)|X_{t_n}, \ldots X_{t_1}), \ \forall f \in B.$$

The theorem is proved.

2.4 Strong Markov property

The ordinary Markov property tells us that the process (possessing this property) starts at any time t from its new position as if it was a new zero time and new nonrandom state. Here t is *non-random*. This poses a natural question: is it true that, likewise, any Markov process starts from a new position at any stopping time τ? The answer is positive for Markov processes in discrete time, but negative in general.

Definition 15. *The process X is called strong Markov iff for any a.s. finite stopping times $\tau_1 \le \tau_2$, a.s.* $\mathrm{E}(g(X_{\tau_2})|\mathcal{F}^X_{\tau_1}) = \mathrm{E}(g(X_{\tau_2})|X_{\tau_1})$, *and strong Markov with respect to $(\mathcal{F}_t)$ iff almost surely* $\mathrm{E}(g(X_{\tau_2})|\mathcal{F}_{\tau_1}) = \mathrm{E}(g(X_{\tau_2})|X_{\tau_1})$ *for any Borel bounded g.*

One of the key notions to establish strong Markov property is *Feller property* of a Markov process.

Definition 16. *The Markov process X is called Feller iff for any $f \in C_b$ the function* $\mathrm{E}_x f(X_t)$ *is continuous wrt x for any $t \ge 0$*

Under our standard assumptions of Lipschitz and linear growth, the solution of SDE is Feller (follows from the continuity theorem). The following theorem closes the issue.

Theorem 16 (without proof, see (A.D. Wentzel, §9.2))**.** *Any Markov process X with trajectories continuous from the right which is Feller, is also strong Markov with respect to the filtration $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$.*

In other words, $E(X_{\tau_2}|\mathcal{F}_{\tau_1+}) = E(X_{\tau_2}|X_{\tau_1})$ if $\tau_1 \leq \tau_2$ a.s. Here $\mathcal{F}_{\tau+} = \sigma(A : A \cap (\tau \leq t) \in \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s)$.

Corollary 2. *Any Wiener process is strong Markov.*

Indeed, clearly, the expression $\mathrm{E} f(x + W_t)$ is continuous in the variable t for $f \in C_b(\mathbb{R}^d)$.

2.5 Averaging property of a Wiener process

Strong Markov property is indispensable in studying *ergodic* features of SDE solutions. This is beyond the scopes of this course. Yet, we will just show here one simple example where strong Markov property is also helpful.

Denote $|B(r)|$ = d-volume of the ball $B(r)$, $|\Gamma(r)|$ = $d-1$-area of the $(d-1)$-surface $\Gamma(r) = \partial B(r)$ (the boundary of $B(r)$). Assume that D is a bounded open domain, $x \in D$, and let $B_x(r)$ denote the open ball of radius r with the center at x. Suppose $r > 0$ is small enough, so that $B_x(r) \subset\subset D$ (is a compact subset in D). Denote $u(x) := \mathrm{E}\varphi(x + W_\tau)$ for some $\varphi \in C(\partial D)$ and also $\tau^r := \inf(t \geq 0 : x + W_t \notin B_x(r))$.

Corollary 3. *We have,*

$$u(x) = \mathrm{E}u(x + W_{\tau^r}) = \frac{1}{|\Gamma(r)|} \int_{\Gamma_x(r)} u(y)dy,$$

and also

$$u(x) = \mathrm{E}u(x + W_{\tau^r}) = \frac{1}{|B(r)|} \int_{B_x(r)} u(z)dz. \tag{29}$$

Proof. We show just the first one, because the second one follows by one more integration from it. Due to the strong Markov property, we have, denoting $X_t =$

$x + W_t$,

$$u(x) = \mathrm{E}\varphi(x + W_\tau) = \mathrm{E}(\mathrm{E}\varphi(x + W_\tau)|\mathcal{F}_{\tau^r})$$

$$= \mathrm{E}(\mathrm{E}\varphi(x + W_\tau)|x + W_{\tau^r}) = \mathrm{E}(\mathrm{E}\varphi(X_\tau)|X_{\tau^r})$$

$$= \mathrm{E}u(X_{\tau^r}) = \frac{1}{|\Gamma(r)|}\int_{\Gamma_x(r)} u(y)dy,$$

the last equality *by the symmetry of* W*:* for the WP starting from x, to hit any area on $\Gamma_x(r)$ at stopping time τ^r is proportional to the $d-1$-dimensional volume of this area.

It may be shown that any Borel measurable bounded function satisfying (77) is continuous in x*, and, moreover, is, in fact, in* C^2 *and satisfies the Laplace equation in* D*,*

$$\Delta u(x) = 0, \quad x \in D. \tag{30}$$

Vice versa, by applying Ito's formula to any $u(x + W_t)$ *satisfying (30) with* $u|_\Gamma = \varphi$ *and integrating from* 0 *to* $\tau := \inf(t \geq 0 : x + W_t \notin D)$*, we get the representation* $u(x) = \mathrm{E}\varphi(x + W_\tau)$*. Note that* $\mathrm{E}_x\tau < \infty$*, which may also be rigorously proved.*

Список литературы

[1] I.I. Gikhman, A.V. Skorokhod, Theory of random processes, vol. I – III, Rusian edition, Moscow, Nauka,

[2] I.I. Gikhman, A.V. Skorokhod, Stochastic differential equations, Russian edition, Kiev, Naukova dumka, 1968.

2.6 Another pathwise uniqueness

This subsection presents a rather simplified version of two independent results, by S. Nakao (1972) and by A.K. Zvonkin (1974). The case is $d = 1$.

Assumption. We assume that the functions b, σ, σ^{-1} are bounded, & b, σ' continuous (continuity may be dropped, but this assumption helps to find accurate references).

Consider a 1D SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x. \tag{31}$$

It is important (for the proof) that the coefficients do not depend on time, as in Nakao's paper (not in Zvonkin's one).

Theorem 17 (simplified Nakao / Zvonkin)**.** *Suppose in addition to the Assumption above that $b, \sigma' \in C_b$. Then the equation (31) is pathwise unique.*[20]

Proof. Let $u(x)$ be a (any!) solution of an ODE

$$(Lu(x) =) \qquad \frac{\sigma^2(x)}{2}u''(x) + b(x)u'(x) = 0.$$

Let us find one suitable solution of this equation. Denote $u' = v$, then assuming $v \neq 0$ and, moreover, $v > 0$, we get

$$\frac{\sigma^2(x)}{2}v'(x) + b(x)v(x) = 0 \ \sim \ \frac{v'}{v}(x) \equiv (\ln v)'(x) = -2\tilde{b}(x),$$

where

$$\tilde{b}(x) = b(x)/\sigma^2(x).$$

Hence,

$$\ln v(x) = -2\int_0^x \tilde{b}(y)dy + C, \text{ but we take } C = 0;$$

then,

$$v(x) = \exp(-2\int_0^x \tilde{b}(y)dy) = u'(x).$$

So, one of solutions (which suffices for our aim) has a form

$$u(x) = \int_0^x \exp(-2\int_0^y \tilde{b}(z)dz)dy + C,$$

and again we take $C = 0$ here. Note that $u \in C^2$. Denote now

$$Y_t = u(X_t).$$

[20] As it will be clear in what follows, it implies the existence of strong solution, but not in this lecture.

By Ito's formula,

$$dY_t = du(X_t) = u'(X_t)\sigma(X_t)dW_t = v(X_t)\sigma(X_t)dW_t.$$

Since $u' = v > 0$, then the mapping $x \mapsto u(x)$ is strictly increasing; hence, it is 1-1. Denote by u^{-1} its inverse (also in C^2), so that $X_t = u^{-1}(Y_t)$. Then

$$dY_t = v(u^{-1}(Y_t))\sigma(u^{-1}(Y_t))dW_t.$$

Note that with $x = u^{-1}(y)$ we have,

$$\begin{aligned}\frac{d}{dy}v(u^{-1}(y)) &= \frac{v'(u^{-1}(y))}{u'(u^{-1}(y))} = \frac{v'(x)}{u'(x)}\\ &= \frac{-2\tilde{b}(x)v(x)}{v(x)} = -2\tilde{b}(x);\end{aligned}$$

$$\begin{aligned}v(u^{-1}(y))\frac{d}{dy}\sigma(u^{-1}(y)) &= v(u^{-1}(y))\frac{\sigma'(u^{-1}(y))}{u'(u^{-1}(y))}\\ &= v(x)\frac{\sigma'(x)}{u'(x)} = v(x)\frac{\sigma'(x)}{v(x)} = \sigma'(x).\end{aligned}$$

In other words, the new diffusion coefficient for Y_t is C^1_b; in particular, it is Lipschitz. So, Y_t is a **pathwise unique** solution of the SDE for Y. Equivalently, solution X_t of the initial SDE is also pathwise unique. QED.

Exercise 11. *Continuing this line of arguments, show that X_t is a strong solution not using the* **Yamada – Watanabe principle** *presented in the next lecture 3 (not yet in the file).*

2.7 Stochastic exponents, beginning

Let b_t be an adapted bounded stochastic vector-valued d-dimensional process[21]. Denote

$$\rho_t = \rho_t[b] := \exp(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds), \quad t \le T, \tag{32}$$

where b^2 is understood as a scalar product $b^2 := (b, b)$. The process b_t is assumed from $H_2[0,T]$. By Ito's formula

$$d\rho_t = b_t \rho_t dW_t. \tag{33}$$

[21] Always assume $d = 1$ for the first reading.

This is a linear SDE, where bdW_t is also a scalar product if $d > 1$. In other words, we have the following

Micro-theorem 1. *The process ρ_t is a* **solution of an SDE** with a random diffusion coefficient $b_t x$,

$$dX_t = b_t X_t dW_t, \quad 0 \le t \le T, \quad X_0 = 1,$$

In the integral form we have,

$$\rho_t = 1 + \int_0^t b_s \rho_s dW_s. \tag{34}$$

We see that solutions of linear SDEs (so far, without drift) are stochastic exponentials. *Note the difference with deterministic exponentials solving deterministic linear ODEs.*

Exercise 12. *If b_t is adapted to the Wiener process filtration $(\mathcal{F}_t^W)$, then this solution ρ_t is* **strong**.

Exercise 13. *This solution ρ_t is also* **pathwise unique**.

3 Lecture 3, 17.09.2021

3.1 Stochastic exponents, ctd.

We continue studying stochastic exponents (32)

$$\rho_t = \rho_t[b] := \exp(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds)$$

which solve linear SDEs (33) with initial value $\rho_0 = 1$ at $t = 0$. Their integral form (34) which does not invlove the Lebesgue integral but is a constant (i.c.) plus a stochastic integral, gives us a hope that possibly ρ_t may be a martingale, as it is usual for stochastic integrals under “normal” conditions that the integrand is in H_2. If this is true, we would have, in particular, $\mathrm{E}\rho_t = 1$. In turn, any object which is non-negative and integrates to one can serve as a density. So, in this case ρ_t may serve as a probability density. *What is a problem with the SI $\int_0^t b_s\rho_s dW_s$? In fact, we do not know yet whether or not* $\mathrm{E}\int_0^t |b_s\rho_s|^2 ds < \infty$*;* i.e., that $b\rho \in H_2[0,T]$ for any $T > 0$, which is used to usually required to justify the martingale property.

Theorem 18 (Girsanov)**.** *If b_t is bounded, then $\rho_t[b]$ is a martingale and* $\mathrm{E}\rho_t = 1$.

Proof. Consider the stopping time $\tau_N := \inf(t \geq 0 : \rho_t \geq N)$. Clearly $\int_0^t 1(s \leq \tau_N) b_s\rho_s dW_s$ is a martingale ("mart") because

$$\mathrm{E}\int_0^t 1^2(s \leq \tau_N) b_s^2\rho_s^2 ds \leq t\|b\|_B^2 N^2 < \infty.$$

So,

$$\mathrm{E}\rho_{t\wedge\tau_N} = 1 + E\int_0^{t\wedge\tau_N} b_s\rho_s dW_s = 1.$$

Moreover, by virtue of the martingale property for $\rho_{t\wedge\tau_N}$,

$$\mathrm{E}(\rho_{t\wedge\tau_N}|\mathcal{F}_s) = \rho_{s\wedge\tau_N}, \quad s < t.$$

Here due to the continuity of ρ_t,

$$\tau_N \to \infty, \quad N \to \infty.$$

Therefore, the right hand side here tends to ρ_s as $N \to \infty$. What happens with the left hand side? We would show the martingale property of ρ_t if we knew that $\rho_{t\wedge\tau_N}$ is uniformly integrable (U.I.). Indeed, uniform integrability allows to use Lebesgue's analogue of the dominated convergence theorem *for conditional expectations*, under the U.I. condition instead of the domination assumption. *[This is a material for your homework: to repeat all limit theorems for conditional expectations.]*

So, it suffices to show that

$$\mathrm{E}\rho^2_{t\wedge\tau_N} \le C$$

with some C, independent of N. We estimate,

$$\begin{aligned}
\mathrm{E}\rho^2_{t\wedge\tau_N}[b] &= \mathrm{E}\exp(2\int_0^{t\wedge\tau_N} b_s dW_s - \int_0^{t\wedge\tau_N} b_s^2 ds) \\
&= \mathrm{E}\exp(\int_0^{t\wedge\tau_N} 2b_s dW_s - \frac{1}{4}\int_0^{t\wedge\tau_N} (2b)_s^2 ds) \\
&= \mathrm{E}\exp(\int_0^{t\wedge\tau_N} 2b_s dW_s - \frac{1}{2}\int_0^{t\wedge\tau_N} (2b)_s^2 ds + \frac{1}{4}\int_0^{t\wedge\tau_N} (2b)_s^2 ds) \\
&\le \exp(\frac{1}{4}t\|(2b)^2\|_B)\mathrm{E}\rho_{t\wedge\tau_N}[2b] = \exp(\frac{1}{4}t\|(2b)^2\|_B) < \infty.
\end{aligned}$$

Note that the right hand side here does not depend on N. Thus, for any bounded b the stochastic exponential ρ_t is a (non-negative) martingale with $E\rho_t = 1$. Hence, this exponential may serve as a probability density. Let us define a new measure on $\mathcal{F}$,

$$\tilde{\mathbb{P}}(A) = \mathbb{P}^{\rho_t}(A) := \mathrm{E}\rho_t 1(A).$$

[Homework: check that $\tilde{\mathbb{P}}$ is, indeed, a probability measure.] The theorem is proved.

Let us pose a question: can the boundedness of b for the martingale property of ρ be relaxed and how far? The most well-known is Novikov's condition

$$\mathrm{E}\exp(\frac{1}{2}\int_0^t b_s^2 ds) < \infty.$$

There were preceding conditions by Gikhman and Skorokhod, and there are more recent extensions due to Krylov. We will learn one small step towards these weaker conditions. Namely, assume

$$\mathrm{E}\exp(C\int_0^t b_s^2 ds) < \infty.$$

We claim that if $C > 0$ is large enough, then this condition implies the U.I. of the exponential, and, hence, it is a martingale. To show it, let us return to the calculus establishing the uniform integrability of $\rho_{\cdot\wedge\tau_N}$; we will try to improve it a bit. We have,

$$\begin{aligned}
\mathrm{E}\rho^2_{t\wedge\tau_N}[b] &= \mathrm{E}\exp(2\int_0^{t\wedge\tau_N} b_s dW_s - \int_0^{t\wedge\tau_N} b_s^2 ds) \\
&= \mathrm{E}\exp(\int_0^{t\wedge\tau_N} 2b_s dW_s - (4-3)\int_0^{t\wedge\tau_N} b_s^2 ds) \\
&\stackrel{CBS}{\le} \left(\mathrm{E}\exp(\int_0^{t\wedge\tau_N} 4b_s dW_s - 8\int_0^{t\wedge\tau_N} b_s^2 ds\right)^{\frac12} \\
\times \left(\mathrm{E}\exp(6\int_0^{t\wedge\tau_N} b_s^2 ds)\right)^{\frac12} &\le \left(\mathrm{E}\exp(6\int_0^{t} b_s^2 ds)\right)^{\frac12}.
\end{aligned}$$

Therefore, we conclude that the condition

$$\mathrm{E}\exp(6\int_0^t b_s^2 ds) < \infty$$

suffices. Better results in this directions may be obtained by using more accurate Hölder inequalities, and there are some further tricks. See [1]:

Список литературы

[1] N.V. Krylov, https://arxiv.org/abs/1903.00759

It is very useful to know that even without any condition of this sort, the stochastic exponential

$$\rho_t = \exp(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds)$$

possesses a supermartingale property. Assume just $P(\int_0^t b_s^2 ds < \infty) = 1$, so that a SI $\int_0^t b_s dW_s$ is defined.

Theorem 19 (recall that $\rho_t = 1 + \int_0^t b_s \rho_s dW_s$). *Under the assumption* $P(\int_0^t b_s^2 ds < \infty) = 1$ *the process* ρ_t *is a supermartingale:* $\rho_{t_1} \ge \mathrm{E}(\rho_{t_2}|\mathcal{F}_{t_1})$, $\forall t_1 < t_2$, *&* $\mathrm{E}\rho_t \le 1$ $\forall t$.

Proof. Return to the beginning of the proof of the last theorem. With a stopping time $\tau_N := \inf(t \ge 0 : \rho_t \ge N)$, the process $\int_0^t 1(s \le \tau_N) b_s \rho_s dW_s$ is a martingale, so,

$$1 + \mathrm{E}(\int_0^{t_2 \wedge \tau_N} b_s \rho_s dW_s | \mathcal{F}_{t_1}) = 1 + \int_0^{t_1 \wedge \tau_N} b_s \rho_s dW_s.$$

In other words, $\mathrm{E}(\rho_{t_2 \wedge \tau_N} | \mathcal{F}_{t_1}) = \rho_{t_1 \wedge \tau_N}$. The supermart inequality follows from the Fatou lemma for conditional expectations $\mathrm{E}(\liminf_{N \to \infty} \rho_{t_2 \wedge \tau_N} | \mathcal{F}_{t_1}) \le \liminf_{N \to \infty} \rho_{t_1 \wedge \tau_N}$, since $\rho_{s \wedge \tau_N} \to \rho_s$ due to continuity of ρ. The Theorem is proved.

Corollary 4. *For any bounded adapted process b_t,*

$$\mathrm{E} \exp(\int_0^t b_s dW_s) < \infty.$$

Further, assume that ρ_t is a martingale on $[0, t]$; define $\tilde{W}_s := W_s - \int_0^s b_u du,\ s \le t$. The next question is quite natural: we changed our measure; it is likely that W_s is no more a WP under this new measure; but is there a new WP instead? Igor Vladimirovich Girsanov proposed a new WP to be:

$$\tilde{W}_s := W_s - \int_0^s b_u du, \quad 0 \le s \le t.$$

Theorem 20 (Girsanov)**.** *Let b_t be bounded. Then $\tilde{W}_s$ is a Wiener process on $[0, t]$ under the measure $\tilde{\mathbb{P}} : d\tilde{\mathbb{P}}/d\mathbb{P} = \rho_t$.*

This theorem will be proved a few pages later.

Corollary 5 (Girsanov)**.** *For any bounded Borel drift $b(\cdot)$ there exists a* weak *solution of an SDE*

$$dX_t = b(X_t)dt + dW_t, \quad X_0 = x. \tag{35}$$

Proof of Corollary. Denote $X_s = W_s + x$ and

$$\tilde{W}_s = W_s - \int_0^s b(x + W_u) du, \quad s \le t.$$

This is a new WP under the new probability measure

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \rho_t := \exp(\int_0^t b(x + W_s) dW_s - \frac{1}{2} \int_0^t b^2(x + W_s) ds).$$

Then we have,

$$W_s = \tilde{W}_s + \int_0^s b(x + W_u)du, \quad s \le t,$$

and hence, clearly,

$$W_s + x = x + \tilde{W}_s + \int_0^s b(x + W_u)du, \quad s \le t,$$

Therefore, $X = x + W$ is a solution to the SDE with a new WP on $[0, t]$,

$$X_s = x + \tilde{W}_s + \int_0^s b(X_u)du, \quad s \le t.$$

Corollary is proved[22].

Exercise 14. *Show that this is a pathwise unique strrong solution, using one of the previous theorems on pathwise uniqueness.*

Remark 4. *In the next lecture there will be definitions of* **strong** *and* **weak** *solutions of SDEs: strong is any solution X_t which is adapted to the Wiener filtration $(\mathcal{F}_t^W)$ with $\mathcal{F}_t^W = \sigma(W_s : s \le t)$, and weak ones are all other solutions (they are adapted to some bigger filtration, not necessarily generated by the Wiener process in the equation). In Girsanov's construction it is evident that $\tilde{W}_t$ is adapted to $(\mathcal{F}_t^W)$ – that is, the new Wiener process $\tilde{W}$ is adapted to the filtration generated by the* **solution** *$(X_t = x + W_t)$ – but totally unclear whether or not W_t is adapted to $(\mathcal{F}_t^{\tilde{W}})$. So, this theorem does not guarantee that the solution that we found is strong*[23].

Remark 5. *For the SDE (35) the claim of Girsanov's theorem remains valid for unbounded coefficients b satisfying the condition*

$$|b(s, x)| \le C(1 + |x|).$$

[22]It is known since 1974(1979) (for $d = 1$ ($d \ge 1$)) that, actually, this solution is strong. However, this fact follows from quite a different technique.

[23]However, in this case the claim is correct: this solution **is** strong, which was discovered in 1974 by A.K.Zvonkin in the case $d = 1$ and in 1979 in the case $d > 1$ by the lecturer; to this date there are similar results for certain classes of unbounded integrable drift functions b, and certain likewise results for SDEs in Hilbert spaces. All extensions after 1979 were developed after year 2000.

Список литературы

[1] *Веретенников А.Ю., О слабых решениях сильно вырожденных СДУ, Автомат. и телемех., 2020, 3, 28–43. DOI: https://doi.org/10.1134/S0005231019030034*

Definition 17. *We call solution of an SDE weakly unique iff all solutions on any probability space with any Wiener process have the same distribution. Solution is strongly (aka pathwise) unique iff on any probability space with any Wiener process if there are solutions X_t and Y_t, then $\mathbb{P}(X_t = Y_t, \forall t) = 1$.*

Strong uniqueness implies weak uniqueness. Both notions are extremely important. Now, the corollary 5 poses the question: is solution of the equation (35) unique, strongly or, at least, weakly?

Proof of Girsanov's theorem 20. First of all, recall that under the assumption of boundedness of the drift b the random variable ρ_t may serve as a probability density. Further, for the proof we need one lemma and the reformulation of the definition of a WP via its characteristic function, namely, for any $0 = t_0 < t_1 < t_2 \ldots < t_N$ and real values $\lambda_j, 1 \le j \le N$,

$$\tilde{\mathrm{E}} \exp(\sum_{j=0}^{N-1} i\lambda_j (\tilde{W}_{t_{j+1}} - \tilde{W}_{t_j})) = \exp(-\frac{1}{2}\sum_j \lambda_j^2 (t_{j+1} - t_j)).$$

Lemma 7. *Let $\beta_t = \beta_t^1 + i\,\beta_t^2$ be a bounded adapted random process, where $i = \sqrt{-1}$. Then the (complex-valued) process*

$$\rho_t[\beta] := \exp(\int_0^t \beta_s dW_s - \frac{1}{2}\int_0^t \beta_s^2 ds)$$

is a (complex-valued) martingale.

To *use* this lemma, let $\lambda_s := i\lambda_j$ on $[t_j, t_{j+1})$, and $B_s = b_s + \lambda_s$. Then

$$\tilde{\mathrm{E}} \exp(\sum_{j=0}^{N-1} i\lambda_j(\tilde{W}_{t_{j+1}} - \tilde{W}_{t_j})) = \mathrm{E}\rho_t[b] \exp(\sum_{j=0}^{N-1} i\lambda_j(\tilde{W}_{t_{j+1}} - \tilde{W}_{t_j}))$$
$$= \mathrm{E} \exp(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds)$$
$$\times \exp(\sum_{j=0}^{N} i\lambda_j(W_{t_{j+1}} - W_{t_j} - \int_{t_j}^{t_{j+1}} b_u du))$$
$$= \mathrm{E} \exp(\int_0^t B_s dW_s - \frac{1}{2}\int_0^t B_s^2 ds + \frac{1}{2}\int_0^t \lambda_s^2 ds).$$

Indeed,

$$B_s^2 = (b_s + \lambda_s)^2 = b_s^2 + \lambda_s^2 + 2b_s\lambda_s.$$

But due to the lemma

$$\mathrm{E} \exp(\int_0^t B_s dW_s - \frac{1}{2}\int_0^t B_s^2 ds) = 1$$

while $\frac{1}{2}\int_0^t \lambda_s^2 ds$ is non-random and equals

$$\frac{1}{2}\int_0^t \lambda_s^2 ds = -\frac{1}{2}\sum_j \lambda_j^2(t_{j+1} - t_j).$$

Therefore,

$$\tilde{\mathrm{E}} \exp(\sum_{j=0}^{N-1} i\lambda_j(\tilde{W}_{t_{j+1}} - \tilde{W}_{t_j})) = \exp(-\frac{1}{2}\sum_j \lambda_j^2(t_{j+1} - t_j)),$$

as required. Girsanov's theorem 20 about a new WP under the (Girsanov's) change of measure is proved.

Proof of lemma 7. It suffices to check for any $A \in \mathcal{F}_{t_1}$, $t_2 > t_1$, and a complex value z,

$$\mathrm{E}1(A) \exp(\int_0^{t_2} (\beta_s^1 + z\beta_s^2) dW_s - \frac{1}{2}\int_0^{t_2} (\beta_s^1 + z\beta_s^2)^2 ds)$$
$$= \mathrm{E}1(A) \exp(\int_0^{t_1} (\beta_s^1 + z\beta_s^2) dW_s - \frac{1}{2}\int_0^{t_1} (\beta_s^1 + z\beta_s^2)^2 ds).$$

We already know the equality for all real-valued z,

$$\begin{aligned}&\mathrm{E}1(A)\exp\Big(\int_0^{t_2}(\beta_s^1+z\beta_s^2)dW_s-\frac{1}{2}\int_0^t(\beta_s^1+z\beta_s^2)^2ds\Big)\\ &=\mathrm{E}1(A)\exp\Big(\int_0^{t_1}(\beta_s^1+z\beta_s^2)dW_s-\frac{1}{2}\int_0^t(\beta_s^1+z\beta_s^2)^2ds\Big).\end{aligned}$$

Hence, the claim will be proved if we show that both sides are analytic functions of z. For the latter, it suffices to show that both sides are continuous in z and that their integrals along any closed bounded contours are equal to zero (Morera's theorem). Because of the analyticity of the expressions **under** the expectations and by Fubini's theorem (i.e., we can change the order of expectation and integration over the contour), we only need to show that for any $R>0$ and $|z|\le R$, these expressions are bounded by an integrable r.v. independently of z.

Clearly, to show such domination we only need to care about the stochasic integrals (since Lebesgue's ones are bounded for $|z|\le R$). By virtue of the clever inequality for any $\alpha,\beta\in R$ with $|\alpha|\le|\beta|$,

$$\exp(\alpha)\le\exp(\alpha)+\exp(-\alpha)\le\exp(\beta)+\exp(-\beta),$$

we have,

$$\begin{aligned}&|\exp\Big(\int_0^{t_2}(\beta_s^1+z\beta_s^2)dW_s\Big)|=\exp\Big(\int_0^{t_2}(\beta_s^1+Re(z)\beta_s^2)dW_s\Big)\\ &\le\exp\Big(\int_0^{t_2}(\beta_s^1+R\beta_s^2)dW_s\Big)+\exp\Big(\int_0^{t_2}(\beta_s^1-R\beta_s^2)dW_s\Big).\end{aligned}$$

The latter expression is integrable independently of z (of course, for $|z|\le R$). Hence, the left hand side is uniformly integrable. The lemma is proved.

Theorem 21 (Gikhman & Skorokhod)**.** *Under the condition of the theorem 20 the solution of the equation (35) is weakly unique*[24].

Proof. It suffices to show uniqueness of the expressions of the form

$$\mathrm{E}\prod_{k=1}^n f_k(X_{t_k})$$

[24]It is my personal opinion that originally this result was established by these two authors in their monograph "Theory of random processes" in 3 volumes. Nowadays it is known that this equation possesses the property of strong uniqueness, which implies weak uniqueness automatically. However, this theorem shows a useful way to prove results of tis kind.

for any n, any $0 \le t_1 < \ldots < t_n$, and any set of functions[25] $f_1, \ldots f_n \in B(\mathbb{R}^d)$ (bounded Borel measurable). Moreover, it suffices to assume $f_1, \ldots f_n \in C_b(\mathbb{R}^d)$, see theorem 2.

In the proof of theorem20 we started with a Wiener process W_t, changed measure according to Girsanov, and found out that $X_t = x + W_t$ is a soution on the new probability space with a new Wiener process $\tilde{W}_t = W_t - \int_0^t b(x+W_s)ds$ under the new measure $\tilde{P}$. Now let us perform the reverse transformation. We have a solution X_t ($dX_t = dW_t + b(X_t)dt$, $X_0 = x$) on the probability space with a Wiener process W_t. Let

$$\tilde{W}_t = W_t + \int_0^t b(X_s)ds, \quad t \le T,$$

and

$$d\tilde{\mathbb{P}} = \rho_T d\mathbb{P},$$

where

$$\rho_T = \exp(-\int_0^T b(X_s)dW_s - \frac{1}{2}\int_0^T b^2(X_s)ds).$$

Note that here

$$X_t = x + \tilde{W}_t, \quad t \le T,$$

and that $\tilde{W}_t$ is a Wiener process on $[0,T]$ under the probability measure $\tilde{\mathbb{P}}$.

We have, using that $\rho_T > 0$ everywhere,

$$\mathrm{E}\prod_{k=1}^n f_k(X_{t_k}) = \mathrm{E}\rho_T\rho_T^{-1}\prod_{k=1}^n f_k(X_{t_k}) = \tilde{\mathrm{E}}\rho_T^{-1}\prod_{k=1}^n f_k(X_{t_k})$$

$$= \tilde{\mathrm{E}}\exp(\int_0^t b(x+\tilde{W}_s)dW_s + \frac{1}{2}\int_0^t b^2(x+\tilde{W}_s)ds)\prod_{k=1}^n f_k(x+\tilde{W}_{t_k}).$$

Recall that according to theorem 20 under the measure $d\tilde{\mathbb{P}} = \rho_t d\mathbb{P}$ the process $X_s - x$ is a new Wiener process. It is true that

$$d\tilde{W}_t = dW_t + b(x+W_t)dt \quad \sim \quad dW_t = d\tilde{W}_t - b(x+W_t)dt.$$

So,

$$\int_0^t b(x+\tilde{W}_s)dW_s + \frac{1}{2}\int_0^t b^2(x+\tilde{W}_s)ds = \int_0^t b(x+\tilde{W}_s)d\tilde{W}_s - \frac{1}{2}\int_0^t b^2(x+\tilde{W}_s)ds$$

[25] Recall that for the first reading $d = 1$.

Hence,

$$\mathrm{E}\prod_{k=1}^{n} f_k(X_{t_k}) = \tilde{\mathrm{E}}\exp(\int_0^t b(x+\tilde W_s)d\tilde W_s - \frac{1}{2}\int_0^t b^2(x+\tilde W_s)ds)\prod_{k=1}^{n} f_k(x+\tilde W_{t_k}).$$

Note that we obtained in the right hand side an expectation of a functional of a (new) Wiener process under the measure $\tilde{\mathbb{P}}$. All Wiener processes on all probability spaces have the same distributions. Therefore, the rights hand side here, and, whence, the left hand side as well, have both the same value irrespectively on what probability space and with which Wiener process W_t we consider the original solution X_t. The theorem 21 is proved.

3.2 Exponential bound for SI

An exponential inequality[26] via stochastic exponent $\rho_t[b] = \exp(\int_0^t b_s dW_s - \frac{1}{2}\int_0^t b_s^2 ds)$.

Theorem 22 (SI exp bounds)**.** *Let the adapted process b. be bounded. Then there exist C_1, C_2 such that for any $a > 0$ and for any $T > 0$,*

$$\mathbb{P}(\sup_{0\le t\le T} |\int_0^t b_s dW_s| \ge a) \le C_1 \exp(-a^2/(C_2 T)).$$

The setting is d-dimensional[27]; b here is a vector. It is also true for b matrices with some slight changes in constants.

Corollary 6. *Under the same assumptions, for any $T > 0$ $\exists \alpha > 0$ such that*

$$E\exp(\alpha \sup_{0\le t\le T} |\int_0^t b_s dW_s|^2) < \infty.$$

Proof of theorem. For any λ, $\rho_t[\lambda b]$ is a continuous martingale. So, with any $\lambda > 0$

[26] May be skipped for the first reading; it is recommended to solve, at least, the exercise 12 for day 2.

[27] For the first reading $d = 1$.

by Bienaymé–Chebyshev–Markov's inequality we have,

$$\mathbb{P}(\sup_{0\le t\le T}|\int_0^t b_s dW_s| \ge a)$$

$$\le \mathbb{P}(\sup_{0\le t\le T}\int_0^t \lambda b_s dW_s \ge \lambda a) + \mathbb{P}(\sup_{0\le t\le T}\int_0^t (-\lambda) b_s dW_s \ge \lambda a)$$

$$\le e^{-\lambda a}\mathrm{E}\exp(\sup_{0\le t\le T}\int_0^t \lambda b_s dW_s) + e^{-\lambda a}\exp(\sup_{0\le t\le T}\int_0^t (-\lambda) b_s dW_s).$$

Consider each term separately and in the same manner. In the middle of the calculus we use Doob's inequality:

$$e^{-\lambda a}\mathrm{E}\exp(\sup_{0\le t\le T}\int_0^t \lambda b_s dW_s) = e^{-\lambda a}\mathrm{E}\sup_{0\le t\le T}\exp(\int_0^t \lambda b_s dW_s)$$

$$= e^{-\lambda a}\mathrm{E}\sup_{0\le t\le T}\rho_t[\lambda b]\exp(+\frac{1}{2}\int_0^t (\lambda b_s)^2 ds) \le \exp(-\lambda a + Ct\lambda^2)\mathrm{E}\sup_{0\le t\le T}\rho_t[\lambda b]$$

$$\stackrel{CBS}{\le} \exp(-\lambda a + Ct\lambda^2)\left(\mathrm{E}\sup_{0\le t\le T}\rho_t^2[\lambda b]\right)^2 \stackrel{\text{Doob's ineq}}{\le} \exp(-\lambda a + Ct\lambda^2)\sqrt{4\mathrm{E}\rho_T^2[\lambda b]}$$

$$= 2e^{-\lambda a + Ct\lambda^2}\left(\mathrm{E}\rho_T[2\lambda b]\exp(\int_0^t (\lambda b_s)^2 ds)\right)^{1/2}$$

$$\stackrel{\text{new } C}{\le} 2e^{-\lambda a + Ct\lambda^2}\left(\mathrm{E}\rho_T[2\lambda b]\right)^{1/2} = 2\exp(-\lambda a + Ct\lambda^2).$$

Taking $\inf_{\lambda>0}$, obtain with $\lambda = a/(2Ct)$ the bound

$$e^{-\lambda a}\mathrm{E}\exp(\sup_{0\le t\le T}\int_0^t \lambda b_s dW_s) \le 2\exp(-a^2/(4Ct)).$$

The other term

$$e^{-\lambda a}\exp(\sup_{0\le t\le T}\int_0^t (-\lambda) b_s dW_s),$$

clearly, admits the same bound,

$$e^{-\lambda a} \mathrm{E} \exp(\sup_{0 \le t \le T} \int_0^t (-\lambda b_s) dW_s) \le 2 \exp(-a^2/(4Ct)).$$

Overall, we obtain, as required,

$$\mathbb{P}(\sup_{0 \le t \le T} | \int_0^t b_s dW_s| \ge a) \le e^{-\lambda a} \mathrm{E} \exp(\sup_{0 \le t \le T} \int_0^t \lambda b_s dW_s)$$

$$+ e^{-\lambda a} \exp(\sup_{0 \le t \le T} \int_0^t (-\lambda) b_s dW_s) \le 4 \exp(-a^2/(4Ct)).$$

Proof of corollary. The idea is to use the bound with $a^2 = z \ge 0$

$$\mathbb{P}(\sup_{0 \le t \le T} | \int_0^t b_s dW_s|^2 \ge a^2) = \mathbb{P}(\sup_{0 \le t \le T} | \int_0^t b_s dW_s| \ge a)$$

$$\le 4 \exp(-a^2/(4Ct)).$$

Now integrate (in the middle by parts) with $\alpha < (4Ct)^{-1}$:

$$\mathrm{E} \exp(\alpha \sup_{0 \le t \le T} | \int_0^t b_s dW_s|^2)$$

$$= \int_0^\infty \exp(\alpha z) d(-P(\sup_{0 \le t \le T} | \int_0^t b_s dW_s|^2 \ge z))$$

$$= 1 + \int_0^\infty \mathbb{P}(\sup_{0 \le t \le T} | \int_0^t b_s dW_s|^2 \ge z) d \exp(\alpha z)$$

$$\le 1 + \alpha \int_0^\infty 4 \exp(-z[(4Ct)^{-1} - \alpha]) dz < \infty.$$

The corollary is proved.

3.3 Further strong solutions

Principle 1 (T. Yamada & S. Watanabe)**.** *Let SDE (37) possess a (weak) solution. Let this solution be pathwise unique. Then this solution is strong, and there exists a (strong) solution on any probability space with any Wiener process.*

Proof (a bit non-rigorous). The idea itself is simple: if for any WP trajectory W_t, $0 \le t \le T$ there might be only one solution (in fact, no more than one), then this one solution X should be uniquely determined by this WP; because if not, then it seems that there are at least two solutions, which is forbidden by the assumption of the uniqueness. And if the mapping $W \mapsto X$ is uniquely determined, then "clearly" X is a function of W, hence, it should be that $X_t \in \mathcal{F}_t^W$, which is the definition of strong solution. *Yet, of course, this consideration is not rigorous.*

Slightly more rigorously
Let X, W and X', W' be two independent solutions of the same SDE. Let Q and Q' be their measures on the space of trajectory pairs x, w and x', w'. Let $\pi : (x, w) \mapsto w$; then the projections πQ and $\pi Q'$ both equal the Wiener measure P^W. Let $Q^w(dx)$ and $Q^{w'}(dx')$ be the *regular conditional probabilities* of the first components given the second (WP). Consider the triple (x, x', w) with the measure

$$Q(dx, dx', dw) := Q^w(dx)Q^w(dx')P^W(dw).$$

With respect to this measure the pair (x, w) and the pair (x', w) are both solutions of our SDE. Due to the assumed pathwise uniqueness we have $x_t = x'_t$ a.s. with respect to the measure $Q^w(dx) \times Q^w(dx')$. "Hence", there exists a function $F_t : w \mapsto x_t$. A non-trivial lemma is that this function is measurable. We leave the reader to read the followng two sorces:

Список литературы

[1] N. Ikeda, S. Watanabe, Stochastic differential equations and diffusion processes, North Holland Publ. Co., Amsterdam – Oxford – New York, 1981; 2nd edition 1989. [Lemma 4.1.1]

[2] A.K.Zvonkin, N.V.Krylov: On the strong solutions of stochastic differential equations, Proceedings of the school-seminar on the theory of random processes (Druskininkai, 1974), Part II, Vilnius, 1975, pp.9-88 (in Russian).

Further strong solution results were established, in particular, in the following theorems, and later in the 21th century. Due to the Yamada and Watanabe principle, pathwise uniquess implies existence of strong solutions, once the presence of weak solutions is already established (in all cases below it is the case).

Theorem 23 (T. Yamada and S. Watanabe I (1971))**.** *Let the drift coefficient $b(t,x)$ in dimension*[28] *$d=1$ be Borel measurable, and bounded, and possessing a modulus of continuity $\rho_b(u) = \sup_{t;|x-x'|\le u} |b(t,x)-b(t,x')|$ satisfying*

$$\int_0^1 \frac{du}{\rho_b(u)} = +\infty,$$

let diffusion $\sigma(t,x)$ have also a modulus of continuity $\rho_\sigma = \sup_{t;|x-x'|\le u} |\sigma(t,x) - \sigma(t,x')|$ satisfying

$$\int_0^1 \frac{du}{\rho_\sigma^2(u)} = +\infty, \tag{36}$$

Then the SDE

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \quad X_0 = x \in \mathbb{R} \tag{37}$$

has a pathwise unique solution.

Hölder 1/2 for σ and Lipschitz condition for b in x variable is a particular case of this result. The equation (37) itself can be regarded in $\mathbb{R}^d$ with any $d\ge 1$; only in this (and in some other places) theorem $d=1$.

Proof sketch.
Assume there are two solutions X and X'. Let $1 > a_1 > \ldots > a_n \to 0$, $n\to\infty$, so that $\int\limits_{a_n}^{a_{n-1}} \rho_\sigma^{-2}(u)du = n$, $\forall\, n$. Then there exist smooth functions $\psi_n \in C_b^2([0,\infty))$ such that $\psi_n(u)\uparrow u$,

$$\psi_n''(u) = \left\{ \begin{array}{cc} 0, & u\le a_n, \\ \text{between } 0 \text{ and } \frac{2}{n}\rho_\sigma^{-2}(u), & a_n \le u \le a_{n-1}, \\ 0, & u \ge a_{n-1}, \end{array} \right.$$

[28] Actually, the authors consider the case $d\ge 1$; however, only for diagonal matrices σ with elements on the diagonal all satisfying the same modulus of continuity ρ_σ under the condition stated in the theorem.

and

$$\psi_n'(u) = \begin{cases} 0, & u \le a_n, \\ \text{between 0 and 1}, & a_n \le u \le a_{n-1}, \\ 1, & u \ge a_{n-1}, \end{cases}$$

Let $C^2_{b,loc} \ni \psi_n(x) := \psi_n(|x|)$, $x \in \mathbb{R}$ (i.e., the function ψ_n is extended on the whole line $\mathbb{R}$ by symmetry); then $\psi_n(x) \uparrow |x|$, $n \to \infty$. Applying Ito's formula to $\psi_n(X_t - X_t')$ and taking expectations we obtain

$$E\psi_n(X_t - X_t') \le C \int_0^t E1(X_s \ne X_s')\rho_b(X_s - X_s')ds$$
$$+C \int_0^t E1(X_s \ne X_s')\psi_n''(X_s - X_s')\rho_\sigma^2(X_s - X_s')ds.$$

Here

$$(0 \le) \quad 1(X_s \ne X_s')\psi_n''(X_s - X_s')$$
$$\le 1(a_n \le |X_s - X_s'| \le a_{n-1})\frac{2}{n}\rho_\sigma^{-2}(X_s - X_s'),$$

so,

$$\sup_s E1(X_s \ne X_s')\psi_n''(X_s - X_s')\rho_\sigma^2(X_s - X_s') \le \frac{2}{n} \to 0,\ n \to \infty.$$

Therefore, we get due to the monotone convergence in the lhs

$$E|X_t - X_t'| \le C \int_0^t E1(X_s \ne X_s')\rho_b(X_s - X_s')ds$$
$$= C \int_0^t E\rho_b(|X_s - X_s'|)ds \overset{\text{Jensen's ineq}}{\le} C \int_0^t \rho_b(E|X_s - X_s'|)ds.$$

The condition of concavity of $\bar\rho(u)$ was used while applying Jensen's inequality. Since $\bar\rho(u)$ is non-decreasing on $[0, \infty)$ and due to the assumption $\bar\rho(0) = 0$ and $\int_0^1 \frac{du}{\rho_b(u)} = +\infty$, by the **Bihari** inequality[29] it is known that in this case

$$E|X_t - X_t'| = 0, \quad t \ge 0,$$

as required. QED

A short explanation of this last step is given in the end of this subsection.

[29] The generalisation of Grönwall's inequality.

Theorem 24 (T. Yamada and S. Watanabe II (1971)). *Let the drift coefficient*[30] $b(x)$ *in dimension* $d > 1$ *be Borel measurable, and bounded, and possessing a modulus of continuity* $\rho_b(u)$ *satisfying*

$$\int_0^1 \frac{du}{\rho_b(u)} = +\infty,$$

let diffusion $\sigma(x)$ *have also a modulus of continuity* $\rho_\sigma(u)$ *such that the function* $u^{-1}\rho^2_\sigma(u)$ *is concave on* $(0,\infty)$ *and*

$$\int_0^1 \frac{u\,du}{\rho^2_\sigma(u)} = +\infty,$$

Then the SDE (37) has a pathwise unique solution.

Lipschitz condition both for b and σ in x variable is a particular case of this result. *The proof* is along the same lines as the proof of the theorem 23 and will not be shown here.

Theorem 25 (S. Nakao (1972)). *Let the drift coefficient* $b(x)$ *in dimension* $d = 1$ *be Borel measurable and bounded; let diffusion* $\sigma(x)$ *have a finite variation, bounded, and bounded away from zero. Then the SDE in dimension one*[31]

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x \in \mathbb{R}$$

has a pathwise unique solution[32].

Theorem 26 (A.K. Zvonkin (case $d = 1$) and A.Yu. Veretennikov (case $d > 1$)). *Let the drift coefficient* $b(t,x)$ *be Borel measurable and bounded*[33]*; let diffusion* $\sigma(t,x)$ *be bounded, uniformly nondegenerate, and let it satisfy uniform Lipschitz condition*[34] *in* x *in case* $d > 1$*, or Hölder-*$1/2$ *condition*[35] *in case* $d = 1$*. Then the SDE*

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t, \quad X_0 = x \in \mathbb{R} \quad (X_0 = x \in \mathbb{R}^d)$$

has a pathwise unique solution.

---

[30] *In fact, in the main part of their paper Yamada and Watanabe consider for simplicity the case of* $b \equiv 0$.

[31] Note that here it is important that b and σ may not depend on t; there are extensions of this theorem on the case where b and σ, actually, *may* depend on t; dimension 1 is also very important.

[32] It is important here that b and σ do not depend on time. There is some extension to the case with $b(t,x)$ and $\sigma(t,x)$

[33] In both results linear growth in x is allowed, too.

[34] That is, $|\sigma(t,x) - \sigma(t,x')| \le C|x - x'| \ \forall (t,x)$.

[35] That is, $|\sigma(t,x) - \sigma(t,x')| \le C|x - x'|^{1/2} \ \forall (t,x)$.

For the *proof of theorem 26* Zvonkin's transformation of space is used to remove the drift, which is a solution of a parabolic *system*

$$u_t(t,x) + Lu(t,x) = 0, \quad u(T,x) = x.$$

Denote $Y_t = u(t, X_t)$. By applying Ito-Krylov's formula (to be presented later in the course) we get

$$dY_t = \tilde{\sigma}(t, Y_t) dW_t,$$

with $\tilde{\sigma}(t,y)$ that is not Lip, but has a derivative integrated in any L_p locally. This suffices to show the pathwise uniqueness; however, the details are a bit involved and will not be shown here.

Список литературы

[1] T. Yamada and S. Watanabe I, II (1971)

[2] S. Nakao

[3] A.K.Zvonkin,

[4] A.Yu.Veretennikov

[5] Yamada

Lemma 8 (simplified Bihari lemma)**.** *If* $(v(x),\ x \geq 0)$ *is a finite continuous function satisfying*

$$0 \leq v(x) \leq \int_0^x \varphi(v(s)) ds, \quad x \geq 0,$$

with a nonnegative nonstrictly increasing function φ *such that* $\varphi(0) = 0$ *and* $\displaystyle\int_0^1 \frac{ds}{\varphi(s)} = \lim_{a \downarrow 0} \int_a^1 \frac{ds}{\varphi(s)} = \infty$, *then*

$$v(x) \equiv 0.$$

Proof.
Denote

$$F(x) := \int_0^x \varphi(v(s)) ds, \quad x \geq 0,$$

so that the condition of the lemma can be rewritten as

$$0 \leq v(x) \leq F(x),$$

and it also follows that
$$0 \le v(x) \le \int_0^x \varphi(F(s))ds.$$
If $\varphi(F(x)) = 0$ for any $x \ge 0$, then clearly $v(x) = 0$ from the condition and from the last inequality. Assume that $\varphi(F(x)) > 0$ for some values of x, starting from x^*. Note that F is differentiable and
$$0 \le F'(x) = \varphi(v(x)) \overset{\text{by the condition}}{\le} \varphi(F(x)), \quad \forall\, x \ge 0.$$
Then for any $x > x^*$ we have
$$0 \le \frac{F'(x)}{\varphi(F(x))} \le 1.$$
So, for any $\epsilon > 0$ and $x \ge \epsilon$ we estimate
$$0 \le \int_{x^*+\epsilon}^{x} \frac{F'(s)ds}{\varphi(F(s))} \le x - x^* - \epsilon.$$
Denoting $F(s) =: z$ and changing variables, the latter inequality may be presented as
$$0 \le \int_{F(x^*+\epsilon)}^{F(x)} \frac{dz}{\varphi(z)} \le x - x^* - \epsilon.$$
However, $F(x^*) = 0$ by our assumption, and so $F(x^* + \epsilon) \to 0$ as $\epsilon \to 0$. Therefore, by virtue of the condition $\int_0^1 ds/\varphi(s) = \infty$ we get in the limit
$$\infty \le x - x^*,$$
which is not possible. This shows that the assumption that $\varphi(F(x)) > 0$ for $x > x^*$ for some $x^* \ge 0$ is not possible either, which means that $\varphi(F(x)) = 0\, \forall\, x \ge 0$. We have already seen that in this case
$$v(x) = 0, \quad x \ge 0,$$
as required[36]. QED.

[36] The "full" Bihari – LaSalle inequality states some bound under the assumption that $v(0) \ge 0$.

3.4 Comparison theorem (case $d = 1$)

In the case $d = 1$ it turns out that under good regularity conditions not only the solution of the SDE is pathwise unique, but also they may be ordered respectively to the initial conditions: if $x_0^1 \le x_0^2$ (for the i.c. x_0^1, x_0^2) then for the corresponding solutions *of the same SDE* the inequality $X_t^1 \le X_t^2$ holds for any t. We will not prove exactly this claim, but something close to it. there is a big series of papers on this topic, and I have chosen one of the most simple. The first result of this sort was established by A.V.Skorokhod in the 60th. It was followed by the paper by T. Yamada in 1973, one theorem from which will be presented now. *It relates to solutions of two different SDEs, with different drifts.*

Theorem 27 (T.Yamada, 73)**.** *Let functions $\sigma(t,x), b_1(t,x), b_2(t,x)$ be bounded and continuous in (t,x), and let σ satisfy the condition (36). Let X_t^1 & X_t^2 be two solutions of the equations* on the same probability space with the same W_t, but with different drifts

$$dX_t^1 = b_1(t, X_t^1)dt + \sigma(t, X_t^1)dW_t, \tag{38}$$
$$dX_t^2 = b_2(t, X_t^2)dt + \sigma(t, X_t^2)dW_t \tag{39}$$

with the same initial condition

$$X_0^1 = X_0^2 = x. \tag{40}$$

Then, the strict inequality for each (t,x)

$$b_1(t,x) < b_2(t,x) \tag{41}$$

implies that

$$\mathbb{P}(X_t^1 \le X_t^2,\, t \ge 0) = 1. \tag{42}$$

Proof. Let

$$\tau := \inf(t \ge 0 :\ b_1(t, X_t^1) \ge b_2(t, X_t^2)) \qquad (\inf(\emptyset) = +\infty).$$

Clearly, $b_1(0,x) < b_2(0,x)$. So,

$$\mathbb{P}(\tau > 0) = 1.$$

Denote $\tau_t = \tau \wedge t$. We have (since expectations of stochastic integrals equal zero)

$$\mathrm{E}(X_{\tau_t}^2 - X_{\tau_t}^1) = \mathrm{E}\int_0^{\tau_t} (b_2(s, X_s^2) - b_1(s, X_s^1))ds. \tag{43}$$

Now use the sequence of functions ψ_n from the proof of the theorem 23 such that $\psi_n(u) \uparrow u$ for any $u \geq 0$ (revise what is the sequence (a_n) in that proof),

$$\psi_n''(u) = \begin{cases} 0, & u \leq a_n, \\ \text{between } 0 \text{ and } \frac{2}{n}\rho_\sigma^{-2}(u), & a_n \leq u \leq a_{n-1}, \\ 0, & u \geq a_{n-1}, \end{cases}$$

and

$$\psi_n'(u) = \begin{cases} 0, & u \leq a_n, \\ \text{between } 0 \text{ and } 1, & a_n \leq u \leq a_{n-1}, \\ 1, & u \geq a_{n-1}, \end{cases}$$

Let $C_b^2 \ni \psi_n(x) := \psi_n(|x|)$, $\forall\, x \in \mathbb{R}$; then $\psi_n(x) \uparrow |x|$, $n \to \infty$ and $\psi_n'(x) \to \operatorname{sign}(x)$, $n \to \infty$, and $|\psi_n'(x)| \leq 1$. Applying Ito's formula to $\psi_n(X_t^2 - X_t^1)$, we obtain

$$\mathrm{E}\psi_n(X_{\tau_t}^2 - X_{\tau_t}^1) = \mathrm{E}\int_0^{\tau_t} \psi_n'(X_s^2 - X_s^1)(b_2(s, X_s^2) - b_1(s, X_s^1))ds$$

$$+\frac{1}{2}\mathrm{E}\int_0^{\tau_t} \psi_n''(X_s^2 - X_s^1)(\sigma_2(s, X_s^2) - \sigma_1(s, X_s^1))^2 ds.$$

By definition,

$$(b_2(s, X_s^2) - b_1(s, X_s^1)) \geq 0$$

for any $s < \tau$, and $|\psi_n'| \leq 1$, so,

$$\mathrm{E}\int_0^{\tau_t} \psi_n'(X_s^2 - X_s^1)(b_2(s, X_s^2) - b_1(s, X_s^1))ds \leq \mathrm{E}\int_0^{\tau_t} (b_2(s, X_s^2) - b_1(s, X_s^1))ds$$

As in the proof of the theorem 23 we have

$$\sup_s E\psi_n''(X_s^2 - X_s^1)\rho_\sigma^2(X_s^2 - X_s^1) \leq \frac{2}{n} \to 0,\ n \to \infty.$$

Thus, we get in the limit as $n \to \infty$ and by virtue of

$$\mathrm{E}|X_{\tau_t}^2 - X_{\tau_t}^1| \leq \mathrm{E}\int_0^{\tau_t} (b_2(s, X_s^2) - b_1(s, X_s^1))ds.$$

Comparing this inequality with (43), we see that $\mathrm{E}|X_{\tau_t}^2 - X_{\tau_t}^1| \leq \mathrm{E}(X_{\tau_t}^2 - X_{\tau_t}^1)$, which can only be true if $\mathbb{P}(X_t^2 - X_t^1 \geq 0,\, \forall\, t < \tau) = 1$. *This implies the theorem.* QED

Exercise 15. *Justify yourself independently the last sentence of the proof. Explain where the property $\mathbb{P}(\tau > 0) = 1$ was needed in the proof.*

4 Lecture 4, 24.09.2021, Krylov's estimates, Sobolev derivatives and Weak solutions

4.1 Weak solutions: statements of the main results

Girsanov's theorem provides one way to obtain a solutions for a certain class of coefficients of drift and diffusion. A priori, this solution is weak: at least, we do not know whether it is strong (unless the coefficients satisfy the conditions of Ito's theorem). There are other theorems about weak solutions: we will state results due to Skorokhod and due to Krylov.

Theorem 28 (A.V. Skorokhod)**.** *Let coefficients $b(t,x)$ and $\sigma(t,x)$ be bounded*[37] *and continuous in (t,x). Then there exists a probability space with a Wiener process (W_t) adapted to a certain filtration $(\mathcal{F}_t)$ such that on this space there exists a solution of the equation (180)*

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t,\ t\ge 0, \qquad X_0 = x_0.$$

Theorem 29 (N.V. Krylov)**.** *Let coefficients $b(t,x)$ and $\sigma(t,x)$ be bounded*[38]*, Borel measurable, and let $\sigma\sigma^*$ nondegenerate, that is, there exists a nonrandom value $\nu>0$ such that*

$$\inf_{t,x}\inf_{\lambda\in\mathbb{R}^d}\lambda^*\sigma(t,x)\sigma^*(t,x)\lambda \ge \nu\,|\lambda|^2.$$

Then there exists a probability space with a Wiener process (W_t) adapted to a certain filtration $(\mathcal{F}_t)$ such that on this space there exists a solution of the equation (180).

Both theorems use Skorokhod's technique of the possibility of realisation of weak convergence of measures as a convergence in probability (or a.s.) of random variables with the same distributions on some new probability space.

Theorem 30 (Weak uniqueness: Krylov, Stroock–Varadhan)**.** *If b and σ are bounded, Borel measurable, and σ is uniformly nondegenerate & uniformly continuous, then solution of the SDE (180) is weakly unique.*

Example 4 (H. Tanaka: no pathwise uniqueness; hence, no strong solution)**.** *Let $d=1$, $b\equiv 0$, $\sigma(x) = sign(x)$ ($\sigma(0)=1$), $X_0=0$. Then solution of the equation (180) is not unique and (180) may not have a strong solution*[39]*.*

[37] or satisfying a linear growth condition in x uniformly with respect to t.

[38] or satisfying a linear growth condition in x uniformly with respect to t.

[39] The fact that there may be no strong solution in tghis case follows from the Yamada & Watanabe principle, see the end of page 33. Note that $\mathbb{P}(X_t=0)=0$, $\forall\, t$; this is why Y_t is also a solution.

Indeed, $X_t = \int_0^t \mathrm{sign}(X_s)dW_s \Longrightarrow Y_t := -X_t = \int_0^t \mathrm{sign}(Y_s)dW_s$ a.s. (another solution).

NB: There is another useful technique not based on Skorokhod's unique probability space method for establishing existence of weak solutions, called the technique of a **martingale problem**. Its basis is a famous P. Lévy characterisation of Wiener process as a continuous square integrable martingale M_t with the compensator (aka bracket, or triangular bracket) which equals t: $\langle M\rangle_t = t, \forall\, t \geq 0$. *The latter means that $M_t^2 - t$ is a martingale as well.* There is a multidimensional version of this result, too.

Example 5 (N.V.Krylov)**.** *Let $d = 2$, $b \equiv 0$, ... (to be completed) Then solution of the equation (180) is not unique and (180) may not have a strong solution*[40] *(the claim relates* only *to the i.c. $X_0 = 0$).*

4.2 Skorokhod's theorems

Theorem 31 (Skorokhod, Studies in the theory of random processes, §1.6)**.** *Let $\{\xi_t^n,\ t \geq 0,\ n \geq 0\}$ be some d-dimensional stochastic processes on some probability spaces and $\forall\, T > 0, \varepsilon > 0$*

$$\lim_{c\to\infty} \sup_n \sup_{t\leq T} \mathbb{P}(|\xi_t^n| > c) = 0,$$
$$\lim_{h\downarrow\infty} \sup_n \sup_{t,s\leq T;\, |t-s|\leq h} \mathbb{P}(|\xi_t^n - \xi_s^n| > \varepsilon) = 0,$$

Then for any sequence $n' \to \infty$ a new probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ can be constructed with the processes $\widetilde{\xi}_t^{n'}$, $t \geq 0$ and $\widetilde{\xi}_t$, $t \geq 0$, such that all finite-dimensional distributions of $\widetilde{\xi}_\cdot^{n'}$ coincide with those of $\xi_\cdot^{n'}$ and $\exists$ a subsequence $n'' \to \infty$ such that for any $t \geq 0$

$$\widetilde{\xi}_t^{n''} \overset{\mathbb{P}'}{\to} \widetilde{\xi}_t, \quad n \to \infty.$$

Remark 6. *In the literature this result is known, usually, in a slightly different formulation: weak convergence of measures may be realised as an a.s. convergence of the sequence of random variables on some probability space, see [P.Billingsley, Weak convergence of measures], or [A.V.Bulinsky and A.N.Shiryaev, Theory of random processes (in Russian)].*

[40] The fact that there may be no strong solution in this case follows from the Yamada & Watanabe principle, see the end of page 33. Note that $\mathbb{P}(X_t = 0) = 0$, $\forall\, t$; this is why Y_t is also a solution.

Theorem 32 (Skorokhod, convergence of SI, simplified)**.** *Let $f^n : \mathbb{R}\times\Omega \to \mathbb{R}\,(n \ge 0)$ be uniformly bounded random processes on some probability space; let $(W^n\,(n\ge 0))$ be a sequence of (d-dimensional) Wiener processes on the same probability space, and let all Itô's stochastic integrals $\displaystyle\int_0^T f^n_s\,dW^n_s, n \ge 0$ be well-defined, and for any $\varepsilon > 0$,*

$$\lim_{h\to 0}\sup_n \sup_{|s-t|\le h} \mathbb{P}\{|f^n_s - f^n_t| > \varepsilon\} = 0, \tag{44}$$

$$\textit{and}\ \forall\ s \in [0,T] \quad (f^n_s, W^n_s) \xrightarrow{\mathbb{P}} (f^0_s, W^0_s).$$

Then

$$\int_0^T f^n_s\,dW^n_s \xrightarrow{\mathbb{P}} \int_0^T f^0_s\,dW^0_s.$$

The proof is not difficult: the SI for $n = 0$ may be approximated by finite sums; then these finite sums are approximated by finite sums for n.

Proof of theorem 28

Assume b, σ bounded. Consider mollified by convolutions Lipschitz in (s,x) coefficients b^n, σ^n which converge pointwise to b, σ, respectively. Wlog they may be assumed all uniformly bounded. Denote by (X^n_t, W^n_t) corresponding (strong) solutions of (180) with b^n, σ^n:

$$X^n_t = x_0 + \int_0^t b^n(s, X^n_s)ds + \int_0^t \sigma^n(s, X^n_s)dW^n_s.$$

We are going to pass to the limit in this equation over some subsequence $n' \to \infty$. Let us check that conditions of the two auxiliary Skorokhod's theorems on $\xi^n = (X^n, W^n)$ (for the first theorem) and on $\tilde{\xi}^n_t = \sigma^n(t, X^n_t)$ (for the second one) are satisfied. Firstly, we have

$$\sup_{t\le T} E|X^n_t|^2 \le 3|x_0|^2 + 3E(\int_0^T \|\sigma^n(s, X^n_s)\|^2 ds)$$
$$+3E(\int_0^T \|b^n(s, X^n_s)\|ds)^2 \le C(|x_0|^2 + T + T^2)$$

From here and from Chebyshev – Markov inequality we get

$$\lim_{c\to\infty}\sup_n\sup_{t\le T} P(|X^n_t| > c) \le \lim_{c\to\infty} c^{-1}C(|x|^2 + T + T^2) = 0.$$

In addition,

$$\lim_{c\to\infty}\sup_n\sup_{t\le T} P(|W^n_t| > c) \le \lim_{c\to\infty} c^{-2}T = 0.$$

Secondly,

$$E|X_t^n - X_{t+h}^n|^2 \le 2E(\int_t^{t+h} \|\sigma^n(s, X_s^n)\|^2 ds)$$
$$+2E(\int_t^{t+h} \|b^n(s, X_s^n)\| ds)^2 \le C(h + h^2),$$

from where we obtain

$$\lim_{h \downarrow 0} \sup_n \sup_{t_1, t_2 \le T; |t_1 - t_2| \le h} E|X_{t_1}^n - X_{t_2}^n|^2 \le \lim_{h \downarrow 0} C(h + h^2) = 0.$$

In addition,

$$\lim_{h \downarrow 0} \sup_n \sup_{t_1, t_2 \le T; |t_1 - t_2| \le h} E|W_{t_1}^n - W_{t_2}^n|^2 = \lim_{h \downarrow 0} h = 0.$$

Hence, the assumptions of the first Skorokhod's theorem 31 are met.

The assumptions of the second theorem 32 are satisfied due to the statement of the first one and because of the uniform boundedness of σ^n. So, there are equivalent processes $(\tilde{X}^n, \tilde{W}^n) \xrightarrow{P} (\tilde{X}^0, \tilde{W}^0)$. "Clearly $\tilde{W}^0$ is a WP wrt a filtration generated by all $(\tilde{X}^n, \tilde{W}^n, n \ge 1)$, and $\tilde{X}^0$ is adapted to it.

Note that the SI $\int_0^t \sigma^n(s, \tilde{X}^0) d\tilde{W}_s^0$ is well-defined.

First of all,

$$X_t^n = x_0 + \int_0^t b^n(s, X_s^n) ds + \int_0^t \sigma^n(s, X_s^n) dW_s^n$$

and assumed (or established) equivalence $(\tilde{X}, \tilde{W}^n) \sim (X^n, W^n)$ implies that

$$\tilde{X}_t^n = x_0 + \int_0^t b^n(s, \tilde{X}_s^n) ds + \int_0^t \sigma^n(s, \tilde{X}_s^n) d\tilde{W}_s^n.$$

Indeed, we have to show

$$E\left|\tilde{X}_t^n - x_0 - \int_0^t b^n(s, \tilde{X}_s^n) ds - \int_0^t \sigma^n(s, \tilde{X}_s^n) d\tilde{W}_s^n\right|^2 = 0.$$

Recall that both integrals in this expression may be approximated by integrals with step functions as integrands, that is, for a.e. $a \in [0,1]$

$$E\left|\int_0^t b^n(s, \tilde{X}_s^n) ds - \int_0^t b^n(\rho_k^a(s), \tilde{X}_{\rho_k^a(s)}^n) ds\right|^2 \to 0, \quad k \to \infty,$$

over *some* sequence $k \to \infty$ (not necessarily over all integers), where

$$\rho_k^a(s) := \rho_k(s+a) - a = [2^k(s+a)]2^{-k} - a.$$

Similarly, for a.e. $a \in [0,1]$

$$E\left|\int_0^t \sigma^n(s, \tilde X_s^n) d\tilde W_s^n - \int_0^t \sigma^n(\rho_k^a(s), \tilde X^n_{\rho_k^a(s)}) d\tilde W_s^n\right|^2 \to 0, \ k \to \infty,$$

where

$$\rho_k^a(s) := \rho_k(s+a) - a = [2^k(s+a)]2^{-k} - a.$$

Hence, we have

$$\begin{aligned}
&E\left|\tilde X_t^n - x_0 - \int_0^t b^n(s, \tilde X_s^n) ds - \int_0^t \sigma^n(s, \tilde X_s^n) d\tilde W_s^n\right|^2 \\
&= \lim_{k\to\infty} E\left|\tilde X_t^n - x_0 - \int_0^t b^n(\rho_k^a(s), \tilde X^n_{\rho_k^a(s)}) ds - \int_0^t \sigma^n(\rho_k^a(s), \tilde X^n_{\rho_k^a(s)}) d\tilde W_s^n\right|^2 \\
&\stackrel{!}{=} \lim_{k\to\infty} E\left|X_t^n - x_0 - \int_0^t b^n(\rho_k^a(s), X^n_{\rho_k^a(s)}) ds - \int_0^t \sigma^n(\rho_k^a(s), X^n_{\rho_k^a(s)}) dW_s^n\right|^2 \\
&= E\left|X_t^n - x_0 - \int_0^t b^n(s, X_s^n) ds - \int_0^t \sigma^n(s, X_s^n) dW_s^n\right|^2 = 0.
\end{aligned}$$

We may assume that $b^{n'} \rightrightarrows b$ over some subsequence. Now – only now – we may pass to the limit in

$$\tilde X_t^n = x_0 + \int_0^t b^n(s, \tilde X_s^n) ds + \int_0^t \sigma^n(s, \tilde X_s^n) d\tilde W_s^n.$$

(Wlog, we assume that $(n') \equiv (n)$.) For the Lebesgue's integral let us fix some $n_0 \gg 1$ and let $n > n_0$;

$$\begin{aligned}
&\int_0^t b^n(s, \tilde X_s^n) ds - \int_0^t b(s, \tilde X_s^0) ds \\
&= \int_0^t b^n(s, \tilde X_s^n) ds - \int_0^t b^{n_0}(s, \tilde X_s^n) ds \quad (J^1) \\
&+ \int_0^t b^{n_0}(s, \tilde X_s^n) ds - \int_0^t b^{n_0}(s, \tilde X_s^0) ds \quad (J^2) \\
&+ \int_0^t b^{n_0}(s, \tilde X_s^0) ds - \int_0^t b(s, \tilde X_s^0) ds \quad (J^3)
\end{aligned}$$

Here $J^1 \to 0$ and $J^3 \to 0$ as $n, n_0 \to \infty$ since $b^{n'} \rightrightarrows b$; also for any fixed n_0 we have $J^2 \to 0$ as $n \to \infty$ due to the smoothness of b^{n_0}. Similarly let us consider the stochastic integrals:

$$\begin{aligned}
&\int_0^t \sigma^n(s, \tilde{X}_s^n) d\tilde{W}_s^n - \int_0^t \sigma(s, \tilde{X}_s^0) d\tilde{W}_s^0 \\
&= \int_0^t \sigma^n(s, \tilde{X}_s^n) d\tilde{W}_s^n - \int_0^t \sigma^{n_0}(s, \tilde{X}_s^n) d\tilde{W}_s^n \quad (J^1) \\
&+ \int_0^t \sigma^{n_0}(s, \tilde{X}_s^n) d\tilde{W}_s^n - \int_0^t \sigma^{n_0}(s, \tilde{X}_s^0) d\tilde{W}_s^0 \quad (J^2) \\
&+ \int_0^t \sigma^{n_0}(s, \tilde{X}_s^0) d\tilde{W}_s^0 - \int_0^t \sigma(s, \tilde{X}_s^0) d\tilde{W}_s^0 \quad (J^3)
\end{aligned}$$

We may assume that $\sigma^{n'} \rightrightarrows \sigma$ We have

$$E(J^1)^2 = E\int_0^t (\sigma^n(s, \tilde{X}_s^n) - \sigma^{n_0}(s, \tilde{X}_s^n))^2 ds \to 0,\ n, n_0 \to \infty$$

due to $\sigma^n \rightrightarrows \sigma$; further, for each n_0

$$\int_0^t \sigma^{n_0}(s, \tilde{X}_s^n) d\tilde{W}_s^n - \int_0^t \sigma^{n_0}(s, \tilde{X}_s^0) d\tilde{W}_s^0 \to 0,\ n \to \infty,$$

by virtue of Skorokhod's theorem 32; and finally

$$E(J^3)^2 \to 0$$

by the same reason as for J^1 – due to $\sigma^n \rightrightarrows \sigma$.

Hence, from the equality

$$\tilde{X}_t^n = x_0 + \int_0^t b^n(s, \tilde{X}_s^n) ds + \int_0^t \sigma(s, \tilde{X}_s^n) d\tilde{W}_s^n$$

we obtain (passing to the limit over a subsequence)

$$\tilde{X}_t^0 = x_0 + \int_0^t b(s, \tilde{X}_s^0) ds + \int_0^t \sigma(s, \tilde{X}_s^0) d\tilde{W}_s^0,$$

as required.

NB: there is no reason to think that this solution is strong, i.e. that $\tilde{X}_t^0 \in \mathcal{F}_t^{\tilde{W}^0}$. There is a (difficult) 1D counterexample by Barlow for the SDE with a continuous non-degenerate bounded σ with two weak solutions (yet, with equal distributions!).

4.3 Krylov's SI estimates & weak solutions of SDEs

Let $dX_t = \sigma_t dW_t + b_t dt$ and consider a bounded domain $D \subset R^d$. Assume bounded b, σ and uniformly non-degenerate $\sigma\sigma^*$.

Theorem 33 (Krylov's bound 1)**.**

$$\mathrm{E} \int_0^{\tau_D} f(t, X_t) dt \le N_p \|f\|_{L_p((0,\infty)\times D)}$$

with any $p \ge d+1$*, where* $\tau_D := \inf(t \ge 0 : X_t \notin D)$*. The constant* N *depends on* p*,* d*, diam*(D)*, and sup-norms of* $b, \sigma\sigma^*$ *and* $(\sigma\sigma^*)^{-1}$*. Also,*

$$\mathrm{E} \int_0^{\tau_D} g(X_t) dt \le N_p \|g\|_{L_p(D)}$$

with any $p \ge d$*.*

Theorem 34 (Krylov's bound 2)**.**

$$E \int_0^{T} f(t, X_t) dt \le N_p \|f\|_{L_p((0,\infty)\times R^d)}$$

The constant N_p *depends on* p*,* d, T*, and sup-norms of* $b, \sigma\sigma^*$ *and* $(\sigma\sigma^*)^{-1}$*. Also,*

$$\mathrm{E} \int_0^{T} g(X_t) dt \le N_p \|g\|_{L_p(\mathbb{R}^d)}$$

with any $p \ge d$*.*

Theorem 35 (Krylov)**.** *Let* b, σ *be of no more than a linear growth in* x *uniformly wrt* $t \ge 0$*,*

$$|b(t,x)| + \|\sigma(t,x)\| \le C(1+|x|),$$

and let $\sigma\sigma^*$ *be non-degenerate:*

$$\inf_{t,x} \inf_{|\lambda|=1, \lambda\in R^d} (\sigma\sigma^*(t,x)\lambda, \lambda) = \nu > 0. \tag{45}$$

Then the SDE (180) has a weak solution on some probability space with some Wiener process (WP).

Additionally for the proof assume b and σ bounded (as in his book 1977), that $\sigma = \sigma^*$ (symmetry), and

$$\inf_{t,x} \inf_{|\lambda|=1, \lambda \in R^d} (\sigma(t,x)\lambda, \lambda) = \nu > 0 \tag{46}$$

(nondegeneracy, or "ellipticity" of sigma).

Consider the mollified by convolutions locally Lipschitz in (s,x) coefficients b^n, σ^n which converge to b, σ, respectively, in $L_{d+1}([0,T] \times R^d)$ and $L_{2d+2}([0,T] \times R^d)$. Wlog they may be assumed all uniformly bounded. Denote by (X_t^n, W_t^n) corresponding (strong) solutions of (180) with b^n, σ^n:

$$X_t^n = x_0 + \int_0^t b^n(s, X_s^n) ds + \int_0^t \sigma^n(s, X_s^n) dW_s^n.$$

We are going to pass to the limit in this equation over some subsequence $n' \to \infty$. Let us check that conditions of the two auxiliary Skorokhod's theorems on $\xi^n = (X^n, W^n)$ (for the first theorem) and on $\tilde{\xi}_t^n = \sigma^n(t, X_t^n)$ (for the second one) are satisfied. Firstly, we have

$$\sup_{t \le T} E|X_t^n|^2 \le 3|x|^2 + 3E(\int_0^T \|\sigma^n(s, X_s^n)\|^2 ds)$$
$$+3E(\int_0^T \|b^n(s, X_s^n)\| ds)^2 \le C(|x|^2 + T + T^2)$$

From here and from Chebyshev – Markov inequality we get

$$\lim_{c \to \infty} \sup_n \sup_{t \le T} P(|X_t^n| > c) \le \lim_{c \to \infty} c^{-1} C(|x|^2 + T + T^2) = 0.$$

In addition,

$$\lim_{c \to \infty} \sup_n \sup_{t \le T} P(|W_t^n| > c) \le \lim_{c \to \infty} c^{-2} T = 0.$$

Secondly,

$$E|X_t^n - X_{t+h}^n|^2 \le 2E(\int_t^{t+h} \|\sigma^n(s, X_s^n)\|^2 ds)$$
$$+2E(\int_t^{t+h} \|b^n(s, X_s^n)\| ds)^2 \le C(h + h^2),$$

from where we obtain (The beginning is identical to Skorokhod's proof)

$$\lim_{h\downarrow 0}\sup_n \sup_{t_1,t_2\le T;|t_1-t_2|\le h} E|X^n_{t_1}-X^n_{t_2}|^2 \le \lim_{h\downarrow 0} C(h+h^2)=0.$$

In addition,

$$\lim_{h\downarrow 0}\sup_n \sup_{t_1,t_2\le T;|t_1-t_2|\le h} E|W^n_{t_1}-W^n_{t_2}|^2 = \lim_{h\downarrow 0} h=0.$$

Hence, the assumptions of the first Skorokhod's theorem are met.

The assumptions of the second theorem are satisfied due to the statement of the first one and because of the uniform boundedness of σ^n. So, there are equivalent processes $(\tilde X^n,\tilde W^n)\xrightarrow{P}(\tilde X^0,\tilde W^0)$. "Clearly $\tilde W^0$ is a WP wrt a filtration generated by all $(\tilde X^n,\tilde W^n,n\ge 1)$, and $\tilde X^0$ is adapted to it.

Note that the SI $\int_0^t \sigma(s,\tilde X^0)d\tilde W^0_s$ is well-defined.

First of all,

$$X^n_t = x_0+\int_0^t b^n(s,X^n_s)ds+\int_0^t \sigma(s,X^n_s)dW^n_s$$

and assumed (or established) equivalence $(\tilde X,\tilde W^n)\sim(X^n,W^n)$ implies that

$$\tilde X^n_t = x_0+\int_0^t b^n(s,\tilde X^n_s)ds+\int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s.$$

Indeed, we have to show

$$E\left|\tilde X^n_t - x_0-\int_0^t b^n(s,\tilde X^n_s)ds-\int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s\right|^2=0.$$

Recall that both integrals in this expression may be approximated by integrals with step functions as integrands.

The approximations via step functions are integral sums; hence, the fact that finite-dimensional distributions of (X^n,W^n) and $(\tilde X^n,\tilde W^n)$ guarantee that the last formula, indeed, has zero in the right hand side and so the equation holds

$$\tilde X^n_t = x_0+\int_0^t b^n(s,\tilde X^n_s)ds+\int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s,$$

as required.

We skip the details of this part which are identical to those in the Skorokhod's proof. Just note (I probably forgot to mention it earlier) that the equality to zero is established *for any particular* t, and this extends to all $t \geq 0$ because the expression under the expectation is continuous a.s.

(I do not drop the calculus since the repetition is the mother of the learning)

We have for a.e. $a \in [0,1]$

$$E\left|\int_0^t b^n(s, \tilde{X}_s^n)ds - \int_0^t b^n(\rho_k^a(s), \tilde{X}^n_{\rho_k^a(s)})ds\right|^2 \to 0, \quad k \to \infty,$$

and

$$E\left|\int_0^t \sigma^n(s, \tilde{X}_s^n)dW_s^n - \int_0^t \sigma^n(\rho_k^a(s), \tilde{X}^n_{\rho_k^a(s)})d\tilde{W}_s^n\right|^2 \to 0, \ k \to \infty,$$

over *some* sequence $k \to \infty$ (not necessarily over all integers), where

$$\rho_k^a(s) := \rho_k(s+a) - a = [2^k(s+a)]2^{-k} - a.$$

Hence, we have

$$\begin{aligned}
&E\left|\tilde{X}_t^n - x_0 - \int_0^t b^n(s, \tilde{X}_s^n)ds - \int_0^t \sigma(s, \tilde{X}_s^n)d\tilde{W}_s^n\right|^2 \\
&\quad = \lim_{k\to\infty} E\left|\tilde{X}_t^n - x_0 - \int_0^t b^n(\rho_k^a(s), \tilde{X}^n_{\rho_k^a(s)})ds \right.\\
&\qquad\qquad \left. - \int_0^t \sigma^n(\rho_k^a(s), \tilde{X}^n_{\rho_k^a(s)})d\tilde{W}_s^n\right|^2 \\
&\quad \overset{!}{=} \lim_{k\to\infty} E\left|X_t^n - x_0 - \int_0^t b^n(\rho_k^a(s), X^n_{\rho_k^a(s)})ds \right.\\
&\qquad\qquad \left. - \int_0^t \sigma^n(\rho_k^a(s), X^n_{\rho_k^a(s)})dW_s^n\right|^2 \\
&= E\left|X_t^n - x_0 - \int_0^t b^n(s, X_s^n)ds - \int_0^t \sigma(s, X_s^n)dW_s^n\right|^2 = 0,
\end{aligned}$$

and by continuity this holds true a.s. for all $t \geq 0$ *simultaneously.* The equality "!" holds since finite-dimensional distributions with tilde and without are the same.

Now we may pass to the limit in

$$\tilde{X}_t^n = x_0 + \int_0^t b^n(s, \tilde{X}_s^n)ds + \int_0^t \sigma(s, \tilde{X}_s^n)d\tilde{W}_s^n.$$

(Wlog, we assume that $(n') \equiv (n)$.) For the Lebesgue's integral let us fix some $n_0 \gg 1$ and let $n > n_0$;

$$\begin{aligned} &\int_0^t b^n(s, \tilde{X}_s^n)ds - \int_0^t b(s, \tilde{X}_s^0)ds \\ = &\int_0^t b^n(s, \tilde{X}_s^n)ds - \int_0^t b^{n_0}(s, \tilde{X}_s^n)ds \quad (J^1) \\ + &\int_0^t b^{n_0}(s, \tilde{X}_s^n)ds - \int_0^t b^{n_0}(s, \tilde{X}_s^0)ds \quad (J^2) \\ + &\int_0^t b^{n_0}(s, \tilde{X}_s^0)ds - \int_0^t b(s, \tilde{X}_s^0)ds \quad (J^3) \end{aligned}$$

Assume for the moment in addition $\|b^n - b\|_{L_{d+1}} + \|\sigma^n - \sigma\|_{L_{2d+2}} \to 0,\ n, n_0 \to \infty$

Here $E|J^1| \to 0$ due to Krylov's bound:

$$E|\int_0^t (b^n(s, \tilde{X}_s^n) - b^{n_0}(s, \tilde{X}_s^n))ds| \le N\|b^n - b^{n_0}\|_{L_{d+1}} \to 0,\ n, n_0 \to \infty. \qquad (47)$$

Also for any fixed n_0 we have $J^2 \to 0$ as $n \to \infty$ due to the smoothness of b^{n_0} in x and due to Lebesgue's bounded convergence theorem.

The term J^3 is a little more tricky, but eventually also tends to zero due to Krylov's bound. There is a natural idea to use the same trick as for J^1. However, *so far we are not aware that $\tilde{X}^0$ has a stochastic differential (hopefully wrt $\tilde{W}^0$) which allows to apply Krylov's bound!* Indeed, we only know that it is a limit of $(\tilde{X}^n)$. To overcome this difficulty we "just"pass to the limit in the Krylov bound itself.

Assume first $f(s,x) \ge 0$ continuous. We now want to pass to the limit in $E\int_0^T f(s, \tilde{X}_s^n)ds \le N\|f\|_{L_{d+1}}$ This is easy due to the Fatou lemma: by virtue of this lemma we have

$$E\int_0^T f(s, \tilde{X}_s^0)ds \le N\|f\|_{L_{d+1}}. \qquad (48)$$

In the left hand side here we have some *measure* in (s,ω), and the analogue of Krylov's bound is established for continuous f. We want to extend it to all Borel bounded and then to all Borel from L_{d+1}. Here is the way to make it.

By continuous functions it is possible to approximate *monotonically* any indicator of a compact. The "Krylov inequality" with $\tilde X^0$ will still hold by the Beppo Lévy theorem about a monotone convergence. By a sequence of such indicators it is possible to approximate monotonically any Borel indicator; by finite sums of such Borel indicators we can approximate monotonically any bounded Borel f.

The space of such step functions is tight in L_{d+1}, and again the approximation (of any $f \geq 0$) may be done monotonically. All these operations keep the "Krylov inequality" with $\tilde X^0$ valid. *(This is the original Krylov's method; just before it is established that $\tilde X^0$ is an Ito process, there is no habit to call it Krylov's bound.)* Hence, (48) holds true for any $f \in L_{d+1}$. Now, $E|J^3| \to 0$ because

$$E|\int_0^t (b^n(s,\tilde X^0_s) - b^{n_0}(s,\tilde X^0_s))ds| \leq N\|b^n - b^{n_0}\|_{L_{d+1}} \to 0,\ n, n_0 \to \infty.$$

Therefore, we get, as desired,

$$E|\int_0^t b^n(s,\tilde X^n_s)ds - \int_0^t b(s,\tilde X^0_s)ds| \to 0,\ n \to \infty.$$

It remains to tackle similarly the stochastic integral.

We again assume first $f(s,x) \geq 0$ continuous. Let us consider the stochastic integrals:

$$\begin{aligned}
&\int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s - \int_0^t \sigma(s,\tilde X^0_s)d\tilde W^0_s \\
&= \int_0^t \sigma^n(s,\tilde X^n_s)d\tilde W^n_s - \int_0^t \sigma^{n_0}(s,\tilde X^n_s)d\tilde W^n_s \quad (J^1) \\
&+ \int_0^t \sigma^{n_0}(s,\tilde X^n_s)d\tilde W^n_s - \int_0^t \sigma^{n_0}(s,\tilde X^0_s)d\tilde W^0_s \quad (J^2) \\
&+ \int_0^t \sigma^{n_0}(s,\tilde X^0_s)d\tilde W^0_s - \int_0^t \sigma(s,\tilde X^0_s)d\tilde W^0_s \quad (J^3)
\end{aligned}$$

We have

$$\begin{aligned}
E(J^1)^2 &= E\int_0^t (\sigma^n(s,\tilde X^n_s) - \sigma^{n_0}(s,\tilde X^n_s))^2 ds \\
&\leq N\|\sigma^n - \sigma^{n_0}\|^2_{L_{2(d+1)}} \to 0,\ n, n_0 \to \infty
\end{aligned}$$

Further, for each n_0

$$\int_0^t \sigma^{n_0}(s,\tilde X^n_s)d\tilde W^n_s - \int_0^t \sigma^{n_0}(s,\tilde X^0_s)d\tilde W^0_s \to 0,\ n\to\infty,$$

by virtue of Skorokhod's lemma. Finally,

$$E(J^3)^2 \to 0$$

by nearly the same reason as for J^1 – due to "Krylov's estimate"(48) with $\tilde X^0$.

$$E|\int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s - \int_0^t \sigma(s,\tilde X^0_s)d\tilde W^0_s|^2 \to 0, \quad n\to\infty.$$

Hence, from the equality

$$\tilde X^n_t = x_0 + \int_0^t b^n(s,\tilde X^n_s)ds + \int_0^t \sigma(s,\tilde X^n_s)d\tilde W^n_s$$

we obtain a weak solution (passing to the limit over a subsequence)

$$\tilde X^0_t = x_0 + \int_0^t b(s,\tilde X^0_s)ds + \int_0^t \sigma(s,\tilde X^0_s)d\tilde W^0_s,$$

as required.

Recall that it was **assumed in addition** that the mollified coefficients converge in L_p:

$$\|b^n - b^{n_0}\|_{L_{d+1}} + \|\sigma^n - \sigma^{n_0}\|^2_{L_{2(d+1)}} \to 0,\ n, n_0\to\infty$$

Let us show how to relax this. It is known that for any R the mollified by convolution functions converge to the limiting ones in L_p spaces. Hence, let us introduce a continuous non-negative bounded function $\psi(x)\le 1$ so that

$$\psi(x) = \left\{\begin{array}{ll} 0, & |x|\ge 2 \\ 1, & |x|\le 1, \end{array}\right.$$

and for any $R>0$ let

$$\psi_R(x) := \psi(x/R).$$

Returning to the term J^1 "for b" – see (47) – let us rewrite it as follows,

$$
\begin{aligned}
& E|\int_0^t (b^n(s,\tilde X^n_s) - b^{n_0}(s,\tilde X^n_s))ds| \\
& \le E\int_0^t |b^n(s,\tilde X^n_s) - b^{n_0}(s,\tilde X^n_s)|\psi_R(\tilde X^n_s)ds \\
& +E\int_0^t |b^n(s,\tilde X^n_s) - b^{n_0}(s,\tilde X^n_s)|(1-\psi_R(\tilde X^n_s))ds \\
& \le N\|b^n - b^{n_0}\|_{L_{d+1}([0,T]\times B_{2R})} + CP(\sup_{0\le s\le T}|\tilde X^n_s| > R).
\end{aligned}
$$

Here the latter term is uniformly small if R is chosen large enough,

$$
P(\sup_{0\le s\le T}|\tilde X^n_s| > R) \to 0, \quad R\to\infty,
$$

because the processes are uniformly bounded in probability, while the former term tends to zero for any fixed R when $n, n_0 \to \infty$,

$$
\|b^n - b^{n_0}\|_{L_{d+1}([0,T]\times B_{2R})} \to 0, \quad n, n_0 \to \infty.
$$

Hence, indeed,

$$
E|\int_0^t (b^n(s,\tilde X^n_s) - b^{n_0}(s,\tilde X^n_s))ds| \to 0, \quad n, n_0\to\infty.
$$

Similarly we tackle the diffusion term J^1, and similarly with J^3 for both coefficients. The theorem is proved. QED

Remark 7. *Существуют варианты объединенных условий типа Скорохода (непрерывности) по одним переменным и Крылова (невырожденности диффузии) по другим, достаточных для существования слабого решения: [Nisio M., ON THE EXISTENCE OF SOLUTIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS, Osaka J. Math. 10 (1973), 185-208].*

4.4 Sobolev derivatives & Ito-Krylov's formula

Due to Krylov's bounds Ito's formula can be applied to the functions with two Sobolev derivatives. Hence, it is called Ito-Krylov's formula. The details will be presented in the next lecture.

5 Lecture 5, 01.10.2021, Ito-Krylov's formula; Lévy characterisation of WP; random time change and weak solutions

5.1 Ito-Krylov's formula

Definition 18. *The function f^1 is called the first Sobolev derivative (in the L_p sense) of the function f in the variable x^i in the domain $D \subset \mathbb{R}^d$ (possibly $D = \mathbb{R}^d$) iff there exists a sequence of function $(f_n) \in C^1(D)$ such that*

$$\|f_n - f\|_{L_p(D)} + \|\partial_{x^i} f_n - f^1\|_{L_p(D)} \to 0, \quad n \to \infty.$$

We highlight that f^1 is not a generalised, but a "normal" Borel measurable function on d variables on D.

The full meaning of Sobolev derivatives emerged in relation to elliptic and parabolic PDEs, which solutions are rarely classic, but more frequently with Sobolev derivatives. Unfortunately, this is not a combined course of SDEs and PDEs, so the reader is just referred to PDEs textbooks such as [Ladyzhenskaya, Solonnikov, Ural'tseva, Parabolic equations of the second order]. The following result is one of the variations of the Ito-Krylov formula, which can be aslo localised, as all other Ito's formulae.

Theorem 36 (Ito–Krylov's formula)**.** *Let $u \in W_p^{1,2}([0,T] \times \mathbb{R}^d)$ with some $p \geq d+1$, and let ξ_t be an Ito process in $\mathbb{R}^r$, that is, a process possessing a stochastic differential $d\xi_t = \sigma_t dW_t + b_t dt$ with a d_1-dimensional WP wrt some filtration $(\mathcal{F}_t)$ and $d \times d_1$ matrix adapted process σ_t and d-dimensional vector adapted process b_t so that σ_t and b_t are bounded and σ_t is uniformly nondegenerate,*

$$\inf_{\omega,t} \inf_{|\lambda|=1, \lambda \in \mathbb{R}^d} \lambda^* \sigma_t \sigma_t^* \lambda > 0.$$

Then

$$du(t,\xi_t) = \nabla_x u(t,X_t)\sigma_t dW_t + \left(u_t(t,\xi_t) + \frac{1}{2} Tr(a_t u_{xx})(t,X_t) + \nabla_x u(t,X_t) b_t\right) dt.$$

Here $a = \sigma\sigma^*$. A similar formula is valid for a function $u \in W_p^2(\mathbb{R}^d)$ of the variable x only with any $p \geq d$.

In other words, under the conditions of boundedness and nondegeneracy the Ito formula is valid with Sobolev derivatives as if they were just classical ones. Also, similar formulae hold true for functions with corresponding Sobolev derivatives in some domain $D \subset \mathbb{R}^d$.

Proof sketch. Denote by ∇u the vector consisting of Sobolev partial derivatives of u wrt x, by u_t its Sobolev partial derivative wrt t, and by u_{xx} its Sobolev Hessian wrt x. Consider the approximations $u^n \in C^2$ due to the definition:

$$\|u^n - u\|_{L_p} + \|u^n_t - u_t\|_{L_p} + \|\nabla_x u^n - \nabla_x u\|_{L_{p'}} + \|u^n_{xx} - u_{xx}\|_{L_p} \to 0, \quad n \to \infty$$

with $p \geq d+1$, $p' \geq 2(d+1)$. (The latter inequality can be satisfied, at least, locally due to the embedding theorems which, in particular, state that $\nabla_x u$ is locally bounded; hence, locally it belongs to $L_{p'}$ with any p'.) Note that these approximations **may** be regarded as functions from C^2_b, as this class is dense in L_p. (This is one of the parts which remains not fully rigorously justified.) By Ito's formula

$$du^n(t,\xi_t) = \nabla_x u^n(t,\xi_t)\sigma_t dW_t + \left(u^n_t(t,\xi_t) + \frac{1}{2}\mathrm{Tr}(a_t u^n_{xx})(t,\xi_t) + \nabla_x u^n(t,\xi_t) b_t \right) dt,$$

that is,

$$u^n(t,\xi_t) = u^n(0,x) + \int_0^t \nabla_x u^n(s,\xi_s)\sigma_s dW_s$$

$$+ \int_0^t \left(u^n_s(s,\xi_s) + \frac{1}{2}\mathrm{Tr}(a_s u^n_{xx})(s,\xi_s) + \nabla_x u^n(s,\xi_s) b_t \right) ds.$$

In this equation it is possible to pass to the limit as $n \to \infty$: for the integral terms we will use Krylov's inequalities, while *for the non-integral terms we will apply* **embedding theorems** *(they will also help with the gradient in the stochastic integral).*

5.2 Martingales from the space $\mathcal{M}_2^c$; brackets

Let $(M_t, \mathcal{F}_t)$ be a continuous martingale with $\mathrm{E}|M_t|^2 < \infty$ for any $t \ge 0$.

Definition 19. *The process m_t is called a triangular bracket for $M_t \in \mathcal{M}_2^c$ iff $M_t^2 - m_t$ is an $\mathcal{F}_t$-martingale. The notation is $m_t = \langle M \rangle_t$.*

Theorem 37 (Lévy characterisation - I, $d = 1$)**.** *If $M_t \in \mathcal{M}_2^c$ with the triangular bracker $\langle M \rangle_t \equiv t$, then M_t is a Wiener process.*

A **particular case** of this theorem will be established in the next subsection, and a bit later a similar partial result will be proved in the case $d \ge 1$.

5.3 Random time change, beginning

Consider the case $d = 1$; σ^2 is non-degenerate Let Y be an Ito process with a stochastic differential

$$dY_t = \sigma_t dW_t + b_t dt.$$

Let $t'(t)$ denote the inverse function $\varphi^{-1}(t)$ for the mapping $t \mapsto \varphi(t) := \displaystyle\int_0^t \sigma_s^2\, ds$, and let

$$\bar{Y}_t := Y_{t'(t)} \qquad \sim \qquad Y_t = \bar{Y}_{\varphi(t)}.$$

Lemma 9. *After this random time change $t \mapsto t'(t) = \varphi^{-1}(t)$ – see, for example, [Gikhman & Skorokhod, SDEs, Theorem 15.5], the process $\bar{Y}_t$ has a differential*

$$d\bar{Y}_t = \bar{b}_t\, dt + d\bar{W}_t, \quad \bar{Y}_0 = x, \tag{49}$$

with a new Wiener process[41]

$$\bar{W}_t = \int_0^{t'(t)} \sigma_s\, dW_s, \tag{50}$$

and with[42] *the new drift*

$$\bar{b}_t = b_{t'(t)} \sigma^{-2}_{t'(t)}.$$

---

[41] This is a particular case of theorem 37; this particular case is rigorously proved on the next pages.

[42] In a deterministic case, e.g., for $\sigma = 2$, we have $t'(t) = t/4$, $\bar{W}_t = \displaystyle\int_0^{t/4} 2\, dW_s = 2W_{t/4}$; it is well-known that the latter is, indeed, a Wiener process, as expected. Then $\bar{b}_t = \frac{1}{4} b_{t/4}$, and $\bar{Y}_t = Y_0 + \bar{W}_t + \int_0^t \bar{b}_s ds$, as required.

Proof. We have,

$$Y_t = Y_0 + \int_0^t \sigma_s dW_s + \int_0^t b_s ds,$$

and by the change of variables $\tilde{s} = \varphi(s)$, $d\tilde{s} = \varphi'(s)ds = \sigma_s^2 ds$ also

$$\int_0^{t'(t)} b_s ds = \int_0^{t'(t)} b_{t'(\varphi(s))} ds \overset{\tilde{s}=\varphi(s)}{=} \int_0^t b_{t'(\tilde{s})} \frac{d\tilde{s}}{\sigma^2_{t'(\tilde{s})}} = \int_0^t \bar{b}_{\tilde{s}} d\tilde{s} = \int_0^t \bar{b}_s ds, \qquad (51)$$

and due to the definition (50)

$$\bar{Y}_t = Y_0 + \bar{W}_s + \int_0^t \bar{b}_s ds. \qquad (52)$$

Now let us consider the process $\bar{W}_t$. We want to show that it is a Wiener process. The idea is to mimic the proof of the Lévy characterisation, but without involving explicitly stochastic integrals over continuous martingales. Let us try to compute first of all its characteristic function for a fixed value of t:

$$\mathrm{E}\exp(i\lambda \bar{W}_t) = \mathrm{E}\exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s).$$

Note that $\exp(i\lambda \int_0^t \sigma_s \, dW_s + \frac{\lambda^2}{2} \int_0^t \sigma_s^2 \, ds)$ is an $(\mathcal{F}_t)$-martingale. So, $\exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s + \frac{\lambda^2}{2} \int_0^{t'(t)} \sigma_s^2 \, ds)$ is a stopped martingale with respect to the filtration $(\bar{\mathcal{F}}_t) = (\mathcal{F}_{t'(t)})$ by Doob's optional stopping theorem. Therefore, it makes sense to write

$$\mathrm{E}\exp(i\lambda \bar{W}_t) = \mathrm{E}\exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s + \frac{\lambda^2}{2} \int_0^{t'(t)} \sigma_s^2 \, ds - \frac{\lambda^2}{2} \int_0^{t'(t)} \sigma_s^2 \, ds).$$

Here the integral

$$\int_0^{t'(t)} \sigma_s^2 \, ds = \varphi(t'(t)) = \varphi(\varphi^{-1}(t)) = t,$$

that is, it is non-random. Hence, using Girsanov's theorem for stochastic exponentials we get

$$\mathrm{E}\exp(i\lambda \bar{W}_t) = \mathrm{E}\exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s - \frac{(i\lambda)^2}{2} \int_0^{t'(t)} \sigma_s^2 \, ds + \frac{(i\lambda)^2 t}{2}) = \exp(-\frac{\lambda^2 t}{2}).$$

This corresponds to the marginal distribution of a Wiener process at time t.

Further, for a fixed non-random sequence of times $0 = t_0 \le t_1 < \ldots < t_{n-1} < t_n$ we compute:

$$\mathrm{E}\exp(i\sum_{j=0}^{n-1}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j})) = \mathrm{E}\,\mathrm{E}\left(\exp(i\sum_{j=0}^{n-1}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))|\mathcal{F}_{t_{n-1}}\right)$$

$$= \mathrm{E}\exp(i\sum_{j=0}^{n-2}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))\;\mathrm{E}\left(\exp(i\lambda_{n-1}(\bar W_{t_n} - \bar W_{t_{n-1}}))|\mathcal{F}_{t_{n-1}}\right)$$

$$= \mathrm{E}\exp(i\sum_{j=0}^{n-2}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))\;\exp(-(\lambda_{n-1}^2(t_n - t_{n-1})/2)$$

$$\stackrel{\text{induction}}{=} \ldots = \exp\left(-\sum_{j=0}^{n-1}\lambda_j^2(t_{j+1} - t_j)/2\right).$$

This corresponds to the multidimensional characteristic function of the increments of the Wiener process. Since $\bar W_t$ is continuous, the claim that it is a Wiener process is proved.

By virtue of (52) we obtain (49), as required. Lemma 9 is proved. QED

Let us return to the SDEs. Consider, still in the case $d = 1$, the SDE

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt, \quad X_0 = x.$$

The functions σ, b are assumed bounded. We know from Girsanov's theorem that there exists a (weak) solution[43] of the SDE

$$dY_t = dW_t + \bar b(Y_t)dt, \quad Y_0 = x,$$

with $\bar b(y) = b(y)/\sigma^2(y)$. We now wish to solve the SDE

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt, \quad X_0 = x.$$

[43] And according to the theorem this solution is strong, at least, if $\bar b$ is continuous; actually, this continuity assumption may be dropped.

with a given Borel measurable diffusion coefficient σ, which is bounded and uniformly nondegenerate[44]. Let $\zeta(t) = \int_0^t \sigma^{-2}(Y_s)\, ds$, and $\psi(t) = \zeta^{-1}(t)$ (the inverse function, which is a stopping time); let

$$\tilde{W}_t := \int_0^{\psi(t)} \sigma^{-1}(Y_s) dW_s \ \sim \ W_{\psi(t)} = \int_0^t \sigma(Y_{\psi(s)}) d\tilde{W}_s. \tag{53}$$

This $\tilde{W}_t$ is a new WP wrt filtration $\tilde{\mathcal{F}}_t := \mathcal{F}^W_{\psi(t)}$, according to the lemma 9. Note that its bracket equals (by Doob's "optional stopping" theorem)

$$\langle \tilde{W} \rangle_t = \int_0^{\psi(t)} \sigma^{-2}(Y_s) ds = \zeta(\psi(t)) = t.$$

Theorem 38. *The process $X_t := Y_{\psi(t)}$ is a solution of the SDE*

$$dX_t = \sigma(X_t) d\tilde{W}_t + b(X_t) dt, \quad X_0 = x.$$

Proof. Firstly, $X_t \in \tilde{\mathcal{F}}_t$ (yet, note that there is no guarantee that $\tilde{\mathcal{F}}_t = \mathcal{F}^{\tilde{W}}_t$; so, there is no reason to think that it is a strong solution).

We have,

$$Y_t = x + W_t + \int_0^t \bar{b}(Y_s) ds.$$

So,

$$X_t = Y_{\psi(t)} = x + W_{\psi(t)} + \int_0^{\psi(t)} \bar{b}(Y_s) ds.$$

By virtue of the change of variables

$$s = \psi(\tilde{s}) \sim \tilde{s} = \zeta(s); \ d\tilde{s} = \sigma^{-2}(Y_s) ds.$$

in the Lebesgue integral we have

$$\int_0^{\psi(t)} \frac{b(Y_s)}{\sigma^2(Y_s)} ds = \int_0^t b(Y_{\psi(\tilde{s})}) d\tilde{s} = \int_0^t b(X_s) ds.$$

So, according ot (53) this equation may be rewritten as

$$X_t = Y_{\psi(t)} = x + \int_0^t \sigma(X_s) d\tilde{W}_s + \int_0^t b(X_s) ds.$$

The theorem 38 is proved. QED

[44] Aka bounded away from zero.

5.4 Lévy characterisation of WP - II ($d = 1$)

For those who are familiar with the more general stochastic integrals over continuous squared integrable martingales we leave another calculus[45] *for computing the marginal characteristic function of the martingale M_t with the bracket $\langle M \rangle_t = t$. By* Ito's formula for such martingales *we have (firstly for a marginal distribution, then for finite-dimensional)*

$$d\exp(i\lambda M_t) = i\lambda \exp(i\lambda M_t) dM_t - \frac{\lambda^2}{2} \exp(i\lambda M_t) d\langle M \rangle_t,$$

from where we get

$$\varphi_t(\lambda) := E\exp(i\lambda M_t) = 1 - \frac{\lambda^2}{2} E \int_0^t \exp(i\lambda M_s) ds$$

So,

$$\varphi_t'(\lambda) = -\frac{\lambda^2}{2} \varphi_t(\lambda), \quad \varphi_0(\lambda) = 1.$$

Thus,

$$\varphi_t(\lambda) = \exp(-\frac{\lambda^2}{2} t), \qquad \textit{as required.}$$

Similarly, using conditional expectations, for $s < t$ we get

$$\varphi_{s,t}(\lambda) := E\exp(i\lambda(M_t - M_s)) = \exp(-\frac{\lambda^2}{2}(t - s)).$$

Now, for any $0 = t_0 \le t_1 < \ldots < t_n$ by induction we obtain

$$\varphi_{t_1,\ldots,t_n}(\lambda_1, \ldots, \lambda_n) := E\exp(i \sum_{k=1}^{n} \lambda_k (M_{t_k} - M_{t_{k-1}})$$

$$= E\exp(i \sum_{k=1}^{n-1} \lambda_k (M_{t_k} - M_{t_{k-1}}) E(\exp(i\lambda_n (M_{t_n} - M_{t_{n-1}}) | \mathcal{F}_{t_{n-1}})$$

$$= E\exp(i \sum_{k=1}^{n-1} \lambda_k (M_{t_k} - M_{t_{k-1}}) \exp(-\frac{\lambda_n^2}{2}(t_n - t_{n-1}))$$

$$\stackrel{(\ldots)}{=} \exp(-\frac{1}{2} \sum_{k=1}^{n} \lambda_k^2 (t_k - t_{k-1})),$$

as required. *QED*

[45]It is not recommended to use this calculus on the exam!

5.5 Random time change - II (case $d \geq 1$)

Consider the case $d \geq 1$; the matrix σ is assumed uniformly bounded and non-degenerate and of the special form

$$\sigma_t = \alpha_t I,$$

where α_t is a **scalar** adapted process, which is uniformly bounded and uniformly bounded away from zero. **In principle, random time change is a one-dimensional tool; hence, only a scalar variable function may be used for it. The unit matrix I may be replaced by a variable matrix $d \times d$ which is Lipschitz, or likewise, but we do not present any result in this direction.**

Let Y be a d-dimensional Ito process with a stochastic differential

$$dY_t = \sigma_t dW_t + b_t dt,$$

where W_t is a d-dimensional Wiener process. Let $t'(t)$ denote the inverse function $\varphi^{-1}(t)$ for the mapping $t \mapsto \varphi(t) := \int_0^t \alpha_s^2\, ds$, and let

$$\bar{Y}_t := Y_{t'(t)} \qquad \sim \qquad Y_t = \bar{Y}_{\varphi(t)}.$$

Lemma 10. *After this random time change $t \mapsto t'(t) = \varphi^{-1}(t)$ – see, for example, [Gikhman & Skorokhod, SDEs, Theorem 15.5], the process $\bar{Y}_t$ has a differential*

$$d\bar{Y}_t = \bar{b}_t\, dt + d\bar{W}_t, \quad \bar{Y}_0 = x, \tag{54}$$

with a new d-dimensional Wiener process[46]

$$\bar{W}_t = \int_0^{t'(t)} \sigma_s\, dW_s, \tag{55}$$

and with the new drift

$$\bar{b}_t = (\sigma\sigma^*)^{-1}_{t'(t)}\, b_{t'(t)}.$$

Proof. We have,

$$Y_t = Y_0 + \int_0^t \sigma_s dW_s + \int_0^t b_s ds,$$

and by the change of variables $\tilde{s} = \varphi(s)$, $d\tilde{s} = \varphi'(s)ds = \alpha_s^2 ds$ also

[46]This is a particular case of the multi-dimensional version of theorem 37; this particular case is rigorously proved on the next pages.

$$\int_0^{t'(t)} b_s ds = \int_0^{t'(t)} b_{t'(\varphi(s))} ds \stackrel{\tilde{s}=\varphi(s)}{=} \int_0^t b_{t'(\tilde{s})} \frac{d\tilde{s}}{\alpha^2_{t'(\tilde{s})}} = \int_0^t \bar{b}_{\tilde{s}} d\tilde{s} = \int_0^t \bar{b}_s ds, \tag{56}$$

and due to the definition (55)

$$\bar{Y}_t = Y_0 + \bar{W}_s + \int_0^t \bar{b}_s ds. \tag{57}$$

Now let us consider the process $\bar{W}_t$. We want to show that it is a Wiener process. The idea is to mimic the proof of the Lévy characterisation, but without involving explicitly stochastic integrals over continuous martingales. Let us try to compute first of all its characteristic function for a fixed value of t: for $\lambda \in \mathbb{R}^d$

$$\mathrm{E} \exp(i\lambda \bar{W}_t) = \mathrm{E} \exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s) = \mathrm{E} \exp(i\lambda \int_0^{t'(t)} \alpha_s \, I \, dW_s).$$

Note that $\exp(i\lambda \int_0^t \alpha_s \, I \, dW_s + \frac{\lambda^2}{2} \int_0^t \alpha_s^2 \, ds)$ is an $(\mathcal{F}_t)$-martingale. So, $\exp(i\lambda \int_0^{t'(t)} \alpha_s \, dW_s + \frac{\lambda^2}{2} \int_0^{t'(t)} \alpha_s^2 \, ds)$ is a stopped martingale with respect to the filtration $(\bar{\mathcal{F}}_t) = (\mathcal{F}_{t'(t)})$ by Doob's optional stopping theorem. Therefore,

$$\mathrm{E} \exp(i\lambda \bar{W}_t) = \mathrm{E} \exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s - \frac{\lambda^2}{2} \int_0^{t'(t)} \alpha_s^2 \, ds + \frac{\lambda^2}{2} \int_0^{t'(t)} \alpha_s^2 \, ds),$$

where the integral

$$\int_0^{t'(t)} \alpha_s^2 \, ds = \varphi(t'(t)) = \varphi(\varphi^{-1}(t)) = t,$$

that is, it is non-random. Hence, using Girsanov's theorem for stochastic exponentials we get

$$\mathrm{E} \exp(i\lambda \bar{W}_t) = \mathrm{E} \exp(i\lambda \int_0^{t'(t)} \sigma_s \, dW_s - \frac{(i\lambda)^2}{2} \int_0^{t'(t)} \alpha_s^2 \, ds + \frac{(i\lambda)^2 t}{2}) = \exp(-\frac{\lambda^2 t}{2}).$$

This corresponds to the marginal distribution of a d-dimensional Wiener process at time t.

Further, for a fixed non-random sequence of times $0 = t_0 \le t_1 < \ldots < t_{n-1} < t_n$ we compute repeating verbatim the calculus in the one-dimensional case:

$$\mathrm{E}\exp(i\sum_{j=0}^{n-1}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j})) = \mathrm{E}\,\mathrm{E}\left(\exp(i\sum_{j=0}^{n-1}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))|\mathcal{F}_{t_{n-1}}\right)$$

$$= \mathrm{E}\exp(i\sum_{j=0}^{n-2}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))\;\mathrm{E}\left(\exp(i\lambda_{n-1}(\bar W_{t_n} - \bar W_{t_{n-1}}))|\mathcal{F}_{t_{n-1}}\right)$$

$$= \mathrm{E}\exp(i\sum_{j=0}^{n-2}\lambda_j(\bar W_{t_{j+1}} - \bar W_{t_j}))\;\exp(-(\lambda_{n-1}^2(t_n - t_{n-1})/2)$$

$$\stackrel{\text{induction}}{=} \ldots = \exp\left(-\sum_{j=0}^{n-1}\lambda_j^2(t_{j+1} - t_j)/2\right).$$

This corresponds to the multidimensional characteristic function of the increments of the d-dimensional Wiener process. Since $\bar W_t$ is continuous, the claim that it is a d-dimensional Wiener process is proved.

By virtue of (57) we obtain (54), as required. Lemma 10 is proved. QED

Let us return to the SDEs. Consider, now in the case $d \ge 1$, the SDE

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt, \quad X_0 = x.$$

The functions σ, b are assumed bounded, and σ of the special form

$$\sigma(x) = \alpha(x)\,I_{d\times d},$$

where α is a scalar function, $\alpha:\, x \mapsto \mathbb{R}^1$. We know from Girsanov's theorem that there exists a (weak) solution[47] of the SDE

$$dY_t = dW_t + \bar b(Y_t)dt, \quad Y_0 = x,$$

with $\bar b(y) = b(y)/\alpha^2(y)$. We now wish to solve the SDE

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt, \quad X_0 = x.$$

[47] Acually, this solution is strong.

with a given Borel measurable diffusion coefficient σ, which is bounded and uniformly nondegenerate[48]. Let $\zeta(t) = \int_0^t \alpha^{-2}(Y_s)\, ds$, and $\psi(t) = \zeta^{-1}(t)$ (the inverse function, which is a stopping time); let

$$\tilde{W}_t := \int_0^{\psi(t)} \sigma^{-1}(Y_s) dW_s \;\sim\; W_{\psi(t)} = \int_0^t \sigma(Y_{\psi(s)}) d\tilde{W}_s. \tag{58}$$

This $\tilde{W}_t$ is a new WP wrt filtration $\tilde{\mathcal{F}}_t := \mathcal{F}^W_{\psi(t)}$, according to the lemma 10. Note that its bracket equals (by Doob's "optional stopping" theorem)

$$\langle \tilde{W} \rangle_{\psi(t)} = \int_0^{\psi(t)} \alpha^{-2}(Y_s)\, I\, ds = \zeta(\psi(t)) I = I \times t.$$

Theorem 39. *The process $X_t := Y_{\psi(t)}$ is a solution of the SDE*

$$dX_t = \sigma(X_t) d\tilde{W}_t + b(X_t) dt, \quad X_0 = x.$$

Proof. Firstly, $X_t \in \tilde{\mathcal{F}}_t$. We have,

$$Y_t = x + W_t + \int_0^t \bar{b}(Y_s) ds.$$

So,

$$X_t = Y_{\psi(t)} = x + W_{\psi(t)} + \int_0^{\psi(t)} \bar{b}(Y_s) ds.$$

By virtue of the change of variables

$$s = \psi(\tilde{s}) \sim \tilde{s} = \zeta(s);\; d\tilde{s} = \alpha^{-2}(Y_s) ds.$$

in the Lebesgue integral we have

$$\int_0^{\psi(t)} \frac{b(Y_s)}{\alpha^2(Y_s)} ds = \int_0^t b(Y_{\psi(\tilde{s})}) d\tilde{s} = \int_0^t b(X_s) ds.$$

So, according ot (58) this equation may be rewritten as

$$X_t = Y_{\psi(t)} = x + \int_0^t \sigma(X_s) d\tilde{W}_s + \int_0^t b(X_s) ds.$$

The theorem 39 is proved. QED

[48] Aka bounded away from zero.

5.6 Martingale problem & its relation to weak solutions

The SDE (37) in $\mathbb{R}^d$

$$dX_t = \sigma(t, X_t) dW_t + b(t, X_t) dt, \ X_0 = x$$

is considered. As usual, denote

$$L = \frac{1}{2}(\sigma\sigma^*\nabla, \nabla) + (b, \nabla).$$

The idea of the *martingale approach* is to find a process X_t satisfying the following properties:

- X_t is continuous;
- for any $f \in C_0^\infty$ the (complex-valued) process

$$M_t = f(t, X_t) - \int_0^t (f_s + Lf)(s, X_s) ds \tag{59}$$

  is a martingale.

  The space C_0^∞ may be replaced by C_0^2.

An alternative (equivalent) requirement is to find X_t such that

- X_t is continuous;
- the process

$$M_t := X_t - x - \int_0^t b(s, X_s) ds \tag{60}$$

  is a (continuous, squared integrable) martingale with the (multi-dimensional) bracket

$$\langle M, M \rangle_t = \int_0^t \sigma\sigma^*(s, X_s) ds. \tag{61}$$

Definition 20. *Any solution of the problem (59), or (60)–(61) is called a solution of the martingale problem. Often, not a process but a probability measure*[49] *is called a solution of the martingale problem, if under this measure the process X satisfies (59), or (60)–(61).*

[49] It may be a probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$, or on the space of trajectories (X, W), which serves as Ω in this case; then $\mathcal{F}$ could be a Borel sigma-algebra.

For the SDE with the coefficients not depending on t

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt,\ X_0 = x \tag{62}$$

still denote

$$L = \frac{1}{2}(\sigma\sigma^*\nabla, \nabla) + (b, \nabla).$$

The *martingale problem* is to find a process X_t satisfying the following properties:

- X_t is continuous;
- for any $g \in C_0^\infty$ the (complex-valued) process

$$M_t = g(X_t) - \int_0^t Lg(X_s)ds \tag{63}$$

  is a martingale.

  The space C_0^∞ may be replaced by C_0^2.

An alternative (equivalent) requirement is to find X_t such that

- X_t is continuous;
- the process

$$M_t := X_t - x - \int_0^t b(X_s)ds \tag{64}$$

  is a (continuous, squared integrable) martingale with the (multi-dimensional) bracket

$$\langle M, M\rangle_t = \int_0^t \sigma\sigma^*(X_s)ds. \tag{65}$$

Definition 21. *Any solution of the problem (63), or (64)–(65) is called a solution of the martingale problem. Often, not a process but a probability measure*[50] *is called a solution of the martingale problem, if under this measure the process X satisfies (63), or (64)–(65).*

The *Lévy characterisation* of the Wiener process provides a link between solutions of the martingale problem and solutions of the corresponding SDE: it follows from both versions of the martingale problem above and from the *Lévy characterisation* that there exists a Wiener process W such that X is a solution of the SDE (37), or of (62). The advantage is that convergence of SI may be dropped. The disadvantage is that only weak solutions may be constructed in this way. The problem of *uniqueness* of solution of the martingale problem becomes essential.

[50]It may be a probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$, or on the space of trajectories (X, W), which serves as Ω in this case; then $\mathcal{F}$ could be a Borel sigma-algebra.

In what follows the diffusion coefficient $\sigma\sigma^*$ is assumed uniformly non-degenerate, both coefficients are bounded. Let us smooth the coefficients and solve the SDE with the approximated ones

$$dX_t^n = b^n(t, X_t^n)dt + \sigma^n(t, X_t^n)dW_t, \quad X_0^n = x.$$

Then for any $f \in C_0^\infty$ the process

$$M_t^n := f(t, X_t^n) - \int_0^t (f_s + L^n f)(s, X_s^n)ds,$$

where $L^n = \frac{1}{2}\sum_{i,j}(\sigma^n(\sigma^n)^*(s,x))_{ij}\frac{\partial^2}{\partial x^i \partial x^j} + \sum_i (b^n)^i(s,x)\frac{\partial}{\partial x^i}$, **is** a martingale in $\mathcal{M}_c^2$ (continuous, squared integrable); of course, M^n depends on f. The bracket of M^n (which is the process in $\mathbb{R}^1$ since $f : (t,x) \mapsto \mathbb{R}^1$) reads

$$\langle M^n, M^n \rangle_t = \int_0^t \langle (\sigma^n)^* \nabla_x f, (\sigma^n)^* \nabla_x f \rangle (s, X_s^n)ds.$$

We now want to use only the first Skorokhod theorem not involving the problem of convergence of stochastic integrals. The conditions of this theorem (uniform boundedness in probability and uniform continuity in probability of X^n and W_t) are valid. Hence, consider the sequence of equivalent processes $(\tilde{X}_t^n, \tilde{W}_t^n, t \geq 0)$ converging in probability for each t:

$$(\tilde{X}_t^n, \tilde{W}_t^n) \stackrel{\mathbb{P}}{\to} (\tilde{X}_t^0, \tilde{W}_t^0).$$

Then the processes $f(t, \tilde{X}_t^n)$ have the same distributions as $f(t, X_t^n)$ and they also converge in probability,

$$f(t, \tilde{X}_t^n) \stackrel{\mathbb{P}}{\to} f(t, \tilde{X}_t^0).$$

The same is valid for the Lebesgue integrals $\int_0^t L^n f(s, X_s^n)ds$,

$$\int_0^t L^n f(s, \tilde{X}_s^n)ds \stackrel{\mathbb{P}}{\to} \int_0^t Lf(s, \tilde{X}_t^0)ds,$$

by Lebesgue's bounded convergence theorem **and** Krylov's bounds, as in the proof of Krylov's existence theorem.

All convergence claims are extended to $f \in C^\infty$ with a moderate growth in x due to the *a priori* bounds. Hence,

$$\tilde{M}_t^n := f(t, \tilde{X}_t^n) - \int_0^t L^n f(s, \tilde{X}_s^n)ds \stackrel{\mathbb{P}}{\to} f(t, \tilde{X}_t^0) - \int_0^t Lf(s, \tilde{X}_s^0)ds.$$

Theorem 40. *Under the assumed conditions the process* $f(t,\tilde X^0_t) - \int_0^t Lf(s,\tilde X^0_s)ds =: \tilde M^0_t$ *is an* $\mathcal{M}^2_c$*-martingale.*

Hence, under the assumed conditions the martingale problem (59) has a solution.

Proof. The continuity may be derived from Kolmogorov's continuity theorem *(Homework)*; the squared integrabilty of $\tilde M^0_t$ is valid since the process is bounded on any finite interval of time. The martingale property for $\tilde M^0_t$ follows from the uniform integrabilty (or just boundedness), from the martingale property for $\tilde M^n_t$ and from convergence. QED

Exercise 16. *Let* $d = 1$*. Derive the existence of solution of the homogeneous martingale problem in the version (64)–(65).*
Hint: use $g(x) \equiv x$ *and* $g(x) \equiv x^2$*.*

6 Lecture 6, 08.10.2021, SDEs with reflection on the line and in a (multidimensional) domain; Skorokhod's problem

6.1 One-dimensional case

In this lecture we shall see how to construct a solution of the SDE on the half line *with reflection* at zero: the easiest example is $(|W_t|, t \geq 0)$ for $d = 1$.

While WP W_t itself is a solution of the simplest SDE $dX_t = dW_t$, let us ask, is there any SDE for $|W_t|$? The modulus is not a smooth function, so at zero there is a problem, while outside zero – i.e., on the set $|W_t| > 0$ – we may write down

$$d|W_t| = dW_t, \quad \text{if } W_t > 0,$$

and

$$d|W_t| = d(-W_t) = -dW_t, \quad \text{if } W_t < 0.$$

Both formulae can be combined on the set $|W_t| > 0$ as

$$d|W_t| = \text{sign}(W_t)dW_t.$$

But how to tackle what occurs at $W_t = 0$? Define $\text{sign}(a) = 0$ for $a = 0$. It turns out that the right way for an SDE with reflection – as for $|W_t|$ – is as follows (we restrict ourselves to homogeneous equations):

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt + dL_t, \tag{66}$$

where L_t is a special process called "local time at zero" for X_t. In the integral form it looks like

$$X_t = x_0 + \int_0^t \sigma(X_s)dW_s + \int_0^t b(X_s)ds + L_t.$$

The main principle is that $dL_t = 0$ where $X_t \neq 0$, and it should not let the process X to leave the positive half-line. Hence, we assume formally

$$X_t \geq 0.$$

This is not all. We want X_t be continuous, from where L_t must be continuous, too. Its value at $t = 0$ plays no role, so we always assume $L_0 = 0$. Intuitively it should be clear that L_t *increases*, only at times where $X_t = 0$; in particular, L must have a locally bounded variation. It is natural to call *solution* of this equation the pair (X_t, L_t) if it is *strong*, or the triple (X_t, L_t, W_t) if it is weak. It turns out that even in the case if zero is not the absorbing barrier there might be many solutions to this SDE because we may "order" the solution to stay at zero for some random time (exponentially distributed) and then to "leave" this state. We will concentrate on strong solutions where this "pause" at zero is not allowed.

These considerations lead to the following

Definition 22. *The following equation with i.c.* $X_0 = x$

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt + dL_t, \; X_t \ge 0, \; L \text{ is continuous, increasing,}$$
$$L_t = \int_0^t 1(X_s = 0)dL_s, \; \int_0^t 1(X_s = 0)ds \stackrel{a.s.}{=} 0, \; \forall\, t \ge 0, \tag{67}$$

is called the **SDE with a non-sticky reflection at zero.**

Our first goal in this lecture is strong solutions for the SDE (67). However, we start with the simplest example.

Example 6. *In particular, the SDE (67) with $\sigma \equiv 1$, $b \equiv 0$ has a (unique) solution equivalent by distribution to $|W_t|, t \ge 0$.*

$$dX_t = dW_t + dL_t, \; X_t \ge 0, \; L \text{ continuous, increasing,}$$
$$L_t = \int_0^t 1(X_s = 0)dL_s, \; \int_0^t 1(X_s = 0)ds = 0, \; \forall\, t \ge 0,$$

with i.c. $X_0 = 0$. Why equivalent to $|W_t|, t \ge 0$? Let us approximate the modulus by the functions in Yamada and Watanabe's paper: $C^2 \ni \varphi_n(x) \uparrow |x|$, $\varphi_n(0) = 0$, and apply Ito's formula to $\varphi_n(W_t)$: we get

$$d\varphi_n(W_t) = \varphi_n'(W_t)dW_t + \frac{1}{2}\varphi_n''(W_t)dt,$$

or,

$$\varphi_n(W_t) = \int_0^t \varphi_n'(W_s)dW_s + \frac{1}{2}\int_0^t \varphi_n''(W_s)ds. \tag{68}$$

Here

$$\varphi_n(W_t) \uparrow |W_t| \;\; \& \;\; \varphi_n'(W_t) \to sign(W_t), \; n \to \infty,$$

and recall that φ_n' *are uniformly bounded. Hence,*

$$\mathrm{E}\int_0^t (\varphi_n'(W_s) - sign(W_s))^2 ds$$

$$= \mathrm{E}\int_0^t (\varphi_n'(W_s))^2 ds - 2\mathrm{E}\int_0^t \varphi_n'(W_s)\, sign(W_s) ds + t$$

$$\to t - 2t + t = 0, \quad n \to \infty.$$

So, the third term in (68) also has a limit, denote it by L_t. *Clearly,* L_t *is continuous. We have,*

$$\varphi_n(W_t) = \int_0^t \varphi_n'(W_s) dW_s + \frac{1}{2}\int_0^t \varphi_n''(W_s) ds.$$

So, from (68) we get

$$|W_t| = \int_0^t sign(W_s) dW_s + L_t. \tag{69}$$

Note that all φ_n'' *are non-negative. So, clearly,* L_t *is increasing. Further,* $\varphi''(x) \equiv 0$ *for* $|x| \geq a_{n-1}$ *where* $a_{n-1} \to 0$, *so, "clearly",*

$$dL_t = 0 \quad \textit{on the set } W_t \neq 0.$$

Hence,

$$L_t = \int_0^t 1(W_s = 0) dL_s.$$

Finally, $\mathrm{E}\int_0^t 1(W_s = 0) ds = 0$. *It remains to note that* $\bar{W}_t = \int_0^t sign(W_s) dW_s$ *is a new WP (by Lévy characterisation). Therefore,* $|W_t|$ *satisfies the SDE*

$$|W_t| = \bar{W}_t + L_t, \tag{70}$$

which (as will be proved in a minute) has a pathwise unique solution X. *Thus,* $\mathcal{L}(|W|) = \mathcal{L}(X)$.

Assume there are two solutions (X^1, L^1, W) *and* (X^2, L^2, W) *Approximate in the same manner as Yamada – Watanabe not* $|x|$, *but the function*

$$\psi(x) = x \times 1(x \geq 0) \equiv x_+ :$$

$$\psi_n(x) \uparrow \psi(x),\ \psi_n'(x) \to 1(x > 0),\ 0 \le \psi_n'(x) \le 1.$$

By Ito's formula

$$d\psi_n(X_t^1 - X_t^2) = \psi_n'(X_t^1 - X_t^2)(dX_t^1 - dX_t^2) + \frac{1}{2}\psi_n''(X_t^1 - X_t^2)(dX_t^1 - dX_t^2)^2$$

$$= \psi_n'(X_t^1 - X_t^2)(dL_t^1 - dL_t^2) + \frac{1}{2}\psi_n''(X_t^1 - X_t^2)(dL_t^1 - dL_t^2)^2.$$

Note that $X_t^1 - X_t^2 = L_t^1 - L_t^2$ *and that*

$$(dL_t^1 - dL_t^2)^2 = 0.$$

The latter follows from the considerations about Riemann–Stiltjes integrals. Indeed, both L_t^1 *and* L_t^2 *are continuous and of (locally) bounded variation.*

$$d\psi_n(X_t^1 - X_t^2) = \psi_n'(X_t^1 - X_t^2)(dL_t^1 - dL_t^2) + \frac{1}{2}\psi_n''(X_t^1 - X_t^2)(dL_t^1 - dL_t^2)^2.$$

In the integral form,

$$\psi_n(X_t^1 - X_t^2) = \int_0^t \psi_n'(X_s^1 - X_s^2)(dL_s^1 - dL_s^2) + \frac{1}{2}\int_0^t \psi_n''(X_s^1 - X_s^2)(dL_s^1 - dL_s^2)^2$$

$$= \int_0^t \psi_n'(X_s^1 - X_s^2)(dL_s^1 - dL_s^2).$$

As $n \to \infty$, *we get*

$$(X_t^1 - X_t^2)_+ = \int_0^t sign(X_s^1 - X_s^2)(dL_s^1 - dL_s^2).$$

We have the representation with the Lebesgue – Stiltjes integral in the rhs:

$$(X_t^1 - X_t^2)_+ = \int_0^t sign(X_s^1 - X_s^2)(dL_s^1 - dL_s^2)$$

$$= \int_0^t sign(X_s^1 - X_s^2)1(X_s^1 - X_s^2 \neq 0)(dL_s^1 - dL_s^2).$$

Note that where $X^1_s - X^2_s \neq 0$ and $X^1_s > X^2_s$, we have $dL^1 = 0$, and where $X^1_s - X^2_s \neq 0$ and $X^1_s < X^2_s$, we have $dL^2 = 0$. Also recall that both functions L^i increase. Hence, in any case, a.s.

$$sign(X^1_s - X^2_s)1(X^1_s - X^2_s \neq 0)(dL^1_s - dL^2_s) \leq 0.$$

Thus

$$(X^1_t - X^2_t)_+ = \int_0^t sign(X^1_s - X^2_s)1(X^1_s - X^2_s \neq 0)(dL^1_s - dL^2_s)$$

and

$$sign(X^1_s - X^2_s)1(X^1_s - X^2_s \neq 0)(dL^1_s - dL^2_s) \leq 0$$

imply

$$(X^1_t - X^2_t)_+ \leq 0.$$

So, clearly,

$$(X^1_t - X^2_t) \leq 0.$$

Similarly, $(X^2_t - X^1_t)_+ \leq 0$, so $(X^2_t - X^1_t) \leq 0$, too, which immediately implies also

$$L^1_t - L^2_t = X^1_t - X^2_t = 0,$$

a.s., as required.

Remark 8. *The problem of finding a solution $(\xi_t, \varphi_t), t \geq 0$ of the equation*

$$\xi_t = W_t + \varphi_t, \qquad \varphi \in C, \text{ and is increasing, and } \varphi_t = \int_0^t 1(W_s = 0)\, d\varphi_s$$

(that is, φ only increases at moments s where $W_s = 0$) is called **Skorokhod's problem.**

Its solution exists and is unique[51] for any $\omega \in \Omega$ and any trajectory $W_t(\omega)$ (which is continuous!). More than that, this solution exists and is unique for any continuous

[51] Exercise.

function $w(t)$ (instead of a trajectory of the Wiener process) with i.c. $w(0) \geq 0$. The solution is given by Skorokhod's representation

$$\xi_t = w_t \, 1(t \leq \tau) + (w_t - \inf(w_s: \, s \leq t))1(t > \tau),$$

where

$$\tau := \inf(t \geq 0 : w_t < 0),$$

and

$$\varphi_t = \xi_t - w_t.$$

Exercise 17. *Verify it!*

Now return to the general SDE with reflection (67). Even we allow the coefficients to depend on time:

$$dX_t = \sigma(t, X_t)dW_t + b(t, X_t)dt + dL_t, \ X_t \geq 0, \ L \text{ is continuous, increasing,}$$
$$\& \ \ L_t = \int_0^t 1(X_s = 0)dL_s, \ \ \int_0^t 1(X_s = 0)ds = 0, \ \forall \, t \geq 0, \ \text{ a.s.} \tag{71}$$

The coefficient σ is assumed non-degenerate; also assume

$$b(s, -x) = b(s, x), \ \ \sigma(s, -x) = \sigma(s, x),$$

that is, b and σ are extended on the negative half-line $x < 0$ in the even way. Let us define new coefficients in the even way

$$\bar{\sigma}(s, x) = \sigma(s, x)\text{sign}(x), \ \bar{b}(s, x) = b(s, x)\text{sign}(x), \ \text{for } x \neq 0,$$

and[52]

$$\bar{\sigma}(s, 0) = \sigma(s, 0), \quad \bar{b}(s, 0) = b(s, 0).$$

Theorem 41 (Existence)**.** *Suppose σ is non-degenerate. Let (Y_t, W_t) be a (weak or strong) solution of the SDE*

$$dY_t = \bar{\sigma}(t, Y_t)dW_t + \bar{b}(t, Y_t)dt, \quad Y_0 = x. \tag{72}$$

Then (X_t, W_t, L_t) with

$$X_t = |Y_t|, \ \ L_t = |X_t| - x - \int_0^t b(s, |X_s|)ds + \int_0^t \sigma(s, |X_s|)dW_s$$

is a solution of (71) (weak or strong[53]).

[52] We want to avoid the degeneracy of σ at zero.
[53] Most likely weak.

The disadvantage, or weakness of this result is that if $\sigma(t,\cdot)$ is not smooth at $x=0$, then we cannot use directly any earlier results on strong solutions. Indeed, σ is not Lipschitz and we cannot mollify it in the nondegenerate way so as to use smooth approximations. Yet, this suffices for a weak solution, e.g., using Krylov's weak existence theorem if σ is non-degenerate (but not Skorokhod's weak existence theorem).

Proof. Let us use the same approximation sequence $\varphi_n(x)\uparrow|x|$ from Yamada – Wanatabe theorem on pathwise uniqueness with the properties $0\le\varphi_n\in C^2$, $\varphi_n'(x)\to \operatorname{sign}(x)$, $|\varphi_n'(x)|\le 1$, $\varphi_n(x)=\varphi_n(-x)$, for each $x\neq 0$ $\varphi_n'(x)=\operatorname{sign}(x)$ if n is large enough, $\varphi_n''\ge 0$. Applying Ito's formula to $\varphi_n(Y_t)$ we get

$$\varphi_n(Y_t)=\varphi_n(x)+\int_0^t 2\varphi_n'(Y_s)\bar\sigma(s,Y_s)dW_s$$

$$+\int_0^t 2\varphi_n'(Y_s)\bar b(s,Y_s)ds+\int_0^t \varphi_n''(Y_s)\sigma^2(s,Y_s)ds.$$

Here all terms except possibly the last integral converge to some limits. Hence, the last term has a limit, too; denote it by L_t. Clearly[54], it is a non-negative continuous function which is increasing only on the set where $Y_s=0$. **Its name is a "local time at zero" for the process Y.** So, in the limit we obtain (recall that $x\ge 0$)

$$|Y_t|=x+\int_0^t 2\operatorname{sign}(Y_s)\bar\sigma(s,Y_s)dW_s+\int_0^t 2\operatorname{sign}(Y_s)\bar b(s,Y_s)ds+L_t$$

$$=x+\int_0^t 2\,\sigma(s,|Y_s|)dW_s+\int_0^t 2\,b(s,|Y_s|)ds+L_t.$$

For completing the proof of the theorem it remains to notice that (for example, due to Krylov's estimates)

$$\mathrm{E}\int_0^t 1(|Y_s|=0)ds=\mathrm{E}\int_0^t 1(Y_s=0)ds=0. \qquad \text{QED}$$

Although we did not use any extension of Ito's formula in the proof above, nevertheless, it exists and may be very useful in some cases.

[54] All claims in this sentence are exercises!

Lemma 11 (Ito's formula with a local time)**.** *Let the process η_t have a (generalised*[55]*) stochastic differential*

$$d\eta_t = \sigma_t dW_t + b_t dt + dL_t,$$

where L_t is a continuous and locally bounded variation adapted process (actually, not necessarily a local time). Then for any $g \in C_b^2(\mathbb{R})$

$$\begin{aligned} dg(\eta_t) &= g'(\eta_t) d\eta_t + \frac{1}{2}\, g''(\eta_t)(d\eta_t)^2 \\ &= g'(\eta_t)(\sigma_t dW_t + b_t dt + dL_t) + \frac{1}{2}\, g''(\eta_t) dt. \end{aligned} \tag{73}$$

Formally, the Ito calculus rules may be extended as follows:

$$\boxed{dL_t \times dt = dL_t \times dW_t = (dL_t)^2 = 0.}$$

Theorem 42. *Under the assumptions above (σ non-degenerate, both coefficients Borel and bounded) solution of the equation (180) exists and is unique in distribution.*

Theorem 43. *Let in addition there exist $C > 0$ such that for all s, x, x'*

$$|\sigma(s,x) - \sigma(s,x')|^2 \le C|x - x'|.$$

Then solution of the equation (180) is pathwise unique and all solutions are strong.

This is a simplified analogue of A.K. Zvonkin's result for Ito equations *without reflection* on $\mathbb{R}^1$.

Список литературы

[1] A. Veretennikov, TViP[56] (PTA in Eng.), 1981, 26, 4, 685-701.

NB: Zvonkin's transformation[57] *was not used in the proof of this paper, although, this transformation is one of the main tools in this area.*

[55] This word "generalised" is normally dropped.

[56] This paper is neither the only one, nor the first result in the area; probably, not the last one either. Just it is easier for the lecturer to refer to his own publications. Also, in a way, the full statement of its result combines (and generalises) several preceding results, mainly by Japanese probabillists.

[57] This is the name of the change of variables eliminating the drift: e.g., $x \mapsto u(x)$ where u is a solution of the equation $\frac{1}{2}\,\sigma^2 u'' + bu' = 0$ in the case of $\mathbb{R}^1$ and a homogeneous SDE (of course, if this function $u(x)$ is strctly monotone).

6.2 Multidimensional case

The following setting is taken with some simplifications from the papers [2], [3] (see below on this page). The preprint [4] – which tackles the case of a unit diffusion and of an arbitrary bounded Borel measurable drift – is included so as to show that still some important problems in the area are under investigation right now, literally nowadays: the paper based on this preprint is submitted and is waiting to be published; also the reader can see that some old problem in the area was solved just one year ago, and some adjacent open questions remain.

Список литературы

[1] Гихман И.И., Скороход А.В., Стохастические дифференциальные уравнения, Киев, Наукова думка, 1968

[2] Shinzo Watanabe, On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions, J. Math. Kyoto Univ. 11-1 (1971) 169-180

[3] Hiroshi Tanaka, Stochastic differential equations with reflecting boundary condition in convex regions, Hiroshima Math. J. 9(1): 163-177 (1979). DOI: 10.32917/hmj/1206135203

[4] S. Yang, T. Zhang, Strong Solutions to Reflecting Stochastic Differential Equations with Singular Drift, https://arxiv.org/abs/2002.12150 (2020)

[5] A. D. Wentzell, On boundary conditions for multi-dimensional diffusion Processes, Theory of Prob. and Its Appl., 4 (1959), 164-177.
А.Д. Вентцель, О граничных условиях для многомерных диффузионных процессов, Теория вероятн. и ее примен., 4:2 (1959), 172-185. http://mi.mathnet.ru/tvp4877

The following SDE system in a half-space $\mathbb{R}^d \cap \{x^1 \geq 0\} =: D$ with reflection is considered (the symbol L_t is used instead of φ_t as in [2]):

$$
\begin{aligned}
dX_t^1 &= \sigma^1(X_t)dW_t + b^1(X_t)dt + dL_t, \\
dX_t^i &= \sigma^i(X_t)dW_t + b^i(X_t)dt + \bar{\sigma}^i(X_t)dM_t + \bar{b}^i(X_t)dL_t, \ i = 2, \dots, d,
\end{aligned}
\tag{74}
$$

with initial conditions

$$X_0 = x = (x^1, \ldots, x^d) =: (x^1, \bar{x}),$$

under the assumptions that

$$L_t \geq 0, \text{ continuous, increasing, } \& \ L_t = \int_0^t 1(X_s^1 = 0)dL_s, \ \& \ \mathrm{E}\int_0^t 1(X_s^1 = 0)ds = 0. \tag{75}$$

The terms in the equation (74)–(75) are explained as follows. Here W is a d-dimensional Wiener process; L_t is a local time of X^1 at zero ($x^1 = 0$); M_t is another (independent) $(d-1)$-*dimensional Wiener process* $\tilde{W}$ *with the changed time,* $M_t = \tilde{W}_{L_t}$ (this means that the trajectory of M_t is only moving when $X_t^1 = 0$); the all Borel measurable coefficients are: the matrix $d \times d$ σ, the matrix $d \times (d-1)$ $\bar{\sigma}$, the vector-functions b and $\bar{b}$ of dimensions d and $d-1$, respectively.

Remark 9. *The coefficients $\bar{\sigma}$ and $\bar{b}$ are "responsible" for the evolution of the process X* **on the boundary $\partial D = \{(x^1 = 0)\}$, which starts whenever the process hits this boundary and stops as soon as it leaves** *∂D. Due to the "non-sticky" reflection condition $\mathrm{E}\int_0^t 1(X_s^1 = 0)ds = 0$ the set of times where L_t changes is a singular set on the line t with respect to the Lebesgue measure. Naturally, $\bar{\sigma}$ is a diffusion coefficient and $\bar{b}$ is a drift coefficient, both "acting" while the process is on the boundary. This setting is based on the general Wentzell boundary conditions for a general continuous Markov process in a domain [5], which conditions claim that the generator of this Markov process should have a form of some second order operator at ∂D. Naturally, this signifies some evolution* **along the boundary**, *not just an immediate reflection without any change in the position.*

Remark 10. *How an SDE with reflection in a* **bounded domain** *may be considered (although we will not pursue this goal here in this course)? The idea is to localise some part of the boundary of the domain and to apply a transformation (change of variables) so as to make this part of the boundary flat; then the problem is reduced to a localised version of the setting with the half-space $\mathbb{R}^d \cap \{x^1 \geq 0\}$. There are two major cases which can be tackled more or less smoothly: the case with a regular boundary ∂D and the case with a convex domain D. The paper [3] is a classical reference for the second case.*

The **definition of a solution** of the equation (74)–(75) should be stated in a usual way via integral equations which hold with the probability one.

Theorem 44.
1. *Suppose all coefficients are bounded, continuous, and that the matrix $\sigma\sigma^*$ is nondegenerate "in the direction of x^1", that is, along the normal to ∂D,*

$$\inf_x \sum_{j=1}^{d} |\sigma^{1j}(x)|^2 \geq c > 0. \tag{76}$$

Then the equation (74)–(75) has a (weak) solution on some probability space with a filtration and two Wiener processes.
2. *If in addition all coefficients are Lipschitz, then the equation (74)–(75) has a pathwise unique solution.*

Proof sketch (only weak existence[58]*, and you may assume the simplification*[59] *$\bar\sigma \equiv 0$).*

1. Case $\sigma^{11}(x) \equiv 1$, $\sigma^{1j}(x) \equiv 0$ for $j = 2, \ldots, d$, $b^1(x) \equiv 0$. In this case the equation on the first component X^1 reads

$$dX^1_t = dW^1_t + dL_t.$$

The results for the one-dimensional case are applicable here. So, solution for X^1 – in fact, for the pair (X^1, L) – exists (and even is pathwise unique). Now it remains to solve the equation on $\bar X_t = (X^2_t, \ldots, X^d_t)$ with the components X^1 and L_t already found. This can be done by the same strategy as for Ito's theorem on existence if the coefficients are Lipschitz, or as for Skorokhod's weak existence theorem by using Skorokhod's method under only continuity conditions on the coefficients[60] (recall that $\bar\sigma \equiv 0$ is assumed for the simplification of the proof).

2. The case of a variable (bounded, nondegenerate) $\sigma^{11}(x)$ and $\sigma^{1j}(x) \equiv 0$ for $j = 2, \ldots, d$, $b^1(x) \equiv 0$. This case is reduced to the case 1 by the random time change (see the lecture about the random time change); note that it leaves the drift identically zero.

[58] In principle, pathwise uniqueness can be proved similarly to Ito's, or Yamada and Watanabe's pathwise uniqueness case for SDEs in the whole $\mathbb{R}^d$.

[59] In this case there is no need in the second Wiener process $\tilde W$, or in the martingale M.

[60] It could have been performed earlier, just after Ito's and Skorokhod's theorems, respectively. However, such an order is certainly not recommended for the first reading. Note that the measurability issue should be carefully explained for the proof in details.

3. The case of a general $\sigma(x)$, and of $b^1(x) \equiv 0$. This case may be reduced to the previous one ($\sigma^{11}(x)$ variable, $\sigma^{1j}(x) \equiv 0$ for $j = 2, \ldots, d$, $b^1(x) \equiv 0$) by using the following trick: there exists[61] (non-unique, in general) orthogonal Lipschitz (or continuous under the continuity assumptions) matrix p such that

$$(\sigma p^{-1})^{1j}(x) \equiv 0, \qquad j = 2, \ldots, d,$$

with some bounded and nondegenerate $(\sigma p^{-1})^{11}(x)$. The point is that under the transformation

$$\sigma(x) \mapsto \sigma p^{-1}(x) =: \tilde{\sigma}(x), \quad b(x) \mapsto b(x)$$

the SDE (74)–(75) corresponding to the 4-tuple (quadruple) $[\sigma, b, \bar{\sigma}, \bar{b}]$ is transformed into the one corresponding to the 4-tuple $[\tilde{\sigma}, b, \bar{\sigma}, \bar{b}]$ *with a new Wiener process*. The reason for this is that

$$\int_0^t \sigma(X_s)\, dW_s = \int_0^t \underbrace{\sigma(X_s)p^{-1}(X_s)}_{=\tilde{\sigma}(X_s)} \underbrace{p(X_s)dW_s}_{=d\tilde{W}_s} = \int_0^t \tilde{\sigma}(X_s)\, d\tilde{W}_s,$$

where the process

$$\tilde{W}_t = \int_0^t p(X_s)\, dW_s \qquad (\text{i.e., } \tilde{W}_t^i = \int_0^t \sum_j p^{ij}(X_s)\, dW_s^j)$$

is again a Wiener one, which follows from the multidimensional Lévy characterisation of a Wiener process. Indeed, $\tilde{W}_t$ is clearly a continuous, squared integrable martingale and its multidimensional bracket $\langle \tilde{W}, \tilde{W} \rangle$ reads, due to the orthogonality of p,

$$\langle \tilde{W}, \tilde{W} \rangle_t = \int_0^t p(X_s)p^*(X_s)\, ds = \int_0^t I_{d\times d}\, ds = I_{d\times d}\, t,$$

as required for the Lévy characterisation.

4. General case for $\sigma(x)$ and $b(x)$. An arbitrary bounded (and probably under the linear growth condition, too) drift b may be added by using Girsanov's theorem via the transformtion of measure. QED

[61] An exercise in linear algebra.

7 Lecture 7, 15.10.2021, Links between Wiener process and Laplace operator; relations of solutions of general SDEs with general elliptic and parabolic operators

We have already seen such links arising directly from Ito's formula. In this lecture we will see some systematisation of such links, which are important for both sides: PDEs for SDEs and vice versa. We will concentrate on PDE problems with Dirichlet boundary conditions, although, other boundary condition types are also possible.

7.1 Heat equation in $\mathbb{R}_+ \times \mathbb{R}^d$

In the first several subsections we treat examples for the Wiener process, which corresponds to the Laplace operator and to the heat equation.

We start with a (d-dimensional) Wiener process W_t and with a few simplest examples. The first one is about heat equation and its relation to WP.

Example 7. *Let $u(t,x) \in C_b^{1,2}([0,T] \times \mathbb{R}^d)$ be a solution of the heat equation*

$$u_t(t,x) + \frac{1}{2}\Delta u(t,x) = 0, \quad 0 \le t \le T,$$
$$u(T,x) = g(x),$$

with $g \in C_b^2(\mathbb{R}^d)$. Then for any $0 \le t \le T$ the value $u(t,x)$ can be represented in the form

$$u(t,x) = \mathrm{E}g(x + W_{T-t}) = \mathrm{E}g(x + (W_T - W_t)).$$

Proof. For the proof, let us apply Ito's formula to $u(t_0 + s, x + W_s)$ for $0 \le t_0 < T$ (since $u(T,x) = g(x) \equiv \mathrm{E}g(x + W_0)$ without any calculus):

$$du(t_0 + s, x + W_s) = \nabla u(t_0 + s, x + W_s) dW_s$$
$$+ [u_s(t_0 + s, x + W_s) + \frac{1}{2}\Delta u(t_0 + s, x + W_s)] ds.$$

In the integral form with $t_0 + s = T$,

$$u(T, x + W_{T-t_0}) = u(t_0, x) + \int_0^{T-t_0} \nabla u(t_0 + s, x + W_s) dW_s$$
$$+ \int_0^{T-t_0} [u_s(t_0 + s, x) + \frac{1}{2}\Delta u(t_0 + s, x + W_s)] ds.$$

Let us now take expectations from both sides of this equality:

$$\mathrm{E}u(T, x + W_{T-t_0}) = u(t_0, x),$$

because

$$\mathrm{E}\int_0^{T-t_0} \nabla u(t_0 + s, x + W_s) dW_s = 0,$$
$$\& \ \ [u_s(t_0, x) + \frac{1}{2}\Delta u(t_0 + s, x + W_s)] = 0.$$

Remark 11. *The condition $g \in C_b^2(\mathbb{R}^d)$ follows automatically from $u(t,x) \in C_b^{1,2}([0,T] \times \mathbb{R}^d)$. Both conditions may be relaxed.*

Example 8. *Let $u(t,x) \in C_{b,loc}^{1,2}((0,T) \times \mathbb{R}^d) \bigcap C_b([0,T] \times \mathbb{R}^d)$ be a solution of the heat equation*

$$u_t(t,x) + \frac{1}{2}\Delta u(t,x) = 0, \quad 0 < t < T,$$
$$u(T,x) = g(x),$$

with $g \in C_b(\mathbb{R}^d)$. Then for any $0 \le t \le T$ the value $u(t,x)$ can be represented in the form

$$u(t,x) = \mathrm{E}g(x + W_{T-t}) = \mathrm{E}g(x + (W_T - W_t)).$$

The conditions of boundedness of g and u with its derivatives may be further considerably relaxed, too.

Proof. Note that the differential form of Ito's equation remains valid, $0 < t_0 < T$

$$du(t_0 + s, x + W_s) = \nabla u(t_0 + s, x + W_s) dW_s$$
$$+[u_s(t_0 + s, x + W_s) + \frac{1}{2}\Delta u(t_0 + s, x + W_s)]ds.$$

Yet, now we cannot simply integrate it to T, because the derivatives are assumed only on the open interval $(0,T)$. Firstly let us consider $t_0 > 0$. Denote $T_n := T - \frac{1}{n}$. Then, for n such that $t_0 < T_n$ we have,

$$u(T_n, x + W_{T_n - t_0}) = u(t_0, x) + \int_0^{T_n - t_0} \nabla u(t_0 + s, x + W_s) dW_s$$
$$+ \int_0^{T_n - t_0} [u_s(t_0 + s, x) + \frac{1}{2}\Delta u(t_0 + s, x + W_s)]ds.$$

Let us take expectations here: since

$$[u_s(t_0+s,x)+\frac{1}{2}\Delta u(t_0+s,x+W_s)]=0$$

and because

$$E\int_0^{T_n-t_0}\nabla u(t_0+s,x+W_s)dW_s=0,$$

we get

$$\mathrm{E}u(T_n,x+W_{T_n-t_0})=u(t_0,x),$$

or,

$$u(t_0,x)=\mathrm{E}u(T_n,x+W_{T_n-t_0}).$$

Here we can pass to the limit as $T_n\uparrow T$ in the r.h.s.: since the function u is continuous and bounded up to T, and because W is continuous in time, we get by Lebesgue's bounded convergence theorem that again

$$u(t_0,x)=\mathrm{E}u(T,x+W_{T-t_0})\equiv\mathrm{E}g(x+W_{T-t_0}),$$

as required. Moreover, in the latter equation we can again pass to the limit as $t_0\downarrow 0$, to get by the same reasoning

$$u(0,x)=\mathrm{E}g(x+W_T),$$

as required.

Now let us consider the equation with a non-zero r.h.s.

Example 9. *Let $u(t,x)\in C_b^{1,2}([0,T]\times\mathbb{R}^d)$ be a solution of the heat equation*

$$u_t(t,x)+\frac{1}{2}\Delta u(t,x)=-f(t,x),\quad 0\le t\le T,$$
$$u(T,x)=g(x),$$

with $g\in C_b^2(\mathbb{R}^d)$, $f(t,x)\in C_b([0,T]\times\mathbb{R}^d)$. Then for any $0\le t\le T$ the value $u(t,x)$ can be represented in the form

$$u(t,x)=\mathrm{E}\left[\int_0^{T-t}f(t+s,x+W_s)ds+g(x+W_{T-t})\right].$$

Proof. Recall that by Ito's formula,

$$du(t_0+s, x+W_s) = \nabla u(t_0+s, x+W_s)dW_s$$
$$+[u_s(t_0+s, x+W_s) + \frac{1}{2}\Delta u(t_0+s, x+W_s)]ds.$$

Now it can be rewritten as follows,

$$du(t_0+s, x+W_s) = \nabla u(t_0+s, x+W_s)dW_s - f(t_0+s, x+W_s)ds,$$

or, in the integral form,

$$u(T, x+W_{T-t_0}) = u(t_0, x) + \int_0^{T-t_0} \nabla u(t_0+s, x+W_s)dW_s$$
$$- \int_0^{T-t_0} f(t_0+s, x+W_s)ds.$$

Taking expectations from both sides we get,

$$u(t_0, x) = \mathrm{E}u(T, x+W_{T-t_0}) + \mathrm{E}\int_0^{T-t_0} f(t_0+s, x+W_s)ds$$
$$= \mathrm{E}g(x+W_{T-t_0}) + \mathrm{E}\int_0^{T-t_0} f(t_0+s, x+W_s)ds,$$

as required.

Remark 12. *Conditions of the example 9 may also be relaxed, as earlier, assuming derivatives only in the open cylinder* $((0,T)\times\mathbb{R}^d)$ *along with continuity of* u *only in the closed cylinder* $([0,T]\times\mathbb{R}^d)$*. Yet, it is not all that may be relaxed here.*

Some issue is that for heat equations with a non-zero r.h.s. it is not often that solutions are classical, that is, from $C_b^{1,2}$: more often all derivatives are in the Sobolev sense. Yet, is there a way to verify that solution $u \in C_b^{1,2}$ under any suitable conditions? In the case of the Laplacian there is a simple approach for such a verification. Let us take the function

$$v(t,x) = \mathrm{E}\left[\int_0^{T-t} f(t+s, x+W_s)ds + g(x+W_{T-t})\right] =: v^2(t,x) + v^1(t,x).$$

We do not know whether it is a classical solution of the heat equation with the r.h.s. f and a terminal condition g; but after the previous examples we guess that this is likely to be a solution. Can we check that this function belongs to one of the classes we considered in the previous Examples? In other words, the task is as follows: how can we differentiate the function v w.r.t. x and t? From the first sight this looks doubtful because the trajectories of W are only Hölder continuous but not differentiable.

Differentiability of v^1. If we write the expectation in the form (change of variables $z = y + x$)

$$v^1(t,x) = \mathrm{E} g(x + W_{T-t})$$
$$= \int g(x+y) \frac{1}{(2\pi(T-t))^{d/2}} \exp(-\frac{1}{2(T-t)} y^2) dy$$
$$= \int g(z) \frac{1}{(2\pi(T-t))^{d/2}} \exp(-\frac{1}{2(T-t)} (x-z)^2) dz,$$

it becomes clear that this expression, indeed, is differentiable both in x and in t. *It is a good exercise to check that* $v^1 \in C_b^{1,2}([0,T) \times \mathbb{R}^d)$ *if* $g \in C_b^2$ *(which follows from* $u \in C_b^{1,2}([0,T) \times \mathbb{R}^d)$*)*, and that the equation holds true, not using WP:

$$v_t^1 + \frac{1}{2} \Delta v^1 = 0, \quad \& \quad v^1(T,x) = g(x).$$

Differentiability of v^2. Assume f_x bounded, and for simplicity even that f_{xx} bounded. For the other term (we use Fubini theorem & change of variables $t + s = r$, so that $s = r - t$)

$$v^2(t,x) = \int_0^{T-t} \mathrm{E} f(t+s, x+W_s) ds,$$

we can also use the density of WP, hence, rewriting it as

$$v^2(t,x) = \int_t^T \mathrm{E} f(r, x + W_{r-t}) dr$$
$$= \int_t^T dr \int_{R^d} \frac{f(r, x+y)}{(2\pi(r-t))^{d/2}} \exp(-\frac{y^2}{2(r-t)}) dy$$
$$= \int_t^T dr \int_{R^d} \frac{f(r,z)}{(2\pi(r-t))^{d/2}} \exp(-\frac{1}{2(r-t)} (x-z)^2) dz.$$

It is a very recommended exercise to differentiate this expression in t and (twice) in x, and to check the corresponding equation.

In the next example c is a constant, but it may be made variable!

Example 10 (Homework). *Let $u(t,x) \in C_b^{1,2}([0,T] \times R^d)$ be a solution of the heat equation with a potential*

$$u_t(t,x) + \frac{1}{2}\Delta u(t,x) - cu(t,x) = -f(t,x), \quad 0 \le t \le T,$$
$$u(T,x) = g(x),$$

with $g \in C_b^2(R^d)$, $f(t,x) \in C_b([0,T] \times R^d)$. Then for any $0 \le t \le T$ the value $u(t,x)$ can be represented in the form

$$u(t,x) = \mathrm{E}\int_0^{T-t} e^{-cs} f(t+s, x+W_s)ds + \mathrm{E}e^{-c(T-t)}g(x+W_{T-t}).$$

7.2 Laplace equation

Let D be a bounded domain in $\mathbb{R}^d$ (by definition open one and connected; condition to be connected can be dropped, it is just for simplicity). Consider the Laplace equation

$$\frac{1}{2}\Delta u(x) = 0, \ x \in D, \quad \& \quad u(x)|_\Gamma = \varphi(x),$$

where $\Gamma = \partial D$ is a smooth boundary of D. Denote $D^c := \mathbb{R}^d \setminus D$. Let

$$\tau := \inf(t \ge 0 : x + W_t \in D^c).$$

Example 11. *Let $u(x) \in C_b^2(\bar{D})$ be a solution of the Laplace with $\varphi \in C(\bar{D})$. Then $u(x)$ can be represented as*

$$u(x) = \mathrm{E}\varphi(x + W_\tau), \quad x \in D.$$

Proof. Let us apply Ito's formula to $u(x + W_t)$:

$$du(x+W_t) = \nabla u(x+W_t)dW_t + \frac{1}{2}\Delta u(x+W_t)dt.$$

In the integral form we have (assuming $u \in C_b^2(\mathbb{R}^d)$),

$$u(x+W_t) - u(x) = \int_0^t \nabla u(x+W_s)dW_s + \frac{1}{2}\int_0^t \Delta u(x+W_s)ds,$$

but **it is not what we need**, because, remember, we do not know anything about u outside $\bar{D}$, or, at most, outside some its neighbourhood. So, we are to use stopping time τ.

It is also true that the left hand side (lhs) here equals the right hand side (rhs) if we integrate from 0 to $t \wedge \tau$ (for which $u \in C_b^2(\bar{D})$ suffices):

$$u(x + W_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} \nabla u(x + W_s) dW_s.$$

Take expectations:

$$\mathrm{E}u(x + W_{t\wedge\tau}) - u(x) = \mathrm{E}\int_0^{t\wedge\tau} \nabla u(x + W_s) dW_s = 0.$$

So,

$$u(x) = \mathrm{E}u(x + W_{t\wedge\tau}).$$

This is true for any $t > 0$. We want now to let $t \to \infty$.

Here is an (unexpected?) issue: why $P(\tau < \infty) = 1$, or even more precisely, why $\mathrm{E}\int_0^\infty 1(s < \tau)|\nabla u(x + W_s)|^2 ds < \infty$? If we knew that $\tau < \infty$ a.s., then in the limit we would have, clearly,

$$u(x) = \mathrm{E}u(x + W_\tau).$$

And since $x + W_\tau \in \Gamma$, and due to the continuity of u, and because of the boundary condition (called Dirichlet's b.c.), we would conclude that, indeed,

$$u(x) = \mathrm{E}\varphi(x + W_\tau),$$

as required. But why this property (or, equivalently, why $E\tau < \infty$) holds? To resolve this difficulty, recall that $x + W_t$ is a Markov and strong Markov process (the material of the earlier lectures: a strong advice is to repeat this stuff).

Note that

$$\inf_x \mathbb{P}_x(x + W_1 \notin D) > 0.$$

Lemma 12. *Let for an MP X_t and a domain D exist $q > 0$ such that*

$$\mathbb{P}_x(X_t \text{ exits from } D \text{ on } [0,1]) \geq q.$$

Then $\sup_x \mathrm{E}_x\tau < \infty$, where $\tau = \inf(t \geq 0 : X_t \notin D)$; in particular, $\mathbb{P}_x(\tau < \infty) = 1$. More than that, $\exists\, \alpha > 0$ such that

$$\sup_x \mathrm{E}_x \exp(\alpha\tau) < \infty.$$

Proof. All these claims follow from the inductive estimate

$$\mathbb{P}_x(\tau > n) = \mathrm{E}_x 1(\tau > n) = \mathrm{E}_x 1(\tau > n-1)\mathrm{E}_{X_{n-1}} 1(\tau > 1)$$

$$\le \mathrm{E}_x (1-q) 1(\tau > n-1) \le \ldots \le (1-q)^n. \qquad \text{QED}$$

Proof of Example 11, ctd. Now we can complete the proof of this Example. We have,

$$\mathbb{P}_x(x + W_1 \notin D) = \int_{R^d \setminus D} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}(y-x)^2) dy$$
$$= 1 - \int_D \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}(y-x)^2) dy \ge q \stackrel{\text{Homework!}}{>} 0.$$

Hence, by Lemma 12, $\mathbb{P}_x(\tau > n) \le (1-q)^n$ uniformly wrt x, and $\sup_x \mathrm{E}_x \tau < \infty$, as required. Therefore, the claim of the Example follows,

$$u(x) = \mathrm{E}\varphi(x + W_\tau).$$

7.3 Averaging property of WP ()

Averaging properties of WP $|B(r)| = d$-volume of $B(r)$, $|\Gamma(r)| = d-1$-area of the $(d-1)$-surface $\Gamma(r)$ Before we start a new topic, Poisson equations, let us discuss some interesting and useful corollary. Assume that $x \in D$, and let $B_x(r)$ denote the open ball of radius r with x the center. Suppose $r > 0$ is small enough, so that $B_x(r) \subset\subset D$. Denote also $\tau^r := \inf(t \ge 0 : x + W_t \notin B_x(r))$.

Corollary 7. *We have,*

$$u(x) = Eu(x + W_{\tau^r}) = \frac{1}{|\Gamma(r)|} \int_{\Gamma_x(r)} u(y) dy,$$

and also

$$u(x) = Eu(x + W_{\tau^r}) = \frac{1}{|B(r)|} \int_{B_x(r)} u(z) dz. \tag{77}$$

Proof of averaging properties We show the first one, because the second one follows by one more integration from it. Due to the strong Markov property, we have, denoting $X_t = x + W_t$,

$$\begin{aligned} u(x) = E\varphi(x+W_\tau) &= E(E\varphi(x+W_\tau)|\mathcal{F}_{\tau^r}) \\ = E(E\varphi(x+W_\tau)|x+W_{\tau^r}) &= E(E\varphi(X_\tau)|X_{\tau^r}) \\ = Eu(X_{\tau^r}) &= \frac{1}{|\Gamma(r)|}\int_{\Gamma_x(r)} u(y)dy, \end{aligned}$$

the last equality *by the symmetry of* W*:* for the WP starting from x, to hit any area on $\Gamma_x(r)$ at stopping time τ^r is proportional to the $d-1$-dimensional volume of this area.

Homework: Show that any Borel measurable bounded function satisfying (77) is continuous in x*, and any* $u \in C$ *is, in fact, in* C^2 *and satisfies the Laplace equation in* D*.*

7.4 Poisson equation in a domain

This is the name for a "Laplace equation with a non-trivial rhs". Let D be a bounded domain in $\mathbb{R}^d$. Consider the Poisson equation

$$\frac{1}{2}\Delta u(x) = -\psi(x),\ x \in D, \quad \& \quad u(x)|_\Gamma = \varphi(x),$$

where $\Gamma = \partial D$ is the boundary of D. Recall that $D^c := R^d \setminus D$, $\tau := \inf(t \geq 0 : x + W_t \in D^c)$.

Example 12. *Let* $u(x) \in C_b^2(\bar{D})$ *be a solution of the Poisson equation with* $\varphi \in C(\Gamma)$*,* $\psi \in C(\bar{D})$*. Then* $u(x)$ *in* D *can be represented as*

$$u(x) = \mathrm{E}[\int_0^\tau \psi(x+W_s)ds + \varphi(x+W_\tau)].$$

Proof. By Ito's formula,

$$du(x+W_t) = \nabla u(x+W_t)dW_t + \frac{1}{2}\Delta u(x+W_t)dt$$

$$= \nabla u(x+W_t)dW_t - \psi(x+W_t)dt.$$

So, in the integral form with a stopping time,

$$u(x+W_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} \nabla u(x+W_s)dW_s - \int_0^{t\wedge\tau} \psi(x+W_s)ds.$$

Taking expectations, we get

$$\mathrm{E}u(x+W_{t\wedge\tau})-u(x)=-\mathrm{E}\int_0^{t\wedge\tau}\psi(x+W_s)ds.$$

Recalling that $\sup_x \mathrm{E}_x\tau<\infty$ and letting $t\to\infty$, we have due to the continuity of u, W and the integral wrt t and by virtue of Lebesgue's dominated convergence theorem,

$$\mathrm{E}u(x+W_{\tau})-u(x)=-\mathrm{E}\int_0^{\tau}\psi(x+W_s)ds,$$

or, equivalently,

$$u(x)=\mathrm{E}\psi(x+W_{\tau})+\mathrm{E}\int_0^{\tau}\psi(x+W_s)ds,$$

as required.

Let D be, again, a bounded domain in $\mathbb{R}^d$. Consider the Poisson equation with a (variable) potential $0\le c(x)\in C(\bar D)$

$$\frac{1}{2}\Delta u(x)-c(x)u(x)=-\psi(x),\ x\in D,\quad \&\quad u(x)|_{\Gamma}=\varphi(x).$$

Denote $\kappa(t):=\int_0^t c(x+W_s)ds$. Recall that $D^c:=R^d\backslash D$, $\tau:=\inf(t\ge 0: x+W_t\in D^c)$.

Example 13 (Poisson equation with a potential $c(\cdot)$)**.** *Let $u(x)\in C_b^2(\bar D)$ be a solution of the Poisson equation with $\varphi\in C(\Gamma)$, $\psi\in C(\bar D)$. Then $u(x)$ in D can be represented as*

$$u(x)=E[\int_0^{\tau}e^{-\kappa(s)}\psi(x+W_s)ds+e^{-\kappa(\tau)}\varphi(x+W_{\tau})].$$

Proof. By Ito's formula,

$$de^{-\kappa(t)}u(x+W_t)=e^{-\kappa(t)}\nabla u(x+W_t)dW_t$$

$$+e^{-\kappa(t)}[\frac{1}{2}\Delta u(x+W_t)-c(x+W_t)u(x+W_t)]dt$$

$$=e^{-\kappa(t)}\nabla u(x+W_t)dW_t-e^{-\kappa(t)}\psi(x+W_t)dt.$$

So, in the integral form with a stopping time,

$$e^{-\kappa(t\wedge\tau)}u(x+W_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} e^{-\kappa(s)}\nabla u(x+W_s)dW_s - \int_0^{t\wedge\tau} e^{-\kappa(s)}\psi(x+W_s)ds.$$

Taking expectations, we get

$$\mathrm{E}e^{-\kappa(t\wedge\tau)}u(x+W_{t\wedge\tau}) - u(x) = -\mathrm{E}\int_0^{t\wedge\tau} e^{-\kappa(s)}\psi(x+W_s)ds.$$

From this equation by letting $t \to \infty$, we obtain, due to the continuity of all terms in t and because of $\sup_x \mathrm{E}\tau < \infty$, and by virtue of the Lebesgue dominated convergence theorem,

$$\mathrm{E}e^{-\kappa(\tau)}u(x+W_\tau) - u(x) = -\mathrm{E}\int_0^{\tau} e^{-\kappa(s)}\psi(x+W_s)ds,$$

or, equivalently,

$$u(x) = \mathrm{E}e^{-\kappa(\tau)}\varphi(x+W_\tau) + \mathrm{E}\int_0^{\tau} e^{-\kappa(s)}\psi(x+W_s)ds,$$

as required. Note that the condition $c \geq 0$ was essential.

Remark 13 (on the case $d = 1$)**.** *From the above it follows why it is useful to solve explicitly one-dimensional second order ODEs (also known as 1D "partial"elliptic differential equations of order 2, although, there is no really partial derivatives in this case; however, for 1D parabolic equations the name "partial"is genuine):*

$$u''(x) + b(x)u'(x) = 0,$$

$$u''(x) + b(x)u'(x) = -f(x),$$

$$u''(x) + b(x)u'(x) - c(x)u(x) = 0,$$

with various boundary conditions. We have already used and (likely) will use in what follows some explicit solutions for such 1D equations later in this course. Explicit formulae (for the "elliptic"1D case) can be found in many sourses, e.g., in [I.I. Gikhman, A.V. Skorokhod, Stochastic differential equations, Kiev, Naukova Dumka, 1968].

Extensions? From WP to SDE solutions; from Δ to elliptic operators of the 2nd order. After we establish Markov and strong Markov property of solutions of SDEs, we will be able to extend this analysis to more general operators of the second order, parabolic and elliptic.

It is not the goal of this short course, but a similar analysis may be extended also to (strong) Markov "diffusions with jumps which is the name used for solutions of SDEs driven by WP and Lévy processes. They correspond to integro-differential equations instead of PDEs; also there is a link to fractional Laplacians.

Also there are differential operators with other boundary conditions; this is a more difficult topic and not in the scope of this course.

7.5 Further examples

Example 8 Particular case of Example 5 Let D be a bounded domain (by definition open one and connected; condition to be connected can be dropped, it is just for simplicity) in R^d. Consider the Laplace equation

$$\frac{1}{2}\Delta u(x) = 0, \; x \in D, \quad \& \quad u|_\Gamma = \varphi(x), \tag{78}$$

where $\Gamma = \partial D$ and $\varphi(x) = 1(x \in A)$, $A \subset \Gamma$. Recall that $D^c := R^d \setminus D$. Let $\tau := \inf(t \geq 0 : x + W_t \in D^c)$.

Example 14 (8 (already proved!))**.** *1. Let $u(x) \in C_b^2(\bar{D})$ be a solution of the Laplace equation (78) with $\varphi \in C(\bar{D})$. Then $u(x)$ can be represented as*

$$u(x) = E\varphi(x + W_\tau) = P(x + W_\tau \in A), \quad x \in D.$$

2. **Vice versa***: the probability $P(x+W_\tau \in A)$ as a function of x satisfies the Laplace equation (78).*

Example 8, Comments Some obstacle may arise form the fact that often an indicator of a part of the boundary turns out to be discontinuous. Indeed, is it possible that the function $\varphi(x) = 1(x \in A) \in C$? And, if it is not, how can we apply the earlier results in such a case?

One option is to consider domains with a disconnected boundary as in the next example 9 of a ring on the plane: in R^2 let $D := \{x : 0 < a < |x| < b\}$, and, say, we are interested in computing the probability $P(|x+W_\tau| = a)$. Here $\varphi(x) = 1(|x| = a)$, while the complement of the boundary is the $(x : |x| = b)$. In this example the function φ is smooth on $\Gamma = \partial D$, not at all discontinuous.

Another option is to study solutions in Sobolev classes which allow discontinuities in boundary conditions. Not in this lecture.

General Theorem $\frac{1}{2}\Delta u(x) = 0,\ x \in D$ (2) *Homework: to prove! (Hint: use stopping times as in Example 2)*

Theorem 45 (variation of K.6.6.6)**.** *Let $u(x) \in C_b^2(D)$ and let $u \in C(\bar{D})$. Assume that $\mathrm{E}\int_0^\tau |\Delta u(x+W_s)|ds < \infty$. Then $u(x)$ can be represented as*

$$u(x) = \mathrm{E}u(x+W_\tau) - \frac{1}{2}\mathrm{E}\int_0^\tau \Delta u(x+W_s)ds, \quad x \in D.$$

Note, we already know that for a bounded domain D, the stopping time τ is a.s. finite, and even $\mathrm{E}\tau < \infty$, due to the Lemma from the previous lecture.

NB: In his lectures N.V. Krylov states Theorem 6.6.6 for a more general diffusion. Hence, he has to <u>assume</u> *that $\mathbb{P}(\tau < \infty) = 1$.*

Next three examples are taken directly from Krylov's textbook with some minor changes.

Example 15 (K.6.7.1)**.** *In $\mathbb{R}^d$ consider an open ball $B_R = \{x : |x| < R\}$ with some $R > 0$, and let*

$$\tau := \inf(t \geq 0 : W_t \notin B_R).$$

Consider the function

$$u(x) := \frac{1}{d}(R^2 - x^2).$$

Then

$$\mathrm{E}\tau = \frac{R^2}{d}.$$

We do not explain how function u was found: an educated guess.

Proof. By Ito's formula,

$$u(W_t) = -t - \int_0^t \frac{2}{d}W_s dW_s + \frac{R^2}{d}.$$

Replace t by $t \wedge \tau$ and take expectations:

$$\mathrm{E}u(W_{t\wedge\tau}) = -\mathrm{E}(t \wedge \tau) - \mathrm{E}\int_0^{t\wedge\tau} \frac{2}{d}W_s dW_s + \frac{R^2}{d},$$

or equivalently

$$\mathrm{E}(t\wedge\tau) = -\mathrm{E}u(W_{t\wedge\tau}) - \mathrm{E}\int_0^{t\wedge\tau}\frac{2}{d}W_s dW_s + \frac{R^2}{d}.$$

We have, $0 \le u(W_{t\wedge\tau}) \le \frac{R^2}{d}$. Since we already know (from the Lemma of the previous lecture) that $\mathrm{E}\tau < \infty$, and as $\mathrm{E}\int_0^t W_s^2 ds < \infty$, then we can pass to the limit here as $t \to \infty$:

$$\mathrm{E}\tau = \frac{R^2}{d} - \mathrm{E}u(W_\tau) = \frac{R^2}{d}.$$

Example 16 (∼K.6.7.1)**.** *Again $B_R = \{x : |x| < R\}$ in $\mathbb{R}^d$, and let*

$$\tau_x := \inf(t \ge 0 : x + W_t \notin B_R).$$

Consider the same function

$$u(x) := \frac{1}{d}(R^2 - x^2).$$

Then

$$\mathrm{E}\tau_x = u(x), \quad |x| \le R.$$

Proof is a **homework**!

Example 17 (K.6.7.2)**.** $D_{\epsilon,R}=\{x\colon \epsilon < |x| < R\}$, $\tau_{\epsilon,R} = \inf(t \ge 0\colon x_0 + W_t \notin D_{\epsilon,R})$,

$$\varphi(x) := \begin{cases} A(|x|^{-(d-2)} - R^{-(d-2)}), & if \quad d \ge 3, \\ A(\ln|x| - \ln R), & if \quad d = 2, \\ A(|x| - R), & if \quad d = 1, \end{cases}$$

$$A := \begin{cases} (\epsilon^{-(d-2)} - R^{-(d-2)})^{-1}, & if \quad d \ge 3, \\ (\ln\epsilon - \ln R)^{-1}, & if \quad d = 2, \\ (\epsilon - R)^{-1}, & if \quad d = 1, \end{cases}$$

Then $\Delta\varphi(x) = 0$ in $D_{\epsilon,R}$, $P_{\epsilon,R} = \mathrm{E}\varphi(x_0 + W_{\tau_{\epsilon,R}})$, and

$$P_\epsilon := \lim_{R\to\infty} P_{\epsilon,R} = \begin{cases} (\epsilon/|x_0|)^{(d-2)}, & d \ge 3, \\ 1, & d \le 2. \end{cases}$$

Homework: check yourself that $\Delta\varphi(x) = 0$ in $D_{\epsilon,R}$, & that the gradient $\nabla\varphi$ is bounded.

Proof of Example 17. Let us apply Ito's formula to $\varphi(x_0 + W_t)$ for $t < \tau_{\epsilon,R}$:

$$d\varphi(x_0 + W_t) = \nabla\varphi(x_0 + W_t)dW_t + \frac{1}{2}\Delta\varphi(x_0 + W_t)dt = \nabla\varphi(x_0 + W_t)dW_t.$$

Hence,

$$\varphi(x_0 + W_{t\wedge\tau_{\epsilon,R}}) = \varphi(x_0) + \int_0^{t\wedge\tau_{\epsilon,R}} \nabla\varphi(x_0 + W_s)dW_s.$$

Since $\nabla\varphi(x_0 + W_s)$ is bounded on $s < \tau_{\epsilon,R}$, we obtain,

$$\mathrm{E}\varphi(x_0 + W_{t\wedge\tau_{\epsilon,R}}) = \varphi(x_0).$$

Taking $t \to \infty$, and using that $\tau_{\epsilon,R} < \infty$ a.s. (by the Lemma), we get due to the Lebesgue bounded convergence theorem,

$$\varphi(x_0) = \mathrm{E}\varphi(x_0 + W_{\tau_{\epsilon,R}}).$$

But at $\tau_{\epsilon,R}$ by the choice of φ, the value of $\mathrm{E}\varphi(x_0 + W_{\tau_{\epsilon,R}})$ coincides with

$$\mathrm{E}\varphi(x_0 + W_{\tau_{\epsilon,R}}) = \mathbb{P}(|x_0 + W_{\tau_{\epsilon,R}}| = \epsilon).$$

Therefore,

$$\mathbb{P}(|x_0 + W_{\tau_{\epsilon,R}}| = \epsilon) = \varphi(x_0) = \begin{cases} A(|x_0|^{-(d-2)} - R^{-(d-2)}),\ d \geq 3, \\ A(\ln|x_0| - \ln R),\ d = 2, \\ A(|x_0| - R),\ d = 1. \end{cases}$$

Recall that

$$A := \begin{cases} (\epsilon^{-(d-2)} - R^{-(d-2)})^{-1}, & if \quad d \geq 3, \\ (\ln\epsilon - \ln R)^{-1}, & if \quad d = 2, \\ (\epsilon - R)^{-1}, & if \quad d = 1, \end{cases}$$

Hence, as $R \to \infty$, we obtain from

$$P(|x_0 + W_{\tau_{\epsilon,R}}| = \epsilon) = \begin{cases} A(|x_0|^{-(d-2)} - R^{-(d-2)}),\ d \geq 3, \\ A(\ln|x_0| - \ln R),\ d = 2, \\ A(|x_0| - R),\ d = 1, \end{cases}$$

that

$$P_\epsilon = \begin{cases} \epsilon^{(d-2)}|x_0|^{-(d-2)},\ d \geq 3, \\ 1,\ d = 2, \\ 1,\ d = 1. \end{cases}$$

Indeed,

$$\lim_{R\to\infty}\frac{(|x_0|^{-(d-2)}-R^{-(d-2)})}{(\epsilon^{-(d-2)}-R^{-(d-2)})}=\frac{|x_0|^{-(d-2)}}{\epsilon^{-(d-2)}}=\frac{\epsilon^{(d-2)}}{|x_0|^{(d-2)}};$$

$$\lim_{R\to\infty}\frac{(\ln|x_0|-\ln R)}{(\ln\epsilon-\ln R)}=1;$$

and

$$\lim_{R\to\infty}\frac{(|x_0|-R)}{(\epsilon-R)}=1,$$

as required.

In words, a 1D WP a.s. returns to any neighborhood – and, hence, to the origin as well – of $0\in\mathbb{R}^1$; a 2D WP a.s. returns to any neighborhood – but, actually, not to the origin! – of the origin in $\mathbb{R}^2$ (but not to the origin!); and any WP in dimension $d\geq 3$ returns to any neighbourhood of the origin in $\mathbb{R}^d$ only with a probability strictly less than one (it is, hence, for $d\geq 3$ it is probably transient[62]).

Example 18 (K.6.7.4)**.** *In $\mathbb{R}^d$ with $d\geq 3$,*

$$\lim_{t\to\infty}|W_t|=+\infty\quad a.s.$$

Proof.[63] Consider the function which slightly differs from φ from the previous Example,

$$f(x)=|x|^{2-d},\quad |x|>0.$$

Let $|x_0|>\epsilon>0$, and define the stopping time

$$\tau_\epsilon:=\inf(t\geq 0:|x_0+W_t|\leq\epsilon).$$

Let

$$X_t:=|x_0+W_{t\wedge\tau_\epsilon}|^{2-d}.$$

We have seen earlier that $\Delta f(x)=0$ at any $|x|>0$. Hence, the process X_t is a *bounded* martingale. (It is bounded because before τ_ϵ the modulus $|x_0+W_{t\wedge\tau_\epsilon}|$ is greater than ϵ, and this value to the negative $2-d$ is bounded by $1/\epsilon^{d-2}$.) By one of the limit theorems for non-negative (super)martingales, such a process has an a.s. limit as $t\to\infty$. Hence, we write,

$$|x_0|^{2-d}=\mathrm{E}X_0=\mathrm{E}X_t=\lim_{t\to\infty}\mathrm{E}X_t=\mathrm{E}\lim_{t\to\infty}X_t,$$

[62] It is, indeed!

[63] The proof is left for the independent study!

the last equality by Lebesgue's bounded convergence. Therefore,

$$|x_0|^{2-d} = \mathrm{E}\lim_{t\to\infty} X_t = \mathrm{E}\lim_{t\to\infty} |x_0 + W_{t\wedge\tau_\epsilon}|^{2-d}$$

$$= \mathrm{E}\lim_{t\to\infty} \frac{1}{|x_0 + W_{t\wedge\tau_\epsilon}|^{d-2}}$$

$$= \mathrm{E}\frac{1}{|x_0 + \lim_{t\to\infty} W_{t\wedge\tau_\epsilon}|^{d-2}} 1(\tau_\epsilon = \infty) + \epsilon^{2-d}\mathrm{E}1(\tau_\epsilon < \infty).$$

We highlight that from the existence of the limit for X_t it follows that there exists a limit a.s. for the term

$$|x_0 + W_{t\wedge\tau_\epsilon}|^{2-d}$$

on the set $(\tau_\epsilon = \infty)$; this set has a positive $\mathbb{P}$-measure (see the previous Example): $\mathbb{P}(\tau_\epsilon = \infty) = 1 - (\epsilon/|x_0|)^{d-2}$. We have (see the previous Example again)

$$\epsilon^{2-d}\mathrm{E}1(\tau_\epsilon < \infty) = \epsilon^{2-d}\mathbb{P}(\tau_\epsilon < \infty)$$

$$= \epsilon^{2-d}\epsilon^{(d-2)}|x_0|^{-(d-2)} = |x_0|^{2-d}.$$

So (we drop the "$\wedge\tau_\epsilon$" sign on the set $\tau_\epsilon = \infty$),

$$\mathrm{E}\frac{1}{|x_0 + \lim_{t\to\infty} W_t|^{2-d}} 1(\tau_\epsilon = \infty) = 0,$$

and this is for any $\epsilon > 0$. This means that for any $\epsilon > 0$

$$\lim_{t\to\infty} |W_t| = +\infty \quad \text{a.s. on the set } (\tau_\epsilon = \infty).$$

But the union of all these sets $(\tau_\epsilon = \infty)$ (which increase as ϵ decreases) equals $\bigcup_{\epsilon>0}(\tau_\epsilon = \infty) = \bigcup_{m\ge1}(\tau_{1/m} = \infty)$, which probability equals

$$\mathbb{P}(\bigcup_{m\ge1}(\tau_{1/m} = \infty)) = \lim_{m\to\infty} \mathbb{P}(\tau_{1/m} = \infty)$$

$$= 1 - \lim_{m\to\infty} \frac{1}{(m|x_0|)^{d-2}} = 1.$$

Therefore,

$$\lim_{t\to\infty} |W_t| = +\infty \quad \text{a.s.},$$

as required.

7.6 Laplace and reflection

Laplace operator and WP with reflection Let X_t be a solution of an SDE with reflection at 0, $dX_t = dW_t + dL_t$

Remark 14 (Ito's formula)**.**

$$dg(X_t) = g'(X_t)1(X_t \neq 0)dW_t + g'(X_t)1(X_t = 0)dL_t \\ + \frac{1}{2}g''(X_t)1(X_t \neq 0)dt.$$

Homework: Proof can be obtained via approximations of $|x|$.

NB: Ito calculus formal rules with local time:

$$\boxed{(dW_t)^2 = dt,\ (dt)^2 = dW_t dt = dW_t dL_t = dt dL_t = (dL_t)^2 = 0.}$$

Laplace operator and WP with reflection X_t is a solution of an SDE with reflection at 0, $dX_t = dW_t + dL_t$, $X_0 = x$

Corollary 8 (operator d^2/dx^2 and reflection)**.** *If $C^2 \ni g : [0, \infty) \mapsto R$ satisfies the equation*

$$1(x > 0)g''(x) = 0, \quad g'(0) = 0,$$

and has a moderate (polynomial) growth at ∞, then

$$Eg(X_t) = g(x).$$

Homework: Proof can be obtained via approximations of $|x|$ (and possibly a localisation).

Homework: Proof by taking expectation in the integral form.

7.7 Link to analytic functions

WP and Cauchy – Riemann's conditions Remark In view of the Cauchy – Riemann conditions in complex valued functions theory, Wiener process also closely relates to this theory. In my (personal) view, this link is not elaborated to the extent that it deserves.

7.8 General PDEs and SDEs: Parabolic equations

Let
$$L = \frac{a_{ij}(t,x)\partial^2}{2\partial x^i \partial x^j} + \frac{b^i(t,x)\partial}{\partial x^i},$$
where $a = \sigma\sigma^*$. Let us inspect the links between solution $X_t = X_t^{0,x_0}$ of the SDE (20) (recall it:)

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t,\ t \ge 0, \quad X_0 = x_0. \tag{20}$$

and parabolic PDEs.

Example 19. *Let $u(t,x) \in C_b^{1,2}([0,T]\times\mathbb{R}^d)$ be a solution of the heat equation*

$$\begin{aligned} u_t(t,x) + Lu(t,x) = 0, \quad 0 \le t \le T, \\ u(T,x) = g(x), \end{aligned}$$

with $g \in C_b^2(\mathbb{R}^d)$. Then for any $0 \le t \le T$ the value $u(t,x)$ can be represented in the form
$$u(t,x) = \mathrm{E} g(X_T^{t,x}) \equiv \mathrm{E}_x g(X_T^{t,x}).$$

Proof. Let us apply Ito's formula to $u(s, X_s^{t_0,x})$ for $0 \le t_0 \le s \le T$ (since $u(T,x) = g(x) \equiv \mathrm{E}g(X_T^{T,x})$):

$$du(s,X_s^{t_0,x}) = \sigma^*\nabla u(s,X_s^{t_0,x})dW_s + [u_s(s,X_s^{t_0,x}) + Lu(s,X_s^{t_0,x})]ds.$$

In the integral form with $t_0 + s = T$,

$$u(T,X_s^{t_0,x}) = u(t_0,x) + \int_{t_0}^T \sigma^*\nabla u(s,X_s^{t_0,x})dW_s + \int_{t_0}^T [u_s(s,x) + Lu(s,X_s^{t_0,x})]ds.$$

Let us now take expectations of both sides of this equality:

$$\mathrm{E}u(T,X_T^{t_0,x}) = u(t_0,x),$$

because

$$\mathrm{E}\int_{t_0}^T \sigma^*\nabla u(s,X_s^{t_0,x})dW_s = 0, \quad \& \quad [u_s(s,x) + Lu(s,X_s^{t_0,x})] = 0.$$

Remark 15. *The condition $g \in C_b^2(\mathbb{R}^d)$ follows automatically from $u(t,x) \in C_b^{1,2}([0,T]\times\mathbb{R}^d)$. Both of them can be relaxed.*

Exercise 18. *Repeat the last example under the relaxed condition that* $u(t,x) \in W^{1,2}_{d+1}([0,T] \times \mathbb{R}^d)$.

Example 20. *Let* $u(t,x) \in C^{1,2}_{b,loc}([0,T) \times \mathbb{R}^d) \bigcap C_b([0,T] \times \mathbb{R}^d)$ *be a solution of the heat equation*

$$u_t(t,x) + Lu(t,x) = 0, \quad 0 \le t \le T,$$
$$u(T,x) = g(x),$$

with $g \in C_b(\mathbb{R}^d)$. *Then for any* $0 \le t \le T$ *the value* $u(t,x)$ *can be represented in the form*

$$u(t,x) = \mathrm{E} g(X_T^{t,x}).$$

The conditions of boundedness of g and u with its derivatives may be further considerably relaxed, too.
Proof.
NB: while $a = \sigma\sigma^*$, we may **recover** the *symmetric positive-definite* square root of the matrix $a(x)$ via the Cauchy – Dunford formula (aka Cauchy – Riesz – Dunford formula, see textbooks[64]).
Note that the differential form of Ito's equation remains valid,

$$du(s, X_s^{t_0,x}) = \sigma^* \nabla u(s, X_s^{t_0,x}) dW_s + [u_s(s, X_s^{t_0,x}) + Lu(s, X_s^{t_0,x})] ds.$$

Yet, now we cannot simply integrate it to T, because the derivatives are assumed only on the semi-open interval $[0,T)$. Hence, let $t_0 \ge 0$, and denote $T_n := T - \frac{1}{n}$. Then, for n such that $t_0 < T_n$ we have,

$$u(T_n, X_{T_n}^{t_0,x}) = u(t_0,x) + \int_{t_0}^{T_n} \sigma^* \nabla u(s, x + W_s) dW_s + \int_{t_0}^{T_n} [u_s(s,x) + Lu(s, x + W_s)] ds.$$

Let us take expectations here: since

$$[u_s(s,x) + Lu(s, x + W_s)] = 0$$

and because

$$\mathrm{E} \int_{t_0}^{T_n} \sigma^* \nabla u(s, X_s^{t_0,x}) dW_s = 0,$$

[64] Assumed known; without proof; recommended to revise, or just to have a look; *not compulsory for the exam.*

we get

$$\mathrm{E}u(T_n, X_{T_n}^{t_0,x}) = u(t_0, x).$$

Equivalently,

$$u(t_0, x) = \mathrm{E}u(T_n, X_{T_n}^{t_0,x}).$$

Here we can pass to the limit as $T_n \uparrow T$ in the r.h.s.: since the function u is continuous and bounded up to T, and because X is continuous in time, we get by Lebesgue's bounded convergence theorem that again

$$u(t_0, x) = \mathrm{E}u(T, X_T^{t_0,x}) \equiv \mathrm{E}g(X_T^{t_0,x}),$$

as required. Recall that here $t_0 \geq 0$.

Example 21. *Let $u(t,x) \in C_b^{1,2}([0,T] \times R^d)$ be a solution of the heat equation*

$$\begin{aligned} u_t(t,x) + Lu(t,x) = -f(t,x), \quad 0 \leq t \leq T, \\ u(T,x) = g(x), \end{aligned}$$

with $g \in C_b^2(R^d)$, $f(t,x) \in C_b([0,T] \times R^d)$. Then for any $0 \leq t \leq T$ the value $u(t,x)$ can be represented in the form

$$u(t_0, x) = \mathrm{E}\left[\int_{t_0}^T f(s, X_s^{t_0,x})ds + g(X_T^{t_0,x})\right].$$

Proof. Recall Ito's formula,

$$du(s, X_s^{t_0,x}) = \sigma^* \nabla u(s, X_s^{t_0,x})dW_s + [u_s(s, X_s^{t_0,x}) + Lu(s, X_s^{t_0,x})]ds.$$

Now it can be rewritten as follows,

$$du(s, X_s^{t_0,x}) = \sigma^* \nabla u(s, X_s^{t_0,x})dW_s - f(s, X_s^{t_0,x})ds,$$

or, in the integral form,

$$u(T, X_T^{t_0,x}) = u(t_0, x) + \int_{t_0}^T \sigma^* \nabla u(s, X_s^{t_0,x})dW_s - \int_{t_0}^T f(s, X_s^{t_0,x})ds.$$

Taking expectations from both sides we get,

$$\begin{aligned} u(t_0, x) &= \mathrm{E}u(T, X_T^{t_0,x}) + \mathrm{E}\int_{t_0}^T f(s, X_s^{t_0,x})ds \\ &= \mathrm{E}g(X_T^{t_0,x}) + \mathrm{E}\int_{t_0}^T f(s, X_s^{t_0,x})ds, \end{aligned}$$

as required.

Remark 16. *Conditions of the example may also be relaxed, as earlier, assuming derivatives only in the semi-open cylinder* $([0,T) \times R^d)$ *along with continuity of* u *only in the closed cylinder* $([0,T] \times R^d)$. *Yet, it is not all that may be relaxed here.*

Remark 17. *For the heat equations with a non-zero r.h.s. it is not often that solutions are classical, that is, from* $C_b^{1,2}$. *Yet, how to verify that solution* $u \in C_b^{1,2}$, *or* $C_{b,loc}^{1,2}$, *if possible?*

$$\begin{aligned} u_t(t,x) + Lu(t,x) = -f(t,x), \quad 0 \le t \le T, \\ u(T,x) = g(x). \end{aligned}$$

In the general case there is no option to differentiate explicit formulae for solutions as for the classical heat equation, since we have no explicit formulae. However, there is another way, to use L_2 *(or* L_p*) directional derivatives of SDEs. In principle, this approach is available* if the coefficients have sufficiently many derivatives with respect to x. *We do not show the details here. Without additional derivatives of coefficients, probabilists are not aware how to show existence of derivatives of expressions like* $\mathrm{E}\left[\int_{t_0}^T f(s, X_s^{t_0,x})ds + g(X_T^{t_0,x})\right]$ *by purely probabilistic tools without PDE techniques* (excluding **the Malliavin calculus**); *PDE techniques provide more.*

Exercise 19. *In the next example c is a constant; it* **may** *be made variable. Show!*

Example 22. *Let* $u(t,x) \in C_b^{1,2}([0,T] \times R^d)$ *be a solution of the heat equation with a potential*

$$\begin{aligned} u_t(t,x) + Lu(t,x) - cu(t,x) = -f(t,x), \quad 0 \le t \le T, \\ u(T,x) = g(x), \end{aligned}$$

with $g \in C_b^2(\mathbb{R}^d)$, $f(t,x) \in C_b([0,T] \times \mathbb{R}^d)$. *Then for any* $0 \le t \le T$ *the value* $u(t,x)$ *can be represented in the form*

$$u(t,x) = \mathrm{E} \int_t^T e^{-c(s-t)} f(s, X_s^{t,x}) ds + \mathrm{E} e^{-c(T-t)} g(X_T^{t,x}).$$

Remark 18. *Normally and usually c is assumed to be non-negative. However, for the parabolic equations this is not important as far as c is a constant, or, in the case of a variable function, if* $c(\cdot)$ *remains bounded.*

7.9 SDEs and elliptic equations

In this section the coefficients of the SDE are homogeneous, i.e., they do not depend on time: $b(x)$ and $\sigma(x)$; as usual, $a(x) = \sigma\sigma^*(x)$. Let the matrix $a(x)$ be uniformly nondegenerate; let, as usual,

$$L = \frac{1}{2}\frac{a_{ij}(x)\partial^2}{\partial x^i \partial x^j} + \frac{b^i(x)\partial}{\partial x^i}.$$

Let D be a bounded domain (by definition open one and connected; condition to be connected can be dropped, it is just for simplicity) in $\mathbb{R}^d$. Consider the elliptic equation

$$Lu(x) = 0,\ x \in D, \quad \& \quad u|_\Gamma = \varphi(x),$$

where $\Gamma = \partial D$ is the boundary of D. Denote $D^c := R^d \setminus D$. Let

$$\tau := \inf(t \geq 0 : X_t^{0,x} \in D^c).$$

Example 23. *Let $u(x) \in C_b^2(\bar{D})$ be a solution of the elliptic equation above with $\varphi \in C(\bar{D})$, $a(x)$ uniformly nondegenerate. Then*

$$u(x) = \mathrm{E}\varphi(X_\tau^{0,x}), \quad x \in D.$$

Proof. Let us apply Ito's formula to $u(X_t)$:

$$du(X_t) = \sigma^* \nabla u(X_t) dW_t + Lu(X_t)dt.$$

In the integral form we have (assuming $u \in C_b^2(R^d)$),

$$u(X_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} \sigma^*\nabla u(X_s)dW_s + \int_0^{t\wedge\tau} Lu(X_s)ds.$$

We have to use stopping time τ here. It does follow from the nondegeneracy of $a(\cdot)$ that $\tau < \infty$ a.s. and, more than that, $\sup_{x\in D} E_x\tau < \infty$. (Recall that domain D is bounded.) However, this claim requires a separate proof, which is offered in the following lemma.

Lemma 13. *Let b and σ be bounded, $\sigma\sigma^*$ uniformly nondegenerate. Then*

$$\sup_{x\in D} \mathrm{E}_x\tau < \infty.$$

Proof [65] consists of three easy steps, but is a bit lengthy. As we know from the previous material (see lemma 12), for a Markov process it suffices to show that there exists $T > 0$ such that

$$\inf_{x\in D} \mathbb{P}_x(\exists t \in [0,T]\,\text{such that}\, X_t \not\in D) > 0.$$

I. Firstly, let us reduce the problem to the case with $b \equiv 0$. This can be done via Girsanov's measure transformation theorem. We will run the whole proof for $D = B_R$. Let

$$\tilde{W}_t = W_t + \int_0^t \tilde{b}(X_s)ds, \quad t \le T, \quad \tilde{b} = \sigma^{-1}b.$$

We have, with some (any) $T > 0$ and $\mathbb{P}^\rho(A) = \mathrm{E}\rho_T 1(A)$,

$$\rho_T = \exp(-\int_0^T \tilde{b}(X_s)dW_s - \frac{1}{2}\int_0^T \tilde{b}^2(X_s)ds),$$

due to the Cauchy – Buniakovskii – Schwarz inequality

$$\mathbb{P}_x(\sup_{t\le T}|X_t| > R) = \mathrm{E}^\rho_x \rho^{-1} 1(\sup_{t\le T}|X_t| > R)$$

$$\ge (\mathrm{E}^\rho_x\rho_T)^{-1}(\mathrm{E}^\rho_x 1(\sup_{t\le T}|X_t| > R))^2.$$

Here (since $+\frac{1}{2}\int_0^T \tilde{b}^2(X_s)ds \le -\frac{1}{2}\int_0^T \tilde{b}^2(X_s)ds + \|\tilde{b}\|^2 T$),

$$\sup_{x\in D}\mathrm{E}^\rho_x\rho_T = \sup_{x\in D}\mathrm{E}^\rho_x \exp(-\int_0^T \tilde{b}(X_s)d\tilde{W}_s + \frac{1}{2}\int_0^T \tilde{b}^2(X_s)ds) < \infty$$

So, to prove Lemma it remains to show that for *some* $T > 0$

$$\inf_{x\in D} \mathrm{E}^\rho_x 1(\sup_{t\le T}|X_t| > R) > 0.$$

Let $\sigma_t = (\sum_j \sigma^2_{1j}(X_t))^{1/2}$ Wanted: $\exists T > 0$ such that

$$\inf_{x\in B_R} \mathrm{E}^\rho_x 1(\sup_{t\le T}|X_t| > R) > 0.$$

[65] Reading this proof is left for your independent study.

Note that under the measure $\mathbb{P}^\rho$ the process X_t satisfies the equation **without a drift** removed by Girsanov, with a new WP $\tilde W$:

$$X_t = x + \int_0^t \sigma(X_s) d\tilde W_s, \; t \ge 0.$$

II. Now, consider the equation for one component of X_t, say, on X_t^1,

$$dX_t^1 = \sigma_{1j}(X_t) d\tilde W_t^j = (\sum_j \sigma_{1j}^2(X_t))^{1/2} d\bar W_t,$$

$$\text{where} \qquad \bar W_t := \int_0^t \frac{\sum_i \sigma_{1i}(X_t) d\tilde W_t^i}{(\sum_j \sigma_{1j}^2(X_t))^{1/2}}.$$

The equation on X_t^1 can be rewritten as

$$dX_t^1 = \sigma_t d\bar W_t, \qquad \sigma_t = (\sum_j \sigma_{1j}^2(X_t))^{1/2}.$$

(Of course, it is not a self-contained equation from which X^1 may be found independently of the other components. Yet, the formula holds true and will be useful.) Recall that a stochastic integral M_t which is a continuous square integrable martingale is, in fact, a Wiener process iff its compensator equals $\langle M\rangle_t = t$. In our case $\bar W_t$ is a continuous martingale (a stochastic integral), and

$$\langle \bar W\rangle_t = \sum_i \int_0^t \frac{\sigma_{1i}^2(X_s) ds}{(\sum_j \sigma_{1j}^2(X_s))} = t,$$

so, $\bar W_t$ is a one-dimensional Wiener process, as required. Moreover, the diffusion coefficient of X_t^1 is nondegenerate: with $\lambda^* = (1, 0, \ldots 0)$,

$$\sigma_t^2 = \sum_j \sigma_{1j}^2(X_t) = \lambda^* \sigma\sigma^*(X_t)\lambda \ge: c_1 > 0,$$

due to the assumption of the uniform nondegeneracy of $\sigma\sigma^*$. We will now verify that $\mathrm{E}\sup_{t\le T} |X_t^1|^2 \to \infty, t \to \infty$. Recall that we are to show that $\exists T > 0$ such that $\inf_{x\in B_R} \mathrm{E}_x^\rho 1(\sup_{t\le T} |X_t^1| > R) > 0$.

III. The last step: we show that for T large enough

$$\inf_{x\in D} \mathrm{E}_x^\rho 1(\sup_{t\le T} |X_t^1| > R) > 0.$$

We have,

$$\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^2 \ge \mathrm{E}^\rho |X_T^1|^2 = x^2 + \int_0^T \mathrm{E}^\rho \sigma_s^2 ds \ge c_1 T. \tag{79}$$

On the other hand,

$$\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^2 = \mathrm{E}^\rho \sup_{t\le T} |X_t^1|^2 1(\sup_{t\le T} |X_t^1|^2 \ge R^2)$$

$$+\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^2 1(\sup_{t\le T} |X_t^1|^2 < R^2)$$

$$\le \mathrm{E}^\rho \sup_{t\le T} |X_t^1|^2 1(\sup_{t\le T} |X_t^1|^2 \ge R^2) + R^2$$

$$\le (\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^4)^{1/2} (\mathbb{P}^\rho (\sup_{t\le T} |X_t^1|^2 \ge R^2))^{1/2} + R^2.$$

Note that if $Y_t = \int_0^t \sigma_s d\tilde{W}_s$, then $dY_t^4 = 4Y_t^3 \sigma_t d\tilde{W}_t + 6Y_t^2 dt$, and $\mathrm{E}^\rho Y_t^4 = 6\int_0^t \mathrm{E} Y_s^2 ds$, where $\mathrm{E}^\rho Y_s^2 \le c_2 s$.
We estimate $\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^4$ via Doob's inequality for continuous martingales ($\mathrm{E}^\rho \sup_{t\le T} |M_t|^p \le C(p) \mathrm{E}^\rho M_T^p$, $p > 1$):

$$\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^4 \le 2^3 x^4 + 24 C(4) (\int_0^T \mathrm{E}^\rho Y_s^2 ds) \le 2^3 (R^4 + 3C(4) c_2 T^2),$$

since $\mathrm{E}^\rho Y_t^4 = 6\int_0^t E^\rho Y_s^2 ds \le 3 c_2 t^2$. Thus, from

$$c_1 T \le (\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^4)^{1/2} (\mathbb{P}^\rho (\sup_{t\le T} |X_t^1|^2 \ge R^2))^{1/2} + R^2,$$

(the first inequality is shown in (79)) we find, for $T > R^2/c_1$,

$$(\mathbb{P}^\rho (\sup_{t\le T} |X_t^1|^2 \ge R^2))^{1/2} \ge \frac{c_1 T - R^2}{(\mathrm{E}^\rho \sup_{t\le T} |X_t^1|^4)^{1/2}}$$

$$\ge \frac{c_1 T - R^2}{(8(R^4 + 3C(4) c_2 T^2))^{1/2}} = c > 0, \qquad \text{as required.}$$

Lemma 13 is proved.

Proof of example 23, ctd. Now we can proceed in a standard way, as for WP and Laplace equation. Recall the equality

$$u(x + X_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} \sigma^*\nabla u(X_s)dW_s.$$

Let us take expectations:

$$\mathrm{E}u(X_{t\wedge\tau}) - u(x) = \mathrm{E}\int_0^{t\wedge\tau} \sigma^*\nabla u(X_s)dW_s = 0.$$

Since $u \in C_b^2(\bar{D})$, we obtain as $t \to \infty$,

$$u(x) = \mathrm{E}_x u(X_\tau) = \mathrm{E}\varphi(X_\tau^{0,x}),$$

as required.

Remark 19. *Similar representations can be established for* <u>*unbounded*</u> <u>*domains*</u>*, in particular, for the complement B_R^c of any ball B_R under the assumption that, due to certain conditions,*

$$\mathrm{E}_x\tau < \infty,$$

where

$$\tau := \inf(t \geq 0 : X_t \in B_R).$$

7.10 Poisson equation in a domain

Again, let $b \,\&\, \sigma$ do not depend on time, $a(x) = \sigma\sigma^*(x)$,

$$L = \frac{1}{2}\frac{a_{ij}(x)\partial^2}{\partial x^i \partial x^j} + \frac{b^i(x)\partial}{\partial x^i}$$

Let D be a bounded domain in $\mathbb{R}^d$. Consider the **Poisson equation**

$$Lu(x) = -\psi(x), \; x \in D, \quad \& \quad u(x)|_\Gamma = \varphi(x),$$

where $\Gamma = \partial D$ is the boundary of D. Recall that $D^c := \mathbb{R}^d \setminus D$, $\tau := \inf(t \geq 0 : X_t \in D^c)$.

Example 24. *Let $u(x) \in C_b^2(\bar{D})$ be a solution of the Poisson equation with $\varphi \in C(\Gamma)$, $\psi \in C(\bar{D})$. Then $u(x)$ in D can be represented as*

$$u(x) = \mathrm{E}_x[\int_0^\tau \psi(X_s)ds + \varphi(X_\tau)].$$

Proof. By Ito's formula, on the set $t < \tau$ we have,

$$du(X_t) = \sigma^* \nabla u(X_t) dW_t + Lu(X_t)dt = \sigma^* \nabla u(X_t) dW_t - \psi(X_t)dt.$$

So, in the integral form with a stopping time,

$$u(X_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} \sigma^* \nabla u(X_s) dW_s - \int_0^{t\wedge\tau} \psi(X_s)ds.$$

Taking expectations, we get

$$\mathrm{E}_x u(X_{t\wedge\tau}) - u(x) = -\mathrm{E}_x \int_0^{t\wedge\tau} \psi(X_s)ds.$$

Since $\sup_x \mathrm{E}_x \tau < \infty$ (see lemma 13) and letting $t \to \infty$, we have due to continuity of u, X and the integral wrt t and by virtue of Lebesgue's dominated convergence theorem,

$$\mathrm{E}_x u(X_\tau) - u(x) = -\mathrm{E}_x \int_0^{\tau} \psi(X_s)ds,$$

or, equivalently,

$$u(x) = \mathrm{E}_x \psi(X_\tau) + \mathrm{E}_x \int_0^{\tau} \psi(X_s)ds,$$

as required.

Similarly a Poisson equation with a potential may be considered. Let D be a bounded domain in R^d. Consider the Poisson equation with a (variable) potential $0 \le c(x) \in C(\bar{D})$

$$Lu(x) - c(x)u(x) = -\psi(x),\ x \in D, \quad \& \quad u(x)|_\Gamma = \varphi(x). \tag{80}$$

Denote $\kappa(t) := \displaystyle\int_0^t c(X_s)ds$. Recall that $D^c := \mathbb{R}^d \setminus D$, $\tau := \inf(t \ge 0 : X_t \in D^c)$.

Example 25. *Let $u(x) \in C_b^2(\bar{D})$ be a solution of the Poisson equation (80) with $\varphi \in C(\Gamma)$, $\psi \in C(\bar{D})$. Then $u(x)$ in D can be represented as*

$$u(x) = E_x[\int_0^{\tau} e^{-\kappa(s)} \psi(X_s)ds + e^{-\kappa(\tau)} \varphi(X_\tau)].$$

Proof. By Ito's formula,

$$de^{-\kappa(t)}u(X_t) = e^{-\kappa(t)}\sigma^*\nabla u(X_t)dW_t + e^{-\kappa(t)}[Lu(X_t) - c(X_t)u(X_t)]dt$$

$$= e^{-\kappa(t)}\sigma^*\nabla u(X_t)dW_t - e^{-\kappa(t)}\psi(X_t)dt.$$

So, in the integral form with a stopping time,

$$e^{-\kappa(t\wedge\tau)}u(X_{t\wedge\tau}) - u(x) = \int_0^{t\wedge\tau} e^{-\kappa(s)}\sigma^*\nabla u(X_s)dW_s - \int_0^{t\wedge\tau} e^{-\kappa(s)}\psi(X_s)ds.$$

Taking expectations, we get

$$\mathrm{E}_x e^{-\kappa(t\wedge\tau)}u(X_{t\wedge\tau}) - u(x) = -\mathrm{E}\int_0^{t\wedge\tau} e^{-\kappa(s)}\psi(X_s)ds.$$

Here by letting $t \to \infty$, we obtain due to continuity of all terms in t, because of $\sup_x \mathrm{E}\tau < \infty$, and by virtue of the Lebesgue dominated convergence theorem,

$$\mathrm{E}_x e^{-\kappa(\tau)}u(X_\tau) - u(x) = -\mathrm{E}\int_0^{\tau} e^{-\kappa(s)}\psi(X_s)ds,$$

or, equivalently,

$$u(x) = \mathrm{E}_x e^{-\kappa(\tau)}u(X_\tau) + \mathrm{E}_x\int_0^{\tau} e^{-\kappa(s)}\psi(X_s)ds,$$

as required. Note that the condition $c \geq 0$ was essential[66].

7.11 Again on Lévy's characterisation (not compulsory; recall)

Recall that in the lemma we dealt with the process X_t satisfying under the probability measure $\tilde{\mathbb{P}} = \mathbb{P}^\rho$ the equation

$$X_t = x + \int_0^t \sigma(X_s)d\tilde{W}_s,\ t \geq 0.$$

Then the equation on X_t^1 reads,

$$dX_t^1 = \sigma_{1j}(X_t)d\tilde{W}_t^j = \underbrace{(\sum_j \sigma_{1j}^2(X_t))^{1/2}}_{=:\sigma_t} d\bar{W}_t,$$

$$\text{where} \qquad \bar{W}_t := \int_0^t \frac{\sum_i \sigma_{1i}(X_t)d\tilde{W}_t^i}{(\sum_j \sigma_{1j}^2(X_t))^{1/2}}$$

[66] *Where about was it used?*

Then $\boxed{\bar{W}_t \text{ is a WP}}$ & the process X_t^1 satisfies the equation

$$dX_t^1 = \sigma_t d\bar{W}_t.$$

Clearly, $\bar{W}_t$ is a continuous martingale,

$$\bar{W}_t := \int_0^t \frac{\sum_i \sigma_{1i}(X_t) d\tilde{W}_t^i}{(\sum_j \sigma_{1j}^2(X_t))^{1/2}};$$

let

$$\psi_t := \exp(i\lambda(\bar{W}_t - \bar{W}_r)),\ t > r.$$

We want to show that $\bar{W}_t$ is, in fact, a Wiener process. Consider a conditional expectation (conditional characteristic function) for $r < t$,

$$\begin{aligned} \varphi(\lambda) &:= \mathrm{E}(\exp(i\lambda(\bar{W}_t - \bar{W}_r))|\mathcal{F}_r) \\ &= \mathrm{E}_{X_r} \exp(i\lambda \int_r^t \frac{\sum_i \sigma_{1i}(X_s) d\tilde{W}_s^i}{(\sum_j \sigma_{1j}^2(X_s))^{1/2}}). \end{aligned}$$

The latter equality is due to the Markov property of the process X. It suffices[67] to show $\varphi(\lambda) = \exp(-\lambda^2(t-r)/2)$.

Denote

$$f_s^i := \frac{\sigma_{1i}(X_s)}{(\sum_j \sigma_{1j}^2(X_s))^{1/2}};\ \text{note that}\ \sum_i (f_s^i)^2 = 1.$$

By Ito's formula we have,

$$\begin{aligned} d\psi_t &= d\exp(i\lambda(\bar{W}_t - \bar{W}_r)) = d\exp(i\lambda \int_r^t \sum_i f_s^i d\tilde{W}_s^i) \\ &= \exp(i\lambda(\bar{W}_t - \bar{W}_r))(i\lambda \sum_i f_t^i d\tilde{W}_t^i - \frac{\lambda^2}{2} \underbrace{\sum_i (f_t^i)^2}_{=1} dt). \end{aligned}$$

Hence,

$$\mathrm{E}_{X_r}\psi_t = 1 - \frac{\lambda^2}{2}\int_r^t \mathrm{E}_{X_r}\psi_s ds \implies \varphi(\lambda) = \mathrm{E}_{X_r}\psi_t = e^{-\lambda^2(t-r)/2}, \qquad \text{QED}$$

[67]It suffices only due to the Markov property, not for an arbitrary process, because in general we have to check the multi-dimensional characteristic functions, while for a Markov process it is enough to verify for the one-dimensional ones. Consider this as an *Exercise*, although, the whole subsection is not compulsory.

8 Lecture 8, 22.10.2021, Ergodic solutions of SDEs, FPK equations & Poisson equation "in the whole space"

8.1 Invariant measures

Consider the SDE with the coefficients which donot depend on t:

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \quad t \geq 0.$$

Under the usual assumptions we know that the solution X_t is a Markov process; its marginal measures are denoted by μ_t. It may occur that μ_t does not change with time; in this case it is called **an invariant measure**.

Lemma 14. *A probability measure μ is invariant for the MP X iff*

$$L^*\mu = 0 \qquad \sim \qquad \int Lg(x)\mu(dx) = \langle Lg, \mu\rangle \overset{\textit{formal pairing}}{=} \langle g, L^*\mu\rangle = 0$$

for any $g \in C_0^2$ (or for any $g \in C_b^2$).

Proof. Let $\mathcal{L}(X_0) = \mu$. We have by Ito's formula and due to Fubini's theorem,

$$\mathrm{E}_\mu g(X_t) = \mathrm{E}_\mu g(X_0) + \mathrm{E}_\mu \int_0^t Lg(X_s)ds$$

$$= \mathrm{E}_\mu g(X_0) + \int_0^t \mathrm{E}_\mu Lg(X_s)ds = \mathrm{E}_\mu g(X_0) + \int_0^t \left(\int Lg(y)\mu_s(dy)\right) ds.$$

I. Assume that $\mu_s \equiv \mu$ for all $s \geq 0$. Then it follows that

$$\int_0^t \left(\int Lg(y)\mu(dy)\right) ds = 0, \quad \forall\, t \geq 0.$$

This means that $\displaystyle\int Lg(y)\mu(dy) = 0$.

II. Let $\displaystyle\int Lg(y)\mu(dy) = 0$ and $\mu_0 = \mu$. It follows from Ito's formula that

$$\partial_t(\mathrm{E}_\mu g(X_t))|_{t=0} = \lim_{\Delta\downarrow 0}\frac{\mathrm{E}_\mu g(X_\Delta) - \mathrm{E}_\mu g(X_0)}{\Delta} = \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\,\mathrm{E}_\mu\int_0^\Delta Lg(X_s)ds$$

$$= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_0^\Delta \mathrm{E}_\mu Lg(X_s)ds = \mathrm{E}_\mu Lg(X_0) = \int Lg(y)\mu(dy) = 0. \tag{81}$$

Similarly, in general, without assuming $\displaystyle\int Lg(y)\mu(dy) = 0$, it holds

$$\partial_t(\mathrm{E}_\mu g(X_t))| = \lim_{\Delta\downarrow 0}\frac{\mathrm{E}_\mu g(X_{t+\Delta}) - \mathrm{E}_\mu g(X_t)}{\Delta} = \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\,\mathrm{E}_\mu\int_t^{t+\Delta} Lg(X_s)ds$$

$$= \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_t^{t+\Delta} \mathrm{E}_\mu Lg(X_s)ds = \mathrm{E}_\mu Lg(X_t) = \int Lg(y)\mu_t(dy) = \int (T_t Lg(x))\mu(dx).$$

Here $T_t h(x) := \mathrm{E}_x h(X_t)$; $T_{t+s} = T_t T_s$. This equation holds true not only for μ, but for any initial measure including delta-measure, that is,

$$\partial_t(\mathrm{E}_x g(X_t)) = \mathrm{E}_x Lg(X_t). \tag{82}$$

The equation (81) means that

$$\partial_t(\int Lg(y)\mu_t(dy))|_{t=0} = 0.$$

We have the identity

$$T_t L = L T_t$$

following formaly from the formula

$$T_t = \exp(Lt),$$

which shows that the operators L and T_t are commutative[68]. So, we have,

[68]Less formally, due to Lebesgue's dominated convergence theorem (e.g., assuming $T_t g(x) \in C^{1,2}$),

$$\partial_t(\mathrm{E}_\mu g(X_t)) = \partial_t(T_t g(X_t)) = \lim_{\Delta\downarrow 0}\frac{\int (T_{t+\Delta}g(x) - T_t g(x))\mu(dx)}{\Delta}$$

$$= \int \lim_{\Delta\downarrow 0}\frac{(T_\Delta T_t g(x) - T_t g(x))}{\Delta}\mu(dx) = \int (LT_t g(x))\mu(dx).$$

$$\partial_t(\int g(y)\mu_t(dy)) = \partial_t(\int T_t g(y)\mu(dy)) = \int \partial_t T_t g(y)\mu(dy)$$

$$= \int T_t L g(y)\mu(dy) = \int L T_t g(y)\mu(dy) = \langle L T_t g, \mu\rangle = 0.$$

Hence, $\mu_t \equiv \mu$, $t \ge 0$, as required. QED

Note that (82) may be accompanied by

$$\partial_t(\mathrm{E}_x g(X_t)) = \mathrm{E}_x L g(X_t). \quad (83)$$

8.2 Poisson equation in the whole space

Starting from the mid 70s of the last century, it was realised the importance of the **Poisson equation in the whole space, i.e., in $\mathbb{R}^d$**

$$Lu(x) = -f(x), \quad x \in \mathbb{R}^d. \quad (84)$$

There are *no boundary conditions* here. Why? How to solve this equation probabilistically? And where about this equation may be useful?

The answer to the latter question is that it is rather useful in establishing functional CLT like Donsker – Prokhorov invariance principle for more general processes than WP. Roughly speaking, it is helpful in studying CLT.

Recall that for the same equation in a domain D, say, with Dirichlet b.c. with $\psi \equiv 0$, the answer is well-known,

$$u(x) = \mathrm{E}_x \int_0^\tau f(X_t)dt,$$

where X_t solves a corresponding SDE with the generator L. How this formula could be extended from D to $\mathbb{R}^d$? Can we write down

$$u(x) = \mathrm{E}_x \int_0^\infty f(X_t)dt\,?$$

Firstly, let us highlight that this equation will be solved not for any generator L, but for those form a certain class (for "ergodic" ones), and **secondly,** not for any function f but only for so called "centered" ones. **Thirdly**, the above formula for $u(x)$ is **never correct**, because the integral diverges a.s. (unless $f \equiv 0$), while a correct way to write down the representation for the function u above is

$$u(x) = \int_0^\infty (\mathrm{E}_x f(X_t))\, dt.$$

(This is an example where Fubini theorem is not working. Yet, note that the integral wrt t is improper one; so, formally there is no logical contradiction here.)

As a minimum (because some more assumptions will be imposed in addition) it is assumed that there exists a unique invariant measure μ for the Markov process X_t (or for its generator L), that is,

$$L^*\mu = 0 \qquad \sim \qquad \int (Lg(x))\mu(dx) = 0$$

for any $g \in C_0^\infty$. If there is an invariant density $\psi(x) = \mu(dx)/dx$, it should satisfy the stationarity equation

$$L^*\psi(x) = \frac{1}{2}\mathrm{div}(a(x)\nabla\psi(x)) - \nabla(b(x)\psi) = 0.$$

The centering condition on the function f reads,

$$\int f(x)\mu(dx) = 0.$$

Also, f will be assumed bounded (which can be relaxed).

8.3 Ergodicity for SDE solutions I

Ornstein – Uhlenbeck process The first important example which may be studied without any additional tool (like the method of coupling) is the **Ornstein–Uhlenbeck process** (the OU process for what follows) X_t defined as a solution of the SDE

$$dX_t = -\theta\, X_t\, dt + \sigma\, dW_t, \quad X_0 = x, \tag{85}$$

with $\theta > 0$, $\sigma \neq 0$. Its *transition density*

$$f(t, x, x') = \frac{\mathbb{P}_x(X_t \in dx')}{dx'}$$

satisfies the backward FPK[69] equation

$$f_t(t,x,x') + L_x f(t,x,x') = 0, \quad f(0,x,x') = \delta(x'-x)$$

and the forward FPK equation

$$f_t(t,x,x') + L^*_{x'} f(t,x,x') = 0, \quad f(0,x,x') = \delta(x'-x).$$

The transition density of the OU process is known explicitly:

$$f(t-s,x;x') = \left(\frac{\theta}{\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{\theta(x'-x\exp(-\theta(t-s)))^2}{\sigma^2(1-\exp(-2\theta(t-s))}\right).$$

Exercise 20. *Show it!*
Hint: solve the equation (85).

At $t \to \infty$ we obtain the expression for the limiting and **invariant density**,

$$f(x') = \left(\frac{\theta}{\pi\sigma^2}\right)^{1/2} \exp\left(-\frac{\theta(x')^2}{\sigma^2}\right). \tag{86}$$

(There is a similar (a bit more involved) expression for the multi-dimensional version.)

Exercise 21.

- *Show that $f(x')$ is invariant.*
- *Using the representation (let $s=0$)*

$$\|\mu^x_t - \mu\|_{TV} = \int |f(t,x,x') - f(x')|dx'$$

estimate the rate of convergence of $\|\mu_t - \mu\|_{TV}$ and show that it is exponentially fast,

$$\|\mu^x_t - \mu\|_{TV} \le C(x)\exp(-\lambda t), \tag{87}$$

for some $C(x)$ and $\lambda > 0$, where x is the i.c. for X_t and μ^x_t is the marginal distribution of X_t started from the (nonrandom) i.c. x. (We may take $C(x) = C \times (1+x^2)$ where C is a constant.)

[69] The standard abbreviation for Fokker – Planck – Kolmogorov.

8.4 Ergodicity for SDE solutions II

Estimates similar to (87) may be obtained for a large class of SDEs without a knowledge of an explicit form of this solution or its density. We will show some easy part of this ergodic theory for SDEs satisfying exponential bounds like (87); however, for a wider class of SDEs there are weaker polynomial bounds, and even exponential bounds are valid for a larger class of equations than those considered in this lecture. Anyway, let us consider the SDE (20) (recall it) with homogeneous b and σ (i.e., not depending explicitly on t), that is,

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,\ t \ge 0, \quad X_0 = x \tag{88}$$

under certain additional assumptions on the coefficients b and σ. In the integral form it reads

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \quad t \ge 0.$$

The solution may be referred to as $X_t^{0,x}$, $t \ge 0$, which signifies that it starts from the state x at time 0; more generally, we allow the initial moment of time t_0, in which case the notation $X_t^{t_0,x}$ ($t \ge t_0$) will be used. To simplify the exposition assume $\sigma =$ const nondegerate matrix $d \times d$. The assumption on b is as follows: b is locally bounded, under the linear growth of its modulus, and there exist $M > 0$ and $r > 0$ such that[70]

$$\boxed{\langle b(x), x\rangle \le -r|x|^2, \qquad |x| \ge M,} \tag{89}$$

which may be equivalently presented as

$$\langle b(x), \frac{x}{|x|^2}\rangle \le -r, \qquad |x| \ge M.$$

For the OU process $r = \theta$ and the condition holds true for all values of x (that is, with $M = 0$).

Remark 20. *Note that the latter condition may be relaxed to*

$$\langle b(x), x\rangle \le -r|x|, \qquad |x| \ge M, \tag{90}$$

or, equivalently,

$$\langle b(x), \frac{x}{|x|}\rangle \le -r, \qquad |x| \ge M,$$

[70] This condition is put in the box so as to emphasize that this is the condition to be used later in the theorem; all other numbered conditions are its possible alternatives.

(in the case $d = 1$ the typical example is $b(x) = -sign(x)$) and still the rate of convergence will remain exponential. For slower polynomial rate the condition (89) may be replaced by

$$\langle b(x), \frac{x}{|x|^2}\rangle \le -r, \qquad |x| \ge M, \tag{91}$$

with $r > 0$ large enough[71]**.**

Theorem 46. *Under the assumption (89) any markovian solution of the equation (88) for any $\alpha > 0$ admits a bound*

$$\mathrm{E}_x \exp(\alpha\tau_R) \le C_1 x^2 + C_2, \tag{92}$$

for[72] *$R \gg M$, where*

$$\tau_R := \inf(t \ge 0 : |X_t| \le R),$$

with some $C_1, C_2 > 0$. Also, there exists a unique invariant measure μ, and

$$\|\mu_t^x - \mu\|_{TV} \le C(1 + x^2)\exp(-\lambda t) \tag{93}$$

with some $C, \lambda > 0$.
Also, this invariant measure integrates any power function, in particular,

$$\int x^2 \mu(dx) < \infty. \tag{94}$$

Proof. **1.** In this part of the proof we will only show a slightly weaker bound than (92); the second part of the claim will require a new technique called **the coupling method**, which will be introduced in what follows; and the last part that μ integrates any power function will be also shown a bit later: it follows from Khasminskii's presentation of the invariant measure and uniqueness (likely not to be presented in this course); one more way is based on comparison theorems, see below.

Consider the *Lyapunov function* x^2 and apply Ito's formula[73] to $\exp(\alpha t)X_t^2$ on the set $t < \tau = \tau_R$ (recall that σ in this proof does not depend on x):

[71] In (89) and in (90) the value r can be any positive
[72] That is, for R large enough, and at least $R \ge M$.
[73] For the first reading assume dimension $d = 1$.

$$de^{\alpha t}X_t^2 = \alpha e^{\alpha t}X_t^2dt + 2X_te^{\alpha t}dX_t + e^{\alpha t}(dX_t)^2$$

$$= \alpha e^{\alpha t}X_t^2dt + 2X_tb(X_t)e^{\alpha t}dt + 2X_t\sigma e^{\alpha t}dW_t + \mathrm{Tr}(\sigma\sigma^*)e^{\alpha t}dt.$$

Hence,

$$\begin{aligned} de^{\alpha t}X_t^2 - 2X_t\sigma e^{\alpha t}dW_t &= 2X_tb(X_t)e^{\alpha t}dt + (\alpha + \mathrm{Tr}(\sigma\sigma^*))e^{\alpha t}dt \\ &\le -2rX_t^2e^{\alpha t}dt + (\alpha + \mathrm{Tr}(\sigma\sigma^*))e^{\alpha t}dt. \end{aligned} \tag{95}$$

In the integral form we have,

$$e^{\alpha(t\wedge\tau)}X_{t\wedge\tau}^2 - 2\int_0^{t\wedge\tau} e^{\alpha s}X_s\sigma dW_s \le x^2 - 2r\int_0^{t\wedge\tau} e^{\alpha s}X_s^2ds + \int_0^{t\wedge\tau} e^{\alpha s}\mathrm{Tr}(\sigma\sigma^*)ds.$$

Let us take expectations: due to the under the linear growth condition on b it is known that

$$\int_0^t \mathrm{E}_x|X_s|^2ds < \infty, \qquad \forall\, t > 0.$$

So, $\mathrm{E}\displaystyle\int_0^{t\wedge\tau} e^{\alpha s}X_s\sigma dW_s = 0$, and we obtain

$$\mathrm{E}_xe^{\alpha(t\wedge\tau)}X_{t\wedge\tau}^2 \le x^2 - 2r\mathrm{E}_x\int_0^{t\wedge\tau} e^{\alpha s}(X_s^2 - \mathrm{Tr}(\sigma\sigma^*))ds.$$

Let us choose $R \ge M$ so that, in addition,

$$R^2 - (\alpha + \mathrm{Tr}(\sigma\sigma^*)) =: \beta > 0.$$

Since $[X_t| \ge R$ on the set $t < \tau$, we get

$$R^2\mathrm{E}_xe^{\alpha(t\wedge\tau)} \le \mathrm{E}_xe^{\alpha(t\wedge\tau)}X_{t\wedge\tau}^2 \le x^2 - 2r\mathrm{E}_x\int_0^{t\wedge\tau} e^{\alpha s}(R^2 - (\alpha + \mathrm{Tr}(\sigma\sigma^*)))ds$$

$$= x^2 - 2r\beta\mathrm{E}_x\int_0^{t\wedge\tau} e^{\alpha s}ds,$$

or, equivalently,

$$R^2 \mathrm{E}_x e^{\alpha(t\wedge\tau)} + 2r\beta \mathrm{E}_x \int_0^{t\wedge\tau} e^{\alpha s} ds \le x^2.$$

As $t \to \infty$, it follows by the monotone convergence theorem that

$$R^2 \mathrm{E}_x e^{\alpha\tau} + 2r\beta \mathrm{E}_x \int_0^{\tau} e^{\alpha s} ds \le x^2,$$

or,

$$R^2 \mathrm{E}_x e^{\alpha\tau} + \frac{2r\beta}{\alpha} \left(\mathrm{E}_x e^{\alpha\tau} - 1\right) \le x^2,$$

or, finally,

$$\left(R^2 + \frac{2r\beta}{\alpha}\right) \mathrm{E}_x e^{\alpha\tau} \le x^2 + \frac{2r\beta}{\alpha}. \qquad \text{QED}$$

2. (At the moment not compulsory!) At this step let us *suppose* that there exists an invariant measure for X to which μ_t^x converges. Let us show the bound (94) for it; note that uniqueness will follow from convergence bounds (to be proved later). We will use the hint based on comparison theorems. Namely, it follows from (95) with $\alpha = 0$ that

$$dX_t^2 - 2X_t \sigma dW_t = 2X_t b(X_t) e^{\alpha t} dt + \mathrm{Tr}(\sigma\sigma^*) dt \le \left(-2rX_t^2 + \mathrm{Tr}(\sigma\sigma^*)\right) dt. \qquad (96)$$

Case $d = 1$.

It follows from the Yamada – Watanabe approximation of the modulus in their theorems that the process $|X_t|$ possesses the following stochastic differential in the case $d = 1$

$$d|X_t| = \mathrm{sign}(X_t)\sigma dW_t + \mathrm{sign}(X_t) b(X_t) dt + dL_t, \qquad (97)$$

where L_t is the local time of $|X_t|$ at zero. Let now

$$M_x := M \vee x,$$

where $x = |X_0|$, and consider now the process on $[M_x, \infty)$ with the non-sticky reflection at M_x satisfying the SDE with reflection

$$
\begin{aligned}
dY_t &= \sigma dW_t + \bar b(Y_t)dt + dL_t^{Y,M}, \quad Y_t \ge M_x, \quad Y_0 \ge M_x, \\
dL_t^{Y,M_x} &= 1(Y_t = M_x)dL_t^{Y,M_x}, \quad \mathbb{P}(\int_0^t 1(Y_s = M_x)ds = 0) = 1 \; \forall\, t \ge 0.
\end{aligned}
\tag{98}
$$

Here

$$
\bar b(y) = -r < 0.
$$

We may take any initial distribution for Y_0 The strong solution Y_t exists and is pathwise unique[74]. The process Y_t on $\mathbb{R}^1$ is ergodic and possesses a unique invariant measure, which can be found from the stationarity equation[75] ($L^*f(y) = 0$, $y > M_x$)

$$
\frac{\sigma^2}{2} f''(y) + r f'(y) = 0, \quad y > M_x, \tag{99}
$$

which has a solution

$$
f'(y) = C_1 \exp(-(2r/\sigma^2)(y - M_x)), \quad y \ge M_x,
$$

$$
f(z) = C_1 \int_{M_x}^z \exp(-(2r/\sigma^2)(y - M_x))dy + C_2 = \frac{C_1\sigma^2}{2r}(-1 + \exp(-(2r/\sigma^2)(z - M_x)) + C_2.
$$

Here the additive constant $C_2 - \frac{C_1\sigma^2}{2r}$ may not be positive (otherwise $\liminf f > 0$ at infinity, then f is not integrable) and may not not be negative (otherwise $\liminf f < 0$ at infinity); hence,

$$
C_2 - \frac{C_1\sigma^2}{2r} = 0 \quad \Longrightarrow \quad f(z) = C\exp(-(2r/\sigma^2)(z - M_x)), \;\; z \ge M_x,
$$

and from the normalizing condition $\int f dz = 1$ we finally find that

$$
f(z) = \frac{\sigma^2}{2r}\exp(-(2r/\sigma^2)(z - M_x)), \quad z \ge M_x.
$$

[74]We have proved earlier corresponding results under slightly different conditions; in our case they also hold true.

[75]Indeed, this equation is $L^*f(y) = 0$ just for all $y > 0$ without any boundary condition because it suffices to check the equality $\int g(y)\mu_t(dy) = \int g(y)\mu(dy)$ only for all $g \in C_0$ with $g(M_x) = 0$.

Note that this measure $f(z)dz$ integrates on $[M_x, \infty)$ any power function and even some exponents; in any case,

$$\int_{M_x}^{\infty} z^2 f(z)dz < \infty.$$

From the comparison arguments[76] it follows that

$$\mathbb{P}(Y_t \geq |X_t|, \ t \geq 0) = 1.$$

From here it follows that **if** there exists an invariant measure for X_t to which μ_t^x converges, then it must satisfy

$$\int x^2 \mu(dx) \leq \int_{M_x}^{\infty} z^2 f(z)dz < \infty.$$

Case $d > 1$ may be considered in a similar way, by applying Ito's formula to $|X_t|$; just the formula (97) should be re-written properly.

3. *Uniqueness of the invariant measure for the Markov process X may be derived by using the coupling method. We postpone it till the next lecture.* QED

8.5 FPK equations II

Recall that we use the notation

$$T_t h(x) := \mathrm{E}_x h(X_t),$$

for $h \in C_b^2$ (or $h \in C_b$, or even $h \in B$ – from the class of bounded Borel measurable functions) where X_t is the solution of the equation (88). Any assumptions which guarantee the existence and pathwise uniqueness are assumed. In addition it may be taken from the C_b^2 class, although, it can be relaxed. The following result is taken from the PDE theory; the proof will be provided under additional assumptions which are designed to make this presentation a bit more self-contained.

Theorem 47 (FPK equation)**.** *If b and σ are bounded, $\sigma\sigma^*$ uniformly non-degenerate and Hölder in x uniformly wrt t, then for any $h \in C_b^2$*

$$\partial_t T_t h(x) \stackrel{FPK}{=} L T_t h(x), \qquad \forall\, x \in \mathbb{R}^d, \ t > 0. \tag{100}$$

If we denote $v(t,x) := T_t h(x)$, then the solution of the equation[77] *(100) is unique in*

[76]To make the conditions closer to what was done for comparison theorems earlier, we may consider the SDE for Y_t with the drift $-r + \epsilon < 0$; it also suffices for a small exponential moment, not even talking of the second moment.

[77]$v_t = Lv$, and $v(0,x) = h(x)$.

the class $C^{1,2}_{moderate}$ (and in $W^{1,2}_p$ with any $p \ge d+1$).

The formal proof of (100) follows from (82) in the subsection 8.1. More rigorous arguments (with a strict proof of existence of all derivatives involved) will require some auxiliary results, see what follows.

Proof of uniqueness. Suppose there are two solutions $v^1(t,x)$ and $v^2(t,x)$. Denote $w(t,x) = v^1(t,x) - v^2(t,x)$. Then due to the linearity the equation holds true for w, too:

$$\partial_t w(t,x) = Lw(t,x), \qquad \forall\, x \in \mathbb{R}^d,\, t > 0,$$

and

$$w(0,x) = 0.$$

Then by Ito's formula (after taking expectations) applied to $w(s, X_{t-s})$ we get

$$w(t,x) = \underbrace{\mathrm{E}_x w(0,X_t)}_{=0} + \mathrm{E}_x \int_0^t (w_s(s,X_s) - Lw(s,X_s))ds = 0.$$

Hence, $w(t,x) \equiv 0$ and $v^1 = v^2$, as required.

For the proof of the equation (100) we will assume in addition that $b, \sigma \in C^2_b$. Before the proof let us state one auxiliary result.

Lemma 15. *Let b and σ be from the class C^2_b, $k = 1, 2, \dots$ Then the solution $X^{0,x}_t$ is twice differentiable wrt i.c. x* in any set of directions[78] *(ℓ_1, ℓ_2) in the sense*[79] *of L_2, that is, by induction, for any $j \le k$ there exist the processes $Y^{(1),0,x,\ell_1}_t$ such that*

$$\mathrm{E}\left|\frac{X^{0,x+\ell_1\Delta}_t - X^{0,x}_t}{\Delta} - Y^{(1),0,x,\ell_1}_t\right|^2 \to 0, \quad \Delta \downarrow 0, \tag{101}$$

and

$$\mathrm{E}\left|\frac{Y^{(1),0,x+\ell_2\Delta,\ell_1}_t - Y^{(1),0,x,\ell_1}_t}{\Delta} - Y^{(2),0,x,\ell_1,\ell_2}_t\right|^2 \to 0, \quad \Delta \downarrow 0; \tag{102}$$

let $Y^{(0),x}_t := X^{0,x}_t$. The equation on all $Y^{(j)}_t$ are obtained by the formal differentiation over the i.c. x of the equation (88): for $j = 1$ it reads

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, \quad t \ge 0.$$

[78] That is, each ℓ_i is a unit vector in some direction. We will need $k = 2$ in the theorem.

[79] Note that this claim is not about classical derivatives, although, the latter also exist but up to the order $k - 1$.

$$Y_t^{(1),0,x,\ell} = \ell + \int_0^t \nabla_x b(X_s) Y_s^{(1),0,x,\ell} ds + \int_0^t \nabla_x \sigma(X_s) Y_s^{(1),0,x,\ell} dW_s, \tag{103}$$

and for $j = 2$

$$\begin{aligned} Y_t^{(2),0,x,\ell_1,\ell_2} &= \int_0^t \nabla_x b(X_s) Y_s^{(2),0,x,\ell_1,\ell_2} ds + \int_0^t \nabla_x^2 b(X_s) Y_s^{(1),0,x,\ell_1} Y_s^{(1),0,x,\ell_2} ds \\ &+ \int_0^t \nabla_x \sigma(X_s) Y_s^{(2),0,x,\ell_1,\ell_2} dW_s + \int_0^t \nabla_x^2 \sigma(X_s) Y_s^{(1),0,x,\ell_1} Y_s^{(1),0,x,\ell_2} dW_s \end{aligned} \tag{104}$$

Strong solutions of the equations (103) and (104) exist and are pathwise unique.

Proof (just[80] *for $j = 1$, and for the first reading assume the dimension $d = 1$.* Existence of solutions and their pathwise uniqueness follows easily because of the linear dependence of the corresponding coefficients with respect to the solutions. From the equation we have,

$$\begin{aligned} X_t^{0,x+\Delta\ell} - X_t^{0,x} = x + \Delta\ell + \int_0^t b(X_s^{0,x+\Delta\ell}) ds + \int_0^t \sigma(X_s^{0,x+\Delta\ell}) dW_s \\ -x - \int_0^t b(X_s^{0,x}) ds + \int_0^t \sigma(X_s^{0,x}) dW_s \end{aligned}$$

$$= \Delta\ell + \int_0^t \left(b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x})\right) ds + \int_0^t \left(\sigma(X_s^{0,x+\Delta\ell}) - \sigma(X_s^{0,x})\right) dW_s.$$

Denoting

$$\frac{X_s^{0,x+\Delta\ell} - X_s^{0,x}}{\Delta} =: Z_s^{\Delta,\ell},$$

we write

$$Z_t^{\Delta,\ell} = \ell + \int_0^t \frac{\left(b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x})\right)}{\Delta} ds + \int_0^t \frac{\left(\sigma(X_s^{0,x+\Delta\ell}) - \sigma(X_s^{0,x})\right)}{\Delta} dW_s.$$

[80] Naturally, for $b, \sigma \in C_b^1$ there only exists a first L_2-derivative, and under the condition $b, \sigma \in C_b^k$ there are mixed L_2-derivatives for any multi-index of the order k.

From Lipschitz conditions on b and σ we find that $\mathrm{E}_x \sup_{s\le t} |Z_s|^2 < \infty$. So,

$$Z^{\Delta,\ell} - Y_t^{(1),0,x,\ell} = \ell + \int_0^t \frac{\left(b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x})\right)}{\Delta} ds + \int_0^t \frac{\left(\sigma(X_s^{0,x+\Delta\ell}) - \sigma(X_s^{0,x})\right)}{\Delta} dW_s$$

$$-\ell + \int_0^t \nabla_x b(X_s) Y_s^{(1),0,x,\ell} ds + \int_0^t \nabla_x \sigma(X_s) Y_s^{(1),0,x,\ell} dW_s$$

$$= \int_0^t \left(\frac{\left(b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x})\right)}{\Delta} - \nabla_x b(X_s) Y_s^{(1),0,x,\ell} \right) ds$$

$$+ \int_0^t \left(\frac{\left(\sigma(X_s^{0,x+\Delta\ell}) - \sigma(X_s^{0,x})\right)}{\Delta} - \nabla_x \sigma(X_s) Y_s^{(1),0,x,\ell} \right) dW_s.$$

We have, by the first theorem of the calculus (aka Newton – Leibniz' formula)

$$b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x}) = \int_0^1 \nabla_x b(X_s^{0,x} + \alpha Z_s^{\Delta,\ell}\Delta) Z_s^{\Delta,\ell} \Delta d\alpha,$$

and a similar representation holds for the difference of sigmas,

$$\sigma(X_s^{0,x+\Delta\ell}) - \sigma(X_s^{0,x}) = \int_0^1 \nabla_x \sigma(X_s^{0,x} + \alpha Z_s^{\Delta,\ell}\Delta) Z_s^{\Delta,\ell} \Delta d\alpha.$$

So,

$$\int_0^t \left(\frac{\left(b(X_s^{0,x+\Delta\ell}) - b(X_s^{0,x})\right)}{\Delta} - \nabla_x b(X_s) Y_s^{(1),0,x,\ell} \right) ds$$

$$= \int_0^t \left(\int_0^1 \nabla_x b(X_s^{0,x} + \alpha Z_s^{\Delta,\ell}\Delta) Z_s^{\Delta,\ell} d\alpha - \nabla_x b(X_s) Y_s^{(1),0,x,\ell} \right) ds$$

$$= \int_0^t \int_0^1 \left(\nabla_x b(X_s^{0,x} + \alpha Z_s^{\Delta,\ell}\Delta) Z_s^{\Delta,\ell} - \nabla_x b(X_s) Y_s^{(1),0,x,\ell} \right) d\alpha ds$$

$$= \underbrace{\int_0^t \int_0^1 \left(\nabla_x b(X_s^{0,x} + \alpha Z_s^{\Delta,\ell}\Delta) - \nabla_x b(X_s) \right) Z_s^{\Delta,\ell} d\alpha ds}_{=o(1),\ \text{due to}\ \Delta \to 0}$$

$$+ \int_0^t \nabla_x b(X_s) \left(Z_s^{\Delta,\ell} - Y_s^{(1),0,x,\ell} \right) ds.$$

Similarly with σ and stochastic integrals. Hence, slightly abusing notations we may write

$$Z_t^{\Delta,\ell} - Y_t^{(1),0,x,\ell} = \int_0^t \nabla_x b(X_s) \left(Z_s^{\Delta,\ell} - Y_s^{(1),0,x,\ell} \right) ds$$

$$+ \int_0^t \nabla_x \sigma(X_s) \left(Z_s^{\Delta,\ell} - Y_s^{(1),0,x,\ell} \right) dW_s + o(1).$$

Using Ito's isometry and Doob's inequality we estimate (again slightly abusing notations)

$$\mathrm{E}_x \sup_{s\le t} |Z^{\Delta,\ell}_s - Y^{(1),0,x,\ell}_s|^2 \le o(1) + C_t\int_0^t \mathrm{E}_x\|\nabla_x b(X_s)\|^2\, |Z^{\Delta,\ell}_s - Y^{(1),0,x,\ell}_s|^2 ds$$

$$+4\int_0^t E_x\|\nabla_x \sigma(X_s)\|^2\, |Z^{\Delta,\ell}_s - Y^{(1),0,x,\ell}_s|^2 ds$$

$$\le o(1) + C_t\int_0^t \mathrm{E}_x|Z^{\Delta,\ell}_s - Y^{(1),0,x,\ell}_s|^2 ds.$$

Here $o(1)\to 0$, $\Delta\to 0$. Hence, by virtue of Grönwall's inequality we obtain

$$\mathrm{E}_x \sup_{s\le t} |Z^{\Delta,\ell}_s - Y^{(1),0,x,\ell}_s|^2 \to 0, \quad \Delta \to 0,$$

as required. The claim about the second derivative follows from a similar calculus. The lemma is proved. QED

Corollary 9. *Under the assumptions*[81] *of the lemma 15 the function* $v(x) = \mathrm{E}_x g(X_t)$ *is differentiable wrt* x *for any* $g \in C^1_b$.

Proof. By the rule of the differentiation of the composite functions, we have (formally)

$$\partial_x v(x) = \partial_x v(x) = \partial_x \mathrm{E} g(X^x_t) = \mathrm{E} g'(X^x_t)\partial_x X^x_t$$

$$= \mathrm{E} g'(X^x_t) Y^{(1),x}_t,$$

where $Y^{(1),x}_t = (Y^{(1),x,\ell_1}_t, \ldots, Y^{(1),x,\ell_d}_t)$.
NB: the property that $\mathrm{E} g'(X^x_t)Y^{(1),x}_t$ is, indeed, a derivative of $v(x)$ should be, of course, rigorously proved.

A rigorous proof of the equation (100) of the theorem 47 may now be completed with the help of the lemma 15. Indeed, now we are able to differentiate the expressions in the probabilistic representations *(NB: not included τ!)* with respect to x just by the rule of the derivative of a composite function. Derivatives of these expressions with respect to t also exist as we know that they satisfy the FPK equations.

[81] In fact, PDE tools provide deeper results: even for $g \in B$ (bounded and Borel measurable) under the additional condition of the non-degeneracy and just of the uniform continuity of $\sigma\sigma^*$, but without the requirement of $b, \sigma \in C^1_b$, for any $t > 0$ the function $v(x)$ has a Sobolev derivative. This lemma (and the corollary from it) serve the only goal to make the presentation more self-contained.

8.6 Poisson equation in the whole space II

In this subsection we will show that the Poisson equation (84) should have a probabilistic solution, assuming that the theorem 46 is completely proved including the bound (93), and assuming one additional condition on the function f in the right hand side of (84). The term "should have" was used because the solutions of the Poisson equation in he bounded domains was established modulus certain PDE results (namely, that a solution with classical, or, at least, with Sobolev derivatives exists and that we can apply Ito's formula to it after a substitution of X_t in this solution).

Assumption 1 (centering condition on f)**.** *The function f must satisfy the condition*

$$\int f(x)\mu^{inv}(dy) = 0, \tag{105}$$

where $\mu^{inv}(dy)$ is the invariant probability measure[82] *for the Markov process X.*

The meaning of this assumption is similar to the subtraction of the expectation in the Central Limit Theorem.

Theorem 48. *Let f be bounded*[83]*. Under the assumptions (89) and (105) the solution of the Poisson equation (84) ($Lu = -f$) exists, satisfies the bound*

$$|u(x)| \le C(1 + x^2), \tag{106}$$

the solution is unique up to any additive constant C in the class of functions growing no faster than some power function ($\le C(1 + x^2)$), and may be presented by the formula

$$u(x) = \int_0^\infty \mathrm{E}_x f(X_t)\, dt + C. \tag{107}$$

For this formula with $C = 0$

$$\int u(x)\mu^{inv}(dx) = 0,$$

i.e., $u(x)$ correspondng to $C = 0$ is centered itself.

---

[82] In other words, we implicitly assume existence (and uniqueness) of this invariant measure. This existence and uniqueness will be proved in what follows *if time allows!* Note that for the OU process this claim is known because of the explicit form of this invariant measure, see (86).

[83] May be considerably relaxed.

Proof.
0. The very first thing is to make sure that the integral in (107) converges, and we start with the special case of the OU-process possesing the bound (87). We estimate

$$\int_0^\infty |\mathrm{E}_x f(X_t)|\, dt = \int_0^\infty |\int f(y)\mu_t(dy)|\, dt$$

$$= \int_0^\infty |\int f(y)\mu_t(dy) - \int f(y)\mu(dy)|\, dt$$

$$\le \|f\|_B \int_0^\infty \|\mu_t(dy) - \mu(dy)\|_{TV}\, dt < \infty,$$

because of the bound (87). Moreover, for the i.c. x we have a bound

$$\int_0^\infty |\mathrm{E}_x f(X_t)|\, dt \le \|f\|_B \int_0^\infty \|\mu_t^x(dy) - f(y)\mu(dy)\|_{TV}\, dt$$

$$\le \|f\|_B \int_0^\infty C(1+x^2)\exp(-\lambda t)\, dt = \|f\|_B C(1+x^2)\lambda^{-1} < \infty.$$

1. The convergence of the integral in (107) under the general assumptions of the theorem **follows from the similar calculus with (93)** in the theorem 46.
Note that the bound (106) (i.e., $|u(x)| \le C(1+x^2)$) holds true for the function u.
2. Let us now show that u is, indeed, a solution of the equation (107). We have formally[84],

$$Lu(x) \stackrel{(!?)}{=} \int_0^\infty LT_s f(x) ds = \lim_{t\to\infty} \int_0^t LT_s f(x) ds$$

$$= \lim_{t\to\infty} \int_0^t \partial_s T_s f(x) ds = \lim_{t\to\infty} (T_t f(x) - T_0 f(x))$$

$$= \lim_{t\to\infty} \int f(y)\mu_t(dy) - f(x) = \int f(x')\mu(dx') - f(x) = -f(x),$$

[84]The first equality (!?) in the calculus must be checked accurately. “Usually” it does hold. In particular, the operator L is applicable to $T_s f$.

due to the identity $T_0 f = f$ and to the centering condition on f

$$\langle f, \mu \rangle = \int f(x')\mu(dx') = 0$$

and because of the convergence (ergodic property)

$$\langle f, \mu_t \rangle \to \langle f, \mu \rangle.$$

3. *Why* $\int x^2 \mu(dx) < \infty$.

3. *Centering condition for $u(x)$ with $C = 0$.* We have, due to $\int_{x \in \mathbb{R}^d} \mu_s^x(dy)\mu(dx) = \mu(dy)$ (since μ is invariant),

$$\int u(x)\mu(dx) = \iint T_s f(x) ds \mu(dx) = \iint T_s f(x)\mu(dx)ds = \iint \mathrm{E}_x f(X_s)\mu(dx)ds$$

$$= \iint \int f(y)\mu_s^x(dy)\mu(dx)ds = \iint f(y)\mu(dy)ds = \int_0^\infty 0\, ds = 0.$$

4. *Uniqueness.* Assume that there are two centered solutions, $u^1(x)$ and $u^2(x)$. Let $w(x) := u^1(x) - u^2(x)$. Then $Lw(x) = 0$. We have,

$$\mathrm{E}_x w(X_t) - w(x) = \mathrm{E}_x \int_0^t Lw(X_s)ds = 0.$$

So,

$$w(x) = \mathrm{E}_x w(X_t) = \lim_{t\to\infty} \mathrm{E}_x w(X_t) = \lim_{t\to\infty} \mathrm{E}_x (v^1 - v^2)(X_t) = \langle v^1, \mu \rangle - \langle v^2, \mu \rangle = 0,$$

as required. QED

Список литературы

[1] N.V. Krylov, Controlled diffusion processes, chapter 2.

[2] A.D. Wentzell, A Course in the Theory of Stochastic Processes.

[3] E. Pardoux, A.Yu. Veretennikov, On Poisson equation and diffusion approximation 2, Annals of Probability, 31(3) (2003), 1166–1192. DOI: 10.1214/aop/1055425774

[4] R. Khasminskii, Stochastic Stability of Differential Equations. 2nd ed. 2011.

9 Lecture 9, 20.10.2021, coupling method, invariant measures

This lecture is a continuation of the previous one with a similar mixture of topics.

9.0 Coupling for finite Markov chains, continued

The continuation will be presented again on the whiteboard, see the jam-files.

9.1 SDEs on the torus, invariant measures, convergence rates

SDEs may be considered on manifolds (as well as in Hilbert spaces, et al.), not only in $\mathbb{R}^d$. Often it is a consiquence of the particular problem, like a mathematical pendulum with a white noise which cannot leave its state space which is embedded into $\mathbb{R}^d$ (the torus T^d in this example). In this subsection let us consider a one-dimensional SDE on the torus $T^1 = [0,1]$ with the identified endpoints "$1 = 0$":

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \pmod 1, \quad t \geq 0;$$

also it is convenient to assume b and σ periodic with the period 1. This means that $x \in [0,1]$, and any value of X_t which in the "usual" $\mathbb{R}^1$ setting would be outside this interval, is mapped into $[0,1]$ by the mapping

$$X_t \mapsto X_t - [X_t].$$

It may seem that it leads to jumps in the trajectory each time when this trajectory crosses some integer value. In fact, this is not true, because in the torus the states 0 *and* 1 *are identified, so, geometrically it is the circle (= the boundary of the disc* $\sqrt{(x^1)^2 + (x^2)^2} \leq (2\pi)^{-1}$*) with a radius* $(2\pi)^{-1}$. In a more general context, torus T^1 (as well as T^d) may be regarded as a manifold (with an atlas and local maps), see, e.g., [Ikeda, Watanabe]. For the torus this approach with the atlas is not necessary.

So, there are questions about existence of invariant measure, its uniqueness, and convergence rates to the stationary regime. In this presentation we assume for simplicity[85] that $\sigma(x) \equiv 1$. Hence, we consider the equation on the torus T^1

$$X_t = x + \int_0^t b(X_s)ds + W_t \pmod 1, \quad t \geq 0.$$

[85] In the more general case of a variable σ under the assumption of its nondegeneracy the crucial tool is Harnack inequality in the parabolic version.

The first issue is to justify the existence of the invariant measure μ. Firstly let us establish one auxiliary bound; here $\mu_t^x(dy) = \mathbb{P}_x(X_t \in dy)$.

Lemma 16. *If b is bounded, then*[86]

$$\kappa := \inf_{x,x' \in T^1} \int \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1 \right) \mu_1^{x'}(dy) > 0. \tag{108}$$

Proof. **1.** Let

$$\rho^x := \exp(-\int_0^1 b(X_s^x) dW_s - \frac{1}{2}\int_0^1 |b(X_s^x)|^2 ds).$$

We have

$$\mu_1^x(A) = \mathrm{E}_x 1(X_1 \in A) = \mathrm{E}_x \rho^x (\rho^x)^{-1} 1(X_1 \in A) = \mathrm{E}_x^{\rho^x} (\rho^x)^{-1} 1(X_1 \in A).$$

Under the measure $\mathbb{P}_x^{\rho^x}(B) = \mathrm{E}_x 1(B) \rho^x$ the process X_t satisfies an SDE

$$X_t = x + \tilde{W}_t^x \pmod 1, \quad 0 \le t \le 1,$$

where $\tilde{W}_t^x = W_t + \int_0^t b(X_s^x) ds$ is a new Wiener process. Hence,

$$\mu_1^x(A) = \mathrm{E}_x^{\rho^x} 1(x + \tilde{W}_1^x \in A \ (\mathrm{mod}\ 1)) \exp(\int_0^1 b(x + \tilde{W}_s^x) dW_s + \frac{1}{2}\int_0^1 |b(x + \tilde{W}_s^x)|^2 ds)$$

$$= \mathrm{E}_x^{\rho^x} 1(x + \tilde{W}_1^x \in A \ (\mathrm{mod}\ 1)) \exp(\int_0^1 b(x + \tilde{W}_s^x) d\tilde{W}_s^x - \frac{1}{2}\int_0^1 |b(x + \tilde{W}_s^x)|^2 ds)$$

$$= \mathrm{E} 1(x + W_1 \in A \ (\mathrm{mod}\ 1)) \exp(\int_0^1 b(x + W_s) dW_s - \frac{1}{2}\int_0^1 |b(x + W_s)|^2 ds). \tag{109}$$

Likewise,

$$\mu_1^{x'}(A) = \mathrm{E} 1(x' + W_1 \in A \ (\mathrm{mod}\ 1)) \exp(\int_0^1 b(x' + W_s) dW_s - \frac{1}{2}\int_0^1 |b(x' + W_s|^2 ds). \tag{110}$$

[86] This constant is called Markov – Dobrushin's ergodic coefficient of the process X.

Both exponentials $\exp(\int_0^1 b(x+W_s)dW_s - \frac{1}{2}\int_0^1 |b(x+W_s|^2 ds)$ and $\exp(\int_0^1 b(x' + W_s)dW_s - \frac{1}{2}\int_0^1 |b(x'+W_s|^2 ds)$ are strictly positive; so, the measures $\mu^{x'}(dy)$ and $\mu^x(dy)$ are equivalent (absolutely continuous one with respect to the other), and they are also both equivalent to the Lebesgue measure on $[0,1]$: indeed, it follows from the formulae above that each value $\mu_1^{x'}(A)$ and $\mu_1^x(A)$ equals zero iff $|A|=0$. Thus, for any pair x, x' the density

$$\frac{\mu_1^{x'}(dy)}{\mu_1^x(dy)} > 0 \quad \text{a.e.}$$

2[87]**.** But why (108) holds true? Let us consider **two independent copies of the pair** (X, W): let it be (X^1, W^1) and (X^2, W^2), the first with $X_0^1 = x$, and the second with $X_0^2 = x'$. Note that the measure $\mu_t^x(dy)$ is the same for each (X^1, W^1) on any probability space (follows from the pathwise uniqueness of solution, which is assumed to be known[88]). On the direct product of the corresponding probability spaces consider the pair (X^1, X^2) (or the quadruple (X^1, W^1, X^2, W^2)). If $X_0^1 \neq X_0^2$, then suppose for the definiteness that $X_0^1 > X_0^2$. There is a strictly positive probability that the process X^1 goes "mostly down" and hits 0 before $t=1$ (estimate this probability from bellow!), while X^2 goes "mostly up" and hits 1 before $t=1$. Why this scenario is beneficial for us: because in this case they inevitably meet on the interval $[0,1]$ (likely many times). Denote

$$\tau := \inf(t \geq 0 : X_t^1 = X_t^2).$$

From what was said above it follows that

$$\inf_{x,x'} \mathbb{P}(\tau \leq 1) \geq \delta > 0. \tag{111}$$

Further, recall that the pair (X^1, X^2) is a strong Markov process and τ is a stopping time for it. Therefore, starting from this moment τ we may *change one of the processes*, say, X^2 by replacing it by X^1. In other words, consider a new process

$$\hat{X}_t^2 := X_t^2 1(t < \tau) + X_t^1 1(t \geq \tau).$$

Due to the pathwise uniqueness, what was said above implies that $\hat{X}_t^2$ is the process equivalent to X_t^2: indeed it solves the same equation as X^2

$$\hat{X}_t^2 = x' + \int_0^t b(\hat{X}_s^2)ds + \hat{W}_t^2 \pmod 1, \quad t \geq 0.$$

[87]In fact, the previous step is just for a better understanding; the next step suffices and is independent of the previous one.

[88]To be absolutely self-contained, you may imagine that $d=1$ and that b is contiuous; in this version it was rigorously proved in one of the earlier lectures.

with a new "combined" Wiener process[89]

$$\hat W_t^2 = W_t^2 1(t<\tau) + (W_t^1 - W_\tau^1) 1(t\ge\tau).$$

In particular,

$$\mu_t^{x'}(dy) = \mathbb{P}(X_t^2 \in dy) = \mathbb{P}(\hat X_t^2 \in dy).$$

Now, in any case,

$$\|\mu_t^x - \mu_t^{x'}\|_{TV} = 2\sup_A (P(X_t^1\in A) - P(X_t^2 \in A))$$

$$= 2\sup_A (P(X_t^1\in A) - P(\hat X_t^2 \in A))$$

$$= 2\sup_A (P(X_t^1\in A) - P(\hat X_t^2 \in A)(1(t<\tau)+1(t\ge\tau))$$

$$\stackrel{(\text{as } X^1 = \hat X^2 \text{ after } \tau)}{=} 2\sup_A |\mathrm{E}(1(X_t^1\in A) - 1(\hat X_t^2 \in A)|1(t<\tau) \le 2\mathrm{E}1(t<\tau).$$

in particular, for $t=1$

$$\boxed{\|\mu_1^x - \mu_1^{x'}\|_{TV} \le 2\mathrm{E}1(1<\tau).} \tag{112}$$

The latter bound is called the **coupling inequality**[90]. Note that

$$\kappa \ge \delta.$$

3. The point is that the value $1-\kappa$ for κ from (108) relates directly to the total variation distance:

$$1-\kappa := 1 - \inf_{x,x'\in[0,1]} \int \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right) \mu_1^{x'}(dy)$$

$$= \inf_{x,x'\in[0,1]} \int \left\{1 - \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right)\right\} \mu_1^{x'}(dy)$$

[89] *Exercise: show that it is a Wiener process!*

[90] Although, usually this term is applied in a slightly different context: we used here an intersection instead of the "full coupling".

is nothing else but[91] $1-\kappa = \frac{1}{2}\,\|\mu_1^x - \mu_1^{x'}\|_{TV}$. So,

$$\frac{1}{2}\,\|\mu_1^x - \mu_1^{x'}\|_{TV} = 1 - \inf_{x,x'\in[0,1]} \int \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right) \mu_1^{x'}(dy)$$

$$= \inf_{x,x'\in[0,1]} \int \left\{1 - \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right)\right\} \mu_1^{x'}(dy) \stackrel{(112)}{\le} \mathbb{P}(1<\tau) \stackrel{(111)}{\le} 1-\delta < 1,$$

as required. The lemma 16 is proved. QED

Remark 21.

Corollary 10. *Under the assumptions of lemma 16 for any $x, x' \in T^1$ (NB: $d=1$)*

$$\|\mu_n^x - \mu_n^{x'}\|_{TV} \le 2(1-\delta)^n; \tag{113}$$

moreover, for any $t>0$

$$\|\mu_t^x - \mu_t^{x'}\|_{TV} \le 2(1-\delta)^{[t]}, \tag{114}$$

where $[t]$ is the integer part of t. The claim is also valid for any two initial distributions μ_0^1 and μ_0^2 in place of x and x'.

Proof of (113) follows by induction from the lemma 16 and from the Markov property[92]. Indeed, if X^1 and $\hat{X}^2$ have not met until $t=1$, we repeat the "attempt" of meeting them on the next unit interval $[1,2]$, and so on.
Proof of (114) is due to the fact that for two Markov processes X^1 and X^2 they do not differ after the moment τ. QED

Corollary 11. *Under the assumptions of the lemma 16 there exists a (unique) invariant measure μ for the Markov process X_n. It is also invariant for X_t.*

Proof.
1. Let

$$\bar{\mu}(A) := \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mu_k^0(A) = \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mathbb{P}_0(X_k \in A),$$

where 0 is the initial state, and the limit is taken over some subsequence of $(n\to\infty)$. The limit over *some* such subsequence does exist for any $A=[r_1,r_2]\subset[0,1]$ because

[91] **Exercise!**
[92] **Exercise!**

the sequence under the limit sign is bounded. Here $r_1 \le r_2$ are any rational values on $[0,1]$. Hence, by the diagonal procedure the limit does exist (over some new subsubsequence) for any couple (r_1, r_2). It is[93] "clear" that the limiting function $\bar\mu(A)$ is a sigma-additive measure on the set of all intervals with rational endpoints; hence, its extension on all Borel sets $\mathcal{B}([0,1])$ is also a sigma-additive measure. It remains to show that this measure is invariant. Note that for any $g \in C_b$ we also have a similar equality

$$\langle g, \bar\mu\rangle = \int g(x)\bar\mu(dx) = \int g(x) \lim_{n'\to\infty} \frac{1}{n'}\sum_{k=1}^{n'} \mu_k^0(dx) = \lim_{n'\to\infty} \int g(x) \frac{1}{n'}\sum_{k=1}^{n'} \mathbb{P}_0(X_k \in dx).$$

Indeed, any $g \in C_b$ can be uniformly approximated by finite linear combinations of indicators of rational intervals for which such an equality is true by definition of $\bar\mu$.

2. It suffices to show that

$$\lim_{n'\to\infty} \frac{1}{n'}\sum_{k=1}^{n'} \mu_k^0(A) \equiv \lim_{n'\to\infty} \frac{1}{n'}\sum_{k=2}^{n'+1} \mathbb{P}_0(X_k \in A) = \bar\mu(A),$$

or, equivalently (since any measure is uniquely determined by its integrals with any $g \in C_b$), that for any $g \in C_b$ we have

$$\int g(x) \lim_{n'\to\infty} \frac{1}{n'}\sum_{k=1}^{n'} \mu_k^0(dx) = \int g(x) \lim_{n'\to\infty} \frac{1}{n'}\sum_{k=2}^{n'+1} \mathbb{P}_0(X_k \in dx).$$

Let the initial measure $\mu_0 = \bar\mu$ be chosen. Then the marginal measure at $n = 1$ is

[93] Here "clear" is a dangerous word; nevertheless, the claim is true.

uniquely determined by its integrals with any $g \in C_b$:

$$E_{\mu_0} g(X_1) = \iint g(x_1)\mu_0(dx)\mathbb{P}_x(X_1 \in dx_1) = \iint g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mu_k^0(dx)\mathbb{P}_x(X_1 \in dx_1)$$

$$= \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \int \mu_k^0(dx)\mathbb{P}_x(X_1 \in dx_1) = \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mu_{k+1}^0(dx_1)$$

$$= \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=2}^{n'+1} \mu_k^0(dx_1) = \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mu_k^0(dx_1)$$

$$+ \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \mu_{n'+1}^0(dx_1) - \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \mu_1^0(dx_1)$$

$$= \lim_{n'\to\infty} \int g(x_1) \frac{1}{n'} \sum_{k=1}^{n'} \mu_k^0(dx_1) = \int g(x_1) \lim_{n'\to\infty} \frac{1}{n'} \sum_{k=1}^{n'} \mu_k^0(dx_1) = E_{\mu_0} g(X_0).$$

Hence, the measure $\bar{\mu}$ is, indeed, invariant. Its uniqueness follows from the corollary 10. QED

9.2 SDEs in T^d, coupling, convergence rates

Here $d \geq 1$. Coupling method for discrete Markov chains was presented in the last lecture and in the beginning of this "lecture-consultation" on the whiteboard (see the jam-files in the Materials at the MsTeams pages of the course). In the dimension one it is natural (although, perhaps, not the most efficient) to use intersections of the trajectories, but in dimensions $d \geq 2$ intersections are hardly available. Yet, coupling method is applicable in any dimension. It uses the idea of induction in the proof of the lemma 16 on deterministic intervals $[n, n+1]$ and another way of justifying the bound (108) for Markov – Dobrushin's coefficient. We will establish the analogue of the lemma 16 in T^d (instead of T^1). Note that the formulae (109) – (110) are valid in the case of $d \geq 1$.

Proof of the lemma 16 in the case $d \geq 1$. We will use the fact already proved in the step 1 of the proof of the lemma that any measure μ^x is equivalent to the Lebesgue measure; let also $\bar\mu^x(dy) = \mathbb{P}(\bar X_1^x \in dy)$, where $\bar X_t^x$ is the **solution of the same SDE on $\mathbb{R}^d$ (not on T^d).** Denote

$$p_1^x(y) := \frac{\mu_1^x(dy)}{dy}, \quad \bar p_1^x(y) := \frac{\bar\mu_1^x(dy)}{dy}.$$

We have for any $A \subset [0,1]$

$$\bar\mu_1(A) = \sum_{k=-\infty}^{\infty} \mu_1(A+k),$$

so, naturally,

$$\bar p_1^x(y) \le p_1^x(y), \quad y \in [0,1].$$

(In fact, even $\bar p_1^x(y) > p_1^x(y)$ a.e., but we will not get any advantage of this strict inequality.) Due to (109) – (110) we have

$$\kappa := \inf_{x,x'\in[0,1]^d} \iint\limits_{T^d} \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right) \mu_1^{x'}(dy) \stackrel{\textbf{Exercise!}}{=} \inf_{x,x'\in[0,1]^d} \iint\limits_{T^d} \left(\frac{\mu_1^x(dy)}{dy} \wedge \frac{\mu_1^{x'}(dy)}{dy}\right) dy$$

$$= \inf_{x,x'\in[0,1]^d} \iint\limits_{T^d} \left(p_1^x(y) \wedge p_1^{x'}(y)\right) dy \ge \inf_{x,x'\in[0,1]^d} \iint\limits_{T^d} \left(\bar p_1^x(y) \wedge \bar p_1^{x'}(y)\right) dy.$$

Denote

$$\rho^x := \exp\left(-\int_0^1 b(\bar X_s^x) dW_s - \frac{1}{2}\int_0^1 \left|b(\bar X_s^x)\right|^2 ds\right),$$

and (the term $x - x'$ is added to the integrand in $\rho^{x'}$ in what follows so as to mimic artifically as if the initial value for $X^{x'}$ were x instead of x' (only valid for $t = 1$))

$$\rho^{x'} := \exp\left(\int_0^1 (-b(\bar X_s^{x'}) + x - x') dW_s - \frac{1}{2}\int_0^1 \left|-b(\bar X_s^{x'}) + x - x'\right|^2 ds\right).$$

These two stochastic exponentials are both probability densities due to Girsanov's theorem; let

$$d\tilde{\mathbb{P}}^x = \rho^x d\mathbb{P}, \quad d\tilde{\mathbb{P}}^{x'} = \rho^{x'} d\mathbb{P},$$

and let $\tilde{\mathrm{E}}^x = \mathrm{E}^{\rho^x}$ denote the expectation with respect to the probability measure $\tilde{\mathbb{P}}^x$. Note that under the measure $\tilde{\mathbb{P}}^x$ the process $\bar X_t^x$ satisfies for $0 \le t \le 1$ the equation

$$\bar X_t^x = x + \tilde W_t^x$$

with a new Wiener process $\tilde{W}_t^x$, $0 \le t \le 1$. Under the measure $\tilde{\mathbb{P}}^{x'}$ we have similarly

$$\bar{X}_t^{x'} = x' + (x - x')t + \tilde{W}_t^{x'}.$$

(So, $\bar{X}_1^{x'} = x + \tilde{W}_1^{x'}$ under the measure $\tilde{\mathbb{P}}^{x'}$.) Further,

$$\bar{\mu}_1^x(dy) = \mathrm{E}1(\bar{X}_1^x \in dy) = \tilde{\mathrm{E}}^x(\rho^x)^{-1}1(\bar{X}_1^x \in dy)$$

$$\stackrel{(\#)}{=} \mathrm{E}1(x + W_1 \in dy)\exp\left(\int_0^1 b(x + W_s)dW_s + \frac{1}{2}\int_0^1 |b(x + W_s)|^2\, ds\right)$$

((#): the law $\mathcal{L}(\tilde{W}^x)$ under $\tilde{\mathbb{P}}^x$ is identical to the law $\mathcal{L}(W)$ under $\mathbb{P}$), and likewise

$$\bar{\mu}_1^{x'}(dy) = \mathrm{E}1(x+W_1 \in dy)\exp\left(\int_0^1 (b(\bar{X}_s^{x'}) - x + x')dW_s + \frac{1}{2}\int_0^1 \left|-b(\bar{X}_s^{x'}) + x - x'\right|^2 ds\right)$$

Denote

$$\gamma := \exp\left(\int_0^1 b(x + W_s)dW_s + \frac{1}{2}\int_0^1 |b(x + W_s)|^2\, ds\right)$$

and

$$\gamma' := \exp\left(\int_0^1 (b(x' + W_s) - x + x')dW_s + \frac{1}{2}\int_0^1 |-b(x' + W_s) + x - x'|^2\, ds\right)$$

Then

$$\bar{\mu}_1^x(dy) = \mathrm{E}1(x + W_1 \in dy)\gamma \quad \& \quad \bar{\mu}_1^{x'}(dy) = \mathrm{E}1(x + W_1 \in dy)\gamma'.$$

Let $K > 0$. We have,

$$\bar{p}_1^x(y) = \frac{\bar{\mu}_1^x(dy)}{dy} = \frac{\mathrm{E}1(x + W_1 \in dy)\gamma}{dy} \ge \frac{\mathrm{E}1(x + W_1 \in dy)\gamma 1(\gamma \ge 1/K)}{dy}$$

$$\ge \frac{1}{K}\,\frac{\mathrm{E}1(x+W_1 \in dy)1(\gamma \ge 1/K)}{dy} = \frac{1}{K}\,\frac{\mathrm{E}1(x + W_1 \in dy)(1 - 1(\gamma < 1/K))}{dy}.$$

Similarly

$$\bar{p}_1^{x'}(y) = \frac{\bar{\mu}_1^{x'}(dy)}{dy} = \frac{\mathrm{E}1(x + W_1 \in dy)\gamma'}{dy} \ge \frac{\mathrm{E}1(x + W_1 \in dy)\gamma' 1(\gamma' \ge 1/K)}{dy}$$

$$\ge \frac{1}{K}\,\frac{\mathrm{E}1(x + W_1 \in dy)1(\gamma' \ge 1/K)}{dy} = \frac{1}{K}\,\frac{\mathrm{E}1(x + W_1 \in dy)(1 - 1(\gamma' < 1/K))}{dy}.$$

So, for $y\in[0,1]$ (using the hint $(a-b)\wedge(a-c)\ge a-b-c$)

$$\bar p_1^{x'}(y)\wedge\bar p_1^{x'}(y)\ge\frac{1}{K}\left(\frac{\mathrm{E}1(x+W_1\in dy)(1-1(\gamma<1/K))}{dy}\wedge\frac{\mathrm{E}1(x+W_1\in dy)(1-1(\gamma'<1/K))}{dy}\right)$$

$$\ge\frac{1}{K}\left(\frac{\mathrm{E}1(x+W_1\in dy)}{dy}-\frac{\mathrm{E}1(x+W_1\in dy)1(\gamma<1/K)}{dy}-\frac{\mathrm{E}1(x+W_1\in dy)1(\gamma'<1/K)}{dy}\right)$$

Therefore,

$$K\times\int_0^1\bar p_1^{x'}(y)\wedge\bar p_1^{x'}(y)\,dy\ge\int_0^1\left(\frac{\mathrm{E}1(x+W_1\in dy)}{dy}\right)dy$$

$$-\int_0^1\frac{\mathrm{E}1(x+W_1\in dy)1(\gamma<1/K)}{dy}dy-\int_0^1\frac{\mathrm{E}1(x+W_1\in dy)1(\gamma'<1/K)}{dy}dy$$

$$\ge\int_0^1\mathrm{E}1(x+W_1\in dy)-\mathbb{P}(\gamma<1/K)-\mathbb{P}(\gamma'<1/K)$$

The first term here is finite and uniformly wrt $0\le x\le 1$ bounded away from 0; it does not depend on K, while the other two negative terms go to zero as $K\to\infty$ **also uniformly wrt x,x' due to**[94] **Markov – Chebyshev's inequality**. Hence, we may choose K so large that

$$\kappa\ge\iint\limits_{T^d}\bar p_1^{x}(y)\wedge\bar p_1^{x'}(y)\,dy\ge\frac{1}{2K}\left(\inf_{0\le x\le 1}\iint\limits_{T^d}\underbrace{\mathrm{E}1(x+W_1\in dy)}_{=(2\pi)^{-d/2}\exp(-(y-x)^2/2)dy}-o_K(1)\right)>0,$$

for K large enough, which implies (108), as required. QED

Remark 22. NB: *Note that this version of the proof of the bound of the lemma 16 is also valid for the SDE in $\mathbb{R}^d$; actually, the main part of the calculus was in $\mathbb{R}^d$. This remark will be exploited in the next subsection along with the recurrence bound (92) ($\mathrm{E}_x\exp(\alpha\tau_R)\le C_1x^2+C_2$).*

[94] **Exercise!**

The coupling method based on the bound (108) ($\kappa > 0$)

Recall that we consider the SDE on the torus in this subsection; however, the next lemma is applicable to any state space, $\mathbb{R}^d$ included.

Lemma 17 (On two random variables)**.** *Let X and X' be two r.v. with distributions μ_1 & μ_2, respectively, and*

$$\kappa = \int \left(\frac{\mu(dy)}{\mu'(dy)} \wedge 1 \right) \mu'(dy) > 0$$

(it is similar to the value κ from (108)). Then there exists a new (except for $\kappa = 1$: then it is the original probability space, no need of a new one) probability space and two random variables $\tilde{X}$ and $\tilde{X}'$ on it such that

$$\mathcal{L}(\tilde{X}) = \mathcal{L}(X) = \mu \quad \& \quad \mathcal{L}(\tilde{X}') = \mathcal{L}(X') = \mu', \tag{115}$$

and

$$\mathbb{P}(\tilde{X} = \tilde{X}') \geq \kappa. \tag{116}$$

Proof. In fact, we may even claim in (116) that

$$\mathbb{P}(\tilde{X} = \tilde{X}') = \kappa := \int \left(\frac{\mu(dy)}{\mu'(dy)} \wedge 1 \right) \mu'(dy).$$

Assume $\kappa < 1$, otherwise the Lemma is trivial, as the random variables are equal with probability one and there is nothing to prove. Denote by $p(y)$ the density of X and by $p'(y)$ the density of X' with respect to any dominating measure (e.g., the Lebesgue one). Consider the case $\kappa > 0$: oterwise there is alos nothing to prove because the bound $\mathbb{P}(\text{any event}) \geq 0$ is trivial. Let r.v. η_1, η_2, ξ, have the following densities:

$$\begin{aligned} p_{\eta^1}(t) &= (1-\kappa)^{-1} \left(p(t) - p(t) \wedge p'(t) \right), \\ p_{\eta^2}(t) &= (1-\kappa)^{-1} \left(p'(t) - p(t) \wedge p(t) \right), \\ p_{\xi}(t) &= (\kappa)^{-1} \left(p(t) \wedge p'(t) \right). \end{aligned}$$

Let ζ be a random variable independent of η^1, η^2 and ξ taking values in $\{0, 1\}$ such that

$$\mathbb{P}(\zeta = 0) = \kappa, \;\; \mathbb{P}(\zeta = 1) = 1 - \kappa.$$

Let

$$\tilde{X} := \eta^1 1(\zeta = 1) + \xi 1(\zeta = 0),$$
$$\tilde{X}' := \eta^2 1(\zeta = 1) + \xi 1(\zeta = 0). \tag{117}$$

Then[95] $\tilde{X} \stackrel{d}{=} X$, $\tilde{X} \stackrel{d}{=} X$, and $\mathbb{P}(\tilde{X} = \tilde{X}') = \kappa$.

Verification. Indeed, from (117), clearly, it follows that

$$\mathbb{P}(\tilde{X} = \tilde{X}') \geq \mathbb{P}(\zeta = 0) = \kappa.$$

Yet, if $\kappa < 1$ then the distributions of η^1 and η^2 are singular, so in fact, we have an equality

$$\mathbb{P}(\tilde{X} = \tilde{X}') = \kappa.$$

If $\kappa = 1$, then we have,

$$\mathbb{P}(\tilde{X} = \tilde{X}') = \kappa = 1.$$

Next, since ζ, ξ and η^1 are independent on $(\Omega, \mathcal{F}, \mathbb{P})$, then for any bounded measurable function g we have (here Λ is Lebesgue's measure),

$$\mathbb{E}g(\tilde{X}) = \mathbb{E}g(\tilde{X})(1(\zeta = 0) + 1(\zeta \neq 0)) = \mathbb{E}g(\tilde{X})1(\zeta = 0) + \mathbb{E}g(\tilde{X})1(\zeta \neq 0)$$

$$= \mathbb{E}g(\xi)1(\zeta = 0) + \mathbb{E}g(\eta^1)1(\zeta \neq 0) = \mathbb{E}g(\xi)\mathbb{E}1(\zeta = 0) + \mathbb{E}g(\eta^1)\mathbb{E}1(\zeta \neq 0)$$

$$= \kappa \int g(y)p^{\xi}(y)\,\Lambda(dy) + (1-\kappa)\int g(y)p^{\eta^1}(y)\,\Lambda(dy)$$

$$= \kappa \int g(x)\frac{p \wedge p'}{\int (p \wedge p')\Lambda(dy)}(x)\Lambda(dx) + (1-\kappa)\int g(x)\frac{p - p \wedge p'}{\int (p - p\wedge p')(y)\Lambda(dy)}(x)\Lambda(dx)$$

$$= \int g(x)p \wedge p'(x)\Lambda(dx) + \int g(x)(p - p\wedge p')(x)\Lambda(dx) = \int g(y)p(y)\,dy = \mathbb{E}g(X).$$

For $\tilde{X}'$ the arguments are similar, so we also have $\mathbb{E}g(\tilde{X}') = \mathbb{E}g(X')$.

[95] Exercise!

3. In the considerations above it was assumed that all the denominators are strictly positive. If any of them equals zero, the claim of the Lemma remains valid and becomes trivial. Yet, for the sequel it makes sense to re-define all four random variables in such cases, too.

In the case $\kappa = 1$, clearly, $p = p'$. Let

$$p^{\eta^1}(x) = p^{\eta^2}(x) = p^{\xi}(x);$$

the definition of ζ does not change, but this random variable is then just a constant $\zeta = 0$ almost surely. The result is that the distributions of X and X' coincide, so the formulae $\tilde{X} = X$ and $\tilde{X}' = X$ can be used.

In the case $\kappa = 0$, the only change needed is for p^{ξ}, because the denominator in the definition of this density equals zero in this case. In fact, p^{ξ} here can be defined arbitrarily and it would not change the result because the two distributions are singular with respect to each other. For the definiteness, we propose $p^{\xi} = p^1$ (however in the application of this lemma in the next subsection it will be re-defined, which will not change the conclusion). The same formula $\tilde{X} = X$ along with $\tilde{X}'$ on some independent probability space can be used; then we can consider $\tilde{X}'$ and $\tilde{X}$ on the product of their probability spaces. Yet, coupling is impossible, which is in agreement with the fact that $\kappa = 0$. The lemma 17 is proved.

Corollary 12. *Under the assumptions of the lemma 16 (in the general case of $d \geq 1$) for any $x, x' \in T^d$*

$$\|\mu_n^x - \mu_n^{x'}\|_{TV} \leq 2(1-\kappa)^n; \tag{118}$$

moreover, for any $t > 0$

$$\|\mu_t^x - \mu_t^{x'}\|_{TV} \leq 2(1-\kappa)^{[t]}, \tag{119}$$

where $[t]$ is the integer part of t. The claim is also valid for any two initial distributions μ_0^1 and μ_0^2 in place of x and x'.

Proof. It is convenient to start with the *independent* Markov chains X_n^x and $X_n^{x'}$ on the direct product of their probability spaces. By induction

$$\frac{1}{2}\|\mu_{n+1}^x - \mu_{n+1}^{x'}\|_{TV} = \sup_A |\mathbb{P}(X_{n+1}^x \in A) - \mathbb{P}(X_{n+1}^{x'} \in A)|;$$

$$|\mathbb{P}(X^x_{n+1}\in A)-\mathbb{P}(X^{x'}_{n+1}\in A)| = |\mathrm{E}1(X^x_{n+1}\in A)-\mathrm{E}1(X^{x'}_{n+1}\in A)|$$

$$= |\mathrm{E}\mathrm{E}(1(X^x_{n+1}\in A)|\mathcal{F}^{X^x}_n)-\mathrm{E}\mathrm{E}(1(X^{x'}_{n+1}\in A)|\mathcal{F}^{X^{x'}}_n)$$

$$= |\mathrm{E}\mathrm{E}(1(X^x_1\in A)|X^x_n)-\mathrm{E}\mathrm{E}(1(X^{x'}_1\in A)|X^{x'}_n)$$

By virtue of the lemma 17, given X^x_n and $X^{x'}_n$ we may construct on some probability space the random variables, say, $\tilde X^x_{n+1}$ and $\tilde X^{x'}_{n+1}$ conditionally (given X^x_n and $X^{x'}_n$) equivalent to X^x_{n+1} and $X^{x'}_{n+1}$, respectively, and such that

$$\mathbb{P}(\tilde X^x_{n+1}\neq \tilde X^{x'}_{n+1}) = 1-\kappa^{X^x_n,X^{x'}_n}\le 1-\kappa.$$

Imagine that by the backward induction[96] all pairs $(X^x_k, X^{x'}_k)), k=1,\dots,n$ are replaced by the equivalent by distribution $(\tilde X^x_k,\tilde X^{x'}_k)), k=1,\dots,n$ using the same hint due to the lemma 17.

Thus,

$$|\mathbb{P}(X^x_{n+1}\in A)-\mathbb{P}(X^{x'}_{n+1}\in A)| = |\mathbb{P}(\tilde X^x_{n+1}\in A)-\mathbb{P}(\tilde X^{x'}_{n+1}\in A)|$$

$$= |\mathrm{E}(1(\tilde X^x_{n+1}\in A)-1(\tilde X^{x'}_{n+1}\in A)| \le \mathrm{E}1(\tilde X^x_{n+1}\neq \tilde X^{x'}_{n+1})$$

$$= \mathrm{E}\mathrm{E}\left(1(\tilde X^x_n\neq\tilde X^{x'}_n))\,1(\tilde X^x_{n+1}\neq\tilde X^{x'}_{n+1})\tilde X^x_n,\tilde X^{x'}_n\right)$$

$$= \mathrm{E}1(\tilde X^x_n\neq\tilde X^{x'}_n))\,\underbrace{\mathrm{E}\left(1(\tilde X^x_{n+1}\neq\tilde X^{x'}_{n+1})|\tilde X^x_n,\tilde X^{x'}_n\right)}_{=:\varphi_n(\tilde X^x_n,\tilde X^{x'}_n)\le 1-\kappa}$$

Clearly, here $0\le\varphi_n(\tilde X^x_n,\tilde X^{x'}_n)\le 1-\kappa(\tilde X^x_n,\tilde X^{x'}_n)\le 1-\kappa$, and

$$\varphi_n(\tilde X^x_n,\tilde X^{x'}_n) = \varphi_n(\tilde X^x_n,\tilde X^{x'}_n)1(\tilde X^x_n\neq\tilde X^{x'}_n)$$

[96] A "forward induction" is also possible instead.

(the latter since $\tilde{X}_n^x = \tilde{X}_n^{x'}$ implies straightforwardly $\tilde{X}_{n+1}^x = \tilde{X}_{n+1}^{x'}$, as $\kappa^{x,x} = 1$). So,

$$|\mathbb{P}(X_{n+1}^x \in A) - \mathbb{P}(X_{n+1}^{x'} \in A)| \le \mathrm{E}1(\tilde{X}_{n+1}^x \neq \tilde{X}_{n+1}^{x'})$$

$$= |\mathbb{P}(\tilde{X}_{n+1}^x \in A) - \mathbb{P}(\tilde{X}_{n+1}^{x'} \in A)| = |\mathrm{E}\varphi_n(\tilde{X}_n^x, \tilde{X}_n^{x'})1(\tilde{X}_n^x \neq \tilde{X}_n^{x'})|$$

$$\le |\mathrm{E}(1-\kappa(\tilde{X}_n^x, \tilde{X}_n^{x'}))1(\tilde{X}_n^x \neq \tilde{X}_n^{x'})| \le (1-\kappa)|\mathrm{E}1(\tilde{X}_n^x \neq \tilde{X}_n^{x'})|.$$

By induction oven n we obtain

$$|\mathbb{P}(X_{n+1}^x \in A) - \mathbb{P}(X_{n+1}^{x'} \in A)| = |\mathbb{P}(\tilde{X}_{n+1}^x \in A) - \mathbb{P}(\tilde{X}_{n+1}^{x'} \in A)|$$

$$\le (1-\kappa)|\mathrm{E}1(\tilde{X}_n^x \neq \tilde{X}_n^{x'})| \le \ldots \le (1-\kappa)^n(1-\kappa(x,x')) \le (1-\kappa)^{(n+1)}.$$

The bound (118) follows.

The first moment when $\tilde{X}_n^1 = \tilde{X}_n^2$ is called the moment of coupling; let us use the notation τ for it.

Proof of (119) is due to the fact that for two Markov processes $\tilde{X}^1$ and $\tilde{X}^2$ they do not differ after the moment of coupling τ. QED

9.3 SDEs in $\mathbb{R}^d$, coupling, convergence rates & uniqueness of invariant measures: the sketch

[97] Coupling method for discrete Markov chains was presented in the last lecture and in the beginning of this "lecture-consultation" on the whiteboard (see the jam-files in the Materials at the MsTeams pages of the course). In the dimension one it is natural (although, perhaps, not the most efficient) to use intersections of the trajectories, but in dimensions $d \ge 2$ intersections are hardly available. Yet, coupling method is applicable in any dimension, as we saw in the previous subsection for SDEs on the torus. For the SDEs in $\mathbb{R}^d$ this approach also works well, just it uses the idea of induction in the proof of the lemma 16, not on deterministic interval $[n, n+1]$, but on random intervals built upon the sequence of stopping times. We will use the fact that the bound (92) of the lemma 16 holds true in $\mathbb{R}^d$, see the remark 22:

$$\mathrm{E}_x \exp(\alpha\tau^X(R)) \le C_1 x^2 + C_2,$$

[97] **The whole subsection is not compulsory.**

where $\tau^X(R) = \inf(t \ge 0 : |X_t| \le R)$. The plan how to use this bound is as follows. Define the sequence of stopping times

$$\tau_1^X < \gamma_1^X < \tau_2^X < \gamma_2^X \ldots < \tau_n^X < \gamma_n^X < \ldots, \quad n \ge 1,$$

where

$$\tau_1^X = \tau^X(R), \quad \gamma_n^X = (\tau_n^X + 1) \wedge \inf(t \ge \tau_n^X : |X_t| \ge R + 1)),$$

$$\tau_{n+1}^X := \inf(t \ge \gamma_n^X : |X_t| \le R).$$

Due to the bound (92) it is possible to prove that for R large enough there is a constant $c > 0$ such that

$$\mathbb{P}(\liminf_{n\to\infty} \frac{\tau_n^X}{n} \ge c) = 1 \tag{120}$$

(in fact, even $\mathbb{P}(\liminf_{n\to\infty} \frac{\tau_n^X}{n} = c) = 1$, but it is a bit more difficult to show and not necessary for our aims at the moment), and, moreover, for any $\epsilon > 0$

$$\mathbb{P}_x(\frac{\tau_n^X}{n} \ge c + \epsilon) \le C(x) \exp(-\lambda n). \tag{121}$$

This signifies that "on average" the sequence (τ_n^X) is growing approximately linearly in n. In other words, with a probability close to one on the interval $[0, t]$ we may expect on average approximately $c^{-1}t$ moments $\tau_1^X, \ldots, \tau_n^X$.

Now consider two *independent* processes $X_t = X_t^x$ (with the initial value x) and $X'_t = X_t^{x'}$ (with the initial value x'). Let

$$\tau_1^{X,X'} < \gamma_1^{X,X'} < \tau_2^{X,X'} < \gamma_2^{X,X'} \ldots < \tau_n^{X,X'} < \gamma_n^{X,X'} < \ldots, \quad n \ge 1,$$

where

$$\tau_1^{X,X'} = \tau^{X,X'}(R) := \inf(t \ge 0 : |X_t| \vee |X'_t| \le R),$$

$$\gamma_n^{X,X'} = (\tau_n^{X,X'} + 1) \wedge \inf(t \ge \tau_n^{X,X'} : |X_t| \vee |X'_t| \ge R + 1)),$$

$$\tau_{n+1}^{X,X'} := \inf(t \ge \gamma_n^{X,X'} : |X_t| \vee |X'_t| \le R).$$

Due to the bound (92) *and by virtue of the independence of X and X'* it is possible to prove[98] that for R large enough there is a constant $c > 0$ (which does not depend on x, x') such that

$$\mathbb{P}(\liminf_{n\to\infty} \frac{\tau_n^{X,X'}}{n} \ge c) = 1 \tag{122}$$

(in fact, even $\mathbb{P}(\liminf_{n\to\infty} \frac{\tau_n^{X,X'}}{n} = c) = 1$), and, moreover, for any $\epsilon > 0$

$$\mathbb{P}_{x,x'}(\frac{\tau_n^{X,X'}}{n} \ge c + \epsilon) \le C(x, x') \exp(-\lambda n). \tag{123}$$

We will arrange **coupling** at the moments $\gamma_n^{X,X'}$. Similarly to the situation of the torus (where all bounds were uniform in space; now it is not the the case), with the only change that in $\mathbb{R}^d$ we use random (stopping) times $\tau_n^{X,X'}$ along with $\gamma_n^{X,X'}$ instead of the deterministic sequence (n) in the case of the torus, at each moment $\tau_n^{X,X'}$ let us regard at the distributions of $X_{\gamma_n^{X,X'}}$ and $X'_{\gamma_n^{X,X'}}$ given the values $X_{\tau_n^{X,X'}}$ and $X'_{\tau_n^{X,X'}}$. These distributions are both equivalent to the Lebesgue measure and to each other; the analogue of the statement of the lemma 16 is also valid in this case in the following version. Denote as earlier

$$\mu_1^x(dy) = \mathbb{P}_x(X_1 \in dy), \quad \mu_1^{x'}(dy) = \mathbb{P}_{x'}(X'_1 \in dy).$$

Lemma 18. *If b is bounded, then for any $R > 0$*

$$\kappa_R := \inf_{|x|,|x'|\le R} \int \left(\frac{\mu_1^x(dy)}{\mu_1^{x'}(dy)} \wedge 1\right) \mu_1^{x'}(dy) > 0. \tag{124}$$

Proof **just repeats the calculus in the subsection 9.2***;* the only little news is the use of any R. Using the lemma 17 we will construct a similar coupling not for the sequence $(X_n^x, X_n^{x'})$ but for $(Z_n, Z'_n) := (X^x_{\gamma_n^{X,X'}}, X^{x'}_{\gamma_n^{X,X'}})$. Denote

$$\hat\mu_n^x(dy) = \mathbb{P}_x(Z_n \in dy), \quad \hat\mu_n^{x'}(dy) = \mathbb{P}_{x'}(Z'_n \in dy).$$

Corollary 13. *Under the assumptions of the lemma 16 (in the general case of $d \ge 1$) for any $x, x' \in \mathbb{R}^d$*

$$\|\hat\mu_n^x - \hat\mu_n^{x'}\|_{TV} \le 2(1 - \kappa_R)^n. \tag{125}$$

[98] The proof for the pair (X, X') is quite similar to the proof for one component; just the dimension is $2d$ instead of d.

Proof. It is convenient to start with the *independent* Markov chains X_n^x and $X_n^{x'}$ on the direct product of their probability spaces. By induction

$$\frac{1}{2}\|\mu^x_{n+1} - \mu^{x'}_{n+1}\|_{TV} = \sup_A |\mathbb{P}(Z_{n+1} \in A) - \mathbb{P}(Z'_{n+1} \in A)|;$$

$$|\mathbb{P}(Z_{n+1} \in A) - \mathbb{P}(Z'_{n+1} \in A)| = |\mathrm{E}1(Z_{n+1} \in A) - \mathrm{E}1(Z'_{n+1} \in A)|$$

$$= |\mathrm{E}\mathrm{E}(1(Z_{n+1} \in A)|\mathcal{F}^Z_n) - \mathrm{E}\mathrm{E}(1(Z'_{n+1} \in A)|\mathcal{F}^{Z'}_n)|$$

$$= |\mathrm{E}\mathrm{E}(1(Z_{n+1} \in A)|Z_n) - \mathrm{E}\mathrm{E}(1(Z'_{n+1} \in A)|Z'_n)|$$

By virtue of the lemma 17, given Z_n and Z'_n we may construct on some probability space the random variables, say, $\tilde{Z}_{n+1}$ and $\tilde{Z}'_{n+1}$ conditionally (given Z_n and Z'_n) equivalent to Z_{n+1} and Z'_{n+1}, respectively, and such that

$$\mathbb{P}(\tilde{Z}_{n+1} \neq \tilde{Z}'_{n+1}) \leq 1 - \kappa^{Z_n, Z'_n} \leq 1 - \kappa.$$

By the backward induction all pairs $(Z_k, Z'_k), k = 1, \ldots, n$ may be replaced by the equivalent by distribution $(\tilde{Z}_k, \tilde{Z}'_k), k = 1, \ldots, n$ using the same hint due to the lemma 17. *(It should be noted that, of course, this replacement also changes the moments (γ_n) (the new $\tilde{\gamma}_n$ are all just on a new probability space!) but they still have the same distributions and, hence, the same exponential moments.)*

Thus,

$$|\mathbb{P}(Z_{n+1} \in A) - \mathbb{P}(Z'_{n+1} \in A)| = |\mathbb{P}(\tilde{Z}_{n+1} \in A) - \mathbb{P}(\tilde{Z}'_{n+1} \in A)|$$

$$= |\mathrm{E}\mathrm{E}(1(\tilde{X}^x_{n+1} \in A)|\tilde{X}^x_n) - \mathrm{E}\mathrm{E}(1(\tilde{X}^{x'}_{n+1} \in A)|\tilde{X}^{x'}_n)|$$

$$= |\mathrm{E}\mathrm{E}(1(\tilde{Z}_{n+1} \in A)|X^x_n) - \mathrm{E}\mathrm{E}(1(\tilde{Z}'_{n+1} \in A)|\tilde{X}^{x'}_n)|$$

$$= |\mathrm{E}\mathrm{E}\left((1(\tilde{Z}_{n+1} \in A) - (1(\tilde{Z}'_{n+1} \in A))|\tilde{Z}_n, \tilde{Z}'_n\right)|$$

$$= |\mathrm{E}1(\tilde{Z} \neq \tilde{Z}'_{n+1})\underbrace{\mathrm{E}\left(1(\tilde{Z}_{n+1} \in A) - 1(\tilde{Z}'_{n+1} \in A))\,|\tilde{Z}_n, \tilde{Z}'_n\right)}_{=:\varphi_n(\tilde{Z}_n, \tilde{Z}'_n)}|$$

Clearly, here $|\varphi_n(\tilde Z_n, \tilde Z'_n)| \le 1 - \kappa(\tilde Z_n, \tilde Z'_n)$, and

$$\varphi_n(\tilde Z_n, \tilde Z'_n) = \varphi_n(\tilde Z'_n, \tilde Z'_n)1(\tilde Z_n \neq \tilde Z'_n)$$

(the latter since $\tilde Z = \tilde Z'_n$ implies straightforwardly $\tilde Z_{n+1} = \tilde Z'_{n+1}$, as $\kappa^{x,x} = 1$). So,

$$|\mathbb{P}(Z_{n+1} \in A) - \mathbb{P}(Z'_{n+1} \in A)| = |\mathbb{P}(\tilde Z_{n+1} \in A) - \mathbb{P}(\tilde Z'_{n+1} \in A)|$$

$$= |\mathrm{E}\varphi_n(\tilde Z_n, \tilde Z'_n)1(\tilde Z_n \neq \tilde Z'_n)| \le |\mathrm{E}(1 - \kappa(\tilde Z_n, \tilde Z'_n))1(\tilde Z \neq \tilde Z'_n)|$$

$$\le (1-\kappa)|\mathrm{E}1(\tilde Z \neq \tilde Z'_n)|.$$

By induction oven n we obtain

$$|\mathbb{P}(Z_{n+1} \in A) - \mathbb{P}(Z'_{n+1} \in A)| = |\mathbb{P}(\tilde Z_{n+1} \in A) - \mathbb{P}(\tilde Z'_{n+1} \in A)|$$

$$\le (1-\kappa)|\mathrm{E}1(\tilde Z_n \neq \tilde Z'_n)| \le \ldots \le (1-\kappa)^n(1-\kappa(x,x')) \le (1-\kappa_R)^{(n+1)}.$$

The bound (118) follows. QED

The first moment when $\tilde Z^1_n = \tilde Z^2_n$ is called the moment of coupling; let us use the notation $\tilde\tau$ for it. In fact, coupling may be also organised on other time intervals, while the components are outside the chosen ball B_R; however, it is less convenient for the calculus.

Once the sequence $(\tilde Z_n, \tilde Z'_n), n = 1, 2, \ldots$ – which is a Markov chain – has been constructed, it may be accomplished to the full continuous trajectory $(\tilde X^x_t, \tilde X^{x'}_t), t \ge 0$, by using the conditional distributions of the path intervals $(\tilde X^x_t, \tilde X^{x'}_t), \tilde\gamma_n \le t \le \tilde\gamma_{n+1}$ given the values $(\tilde X^x_{\tilde\gamma_n}, \tilde X^{x'}_{\tilde\gamma_n}) = (\tilde Z_n, \tilde Z'_n)$ and $(\tilde X^x_{\tilde\gamma_{n+1}}, \tilde X^{x'}_{\tilde\gamma_{n+1}}) = (\tilde Z_{n+1}, \tilde Z'_{n+1})$. As a result, we have the processes equivalent to the original solutions $(X^x_t, X^{x'}_t), t \ge 0$. Note that after the moment of coupling we have the equality

$$\tilde X^x_t = \tilde X^{x'}_t, \quad t \ge \tilde\tau. \tag{126}$$

The bound (125) seems uniform, and this is true. Yet, the consequence for the total variation distance for the measures μ^x_t and $\mu^{x'}_t$, which may be derived from it, is not uniform in the initial values.

Corollary 14. *Under the assumptions of the lemma 16 (in the general case of $d \geq 1$) for any $x, x' \in \mathbb{R}^d$ with $n = [t/(c+\epsilon)]$*

$$\|\mu_t^x - \mu_t^{x'}\|_{TV} \leq 2(1-\kappa_R)^n + 2\mathbb{P}_{x,x'}(\tilde{\gamma}_n > t). \tag{127}$$

Proof.

Finally, the bound (123) allows to justufy the following inequality:

Lemma 19. *Under the assumptions of the lemma 16*

$$\mathbb{P}_{x,x'}(\tilde{\gamma}_n > t) \leq C(1 + x^2 + (x')^2)\exp(-\lambda t) \tag{128}$$

with some $C, \lambda > 0$.

This along with (127) provides the resulting exponential (nonuniform) bound on the distance $\|\mu_t^x - \mu_t^{x'}\|_{TV}$.

Theorem 49. *Under the assumptions of the lemma 16*

$$\|\mu_t^x - \mu_t^{x'}\|_{TV} \leq 2(1-\kappa_R)^{[t/(c+\epsilon)]} + C(1 + x^2 + (x')^2)\exp(-\lambda t). \tag{129}$$

If the initial distributions μ_0 and μ_0' are not necessarily delta-functions, then it transforms into

$$\|\mu_t^{\mu_0} - \mu_t^{\mu_0'}\|_{TV} \leq 2(1-\kappa)^{[t/(c+\epsilon)]} + C\left(1 + \int x^2\mu_0(dx) + \int (x')^2\mu_0(dx')\right)\exp(-\lambda t). \tag{130}$$

Proof.

Corollary 15. *Under the assumptions of the lemma 16 (in the general case of $d \geq 1$) the invariant measure μ for the Markov process X_t is unique.*

Proof.

9.4 Addendum: Proof of Poisson equation in the theorem 48 based on PDEs (not compulsory)

This version of the proof **does not require the continuity** of f; the only assumptions are that f is Borel measurable and bounded, and that it is centered. Recall, the semigroup $T_t f(x) := \mathrm{E}_x f(X_t)$ satisfies

$$\frac{d}{dt}T_t f(x) = LT_t f(x).$$

Accroding to the PDE results [Solonnnikov], [Ladyzhenskaya et al.], the parabolic equation

$$v_s(s,x) - Lv(s,x) = 0, \quad v(0,x) = f(x),$$

with $f \in L_p(R^d)\bigcap C_b(R^d)$, $p \geq d+1$, has a unique solution in the class $W_p^{1,2}((0,T]\times R^d)\bigcap C([0,T]\times R^d)$ to which Ito–Krylov's formula is applicable (under the assumption of bounded coefficients of the SDE and the non-degeneracy of $\sigma\sigma^*$). We have $v(t,x) = E_x f(X_t) = T_t f(x)$. Moreover, in the case of $f \in C_b^2$ and b, σ Hölder in x uniformly wrt s, this solution is of the class $C_b^{1,2}$. The function $v(T-s,x)$ with $0 \leq s \leq T$ has a probabilistic representation

$$v(T-s,x) = \mathrm{E} f(X_T^{s,x}).$$

The parabolic equation

$$v_s(s,x) - Lv(s,x) = g(s,x), \quad v(0,x) = 0,$$

with $g \in L_p([0,T]\times R^d)\bigcap C(R^d)$, $p \geq d+1$, has a unique solution in the class $W_p^{1,2}((0,T]\times R^d)\bigcap C([0,T]\times R^d)$ to which Ito–Krylov's formula is applicable (under the assumption of bounded coefficients of the SDE and the non-degeneracy of $\sigma\sigma^*$). We have

$$v(t,x) = \mathrm{E}_x \int_0^t g(t-s, X_s)ds.$$

If $g \in C_b^2$ & $b, \sigma \in H_x^\alpha$ uniformly wrt s, then $v \in C_b^{1,2}$. We have,

$$\begin{aligned} dv(t-s,X_s) &= -v_s(t-s,X_s)ds + Lv(t-s,X_s)ds + d(mart)_s \\ &= g((t-s,X_s)ds + d(mart)_s; \ -v(t,x) = -\mathrm{E}_x \int_0^t g((t-s,X_s)ds. \end{aligned}$$

Let us show that the function $u(x) = \int_0^\infty \mathrm{E}_x f(X_t)dt$ is continuous. Indeed, for each fixed $t>0$, $u^t(s,x) = \int_s^t \mathrm{E} f(X_r^{s,x})dr$ for $0 \leq s < t$ is a *continuous* solution of the parabolic equation

$$v_s(s,x) - Lv(s,x) = -f(s,x), \quad v(0,x) = 0, \quad 0 \leq s \leq t,$$

[Krylov, Safonov, 1980]; see also [Ladyzenskaja, Solonnikov and Ural'ceva (1968), Chapter 4], [Veretennikov (1982)].

Moreover, $\int_0^t \mathrm{E}_x f(X_r)dr = \int_0^t \mathrm{E} f(X_r^{0,x})dr$ converges to $u(x)$ as $t \to \infty$ locally uniformly in x due to the convergence bounds for the total variation distance between μ_t^x and μ. Hence, the limit $u(x)$ is (locally uniformly) continuous.

Further, $u(x)$ satisfies $u(x) = \mathrm{E}_x u(X_\gamma) + \mathrm{E}_x \int_0^\gamma f(X_t)dt$. For any domain D let $\gamma := \inf(s \ge 0 : X_s \not\in D)$. By the strong Markov property,

$$u(x) = \int_0^\infty \mathrm{E}_x f(X_r^x)(1(r < \gamma) + 1(r \ge \gamma))dr$$

$$= \int_0^\infty \mathrm{E}_x f(X_r^x)(1(r < \gamma)dt + \int_0^\infty \mathrm{E}_x 1(r \ge \gamma)(\mathrm{E}(f(X_r^x)|\mathcal{F}_\gamma)dt$$

$$= \int_0^\infty \mathrm{E}_x f(X_r^x)(1(r < \gamma)dt + \int_0^\infty \mathrm{E}_x 1(r \ge \gamma)(\mathrm{E}_{X_\gamma}(f(X_r^x))dt$$

$$\stackrel{Fubini!}{=} \mathrm{E}_x \int_0^\gamma f(X_r^x)dt + \mathrm{E}_x \int_0^\infty 1(r \ge \gamma)(\mathrm{E}_{X_\gamma}(f(X_r^x))dt$$

$$= \mathrm{E}_x \int_0^\gamma f(X_r^x)dt + \mathrm{E}_x u(X_\gamma). \qquad (131)$$

Let us show that $u(x)$ belongs to Sobolev classes (locally) (it is important that we already know that u is continuous!). Consider any ball D and the Dirichlet problem

$$L\hat{u}(x) = -f(x), \ x \in D, \quad \hat{u}(x) = u(x)|_{\partial D}.$$

This equation has a unique solution $u \in W^2_{p,loc}(D) \cap C(\bar{D})$ for any $p > 1$; see [Gilbarg and Trudinger (1983), Corollary 9.18]. We can then apply Ito - Krylov's formula to $\hat{u}(X_t)$ on the random interval $[0, \gamma]$, where $\gamma := \inf(s \ge 0 : X_s \not\in D)$ [see Krylov (1980), Theorem 2.10.1]. One deduces that

$$\hat{u}(x) = \mathrm{E}_x u(X_\gamma) + \mathrm{E}_x \int_0^\gamma f(X_t)dt \stackrel{!}{=} u(x)$$

(compare with the formula (131)). It follows that u itself is a (Sobolev, or classical under additional assumptions on f and b, σ) solution of the Poisson equation $Lu = -f$ in D, which domain is arbitrary; so, the equation holds in the whole $\mathbb{R}^d$. QED

10 Lecture 10, 12.11.2021, Bellman equation on a finite horizon; Howard – Bellman's algorithm of the cost function improvement; viscosity solutions

10.1 Controlled diffusion processes

Suppose we have an SDE with coefficients depending on a parameter,

$$dX_t = \sigma(a,t,X_t)dW_t + b(a,t,X_t)dt, \quad X_0 = x,$$

and this parameter may be **tuned** at any time, $a = a(t) \in U$ for a given set $U \subset \mathbb{R}$ (for example; maybe $U \subset \mathbb{R}^d$, etc.), and a may also depend on ω. The goal of tuning is to maximise (or minimise) some functional, in the following standard form,

$$v^{a(\cdot)}(x) = \mathrm{E}_x^{a(\cdot)} \left(\int_0^T f(a(t),t,X_t)dt + g(X_T) \right) \to \sup_{(a(\cdot))} . \tag{132}$$

Here U is assumed to be a bounded connected convex domain. Tuning is allowed at any moment based only on the *information available to this moment*, that is, the *strategy* $a(t) \in \mathcal{F}_t$ (given filtration), or *control strategy*; in particular, $a(t, X_t)$ is allowed (this is called a feedback control, or a markovian control).

The *process* $a_t \in \mathcal{F}_t$ is accepted to be **a strategy** iff there exists a (strong or weak[99]) solution of the equation

$$dX_t = \sigma(a_t,t,X_t)dW_t + b(a_t,t,X_t)dt, \quad X_0 = x. \tag{133}$$

For this course we accept strong solutions.

Theorem 50 (Homework). *Let Lipschitz condition hold true,*

$$\|\sigma(a,t,x) - \sigma(a,t,y)\| + |b(a,t,x) - b(a,t,y)| \le L|x-y|,$$

and usual linear growth or boundedness conditions on b and σ hold true. Then the equation (133) has a unique strong solution.

Hint: Just re-read the earlier proofs of to Ito's theorems "without $a(\cdot)$" and add this control strategy process; check that the calculus does not change.

[99]Strong ones are preferred; however, there is a big theory of a weak control, too.

For Markov (=feedback) strategies the issue of the existence of solutions is very important: the (Markov) strategy $a(\cdot)$ is called admissible iff the equation

$$dX_t = \sigma(a(t, X_t), t, X_t)dW_t + b(a(t, X_t), t, X_t)dt, \quad X_0 = x, \tag{134}$$

has a (strong) solution. Recall the following result.

Theorem 51 (simplified Zvonkin-Veretennikov; for $d = 1$ & Lipschitz σ also Nakao)**.** *Assume that*

$$\sigma = \sigma(t, x),$$

that $\sigma\sigma^*$ is uniformly non-degenerate, and that σ and b bounded Borel measurable; let σ be Lipshitz continuous (for $d = 1$ Hölder-$\frac{1}{2}$ continuous). Then the equation (134) has a unique strong solution.

We will often (except for the next subsection) deal with SDEs satisfying $\sigma = \sigma(t, x)$, that is, without a dependence on the control in σ. The reason is that under certain conditions on σ (e.g., nondegeneracy and Lipschitz in x), for any bounded (measurable) drift b there exists a strong solution of the corresponding equation.

Why do we consider Markov strategies: because such strategies suffice to achive the optimum of the functional. This will be one of the main auxiliary results of stochastic control for diffusion processes on the finite horizon that supremum over all admissible strategies equals the supremum over all Markov admissible ones, which is, clearly a more narrow set of strategies.

According to the above, two notions of strategies are defined: admissible (notation $\mathcal{A}$) and Markov admissible (notation $\mathcal{A}_M$).

Definition 23. *An adapted (to the corresponding filtration) process α_s with values in U is called an admissible strategy iff the SDE*

$$dX_t = \sigma(\alpha_t, t, X_t)dW_t + b(\alpha_t, t, X_t)dt, \quad X_0 = x,$$

has a pathwise unique strong solution for any x.

Definition 24. *A Borel measurable function $a(t, x)$ with values in U is called a Markov strategy iff the SDE*

$$dX_t = \sigma(a(t, X_t), t, X_t)dW_t + b(a(t, X_t), t, X_t)dt, \quad X_0 = x,$$

has a pathwise unique strong solution for any x. If $a(\cdot)$ is a Markov strategy, the process $\alpha_t = a(t, X_t)$ is usually also called a Markov strategy[100]*.*

[100] Slightly abusing the notations.

Using the last convention, we have $\mathcal{A}_M \subset \mathcal{A}$.

Denote

$$v(0,x) = \sup_{a(\cdot)\in\mathcal{A}} v^{a(\cdot)}(0,x) = \sup_{a(\cdot)\in\mathcal{A}} \mathrm{E}_x^{a(\cdot)}\Big(\int_0^T e^{-ct} f(a(t),t,X_t)dt + e^{-cT} g(X_T)\Big)$$

and

$$v^M(0,x) = \sup_{a(\cdot)\in\mathcal{A}_M} v^{a(\cdot)}(0,x) = \sup_{a(\cdot)\in\mathcal{A}_M} \mathrm{E}_x^{a(\cdot)}\Big(\int_0^T e^{-ct} f(a(t),t,X_t)dt + e^{-cT} g(X_T)\Big)$$

and, of course, we will need the objects

$$v(s,x) = \sup_{a(\cdot)\in\mathcal{A}} v^{a(\cdot)}(s,x) = \sup_{a(\cdot)\in\mathcal{A}} \mathrm{E}_{s,x}^{a(\cdot)}\Big(\int_s^T e^{-c(t-s)} f(a(t),t,X_t)dt + e^{-c(T-s)} g(X_T)\Big) \tag{135}$$

and

$$v^M(s,x) = \sup_{a(\cdot)\in\mathcal{A}_M} v^{a(\cdot)}(s,x) = \sup_{a(\cdot)\in\mathcal{A}_M} \mathrm{E}_{s,x}^{a(\cdot)}\Big(\int_s^T e^{-c(t-s)} f(a(t),t,X_t)dt + e^{-c(T-s)} g(X_T)\Big) \tag{136}$$

for $0 \le s \le T$. In what follows we will use a short notation a instead of $a(\cdot)$, but in some other formulae a will be used for a state in U. Naturally, we have

$$v^M \le v.$$

It turns out that the optimal cost function (either of them) must satisfy some nonlinear PDE called Bellman's equation. If we manage to solve this equation, then also an optimal (or nearly optimal) strategy may be determined.

10.2 Parabolic Bellman's (aka HJB) equations

The notation will be used:

$$dX_s^{a,t,x} = b(a(s,X_s^{a,t,x}),s,X_s^{a,t,x})ds + \sigma(a(s,X_s^{a,t,x}),X_s^{a,t,x})dW_s, s \ge t,\ X_t^{a,t,x} = x$$

Parabolic HJB equations arise where control is on a non-random finite interval of time. The introduction to parabolic Bellman's equation for diffusions (without jumps) can be read in [Krylov, Controlled diffusion processes, Chapter 1, section 4]. For a function $u(t,x)$ the parabolic Bellman equation reads

$$u_t(t,x) + \sup_{a\in U}[(L^a - c)u(t,x) + f^a(t,x)] = 0,\ u(T,x) = g(x). \tag{137}$$

where (in the case of $d = 1$)

$$L^a h(s,x) = \frac{\sigma^2(a,s,x)}{2} h''(s,x) + b(a,s,x)h'(s,x).$$

Theorem 52. *If there exists a solution u of the equation (137) to which Ito's formula is applicable (after a substitution of a process with a stochastic differential), then*[101]

$$u(0,x) = v(0,x) = v^M(0,x).$$

The proof (see [Krylov, Chapter 1]) usually consists of two steps.

Step 1: $u \geq v$. Suppose that the **solution** $u(t,x)$ **of the HJB equation** has one continuous derivative in t and two in x, so that Ito's formula can be applied. Consider any strategy α_t and apply Ito's formula to $\exp(-c(t-s))u(t, X_t^{\alpha,s,x})$. From the equation

$$u_t(t,x) + \sup_{a \in U}[(L^a - c)u(t,x) + f^a(t,x)] = 0$$

it follows that without sup there will be the **in**equality

$$u_t(t,x) + [(L^a - c)u(t,x) + f^a(t,x)] \leq 0.$$

Hence, Ito's formula tells us that

$$\mathrm{E}^a_{s,x} e^{-c(T-s)} u(T, X_T) - u(s,x)$$

$$= \mathrm{E}^\alpha_{s,x} \int_s^T e^{-c(t-s)} [u_t(t,X_t) + (L^{\alpha_t} - c)u(t,X_t)]dt \leq -\mathrm{E}^\alpha_{s,x} \int_s^T e^{-c(t-s)} f^{\alpha_t}(t,X_t)dt.$$

In other words,

$$u(s,x) \geq \mathrm{E}^\alpha_{s,x} e^{-c(T-s)} g(X_T) + \mathrm{E}^\alpha_{s,x} \int_s^T e^{-c(t-s)} f^{\alpha_t}(t,X_t)dt = v^\alpha(s,x).$$

Thus, the inequality

$$u(0,x) \geq v^\alpha(0,x)$$

[101] The last equality is interpreted as a sufficiency of Markov strategies.

follows. Since this holds true for any strategy $\alpha \in \mathcal{A}$, hence, we conclude

$$u(0,x) \ge \sup_{\alpha\in\mathcal{A}} v^{\alpha}(0,x) = v(0,x),$$

as well.

Step 2: $u \le v^M \le v$.

Now choose a particular Borel measurable Markov strategy $\bar{a}$ such that for each x the maximum is attained in the HJB equation (in its left hand side)[102]. Again applying Ito's formula to $\exp(-ct)u(t, X_t^{\bar{a}})$ and using the HJB equation – equality! – we will get the equality

$$u(0,x) = v^{\bar{a}}(0,x).$$

Which implies

$$u(0,x) = v^{\bar{a}}(0,x) \le \sup_{a\in\mathcal{A}_M} v^{a}(0,x) = v^M(0,x).$$

Hence, overall

$$v(0,x) \le u(0,x) \le v^M(0,x) \le v(0,x) \quad \Longrightarrow \quad u(0,x) = v(0,x). \qquad \text{QED}$$

Exercise 22. It is your task to apply Ito's formula and to verify what is claimed above!

Exercise 23. *Find a possible gap in the latter proof: what is (or may be) missing?*

Exercise 24. *If the supremum in* $\sup_a(L^a u + f^a)$ *is not attainable, so no* $\bar{a}$ *exists which should be optimal, what can be done? Is the equality* $u = v$ *still true?*

Note on the multidimensional case. Let $\sigma\sigma^*/2 =: A$. In the multidimensional case the result is similar to what was shon n $d = 1$; the only difference is that the generators L now have the form

$$L^a = \sum_{i,j=1}^{d} A^a_{ij}(s,x)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^a_i(s,x)\frac{\partial}{\partial x^i},$$

for a fixed tuning parameter $a \in U$, or

$$L^a = L^{a(s,x)} = \sum_{i,j=1}^{d} A^{a(s,x)}_{ij}(s,x)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^{a(s,x)}_i(s,x)\frac{\partial}{\partial x^i},$$

if the Markov strategy $a(s,x)$ has been chosen. The parabolic Bellman equation then still reads as in (137).

[102]It is assumed that such a Borel measurable function exists; there are certain sufficient conditions for that: e.g., all coefficients are continuous in $a \in U$ and U is a compact.

10.3 Another intuition towards Bellman's equation

The intuitive reasoning[103] in this section shows *non-rigorously* that it *should be*

$$u(0, x) = v^M(0, x).$$

(But nothing about the equality $u(0, x) = v(0, x)$.) Forget that we know what Bellman's equation is and start from scratch. Assume $c \equiv 0$ for simplicity. Suppose there exists an optimal Markov strategy $\bar{a} \in \mathcal{A}_M$ such that $v^{\bar{a}(\cdot)} = v^M$. Take any $t \in (0, T)$ and consider the value[104]

$$\varphi(0, x) := \mathrm{E}_x^{\bar{a}} \left(\int_0^t f(X_s) ds + v^M(t, X_t) \right),$$

where $v^M(t, x') = \sup_{a' \in \mathcal{A}_M} \mathrm{E}_{t,x'}^{a'} \left(\int_t^T f(X_s) ds + g(X_T) \right)$. Here $v^M(t, X_t)$ is maximum what can be on average achieved by using the best strategy *after* t. So, it is expected that

$$\varphi(0, x) = v^M(0, x).$$

If it is so, and assuming that $v^M(t, x)$ is smooth enough, let us use Ito's formula as usual[105], subtract and divide by t: then we get

$$\lim_{t \downarrow 0} \frac{\mathrm{E}_{0,x}^{\bar{a}} v^M(t, X_t) - v^M(0, x)}{t} = -\lim_{t \downarrow 0} t^{-1} \mathrm{E}_{0,x}^{\bar{a}} \int_0^t f(X_s) ds.$$

Assuming again that v^M (which is not the solution of Bellman's equation here, but the optimal over all Markov strategies cost function) is smooth enough, we get

$$v_s^M(0, x) + L^{\bar{a}} v^M(0, x) = -f^{\bar{a}}(x). \tag{138}$$

Here we treated $f^{\bar{a}}(x)$ as $f^{\bar{a}(x)}(x)$. On the other hand, for any other, likely non-optimal strategy a, we similarly argue that the optimal strategy after t (for $t \le T$) combined with any a before t gives us less that the supremum, that is,

$$v^M(0, x) \ge \mathrm{E}_{0,x}^{a} \left(\int_0^t f(X_s) ds + v^M(t, X_t) \right).$$

[103] This section is about how to *guess* at all that Bellman's equation is worth studying in stochastic control problems.

[104] The function f here *may* also depend on time.

[105] *Exercise: Show the details!*

Subtracting here and dividing by t, we get

$$\lim_{t\downarrow 0}\frac{\mathrm{E}^a_{0,x}v^M(t,X_t)-v^M(0,x)}{t}\le -\lim_{t\downarrow 0}t^{-1}\mathrm{E}^a_{0,x}\int_0^t f(X_s)ds,$$

which leads to the inequality

$$v^M_s(0,x)+L^a v^M(0,x)\le -f^a(x). \tag{139}$$

From (210) and (139) the equation (137) with $c\equiv 0$ follows for the function v^M.

Exercise 25. *1. Perform the same calculus with $f^a(s,x)$ and with a potential $c(s,x)$. 2. How to tackle problem 4 if the supremum is not attained and no "optimal" $\bar a$ is available? 3. Where is the gap, why the calculus is non-rigorous in this section?*

10.4 Howard's algorithm

Suppose we can solve precisely any linear PDE; then how to get the value $v(0,x)$, how approach to it? – The answer is: by Howard's algorithm. Howard's (or, Howard – Bellman's) algorithm works as follows. Let us start with some (any) initial Markov strategy $a_0(\cdot)$. Denote

$$v^0(t,x)=v^{a_0}(t,x)=\mathrm{E}^{a_0}_{t,x}\left(\int_t^T f(s,X_s)ds+g(X_T)\right),$$

where $X_s=X_s^{t,x}, s\ge t$ is a solution of the SDE with a_0 which starts at x at time t. This function v^0 satisfies a linear PDE (FPK equation)

$$v^0_t(t,x)+L^{a_0}v^0(t,x)+f^{a_0}(t,x)=0,\ v^0(T,x)=g(x).$$

Let us choose a Borel measurable $a_1(\cdot)$ as follows:

$$a_1(t,x):=\operatorname*{argsup}_{a\in U}[v^0_t+L^a v^0(t,x)+f^a(x)]\ (\stackrel{!}{\ge}0)\ \ \&\ \ v^1(t,x):=v^{a_1}(t,x).$$

Further, by induction, let

$$a_{n+1}(t,x):=\operatorname*{argsup}_{a\in U}[v^n_t+L^a v^n(t,x)+f^a(x)]\ \ (\stackrel{!}{\ge}0),$$

and denote

$$v^{n+1}(t,x):=v^{a_{n+1}}(t,x)=\mathrm{E}^{a_{n+1}}_{t,x}\left(\int_t^T f(X_s)ds+g(X_T)\right).$$

In this algorithm we do not care about *convergence of strategies*, but we do care about convergence of the sequence of *cost functions* (v^n). Note that under the assumption of boundedness of f and g, all function v^n are bounded by modulus by $\|f\|\,T + \|g\|$. Further, we will show that the sequence of v^n is pointwise monotone increasing and, hence, has a pointwise limit. There is a natural hope that this limit equals the optimal cost $v(t,x)$. This is true under some conditions. In this course we will only show the improvement property of Howard algorithm, that is, that every next strategy provides a better cost function.

Theorem 53. *The sequence v^n convergence monotonically to some limit*[106]*: $v^n \uparrow \bar v$.*

Proof. Let us compare v^n and v^{n+1}. Both functions are solutions of similar but not identical FPK equations

$$v_t^n(t,x) + L^{a_n} v^n(t,x) + f^{a_n}(t,x) = 0, \quad 0 \le t \le T, \quad v^n(T,x) = g(x),$$

and

$$v_t^{n+1}(t,x) + L^{a_{n+1}} v^{n+1}(t,x) + f^{a_{n+1}}(t,x) = 0, \quad v^{n+1}(T,x) = g(x),$$

and also the inequality

$$\begin{aligned} 0 &= v_t^n(t,x) + \sup_a [L^a v^n(t,x) + f^a(x)] \\ &= v_t^n(t,x) + [L^{a_{n+1}} v^n(t,x) + f^{a_{n+1}}(x)] =: h_n(t,x) \overset{!}{\ge} 0. \end{aligned}$$

Denote $w^n = v^n - v^{n+1}$. Subtracting, we obtain for w^n a differential inequality

$$w_t^n(t,x) + L^{a_{n+1}} w^n(t,x) = h_n(t,x) \ge 0, \qquad w^n(T,x) = 0.$$

Apply Ito-Krylov's (or Dynkin's) formula to $w^n(t, X_t^{a_{n+1}})$:

$$\begin{aligned} -w^n(0,x) &= \mathrm{E}_x \underbrace{w^n(T, X_T^{a_{n+1}})}_{=0} - w^n(0,x) \\ &= \mathrm{E}_x \int_0^T [w_s^n(s, X_s^{a_{n+1}}) + L^{a_{n+1}} w^n(s, X_s^{a_{n+1}})] ds \\ &= \mathrm{E}_x \int_0^T h_n(s, X_s^{a_{n+1}}) ds \ge 0. \end{aligned}$$

[106] The fact that this limit coincides with the optimal cost v is left without a proof in this course.

Thus, $w^n(0,x) \le 0$, so,

$$v^n(0,x) \le v^{n+1}(0,x),$$

as required. QED

Exercise 26. *Show that* $v^n(t,x) \le v^{n+1}(t,x)$, $\forall\, t \le T$.

Remark: Why $\bar v = v$. Suppose we know that $\bar v = \lim v^n$ solves Bellman's equation. (It is a long but simple story based on PDE a priori bounds which we do not show here as it is not the course on PDEs.) These bounds imply that the limit is a solution of Bellman's equation. However, this solution is unique, and we (presumably) already know that v is also its solution. Therefore, $\bar v = v$.

The same reasoning shows that the limit $\bar v$ depends neither on the initial strategy a_0, nor on the particular (measurable) choice of each successive strategy a_n in the case if the sup at each step of the Howard algorithm may be attained not at a unique point. (NB: In principle v^n *may* depend on this non-unique choice, though.)

An example of such rigorous reasoning in the 1D case may be found in the proof of theorem 1.4.5 in [N.V. Krylov, Controlled diffusion processes]. *It is recommended to study this result independently.*

More issues to think about. *The strategy which we use must be admissible, that is, a solution of the corresponding SDE must exist (strong, or, at least, weak). By choosing the next strategy as argsup in Howard's algorithm, we may have a trouble with no solution of the SDE. How to tackle it?*

The solution is to use approximate argsup. You may read [Krylov's Controlled diffusion processes, Chapter 1] about it.)

10.5 Elliptic Bellman's equation

This section tackles some more types of HJB equation for diffusions on the infinite horizon. Firstly, we wish to optimise the functionals (with $c > 0$ and f bounded)

$$v^{a(\cdot)}(x) := \mathrm{E}_x^{a(\cdot)} \int_0^\infty e^{-ct} f(X_t)dt \to \sup_{a(\cdot)\in\mathcal{A}_M} =: v^M(x)$$

and

$$v^{\alpha)}(x) := \mathrm{E}_x^{\alpha} \int_0^\infty e^{-ct} f(X_t)dt \to \sup_{a(\cdot)\in\mathcal{A}} =: v(x).$$

Theorem 54. *Assume f bounded, $c > 0$, b, σ, σ^{-1} bounded. Let $u \in C_b^{1,2}$ solve the HJB equation*

$$\sup_{a \in U}[L^a u(x) - cu(x) + f^a(x)] = 0. \tag{140}$$

Then

$$u(x) = v^M(x) = v(x). \tag{141}$$

Also, the (Borel measurable) strategy $\bar{a}(x) = \underset{a \in U}{argsup}[L^a u(x) - cu(x) + f^a(x)]$ is optimal if it is admissible. Howard's algorithm provides a sequence v_n which converges to v.

NB: If u solves (140), then $u + k$ with $k \neq 0$ is not a solution. Similarly to the parabolic case the latter equality in (141) is interpreted as a sufficiency of Mrkov strategies.

Another version of an elliptic HJB equation is provided by the following setting. We have a bounded domain (an interval in $\mathbb{R}^1$) D and the initial data $x \in \bar{D}$. The optimisation problems are

$$v^{a(\cdot)}(x) := \mathrm{E}_x^{a(\cdot)} \int_0^\tau e^{-ct} f(X_t)dt \to \sup_{a(\cdot) \in \mathcal{A}_M} =: v^M(x)$$

where $\tau = \inf(t \geq 0 : \ X_t \notin D)$, and

$$v^\alpha(x) := \mathrm{E}_x^\alpha \int_0^\tau e^{-ct} f(X_t)dt \to \sup_{\alpha \in \mathcal{A}} =: v(x).$$

Theorem 55. *Assume $c \geq 0$, $f, b, \sigma, \sigma^{-1}$ bounded, D bounded. If $u \in C^{1,2}(\bar{D})$ solves the HJB equation*

$$\sup_{a \in U}[L^a u(x) - cu(x) + f^a(x)] = 0, \, x \in D, \quad u|_{\partial D} = 0, \tag{142}$$

then

$$u(x) = v^M(x) = v(x). \tag{143}$$

The (Borel measurable) strategy $\bar{a}(x) = \underset{a \in U}{argsup}[L^a u(x) - cu(x) + f^a(x)]$ is optimal (if it is admissible). Howard's algorithm provides a sequence v_n which converges to v.

Exercise 27. *Show the heuristics leading to a justification of the equations (140)-(141) and (142)-(143).*

10.6 On viscosity solutions (sketch)

In the case of $d > 1$ even for Sobolev solutions of elliptic and parabolic PDEs and Bellman's equations certain conditions are required. As a minimum, they include the nondegeneracy of the diffusion matrix, and some minimal regularity (like continuity) is also needed. To overcome these restrictions the notion of viscosity solution was introduced. The idea is they we refuse to apply Ito's formula and, hence, are not asking for two derivatives in whatever sense. However, instead we want to be able to write down Dynkin's formula for "solution", or for what should have been a solution in a good situation with two derivatives: we only wish the formula like

$$v(x) = \mathrm{E}_x g(X_\tau)$$

in the case of the elliptic PDE with a zero right hand side, or,

$$v(x) = \mathrm{E}_x^{\bar{a}} g(X_\tau) \tag{144}$$

with some "optimal" Markov strategy $\bar{a}$ in the case of the elliptic Bellman's equation in a domain D. Similar formulae should also be valid for nonzero right hand sides in both problems. Let us talk about the linear PDE, for simplicity. The process X_t does have a generator (possibly a quasi-generator) L. So, how at all the "solution" in whatever sense might be defined if we do not want the derivatives to be involved? Here is the sketch of the answer based on the notions of viscosity sub- and super-solutions.

Let us write down our (elliptic) Bellman's equation in a domain D in the form of

$$F[u](x) := \sup_{a \in U} L^a u(x) = 0, \quad x \in D, \quad u(x)|_\Gamma = g(x). \tag{145}$$

Definition 25. *An upper semicontinuous function u on D is called a subsolution of the equation (145) in the viscosity sense iff for any $x_0 \in D$ and any $\varphi \in C^2(D)$ such that $\varphi(x_0) = u(x_0)$ and $u \le \varphi$ in some (any) neighborhood of x_0, the inequality*

$$F[\varphi](x_0) \le 0$$

holds.

Definition 26. *A lower semicontinuous function u on D is called a supersolution of the equation (145) in the viscosity sense iff for any $x_0 \in D$ and any $\varphi \in C^2(D)$ such that $\varphi(x_0) = u(x_0)$ and $u \ge \varphi$ in some (any) neighborhood of x_0, the inequality*

$$F[\varphi](x_0) \ge 0$$

holds.

Reminder. A lower semicontinuous function u on D is such a function that

$$u(x_0) \le \liminf_{x\to x_0} u(x), \quad \forall\, x_0 \in D.$$

An upper semicontinuous function u on D is such a function that

$$u(x_0) \ge \limsup_{x\to x_0} u(x), \quad \forall\, x_0 \in D.$$

Definition 27. *A continuous function u is a viscosity solution of the equation (145) in the viscosity sense iff it is both a supersolution and a subsolution.*

References: [Crandall, Michael G.; Evans, Lawrence C.; Lions, Pierre-Louis (1984), Some properties of viscosity solutions of Hamilton–Jacobi equations, Transactions of the American Mathematical Society, 282 (2): 487–502, doi:10.2307/1999247, ISSN 0002-9947, JSTOR 1999247]

Proposition 1. *Let D be a bounded domain in $\mathbb{R}^d$ and $u \in C_b(\bar D)$ be a viscosity solution in D of the elliptic Bellman equation (145). Also assume that*

$$\sup_{x\in D} \mathrm{E}_x \tau_D < \infty$$

where τ_D is the exit time of the process X from D uniformly with respect to the strategy applied. Then at any $x \in D$ the representation (144) holds true.

Proof (sketch[107]). 1. Let φ be a subsolution of the equation (144) at x, and let us choose a small enough neighbourhood of x . By Ito's formula with X^a_t corresponding to a Markov strategy $a \in \mathcal{A}^M$ which is "optimal" in this neighbourhood in the sense of (145) (assuming that the supremum in (145) is attained),

$$\mathrm{E}^a_x \varphi(X^a_{t\wedge\tau^1_\epsilon}) - \varphi(x) = \mathrm{E}^a_x \int_0^{t\wedge\tau^1_\epsilon} L^a \varphi(X^a_s) ds \le \delta \mathrm{E}^a_x \tau^1_\epsilon,$$

where τ_ϵ is the exit time from the neighbourhood of x in which $u \le \varphi$. Then,

$$\mathrm{E}^a_x u(X^a_{t\wedge\tau^1_\epsilon}) - u(x) \le \mathrm{E}^a_x \int_0^{t\wedge\tau^1_\epsilon} L^a \varphi(X^a_s) ds \le \delta \mathrm{E}^a_x \tau^1_\epsilon.$$

By virtue of Fatou's lemma we obtain as $t \to \infty$,

$$\mathrm{E}^a_x u(X^a_{\tau^1_\epsilon}) - u(x) \le \delta \mathrm{E}^a_x \tau^1_\epsilon,$$

[107] The proof is not required for the exam, unlike for the definitions, which are necessary!

or, equivalently
$$u(x) \geq \mathrm{E}_x^a u(X^a_{\tau^1_\epsilon}) - \delta \mathrm{E}_x^a \tau^1_\epsilon.$$
Considering now $X_{\tau^1_\epsilon}$ as a new initial state and using induction and a sequence of stopping times $(\tau_n \wedge \tau_D)$, we finally (hope to) obtain the bound

$$u(x) \geq \mathrm{E}_x^{\tilde a} g(X^{\tilde a}_{\tau_D}) - o_\epsilon(1),$$

where $\tilde a$ is a combined "optimal" strategy in a sequence of neighbourhoods. Hence,

$$u(x) \geq \mathrm{E}_x^{\tilde a} g(X^{\tilde a}_{\tau_D}).$$

2. Similarly using supersolutions we get the opposite inequality

$$u(x) \leq E_x^{\tilde a} g(X^{\tilde a}_{\tau_D}).$$

Thus,
$$u(x) = E_x^{\tilde a} g(X^{\tilde a}_{\tau_D}).$$
and so, the representation (144) holds true, as required. QED

NB: *As usual, if supremum is not attained, then it is possible to use nearly optimal strategies.*

11 Lecture 11, 19.11.2021, Ergodic stochastic control & Bellman's equation

11.1 Main result: theorem 56

Consider an SDE in[108] $\mathbb{R}^1$ with a compact set U of controls,

$$dX_t = \sigma(a(X_t), X_t)dW_t + b(a(X_t), X_t)dt, \quad X_0 = x.$$

Assume the process X_t is "ergodic" for each Markov (aka stationary) strategy $a(x) \in \mathcal{A}_M$ and that any Borel measurable strategy a is admissible (in the sense that there exists a weak solution[109] unique in distribution and which is a strong Markov process. Other strategies are not considered in this setting. By ergodicity we mean existence of a unique stationary measure $\mu^a(dx)$ towards which there is a convergence

$$\sup_a \|\mu_t^{x,a} - \mu^a\|_{TV} \le \varphi(x,t) \to 0 \quad \text{"fast enough"}, \quad a \in \mathcal{A}_M. \tag{146}$$

By fast enough we mean some exponential convergence $C(x)\exp(-\lambda t)$ with $C(x) = C\exp(c'|x|)$, or, a convergence faster than any polynomial, that is, $\varphi(x,t) \le C(x)(1+t^k)^{-1}$ with some $C = C_k, m = m_k$ and $C(x) = C(1+|x|^m)$, and

$$(1+t^k)^{-1}\varphi(x,t) \to 0, \quad t \to \infty, \quad \forall\, k > 0.$$

Also, this function $C(x)$ in either case should be integrable with respect to the invariant measure μ^a: $\int C(x)\mu^a(dx) < \infty$. To guarantee the latter convergence we will assume

$$\langle b(x), x\rangle \to -\infty, \quad |x| \to \infty. \tag{147}$$

The condition (89)

$$\langle b(x), x\rangle \le -r|x|^2, \qquad |x| \ge M$$

also suffices.

Recall that $\mathcal{A}_M$ is the class of all Markov strategies depending only on x; σ^{-1} is assumed bounded. The cost function (=aka value function, aka reward function) is

$$\rho^a := \liminf_{T\to\infty} \frac{1}{T} \mathrm{E}_x^{a(\cdot)} \int_0^T f(X_t)dt.$$

[108] In his lecture the dimension $d = 1$ is essential, although, some partial results are available for $d > 1$, too, for example, in the case $\sigma(a,x) \equiv \sigma(x)$, that is, where σ does not depend on the control.

[109] Strong solutions would be preferred, but if we want σ to depend on the control, then strong solutions are just not achievable for **any** Markov strategy; later on it will be assumed that $_s \equiv 1$, then, of course, we will deal with strong solutions.

Lemma 20. *Under the assumption of ergodicity above, ρ^a does not depend on x, and it equals*

$$\rho^a = \lim_{t\to\infty} \frac{1}{T} \mathrm{E}_x^{a(\cdot)} \int_0^T f(X_t)dt = \int f^{a(\cdot)}(y)\mu^a(dy).$$

The problem is to optimise

$$\rho^a \to \sup_a =: \rho.$$

In order to write down Bellman's equation for this problem we need an auxiliary functions $v^a(x)$; however, the cost will be ρ, not $v^a(x)$. Let

$$v^a(x) := \int_0^\infty \mathrm{E}_x^a(f(X_t) - \rho^a)dt, \quad L^a = \frac{(\sigma^a)^2}{2}\frac{d^2}{dx^2} + b^a\frac{d}{dx}. \tag{148}$$

Note that under the ergodicity conditions this integral converges.

Theorem 56. *Let $\sigma, b, f \in C_b^1$ in x; let σ^{-1} be bounded. Then ergodic Bellman's equation reads*

$$\sup_{a\in U}[L^a v(x) + f^a(x) - \rho] = 0. \tag{149}$$

Under our assumptions this equation has a unique solution (v, ρ) with a bounded ρ and with the auxiliary function v which may increase no faster than some polynomial, where the "optimal" auxiliary function v is unique up to an additive constant. Howard's reward improving algorithm (RIA) provides a converging sequence $\rho_n \uparrow \rho$.

For the sequel let us assume for simplicity $\sigma(x) \equiv 1$. However, you may find the complete proof without this simplified assumption at

Список литературы

[1] https://arxiv.org/abs/1812.10665; Svetlana Anulova, Hilmar Mai, Alexander Veretennikov, On iteration improvement for averaged expected cost control for one-dimensional ergodic diffusions, SIAM Journal on Control and Optimization. 2020. Vol. 58. No. 4. P. 2312-2331 DOI: 10.1137/19M1271944

For $\sigma \equiv 1$ (and for $\sigma(x)$ not depending on the control[110]) the proof of a very closed problem (yet, with a somewhat different set of strategies, so called extended, or randomised ones) may be found in the following book:

[110]In this lecture the same assumption is assumed; even more, we will only consider $\sigma \equiv 1$.

Список литературы

[1] A. Arapostathis, V. S. Borkar, and M. K. Ghosh, Ergodic control of diffusion processes, Encyclopedia of Mathem. and its Appl. 143, CUP, Cambridge, 2012.

Note that because of the assumption $d = 1$, under our conditions for any Borel function $\alpha \in \mathcal{A}$ there is a unique stationary measure μ^α, which is *equivalent to the Lebesgue measure* Λ. The latter follows from the formula for the unique stationary density

$$p^\alpha(x) := \frac{d\mu^\alpha(x)}{dx} = \frac{C_\alpha}{\sigma^2(\alpha(x), x)} \exp\left(2\int_0^x \frac{b(\alpha(y), y)}{\sigma^2(\alpha(y), y)}\, dy\right), \tag{150}$$

where C_α is a normalising constant. The fact that p^α is a stationary density can be seen from a substitution to the equation of stationarity $(L^\alpha)^* p = 0$ (see, for example, [Khasminskii, Lemma 4.16, equation (4.70)]); its uniqueness in the class of integrable functions satisfying the normalizing condition $\int p\, dx = 1$ can be justified via the explicit solution of the stationarity equation in the one-dimensional case (& there are some other ways to do it).

Exercise 28. • *Show that the function $p^a(x)$ given in (150) is a stationary density for L^a.*

- *Explain why $\exists K_1, K_2 > 0$ such that*

$$K_1 \le C_a \le K_2, \qquad \forall\, a.$$

- *Show that $\forall k > 0\ \exists C, c, \kappa > 0$ such that*

$$c \exp(-\kappa |x|) \le p^a(x) \le \frac{C}{1 + |x|^k} \tag{151}$$

for all $x \in R^1$.

In particular, it follows that the invariant measure integrates any power function.

Proof of theorem (partial). We will prove rigorously the monotonicity of the RIA and, hence, its convergence to *some* limiting value, and the uniqueness of ρ in the ergodic Bellman equation. Recall that $\sigma \equiv 1$ and that for any strategy $a(x)$

$$\rho^a = \langle f^a, \mu^a \rangle.$$

For any pair (v, ρ) define

$$F[v, \rho](x) := \sup_{u \in U} \left[L^u v(x) + f^u(x) - \rho \right],$$

$$G[v](x) := \sup_{u \in U} \left[L^u v(x) + f^u(x) \right],$$

and

$$F_1[v, \rho](x) := \sup_{u \in U} [b^u v' + f^u - \rho](x).$$

For any strategy $a(x)$ we have (a Poisson equation)

$$L^a v^a(x) + f^a(x) - \rho^a = 0.$$

Let us recall what a reward improvement algorithm (RIA) is. We start with some (any) Markov strategy $\alpha_0 \in \mathcal{A}_M$. Denote the corresponding value "function" (which is, actually, a constant depending on the strategy), the invariant measure, and the auxiliary function $\rho_0 = \rho^{\alpha_0} = \langle f^{\alpha_0}, \mu^{\alpha_0} \rangle$, and $v_0 = v^{\alpha_0}$. If for some $n = 0, 1, \ldots$ the triple (α_n, ρ_n, v_n) is determined, then the strategy α_{n+1} is defined as follows: for a.e. x the function α_{n+1} is chosen so that for each x

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) = G[v_n](x) = \sup_{u \in U} \left[L^u v_n(x) + f^u(x) \right], \tag{152}$$

or, in other words, $\alpha_{n+1}(x) \in \mathrm{Argmax}_{u \in U} \left[L^u v_n(x) + f^u(x) \right]$. We assume that a Borel measurable version of such strategy may be chosen. The unique invariant measure $\mu^{\alpha_{n+1}}$ corresponds to this strategy α_{n+1}, the value $\rho_{n+1} := \langle f^{\alpha_{n+1}}, \mu^{\alpha_{n+1}} \rangle$, and the function $v_{n+1} = v^{\alpha_{n+1}}$.

Let us show that ρ_n increases with n. For almost every (a.e.) $x \in \mathbb{R}$ we have,

$$\rho_n = L^{\alpha_n} v_n(x) + f^{\alpha_n}(x) \le G[v_n](x) = L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x)$$

and also for a.e. $x \in \mathbb{R}$,

$$\rho_{n+1} = L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x).$$

So,

$$\begin{aligned} \rho_n - \rho_{n+1} &\stackrel{a.e.}{\le} (L^{\alpha_{n+1}} v_n + f^{\alpha_{n+1}})(x) - (L^{\alpha_{n+1}} v_{n+1} + f^{\alpha_{n+1}})(x) \\ &= (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(x). \end{aligned} \tag{153}$$

Let us apply Ito – Krylov's formula with expectations (also known as Dynkin's formula) to $(v_n - v_{n+1})(X_t^{\alpha_{n+1}})$: we have for any $x \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}_x \left(v_n(X_t^{\alpha_{n+1}}) - v_{n+1}(X_t^{\alpha_{n+1}})\right) - \left(v_n - v_{n+1}\right)(x) \\
&= \mathbb{E}_x \int_0^t (L^{\alpha_{n+1}} v_n - L^{\alpha_{n+1}} v_{n+1})(X_s^{\alpha_{n+1}})\, ds \\
&\geq \mathbb{E}_x \int_0^t (\rho_n - \rho_{n+1})\, ds = (\rho_n - \rho_{n+1})\, t.
\end{aligned} \tag{154}$$

Since the left hand side in (154) is bounded for a fixed x (because it converges to $\langle v_n, \mu^n \rangle$), we divide all terms of the latter inequality by t and let $t \to \infty$ to get,

$$0 \geq \rho_n - \rho_{n+1}, \qquad \text{as required.}$$

Thus, $\rho_n \leq \rho_{n+1}$, so that $\rho_n \uparrow \tilde{\rho}$ (since the sequence ρ_n is bounded) with some $\tilde{\rho}$. So, the RIA does converge.

Note that clearly $\tilde{\rho} \leq \rho$, since ρ is the supremum over all Markov strategies, while $\tilde{\rho}$ is the supremum over some countable subset of them. Later on we shall show that they do coincide. It will follow from the uniqueness of the component ρ in the solution (v, ρ) of the ergodic Bellman equation and the claim that some subsequence $v_{n'}$ up to some constants $\beta_{n'}$ also converges to some $\tilde{v}$, and that the pair $(\tilde{v}, \tilde{\rho})$ is a solution of this equation. In fact, it is the most involved part of the proof which is currently skipped. Next, it may be proved that over some subsequence $(n' \subset n)$

$$v_{n'} \to \tilde{v} \qquad \text{in } C^1[-N, N],\ \forall N > 0,$$

and, indeed, that the limiting couple $(\tilde{v}, \tilde{\rho})$ is a solution of Bellman's ergodic equation.

Local boundedness of v_n. It follows from the definition (148) and from the assumed rate of convergence (146).

Local boundedness of v_n'. It is explained in the Appendix, step 6.

Local boundedness of v_n''. It follows from the local boundedness of v_n' and from the Poisson equation for the function v_n which reads

$$\frac{1}{2} v_n''(x) = -b^{a_n}(x) v_n'(x) - (f^{a_n}(x) - \rho_n).$$

It suffices for the pre-compactness of the sequence (v_n) in C^1 locally on any bounded interval. Hence, by virtue of the Arzela – Acsoli theorem it is possible to choose a subsequence ($n' \to \infty$) (using the diagonal method) such that $v_{n'}$ converges in C^1 on any bounded interval (i.e., along with its derivative v_n' in $C([-N,N])$ for any N).

Passing to the limit in PDEs

$$\rho_{n+1} = L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x),$$

$$\rho_n \le \rho_{n+1} = L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x),$$

using the fact that $\rho_{n+1} - \rho_n \to 0$. **Denote**

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n =: h_n(x) \ge 0.$$

Note that v_n and v_{n+1} satisfy close Poisson equations (i.e., with close rhs) *with the same generator:*

$$L^{\alpha_{n+1}} v_{n+1}(x) + f^{\alpha_{n+1}}(x) - \rho_{n+1} = 0,$$

$$L^{\alpha_{n+1}} v_n(x) + f^{\alpha_{n+1}}(x) - \rho_n - h_n(x) = 0.$$

Hence, $-f^{\alpha_{n+1}} + \rho_n + h_n(x)$ is *centered* wrt μ_{n+1} (otherwise the second Poisson equation here in the whole $\mathbb{R}^1$ has no solution), and so $\langle h_n, \mu_{n+1}\rangle$ is small, since $\langle f^{\alpha_{n+1}}, \mu_{n+1}\rangle = \rho_{n+1}$ and because $\rho_n - \rho_{n+1} \to 0$ as $n \to \infty$. Then, we may conclude that functions v_n and v_{n+1} are close up to an additive constant (generally speaking, depending on n). This fact follows from the probabilistic representation of the solution of Poisson equation and from the auxiliary bounds (151) *(Exercise).*

Convergence of v_n over a subsequence. Return to the most involved part: $v_n \to \tilde v$, and the limit solves Bellman's equation with a corresponding second component $\tilde\rho$. It was explained earlier that *over some subsequence* ($n' \subset n$)

$$v_{n'} \to \tilde v \qquad \text{in } C^1_{loc}.$$

Now we want to show that the second derivative of this limit $\tilde v''$ exists, **and that the limiting couple $(\tilde v, \tilde \rho)$ is a solution of Bellman's ergodic equation (149).**

Denote $w_n = v_n - v_{n-1}$. Considering a subsequence $v_{n'}$ converging locally in C^1

$$v_{n'} \underset{C^1_{loc}}{\to} \tilde v,$$

we are going to show that

$$w_n'(x) \to 0,\ n \to \infty \quad \text{(locally uniformly)}. \tag{155}$$

Once this is established, we would pass to the limit in the equation

$$\begin{aligned} v_{n'}'(x_1) - v_{n'}'(x_2) = \int_{x_2}^{x_1} v_{n'}''(s)ds \\ = -\int_{x_2}^{x_1} \sup_u [b^u v_{n'-1}'(s) + f^u(s) - \rho_{n'-1}]ds + o_{n'}(1), \end{aligned} \tag{156}$$

to achieve as $n' \to \infty$

$$\tilde{v}'(x_1) - \tilde{v}'(x_2) = -\int_{x_2}^{x_1} \sup_u [b^u \tilde{v}'(s) + f^u(s) - \tilde{\rho}]ds. \tag{157}$$

The equation (157) is equivalent to ergodic Bellman's equation (just differentiate both sides of (157)), i.e., the pair $(\tilde{v}, \tilde{\rho})$ is a solution of (149). Here is how to pass to the limit in (156). We have

$$v_{n'}'(x_1) - v_{n'}'(x_2) = \int_{x_2}^{x_1} v_{n'}''(s)ds = -\int_{x_2}^{x_1} [b^{a_{n'}} v_{n'}'(s) + f^{a_{n'}}(s) - \rho_{n'}]ds$$

$$= -\int_{x_2}^{x_1} \sup_u [b^u v_{n'-1}'(s) + f^u(s) + \rho_{n'}]ds - \int_{x_2}^{x_1} [b^{a_{n'}}(v_{n'}'(s) - v_{n'-1}'(s))]ds$$

$$= -\int_{x_2}^{x_1} \sup_u [b^u v_{n'-1}'(s) + f^u(s) - \rho_{n'}]ds + o_{n'}(1)$$

This is where we need to know that $v_{n'}'(s) - v_{n'-1}'(s) \overset{\text{loc-uni}}{\to} 0$: so as to justify that for any x_2, x_1

$$\int_{x_2}^{x_1} [b^{a_{n'}}(v_{n'}'(s) - v_{n'-1}'(s))]ds = o_{n'}(1).$$

<u>So, why $w_n'(s) = v_n'(s) - v_{n-1}'(s) \overset{\text{loc-uni}}{\to} 0$?</u> Recall that

$$L^{a_n} w_n - \rho_n + \rho_{n-1} - h_{n-1} = 0.$$

Denote

$$-\rho_n + \rho_{n-1} - h_{n-1}(x) =: g_n(x).$$

Function g_n is locally bounded (because h_n is locally bounded and due to the global boundedness of all ρ_n) and it is locally uniformly in L_1 small as $n \to \infty$. Since w_n solves a Poisson equation, it has a representation with some bounded constant β_n,

$$w_n(x) - \beta_n = \int_0^\infty \mathrm{E}_x g_n(X_t^n) dt.$$

(We have to subtract some constant to get a centered solution, since it is known that $\int_0^\infty E_x g_n(X_t^n) dt$ is centered with respect to μ_n.) Hence, it is likely and natural that $w_n(x) - \beta_n$ is small, at least, in C^1_{loc}. For the function w_n we have a Poisson equation in R:

$$\frac{1}{2} w_n''(x) + b^{a_n} w_n'(x) + g_n(x) = 0. \tag{158}$$

The ODE/PDE theory claims that if g_n is small in L_2, or uniformly, or in any L_p on some interval $[-N-1, N+1]$, then w_n' and w_n'' are small in $C[-N,N]$, and w_n' is small in the uniform norm locally, see the Appendix to this lecture in the next subsection. This implies that

$$\sup_{[-N,N]} |v_n' - v_{n-1}'|(x) dx \to 0, \quad n \to \infty,$$

and this suffices for our limiting procedure. Indeed, we have

$$v_{n'}'(x_1) - v_{n'}'(x_2) = \int_{x_2}^{x_1} v_{n'}''(s) ds$$

$$= -\int_{x_2}^{x_1} \sup_u [b^u v_{n'-1}'(s) + f^u(s) - \rho_{n'}] ds + o_{n'}(1)$$

Now we know that $v_{n'}'(s) - v_{n'-1}'(s) \stackrel{\text{loc-uni}}{\to} 0$. In the limit we obtain

$$\tilde{v}'(x_1) - \tilde{v}'(x_2) = -\int_{x_2}^{x_1} \sup_u [b^u \tilde{v}'(s) + f^u(s) - \tilde{\rho}] ds.$$

Here the right hand side is clearly differentiable with respect to x_1 (as well as wrt x_2) So, the left hand side is also differentiable. Hence, differentiating, we get the existence of the second derivative $\tilde{v}''$ and the equation

$$\frac{1}{2} \tilde{v}''(x) + \sup_u [b^u \tilde{v}'(s) + f^u(s) - \tilde{\rho}] = 0.$$

As it was already noticed, this equation is equivalent to the ergodic Bellman's one (149).

Uniqueness for ρ in ergodic Bellman's equation. Assume that there are two solutions of the ergodic (HJB) equation (149), (v^1,ρ^1) and (v^2,ρ^2) with $|v^i(x)|\le C$, $i=1,2$:

$$\sup_{u\in U}(L^u v^1(x)+f^u(x)-\rho^1)=\sup_{u\in U}(L^u v^2(x)+f^u(x)-\rho^2)=0.$$

Both v^1 and v^2 are classical solutions with locally Lipschitz second derivatives. Let $w(x):=v^1(x)-v^2(x)$ and consider two strategies $\alpha_1,\alpha_2\in\mathcal{A}_M$ such that $\alpha_1(x)\in \mathrm{Argmin}_{u\in U}(L^u w(x))$ and $\alpha_2(x)\in\mathrm{Argmax}_{u\in U}(L^u w(x))$, and let X^1_t,X^2_t be solutions of the SDEs corresponding to each strategy α_i, $i=1,2$. Let us denote

$$h_1(x):=\inf_{u\in U}(L^u w(x)-\rho^1+\rho^2),\ h_2(x):=\sup_{u\in U}(L^u w(x)-\rho^1+\rho^2).$$

Then,

$$h_2(x)=\sup_{u\in U}(L^u v^1(x)+f^u(x)-\rho^1-(L^u v^2(x)+f^u(x)-\rho^2))$$

$$\ge\sup_{u\in U}(L^u v^1(x)+f^u(x)-\rho^1)-\sup_{u\in U}(L^u v^2(x)+f^u(x)-\rho^2)=0.$$

Similarly,

$$h_1(x)=-\sup_u(L^u(-v^2)(x)-\rho^2+\rho^1)$$

$$=-\sup_u(L^u v^2(x)+f^u(x)+\rho^2-(L^u v^1(x)+f^u(x)+\rho^1))$$

$$\le-\left[\sup_u(L^u v^2(x)+f^u(x)-\rho^2)-\sup_u(L^u v^1(x)+f^u(x)-\rho^1)\right]=0.$$

We have,

$$L^{\alpha_2}w(x)=h_2(x)-\rho^2+\rho^1,$$

and

$$L^{\alpha_1}w(x)=h_1(x)-\rho^2+\rho^1.$$

Due to Dynkin's formula we have,

$$\mathrm{E}_x w(X^1_t)-w(x)=\mathrm{E}_x\int_0^t L^{\alpha_1}w(X^1_s)\,ds$$
$$=\mathrm{E}_x\int_0^t h_1(X^1_s)\,ds+(\rho^1-\rho^2)\,t\overset{(h_1\le 0)}{\le}(\rho^1-\rho^2)\,t.$$

Since the left hand side here is bounded for a fixed x, we get,

$$\rho^1 - \rho^2 \geq 0.$$

Similarly, considering α_2 we conclude that

$$E_x w(X_t^2) - w(x) = E_x \int_0^t L^{\alpha_2} w(X_s^2)\, ds$$

$$= E_x \int_0^t h_2(X_s^2)\, ds + (\rho^1 - \rho^2)\, t.$$

From here, due to the boundedness of the left hand side we get,

$$\rho^2 - \rho^1 = \liminf_{t\to 0} (t^{-1} E_x \int_0^t h_2(X_s^2)\, ds) \overset{(h_2 \geq 0)}{\geq} 0.$$

Thus, $\rho^1 - \rho^2 \leq 0$ and, hence,

$$\rho^1 = \rho^2.$$

<u>*Why $\rho = \tilde{\rho}$?*</u> Recall that for any initial $\alpha_0 \in \mathcal{A}_M$, the sequence ρ_n converges to the same value $\tilde{\rho}$, which is a unique component of solution of Belman's ergodic equation. Let us take any $\epsilon > 0$ and consider a strategy α_0 such that

$$\rho_0 = \rho^{\alpha_0} > \rho - \epsilon.$$

Since the sequence (ρ_n) increases, the limit $\tilde{\rho}$ must satisfy the same inequality,

$$\tilde{\rho} = \lim_{n\to\infty} \rho_n > \rho - \epsilon.$$

Due to uniqueness of $\tilde{\rho}$ as a component of solution of Bellman's ergodic equation and since $\epsilon > 0$ is arbitrary, we find that

$$\tilde{\rho} \geq \rho.$$

But also $\tilde{\rho} \leq \rho$ since $\tilde{\rho}$ is the sup of the cost function values over a smaller family of strategies. So, in fact, $\boxed{\tilde{\rho} = \rho.}$ The theorem is proved[111]. QED

[111] Up to some arguments promised from the ODE, see the Appendix immediately in the next subsection.

11.2 Appendix: ODE/PDE bounds for the Poisson equation

We are still working in the dimension $d = 1$. Here is the proof that if g_n is small in L_2 (or in any other reasonable norm) on some interval $[-N-1, N+1]$, then w_n' and w_n'' are small in $L_2[-N, N]$, and w_n' is small in the uniform norm locally. We will just solve explicitly the equation (158) (dropping n and a_n in all terms)

$$\frac{1}{2}w''(x) + b(x)w'(x) + g(x) = 0. \tag{159}$$

1. By a standard substitution $w' =: v$ and $v = u\exp(-\int_0^x 2b(z)dz) =: ue_-$, and using also the notation $\exp(+\int_0^x 2b(z)dz) =: e_+$, we get

$$u' = -ge_+, \quad u(x) = C_0 - \int_0^x 2g(x_1)e_+(x_1)dx_1,$$

$$w'(x) = v(x) = u(x)e_-(x) = C_0e_-(x) - e_-(x)\int_0^x 2g(x_1)e_+(x_1)dx_1.$$

Note that the term $e_-(x)\int_0^x 2g(x_1)e_+(x_1)dx_1$ is small for any fixed value of x if g is small in any integral (L_p), or sup-norm locally. A bit later we will explain that the value C_0 must be equal to zero; otherwise the solution w increases too fast at infinity.

2. From the above we get (recall that $e_\pm = \exp(\pm\int_0^x 2b(z)dz)$)

$$w(x) = C_1 + C_0\int_0^x e_-(x_1)dx_1 - \int_0^x e_-(x_1)\int_0^{x_1} 2g(x_2)e_+(x_2)dx_2. \tag{160}$$

The constant C_1 may be dropped at once because any additive constant does not change the fact that w is a solution; hence, we may choose this constant $C_1 = 0$. This does not mean that w will be centered precisely with such a choice; but for the arguments about the derivative w' this constant does not matter:

$$w'(x) = C_0e_-(x) - e_-(x)\int_0^x 2g(x_1)e_+(x_1)dx_1. \tag{161}$$

Denote

$$W(x) = -\int_0^x e_-(x_1)\int_0^{x_1} 2g(x_2)e_+(x_2)dx_2 \quad \& \quad V(x) = C_0\int_0^x e_-(x_1)dx_1.$$

It is easy to verify that

$$W''(x) + 2b(x)W'(x) = -2g(x) \quad \& \quad V''(x) + 2b(x)V'(x) = 0.$$

Recall that the invariant density for the invariant measure μ reads (see (150))

$$f(x) := \frac{\mu(dx)}{dx} = C \exp\left(2\int_0^x b(y)\, dy\right),$$

and that we want the integral $\int w(x)\mu(dx)$ to be finite: we consider solutions in this class, also requiring that solutions should not grow faster than any polynomial. Now, it can be checked that the integral $\int W(x)f(x)dx$ converges[112], while the integral $\int V(x)f(x)dx$ diverges[113], which explains why the constant C_0 should equal zero. Recall that by the standing centering assumption for g

$$\int g(x)f(x)dx = 0 \quad \overset{also}{\&} \quad \int |g(x)|f(x)dx < \infty.$$

3. Now, if we accept that $C_0 = 0$, then the claim that g is small on any fixed interval $[-N, N]$ in any integral or sup-norm implies that w' is also small, follows straightforwardly from the formula (161).

4. Note that even without assuming $C_0 = 0$ this claim may be justified as follows.

(1) If g is small on some fixed interval $[-N, N]$ (which is large enough) and its modulus is growing moderately at infinity, then the function $w - \beta$ is also small, which follows from the representation

$$w(x) - \beta = \int_0^\infty \mathrm{E}_x g(X_t^n) dt \tag{162}$$

and from[114] (146), where β is some bounded constant necessary for centering and for the probabilistic representation (162).

(2) Hence, taking in (160) $x = 0$, we conclude that $C_1 + \beta$ must be small.

(3) Further, taking in the same (160) $x = \epsilon$ with some small enough[115] ϵ and noticing that the term $V(x)$ increases or decreases locally approximately linearly in the neighbourhood of zero where $V(0) = 0$, while the term

[112] (Not elementary)

[113] (Easy)

[114] (146): $\sup_a \|\mu_t^{x,a} - \mu^a\|_{TV} \le \varphi(x,t) \to 0$ "fast enough", $a \in \mathcal{A}_M$.

[115] Small epsilon is required here in order make sure that the term V is not affected too much by the "second order term" $W(x)$ (as the latter is approximately quadratic in x for x small enough).

$W(x)$ changes in a neighbourhood of zero where $W(0) = 0$ approximately quadratically, we may conclude that also C_0 should be small enough. From here it follows that the whole expression for w' in (161) is small enough on a (any) fixed bounded interval (but not uniformly on the whole line), as required.

5. Now if g and w' are small in some interval, then w'' is also small in the same interval due to the equation (159). The finishes the ODE justification that in terms of the previous subsection

$$v'_{n'} - v'_{n'+1} \to 0, \quad n' \to \infty$$

over a (any) subsequence (n') such that $v_{n'}$ converges in $C^1([-N, N])$ for any fixed N.

6. The same considerations show that the derivative $(v^a)'$ of the auxiliary function (v^a) is just locally bounded under the consition (which was assumed) that the integrand $(f(X_t) - \rho^a)$ is bounded[116].

[116] A bit more precisely, f ws assumed bounded, and it implies that ρ^a is also bounded by a constant independent of the strategy a

12 Lecture 12, 26.11.2021, Optimal stopping

12.1 Elliptic Bellman equation in $\mathbb{R}^1$, revisited

In earlier lectures the elliptic Bellman equation was established "conditionally", assuming that there exists a classical (or Sobolev) solution to the corresponding nonlinear elliptic PDE. In this subsection in the case of $\mathbb{R}^1$ the complete result will be provided with sufficient conditions for such an existence (and uniqueness) of a classical solution. This will be used in what follows for the presentation of the optimal stopping problem, the main topic of this lecture.

Hence, let $r_1 < r_2$ and consider the "elliptic" stochastic optimization problem

$$v^a(x) := \mathrm{E}^a_x\left(\int_0^\tau e^{-\varphi_s} f^a(X^x_s)ds + e^{-\varphi_\tau} g(X^x_\tau)\right) \to \sup_{a\in\mathcal{A}} =: v(x),$$

where $\varphi_s = \int_0^s c(X_r)dr$; $c \geq 0$, and

$$\tau := \inf(t \geq 0 : X_t \notin [r_1, r_2]).$$

Recall that $\mathcal{A}$ is the full class of adapted strategies α_t with values in a set U. Also the class of Markov strategies $\mathcal{A}_M$ of the form $\alpha_t = a(X_t)$ plays an important role; in the case of such a strategy the diffusion becomes an equation where X_t shows up in the coefficients twice: as such and in a composite function through the strategy $a(X_t)$.

Theorem 57 (N.V. Krylov, Controlled diffusion processes, Theorem 1.4.5)**.** *Let $\sigma, \sigma^{-1}, b, f,$ be bounded, continuous and Lipschitz*[117] *in (u, x). Then $v \in C^2$, $v(r_i) = g(r_i), i = 1, 2,$ &*

$$\sup_u\left[\frac{\sigma^2(u,x)}{2}v''(x) + b(u,x)v'(x) - c(u,x)v(x) + f(u,x)\right] = 0 \quad \forall\, x \in [r_1, r_2]. \quad (163)$$

Also, v is a unique solution of this equation in C^2 with given boundary conditions.

Lemma 21 (Krylov, Lemma 1.4.6)**.** *Let $\sigma, \sigma^{-1}, b, c, f$ be continuous function of the variable x on $[r_1, r_2]$, $c \geq 0$. Then $\exists!$ solution in C^2 of the equation*

$$\left[\frac{\sigma^2}{2}v'' + bv' - cv + f\right](x) = 0 \quad \forall\, x \in [r_1, r_2], \quad (164)$$

[117](Lipschitz condition may be redundant for the reduced claim of the theorem in comparison to the exact statement in Krylov's book: it is likely that it is needed for $v'' \in Lip$, which claim I dropped.)

with b.c. $v(r_i) = g(r_i), i = 1, 2$. *A priori bounds hold,*

$$\|v''\|_B + \|v'\|_B + \|v\|_B \le N(1 + \|f\|_B), \tag{165}$$

and for $g(r_i) = 0$,

$$\|v\|_B \le N\|f\|_{L_1([r_1,r_2])}. \tag{166}$$

More than that[118]*, in the case of* $g(r_i) = 0$, $i = 1, 2$

$$\|v\|_{L_1} \le C\|u\|_B \le C\|f\|_{L_1}. \tag{167}$$

and

$$\|v'\|_B \le C(\|f\|_{L_1} + \|v\|_{L_1}) \le C\|f\|_{L_1}, \tag{168}$$

and

$$\|v''\|_{L_1} \le C(\|f\|_{L_1} + \|v\|_{L_1}) \le C\|f\|_{L_1}. \tag{169}$$

Sketch of proof[119] *of lemma 21.* Assume[120] $g(r_i) = 0, i = 1, 2$. The equation

$$\frac{\sigma^2(x)}{2} v''(x) + b(x)v'(x) - c(x)v(x) + f(x) = 0$$

is equivalent to

$$v''(x) + \frac{2b(x)}{\sigma^2(x)} v'(x) - \frac{2c(x)}{\sigma^2(x)} v(x) + \frac{2f(x)}{\sigma^2(x)} = 0.$$

Denote $\frac{2b(x)}{\sigma^2(x)} = \hat{b}(x)$, $\frac{2c(x)}{\sigma^2(x)} = \hat{c}(x)$, $\frac{2f(x)}{\sigma^2(x)} = \hat{f}(x)$. The change

$$v(x) := u(x) \exp(-\int_0^x \hat{b}(z)dz) \;\sim\; u(x) = v(x) \exp(\int_0^x \hat{b}(z)dz)$$

leads to the equation on u with $\tilde{c} \ge 0$,

$$u'' - \tilde{c}u + \tilde{f} = 0. \tag{170}$$

[118] I am not sure that these additional bounds (167) – (169) will be used in the proof of the theorem, but they *might* be used, so I leave them here. In any case, their justification just required a couple of lines.

[119] *Exercise: Justify all steps and extend to the case of discontinuous coefficients.*

[120] *Why it suffices to consider only this particular case?*

Wlog, we may assume $r_1 = 0, r_2 = 1$. Firstly consider a simplified equation with a constant $\lambda > 0$:

$$u'' - \lambda u + \tilde{f} = 0. \tag{171}$$

The equation (171) has a unique solution

$$u(x) = \int_0^1 g_\lambda(x, y)\tilde{f}(y)dy,$$

where[121]

$$g_\lambda(x, y) = \frac{\sinh(\lambda^{1/2}(x \wedge y))\sinh(\lambda^{1/2}(1 - x \vee y))}{\lambda^{1/2}\sinh\lambda^{1/2}}.$$

Equation (170) is equivalent to

$$u'' - \lambda u + \tilde{f} + (\lambda - \tilde{c})u = 0. \tag{172}$$

Denote $\bar{f} := \tilde{f} + (\lambda - \tilde{c})u$. Then

$$u'' - \lambda u + \bar{f} = 0.$$

Therefore,

[121] sinh is a hyperbolic sine; cosh is a hyperbolic cosine.

$$u(x) = \int_0^1 g_\lambda(x,y)\bar{f}(y)dy$$
$$= \int_0^1 g_\lambda(x,y)\tilde{f}(y)dy + \int_0^1 g_\lambda(x,y)(\lambda - \tilde{c}(y))u(y)dy.$$

Denote the second integral in the rhs by Tu and choose $\lambda = \|\tilde{c}\|_B$; then $\lambda - \tilde{c}(y) \geq 0$. We have,

$$\|Tu_1 - Tu_2\|_B \leq \lambda \max_x \int_0^1 g_\lambda(x,y)dy \|u_1 - u_2\|_B =: \epsilon\|u_1 - u_2\|_B.$$

It is claimed that[122]

$$\epsilon := \lambda \max_x \int_0^1 g_\lambda(x,y)dy = 1 - \frac{1}{\cosh \frac{1}{2}\lambda^{1/2}} < 1.$$

~~So, T is a contraction in B and $\exists!\, u : u = Tu$.~~[123] Now the bounds follow. We obtain[124],

$$\|u\|_B \leq \max_x g_\lambda \|\tilde{f}\|_{L_1} + \epsilon\|u\|_B \implies \|u\|_B \leq C\|\tilde{f}\|_{L_1} \;\&\; \|u\|_{L_1} \leq C\|\tilde{f}\|_{L_1}.$$

This implies

$$\|u''\|_B \leq C(\|\tilde{f}\|_B + \|u\|_B) \leq C(\|\tilde{f}\|_B + \|\tilde{f}\|_{L_1}) \leq C\|\tilde{f}\|_{B[0,1]}.$$

Also,

$$\|u''\|_{L_1} \leq C(\|\tilde{f}\|_{L_1} + \|u\|_{L_1}) \leq C\|\tilde{f}\|_{L_1}.$$

Take x_0 such that $u'(x_0) = 0$. Then

$$\|u'\|_B \leq \sup_x |\int_{x_0}^x u''(y)dy| \leq C\|\tilde{f}\|_B,$$

and also

$$\|u'\|_B \leq \sup_x |\int_{x_0}^x u''(y)dy| \leq \|u''\|_{L_1} \leq C(\|\tilde{f}\|_{L_1} + \|u\|_{L_1}) \leq C\|\tilde{f}\|_{L_1}.$$

[122] *Exercise!*
[123] True but useless.
[124] *Exercise!*

Returning to $v(x) = u(x)\exp(-\int\limits_0^x \hat{b}(z)dz)$, we get (165) and (166), as well as (167) – (169). The lemma is proved. QED

Proof of theorem 57 is based on the Howard algorithm and on the a priori bounds from the lemma, similarly to the case of the ergodic control.

1. Let a_0 be any initial markovian strategy, and

$$v_0(x) = \mathrm{E}_x^{a_0}\left(\int_0^\tau e^{-\varphi_s} f^{a_0}(X_s^x)ds + e^{-\varphi_\tau} g(X_\tau^x)\right).$$

According to the lemma, $v_0 \in C^2$ and the a priori bounds (165)– (169) hold true. Given that (a_n, v_n) are defined, let

$$a_{n+1}(x) \in \mathrm{Argsup}_{u\in U}[L^u v_n(x) + f^u(x)],$$

and

$$v_{n+1}(x) = \mathrm{E}_x^{a_{n+1}}\left(\int_0^\tau e^{-\varphi_s} f^{a_{n+1}}(X_s^x)ds + e^{-\varphi_\tau} g(X_\tau^x)\right).$$

(NB: $v_{n+1}(x)$ must be, of course, Borel measurable.) This is Howard's algorithm.

2. We have

$$L^{a_n} v_n(x) + f^{a_n}(x) = 0,\ x \in (r_1, r_2), \quad v_n(r_i) = g(r_i),\ i = 1, 2,$$

or,

$$v_n''(x) + \hat{b}^{a_n}(x)v_n'(x) - \hat{c}^{a_n}(x)v_n(x) + \hat{f}^{a_n}(x) = 0,\ x \in (r_1, r_2), \quad v_n(r_i) = g(r_i),\ i = 1, 2,$$

where $\hat{b}^{a_n}(x) = \dfrac{2b^{a_n}}{(\sigma^{a_n})^2}(x)$, $\hat{c}^{a_n}(x) = \dfrac{2c^{a_n}}{(\sigma^{a_n})^2}(x)$, $\hat{f}^{a_n}(x) = \dfrac{2f^{a_n}}{(\sigma^{a_n})^2}(x)$, and

$$L^{a_{n-1}} v_{n-1}(x) + f^{a_{n-1}}(x) = 0,\ x \in (r_1, r_2), \quad v_{n-1}(r_i) = g(r_i),\ i = 1, 2,$$

and

$$L^{a_n} v_{n-1}(x) + f^{a_n}(x) \ge 0,\ x \in (r_1, r_2), \quad v_{n-1}(r_i) = g(r_i),\ i = 1, 2.$$

Hence,

$$g_n(x) := L^{a_n}(v_{n-1} - v_n)(x) \ge 0, \quad (v_{n-1}(x) - v_n)(r_1) = (v_{n-1}(x) - v_n)(r_2) = 0.$$

Dynkin's formula applied to $(v_{n-1} - v_n)(X^{a_n})$ from 0 to τ (this stopping time, of course, also depends on the strategy via X_t) shows that

$$0 - (v_{n-1} - v_n)(x) = \mathrm{E}_x \int_0^\tau L^{a_n}(v_{n-1} - v_n)(X_s^{a_n})ds \geq 0.$$

Thus,

$$v_n(x) \geq v_{n-1}(x).$$

Naturally, a monotone bounded sequence $(v_n(x), n \geq 0)$ has a bounded limit,

$$\tilde{v}(x) := \lim_{n\to\infty} v_n(x). \tag{173}$$

Naturlaly, it follows that for each $x \in [r_1, r_2]$

$$w_n(x) := v_{n-1}(x) - v_n(x) \to 0, \quad n \to \infty.$$

3. Note that due to the uniform boundedness of all second and first derivatives $|v_n''(x)| \leq C$, $|v_n'(x)| \leq C$ as well as of v_n themselves (see the lemma) the sequence (v_n) is pre-compact in C^1. Let $n' \to \infty$ be a subsequence such that

$$(v_{n'}, v_{n'}') \underset{C}{\to} (\hat{v}, \hat{v}^1)$$

for some $\hat{v}, \hat{v}^1 \in C([r_1, r_2])$. This is, of course, only possible if

$$\hat{v} = \tilde{v}.$$

In particular, it follows that $\tilde{v} \in C([r_1, r_2])$ and, hence, the convergence in (173) is uniform:

$$v_n(x) \underset{C}{\to} \tilde{v}(x), \quad n \to \infty. \tag{174}$$

Moreover, from the equality

$$v_{n'}(y) - v_{n'}(x) = \int_x^y v_{n'}'(s)ds$$

in the limit $(\tilde{v}(y) - \tilde{v}(x) = \int_x^y \hat{v}^1(s)ds)$ it follows that $\hat{v}^1(x) = \tilde{v}'(x)$, so that $\tilde{v}$ is differentiable and we have

$$(v_{n'}, v_{n'}') \underset{C}{\to} (\tilde{v}, \tilde{v}'). \tag{175}$$

4. Let us show that

$$v'_{n'-1} \underset{C}{\rightarrow} \tilde{v}', \tag{176}$$

possibly over some further sub-subsequence (which suffices for our aim). Strictly speaking, this does not follow straightforwardly from (175) because we are not aware whether or not the indices $n'-1$ are included in the subsequence (n'). However, the sequence $v_{n'-1}$ is also pre-compact in C^1, so we may choose a sub-subsequence $(n'') \subset (n'-1) \subset (n)$ such that the pair $(v_{n''}, v'_{n''})$ converges in C to some couple $(\bar{v}, \bar{v}^1)$:

$$(v_{n''}, v'_{n''}) \underset{C}{\rightarrow} (\bar{v}, \bar{v}^1).$$

Here naturally

$$\bar{v} \equiv \tilde{v},$$

and

$$\bar{v}^1 = \bar{v}' = \tilde{v}',$$

and so, actually,

$$(v_{n''}, v'_{n''}) \underset{C}{\rightarrow} (\tilde{v}, \tilde{v}'). \tag{177}$$

5. So, we have simultaneously (177) and (175). Now let us pass to the limit in the equation

$$v'_{n''+1}(y) - v'_{n''+1}(x) = \int_x^y v''_{n''+1}(s)ds$$

$$= \int_x^y (-\hat{b}^{a_{n''+1}} v'_{n''} + \hat{c}^{a_{n''+1}} v_{n''} - \hat{f}^{a_{n''+1}})(s)ds = -\int_x^y \sup_{u\in U}(\hat{b}^u v'_{n''} - \hat{c}^u v_{n''} + \hat{f}^u)(s)ds.$$

Due to (177) and (175) and using that

$$\sup_{u\in U}(\hat{b}^u v'_{n''} - \hat{c}^u v_{n''} + \hat{f}^u)(x) \rightrightarrows \sup_{u\in U}(\hat{b}^u \tilde{v}' - \hat{c}^u \tilde{v} + \hat{f}^u)(x),$$

we get in the limit

$$\tilde{v}'(y) - \tilde{v}'(x) = -\int_x^y \sup_{u\in U}(\hat{b}^u \tilde{v}' - \hat{c}^u \tilde{v}' + \hat{f}^u)(s)ds.$$

The right hand side here is differentiable wrt y; hence, the lhs is also differentiable and

$$\tilde{v}''(y) + \sup_{u\in U}(\hat{b}^u \tilde{v}' - \hat{c}^u \tilde{v}' + \hat{f}^u)(y) = 0. \tag{178}$$

The latter equation is equivalent to the desired Bellman equation (163). Hence, solution exists.

5. Suppose there are two solutions v_1 and v_2 of the equation (163), or, equivalently, (178). Denote $w := v_1 - v_2$; note that $w(r_i) = 0$, $i = 1, 2$. Let us determine two (Borel measurable) strategies, $a_1, a_2 \in \mathcal{A}_M$ by the rule

$$a_1(x) \in \mathrm{Argsup}_{u\in U}[L^u v_1 + f^u](x), \quad a_2(x) \in \mathrm{Argsup}_{u\in U}[L^u v_2 + f^u](x).$$

Note that

$$L^1 v_1(x) = 0,\ L^2 v_1(x) \le 0,\ L^1 v_2(x) \le 0,\ L^2 v_2(x) = 0.$$

Therefore,

$$L^1 w(x) = L^1(v_1 - v_2)(x) \ge 0, \ \ L^2 w(x) = L^2(v_1 - v_2)(x) \le 0.$$

Let X^i_t be solutions of the SDEs corresponding to each strategy a_1, a_2 and let $L^i = L^{a_i}$. By Dynkin's formula applied to $\exp(\int_0^t c(X^1_s)ds)w(X^1_t)$ from 0 to τ we have

$$0 - w(x) = \mathrm{E}_x \int_0^\tau L^1 w(X^1_s)ds \ge 0 \iff w(x) \le 0.$$

On the other hand, applying Dynkin's formula to $\exp(\int_0^t c(X^2_s)ds)w(X^2_t)$ from 0 to τ we obtain

$$0 - w(x) = \mathrm{E}_x \int_0^\tau L^2 w(X^2_s)ds \le 0 \iff w(x) \ge 0.$$

Overall, we got the equality $w(x) = 0$, which means uniqueness of solution, as required.

6. Let us finally show that $\tilde v(x) = v(x)$. For this aim let us choose the initial strategy a_0 so that $v_0(x) \ge v(x) - \epsilon$. Then $\tilde v(x) \ge v(x) - \epsilon$ because $\tilde v(x) \ge v_0$ and since the solution of Bellman's equation is unique (so that we get the same $\tilde v$ starting from whatever initial strategy). On the other hand, $\tilde v(x) \le v(x)$ since $\tilde v(x)$ is a supremum over countably many markovian strategies, while $v(x)$ is a supremum over all strategies. Hence,

$$\tilde v(x) = v(x),$$

as required. **QED**

12.2 Optimal stopping on the infinite horizon

Consider an SDE (without control)

$$dX_t = \sigma(X_t)dW_t + b(X_t)dt, \ X_0 = x. \tag{179}$$

The optimal stopping problem for this SDE, or for the process (X_t) is stated as follows: how to choose a *stopping time* τ which would maximize the expectation

$$V^\tau(x) = \mathrm{E}_x \left(\int_0^\tau e^{-cs} f(X_s) ds + e^{-c\tau} g(X_\tau) \right) \to \max_\tau =: \bar{V}(x).$$

Here before we state a theorem, let us agree how to understand the expectation here if $\tau = \infty$. For g bounded and $c > 0$ clearly $\lim_{t\to\infty} e^{-ct} g(X_t) = 0$; so, on ω where $\tau(\omega) = \infty$ we agree that $e^{-c\tau} g(X_\tau) = 0$. There is no problem with the interpretation of the integral term for f bounded and $c > 0$: in the case of $\tau = \infty$, it is just the value of $\int_0^\infty e^{-cs} f(X_s) ds$ without the additive term involving the function g.

Theorem 58. *Let $c > 0$, and let f and g be bounded and continuous. Then the function $\bar{V}$ is a bounded solution of the* **Bellman's equation for optimal stopping problem**

$$\max(LV(x) - cV(x) + f(x), g(x) - V(x)) = 0, \tag{180}$$

where L is the generator of X_t.

In other words, everywhere an **inequality** holds true (in fact, two inequalities)

$$(LV - cV(x) + f(x)) \le 0 \quad \& \quad V(x) \ge g(x);$$

and where $V(x) > g(x)$, we should get an **equality**

$$(LV - cV(x) + f(x)) = 0;$$

vice versa, if $(LV - cV(x) + f(x)) < 0$ at some state x, then we must have

$$V(x) = g(x).$$

Proof of theorem 58 (conditional on the property that $V \in C_b^2$).

0. The stopping problem may be stated in the style of Prince Hamlet: "to stop or not to stop"? Note that if we stop immediately (which is well allowed), then our profit will be exactly $g(x)$. Hence, there should be $\bar V(x) \ge g(x)$. Suppose $\bar V(x) > g(x)$ and assume that we know that the function $\bar V$ is continuous. Then clearly $\mathbb{P}_x(\tau > 0) > 0$. There is a **natural principle**[125]: if it is profitable at some x to continue (= not to stop) with a positive probability, then it is profitable to continue with probability one. Hence, if $\bar V(x) > g(x)$, then we should have

$$\mathbb{P}(\tau > 0) = 1.$$

1. Let us show that $V(x) \ge \bar V(x)$. Let any stopping rule[126] τ be applied with values $0 \le \tau \le +\infty$. Suppose that Ito's or Dynkin's formula is applicable to $\exp(-ct)V(X_t)$ in some neighbourhood of x; we do not stop, at least, until where $V \le g$ (which means, in fact, $V = g$, as $V \ge g$ everywhere by the Bellman stopping equation). We get,

$$d\exp(-ct)V(X_t) = \exp(-ct)(LV(X_t) - cV(X_t))dt$$

$$+ \exp(-ct)V'(X_t)\sigma(X_t)dW_t$$

So, assuming that the expectation of the stochastic integral equals zero (otherwise the same result can be justified by using some auxiliary localising sequence of stopping times), we write

$$\mathrm{E}_x \exp(-c(\tau\wedge t))V(X_{\tau\wedge t}) = V(x) + \mathrm{E}_x \int_0^{\tau\wedge t} \exp(-cs)(LV(X_s) - cV(X_s))ds. \tag{181}$$

Then

$$\mathrm{E}_x \exp(-c(\tau\wedge t))g(X_{\tau\wedge t}) \le \mathrm{E}_x \exp(-c(\tau\wedge t))V(X_{\tau\wedge t})$$

$$= V(x) + \mathrm{E}_x \int_0^{\tau\wedge t} \exp(-cs)(LV(X_s) - cV(X_s))ds$$

$$\le V(x) - \mathrm{E}_x \int_0^{\tau\wedge t} \exp(-cs)f(X_s)ds$$

[125] Although, non-rigorous.

[126] We highlight that τ must be a stopping time for a chosen filtration.

Hence,

$$V(x) \geq \mathrm{E}_x \left(\int_0^{\tau \wedge t} \exp(-cs) f(X_s) ds + \exp(-c(\tau \wedge t)) g(X_{\tau \wedge t}) \right).$$

As $t \to \infty$, we get by Fatou's lemma (recall that f, g are assumed bounded)

$$V(x) \geq \mathrm{E}_x 1(\tau < \infty) \left(\int_0^{\tau} \exp(-cs) f(X_s) ds + \exp(-c\tau)) g(X_\tau) \right)$$

$$+\mathrm{E}_x 1(\tau = \infty) \int_0^{\infty} \exp(-cs) f(X_s) ds$$

$$= \mathrm{E}_x \left(\int_0^{\tau} \exp(-cs) f(X_s) ds + \exp(-c\tau)) g(X_\tau) \right)$$

according to our agreement about $\tau = \infty$. The inequality

$$V(x) \geq \mathrm{E}_x \left(\int_0^{\tau} \exp(-cs) f(X_s) ds + \exp(-c\tau)) g(X_\tau) \right)$$

is true for any stopping rule $0 \leq \tau \leq \infty$; hence,

$$V(x) \geq \sup_\tau \mathrm{E}_x \left(\int_0^{\tau} \exp(-cs) f(X_s) ds + \exp(-c\tau)) g(X_\tau) \right) = \bar{V}(x).$$

2. Let us now show the opposite inequality $V \leq \bar{V}$. We have already seen that

$$\mathrm{E}_x \exp(-c(\tau \wedge t)) V(X_{\tau \wedge t}) \stackrel{(181)}{=} V(x) + \mathrm{E}_x \int_0^{\tau \wedge t} \exp(-cs)(LV(X_s) - cV(X_s)) ds.$$

On the other hand, let us look more attentively at this equation for the stopping time $\tau^* := \inf(t \geq 0 : V(X_t) = g(X_t))$ ($\inf(\emptyset) = \infty$): we still have

$$\mathrm{E}_x \exp(-c(\tau^* \wedge t)) V(X_{\tau^* \wedge t}) = V(x) + \mathrm{E}_x \int_0^{\tau^* \wedge t} \exp(-cs)(LV(X_s) - cV(X_s)) ds.$$

As $t \to \infty$, in the right hand side we have convergence due to the Lebesgue dominated convergence theorem (at least, if f is bounded); that is,

$$\mathrm{E}_x \int_0^{\tau^* \wedge t} \exp(-cs)(LV(X_s) - cV(X_s)) ds \to \mathrm{E}_x \int_0^{\tau^*} \exp(-cs)(LV(X_s) - cV(X_s)) ds.$$

Here in the left hand side due to the continuity of V we have a convergence

$$V(X_{\tau^*\wedge t}) \to V(X_{\tau^*})$$

on the set where $\tau^* < \infty$. Since our solution V is bounded (which is reasonable since clearly $\bar{V}$ is bounded not exceeding the value $c^{-1}\|f\|_B + \|g\|_B$), then we have

$$\mathrm{E}_x 1(\tau^* < \infty) \exp(-c(\tau^* \wedge t)) V(X_{\tau^*\wedge t}) \to \mathrm{E}_x 1(\tau^* < \infty) \exp(-c\tau^*) V(X_{\tau^*})$$

$$= \mathrm{E}_x 1(\tau^* < \infty) \exp(-c\tau^*) g(X_{\tau^*}).$$

What happens on the set $\tau^* = \infty$? There we have

$$\mathrm{E}_x 1(\tau^* = \infty) \exp(-c(\tau^* \wedge t)) V(X_{\tau^*\wedge t}) = \mathrm{E}_x 1(\tau^* = \infty) \exp(-ct) V(X_t) \to 0.$$

So, we obtain

$$\mathrm{E}_x 1(\tau^* < \infty) \exp(-c\tau^*) V(X_{\tau^*}) = V(x) + \mathrm{E}_x \int_0^{\tau^*\wedge t} \exp(-cs)(LV(X_s) - cV(X_s))ds,$$

or, equivalently,

$$\mathrm{E}_x 1(\tau^* < \infty) \exp(-c\tau^*) g(X_{\tau^*}) = V(x) + \mathrm{E}_x \int_0^{\tau^*} \exp(-cs)(-f(X_s))ds.$$

Since $1(\tau^* < \infty)$ may be dropped here, we obtain

$$V(x) = E_x \exp(-c\tau^*) g(X_{\tau^*}) + E_x \int_0^{\tau^*} \exp(-cs)(f(X_s))ds = V^{\tau^*}(x).$$

However, if

$$V(x) = \mathrm{E}_x \left(\int_0^{\tau^*} \exp(-cs) f(X_s) ds + \exp(-c\tau^*) g(X_{\tau^*}) \right),$$

it signifies that $V(x)$ coincides with the profit obtained by using one particular stopping rule, $V^{\tau^*}(x)$. So,

$$V(x) \le \bar{V}(x).$$

Overall, it follows that

$$V(x) = \bar{V}(x),$$

as required. QED

For further reading see the monograph by A.N. Shiryaev.

Список литературы

[1] Ширяев А.Н. Статистический последовательный анализ. Оптимальные правила остановки, М.: ФИЗМАТЛИТ, 1976. - 272 с. 2-е изд., перераб.

12.3 Killing approach to optimal stopping on a finite horison

We study the same process which is a solution of the SDE (179) (e.g., $dX_t = \sigma(X_t)dW_t + b(X_t)dt, \;\; X_0 = x$). "Killing" with a positive potential or rate means stopping with a certain intensity, or probability of stop at any small interval of time; at the moment of killing the particle disappears: it does not just stop and remains at the state where it is, but it goes to the "cemetery", a special remote artificial state, sometimes denoted by ∞, or by some other symbol (e.g., ∂_∞). It turns out that the optimal stopping – the rule to stop, e.g., upon attaining some level – may be regarded as a limiting case of stopping via killing with an indefinite increase of the upper bound for the rate of killing and with some change in the instantaneous cost function f. The rate or intensity of killing is denoted by c^a, and we will need a notation for the itegrated c^a on the trajectory of the process

$$\varphi_{t_0,t} = \int_{t_0}^t c^a_s ds, \quad \text{or} \quad \varphi_{t_0,t} = \int_{t_0}^t c^a(X_s) ds, \quad t \ge t_0.$$

The value of c^a may be any number between c and n; eventually, we will let n go to infinity; the value of c is either strictly positive if the process evloves in the whole space $\mathbb{R}^d$, or even $c = 0$ if the process is stopped on the exit from some bounded domain[127], or if the process is considered on a fixed finite horizon. In this and in the next subsection the latter case is tackled. The value function is given by the formula

$$v(t,x) = \sup_{a \in \mathcal{A}_M} \mathrm{E}^a_{t,x} \left(\int_t^T e^{-\varphi^a_{t,s}} f^a(X^{t,x}_s) ds + e^{-\varphi_{t,T}} g(X^{t,x}_T) \right), \tag{182}$$

where

$$f^a(x) = f(x) + c^a g(x).$$

Note that initially f did not depend on a. In what follows we will accept for simplicity of presentation that $c^a = a$. This is a convention which does not lead to any loss of a generality. With this agreement we write

$$f^a(x) = f(x) + a g(x),$$

[127] In this case we would require the diffusion coefficient nondegenerate.

and

$$\varphi_{t_0,t} = \int_{t_0}^{t} a_s ds, \quad \text{or} \quad \varphi_{t_0,t} = \int_{t_0}^{t} a(X_s) ds, \quad t \geq t_0,$$

and the value function is still given by the formula (182).

Recall that the "usual" (i.e., for a controlled diffusion, not in the optimal stopping setting) parabolic Bellman equation on v reads

$$v_t(t,x) + \sup_{a\in U} [(L^a - a)v(t,x) + f^a(t,x)] = 0, \ v(T,x) = g(x).$$

In the case of "just" a control via stopping, $L^a = L$ does not depend on control; the only way to control is choosing $a = a(x) \in [0, K]$ (or, possibly, $a = a(x) \equiv a(x) \in [K_0, K]$). So, we have a little simplified equation,

$$v_t(t,x) + \sup_a [(L-a)v(t,x) + f(x) + ag(x)] = 0, \ v(T,x) = g(x). \tag{183}$$

If the evolution takes place in a domain D (usually, bounded), then the equation holds for $x \in D$ and the b.c. should be added:

$$v(t,x)|_{x\in\partial D} = g(x).$$

(Here g is **the same** as in the Bellman equation (183).)

The intuitive meaning of the transformation of f to f^a is as follows: without the additional term $+ag(x)$ under the integral, at the moment of killing, if it occurs, we get nothing because the particle disappears and there is no terminal payoff; the whole profit is then just the integral $\int_0^\tau \exp(-cs) f(X_s) ds$. However, we aim to receive $g(X_\tau)$, where τ is the moment of killing (which is, of course, random, being a stopping time). Now with $+ag(x)$ we interpret $a\,dt$ as the infinitesimal probability of killing at t (or, on $(t, t+dt)$), and the multiplier g signifies that we receive $g(x)$, as if killing just meant that the particle stops, not disappears.

In other words[128],

$$\mathrm{E}^a_{t,x} \int_t^T e^{-\varphi_{t,s}} a_s g(X^{t,x}_s) ds = \mathrm{E}^a_{t,x} g(X^{t,x}_\tau) 1(\tau \leq T).$$

[128] Note that $\mathbb{P}(\tau = s) = 0$ for any fixed nonrandom s, given that the intensity c^a is bounded. So, in the next formula $1(\tau \leq T)$ is equivalent to $1(\tau < T)$.

The left hand side here is just a consequence of the tower property of conditional expectations, it uses the conditional density of τ given the trajectory of X. For example, if $a = \text{const}$, then τ does not depend on the evolution of the process X_t , and, as it was expected,

$$\mathrm{E}^a_{t,x} g(X^{t,x}_\tau) 1(\tau \le T) = \int_t^T a e^{-a(s-t)} \mathrm{E}^a_{t,x} g(X^{t,x}_s) ds.$$

The term $f(X_s)\exp(-\varphi_{t,s})ds$ is interpreted differently, as an instantaneous payment (profit or loss depending on the sign of f) on each infinitesimal unit of time ds. This term is multiplied by $\exp(-\varphi_{t,s})$ which is the (conditional) probability of no killing until time s. That is, after killing occurs, the instantaneous cost becomes zero. For example, if $a = \text{const}$, then τ does not depend on X_t, and, as it was expected,

$$\int_t^T a e^{-a(s-t)} \mathrm{E}^a_{t,x} f(X^{t,x}_s) ds$$

which may be regarded as the average profit until T; it already takes into account the possibility of killing earlier that T.

It is a good question how to distinguish those two terms, f and $ag(x)$ in the case of $a \equiv 1$. Both can be interpreted either way: as an instantaneous cost, or as a final payment at the moment of killing. More than that, if $a \not\equiv 1$, then f and $a\tilde f$ with $\tilde f = f/a$ also admit the same two different interpretations.

12.4 Killing approach in a bounded interval, case $d = 1$

We consider the same SDE (179) in the dimension $d = 1$ here[129]. A completely rigorous result will be stated and proved in this subsection. Let $D = (r_1, r_2)$, $T = \infty$, $U \ni a \ge 0$, $\varphi_s = \varphi_{0,s}$,

$$\tau := \inf(t \ge 0 : X_t \notin (r_1, r_2)).$$

Let

$$v^\gamma(x) := \mathrm{E}_x \left(\int_0^{\tau\wedge\gamma} f(X^x_s) ds + g(X^x_{\tau\wedge\gamma}) \right),$$

for any stopping time γ and the **value function** for the optimal stopping problem

$$w(x) := \sup_\gamma v^\gamma(x). \tag{184}$$

[129] $dX_t = \sigma(X_t)dW_t + b(X_t)dt, \ X_0 = x.$

Denote[130] $\mathcal{A}_n = \{\text{Markov strategies with } 0 \le a \le n\}$, $\mathcal{A}_M = \bigcup_n \mathcal{A}_n$,

$$\tilde{v}^a(x) := \mathrm{E}^a_x \left(\int_0^\tau e^{-\varphi^a_s} f^a(X^x_s) ds + e^{-\varphi^a_{t,\tau}} g(X^x_\tau) \right),$$

where $f^a(x) = f(x) + ag(x)$. We will also need auxiliary functions

$$\tilde{v}_n(x) := \sup_{a \in \mathcal{A}_n} \tilde{v}^a(x), \quad \tilde{v}(x) := \lim_{n\to\infty} \tilde{v}_n(x).$$

Lemma 22 (N.V. Krylov, Controlled diffusion processes, Lemma 1.5.2)**.** *Let σ, b be Lipschitz continuous in x and bounded along with σ^{-1}, $f \in C$, $g \in C^2[r_1, r_2]$. Then there exists $N > 0$ such that:*

(a) $w(x) = \tilde{v}(x)$;

(b) $0 \le w(x) - \tilde{v}_n(x) \le n^{-1}N$;

(c) $\tilde{v}_n \in C^2$, $\tilde{v}''_n \in Lip$;

(d) $[\frac{\sigma^2}{2}\tilde{v}''_n + b\tilde{v}'_n + f] + n(g - \tilde{v}_n)_+ = 0$ *on* $[r_1, r_2]$, *and* $\tilde{v}_n(r_i) = g(r_i)$;

(e)

$$\|\tilde{v}''_n\|_B + \|\tilde{v}'_n\|_B + \|\tilde{v}_n\|_B \le N.$$

Proof. Recall the notation $f^a(x) = f(x) + ag(x)$.

1. Bellman's equation for $\tilde{v}_n$ reads:

$$\sup_{0\le u\le n} [(\sigma^2(x)/2)\tilde{v}''_n(x) + b(x)\tilde{v}'_n(x) + f(x) + ug(x) - u\tilde{v}_n(x)] = 0, \ r_1 < x < r_2,$$

with the boundary conditions $\tilde{v}_n(r_1) = g(r_1)$ and $\tilde{v}_n(r_2) = g(r_2)$. Here sup is attained at $u = 0$ if $g < \tilde{v}_n$, and at $u = n$ if $g \ge \tilde{v}_n$. Hence, we get the desired Bellman equation for the point (d) of the lemma:

$$(d) \qquad [\frac{\sigma^2}{2}\tilde{v}''_n + b\tilde{v}'_n + f] + n(g - \tilde{v}_n)_+ = 0, \ \& \ \tilde{v}_n \in C^2.$$

Note that $[\frac{\sigma^2}{2}\tilde{v}''_n + b\tilde{v}'_n + f] \le 0$ for all x.

[130]For the brevity, we skip the difference between Markov and general admissible strategies; as usual, the sufficiency of markovian strategies can be proved in this situation.

In the limit we hope to get from this equation that

$$[\frac{\sigma^2}{2}w'' + bw' + f](x) = 0$$

where $g - w < 0$, and that

$$[\frac{\sigma^2}{2}w'' + bw' + f] \le 0, \quad \forall x,$$

and that $g = w$ where $[\frac{\sigma^2}{2}w'' + bw' + f] < 0$. It will follow from (b). But these statements will be stated in the next theorem after the proof of the lemma.

2. Parts (c) and (e) follow from theorem 57 or from lemma 21 above.

3. Let us establish a priori bounds on $w - \tilde v_n$. Let $g_n := g - (g - \tilde v_n)_+$. We get from the already established part (d) and from Dynkin's formula that

$$\mathrm{E}_x \tilde v_n(X_{\tau\wedge\gamma}) - \tilde v_n(x) = \mathrm{E}_x \int_0^{\tau\wedge\gamma} L\tilde v_n(X_s)ds$$

$$= \mathrm{E}_x \int_0^{\tau\wedge\gamma} [-f(X_t) - n(g - \tilde v_n)_+(X_t))]dt.$$

So,

$$\tilde v_n(x) = \mathrm{E}_x(\int_0^{\tau\wedge\gamma} (f(X_t) + n(g - \tilde v_n)_+(X_t))dt + \tilde v_n(X_{\tau\wedge\gamma})). \tag{185}$$

Here $\tilde v_n \ge g - (g - \tilde v_n)_+ = g \wedge \tilde v_n =: g_n$ (in particular, $g_n \le g$, which will be used a bit later). So, for the stopping time $\gamma^* := \inf(t \ge 0 : \tilde v_n(X_t) = g_n(X_t))$, or, equivalently, $\gamma^* := \inf(t \ge 0 : \tilde v_n(X_t) \le g_n(X_t))$ we have $\tilde v_n(X_t) > g_n(X_t))$, $t < \gamma^*$. Therefore, for this $\gamma^* \le \tau$ we get

$$\tilde v_n(x) = \mathrm{E}_x(\int_0^{\tau\wedge\gamma^*} f(X_t)dt + \tilde v_n(X_{\tau\wedge\gamma^*})) = \mathrm{E}_x(\int_0^{\tau\wedge\gamma^*} f(X_t)dt + g_n(X_{\tau\wedge\gamma^*})). \tag{186}$$

The equality (186) signifies that actually[131]

$$\tilde v_n(x) = \sup_\gamma \mathrm{E}_x \left(\int_0^{\tau\wedge\gamma} f(X^x_s)ds + g_n(X^x_{\tau\wedge\gamma})\right). \tag{187}$$

[131] The comparison of (186) and (187) may look a bit strange, but remember that the definition of g_n invloves $\tilde v_n$.

This provides a tool to estimate the difference $w-\tilde v_n$. Recall that $g \ge g_n$ (see above). So,

$$w(x) \ge \sup_\gamma \mathrm{E}_x \left(\int_0^{\tau\wedge\gamma} f(X_s^x)ds + g_n(X_{\tau\wedge\gamma}^x)\right) = \tilde v_n(x).$$

We have,

$$|g-g_n| = (g-\tilde v_n)_+ \le |g-\tilde v_n|.$$

Hence, due to the elementary hint $\sup(A)-\sup(B) \le \sup(A-B)$, we estimate

$$0 \le w(x) - \tilde v_n(x)$$

$$= \sup_\gamma \mathrm{E}_x \left(\int_0^{\tau\wedge\gamma} f(X_s^x)ds + g(X_{\tau\wedge\gamma}^x)\right) - \sup_\gamma \mathrm{E}_x \left(\int_0^{\tau\wedge\gamma} f(X_s^x)ds + g_n(X_{\tau\wedge\gamma}^x)\right)$$

$$\le \sup_\gamma E_x|g-g_n|(X_{\gamma\wedge\tau}) \le \sup_x |g-\tilde v_n|(x).$$

In order to evaluate $|g-\tilde v_n|$ let us write down the equation on the difference $\tilde v_n - g$: since $[\frac{\sigma^2}{2}\tilde v_n'' + b\tilde v_n' + f] + n(g-\tilde v_n)_+ = 0$ (see (d)), we obtain

$$\sigma^2/2(\tilde v_n - g)'' + b(\tilde v_n - g)' - n(\tilde v_n - g) + F_n = 0, \tag{188}$$

where due to the elementary inequalities $a_+ - a \ge 0$ and[132] $a \ge -a_-$ we have

$$F_n = f + n(g-\tilde v_n)_+ + n(\tilde v_n - g) + Lg$$

$$= f + n(g-\tilde v_n)_+ - n(g-\tilde v_n) + Lg \ge f + Lg \ge -(f+Lg)_-.$$

From the Poisson equation (188) on the interval (r_1, r_2) with the killing n we get

$$(\tilde v_n - g)(x) = \mathrm{E}_x \int_0^\tau e^{-nt} F_n(X_t)dt \ge -\mathrm{E}_x \int_0^\tau e^{-nt}(f(X_t) + Lg(X_t))_- dt$$

$$\ge -\mathrm{E}_x \int_0^\infty e^{-nt} N dt = -\frac{1}{n}N.$$

[132] Here $a_+ = \max(a,0)$ and $a_- = -\min(a,0)$; clearly, $a + a_- = a - \min(a,0) \ge 0$.

It follows that

$$\|w - \tilde{v}_n\| \le \sup_x |g - \tilde{v}_n|(x) \le \frac{1}{n} N,$$

which in particular shows (b). The claim (a) follows straightforwardly. QED

Now we can state and prove the main result of this subsection about the value function w given in (184), which is a simplification of the corresponding theorem given in the reference.

Theorem 59 (N.V. Krylov, Controlled diffusion processes, Theorem 1.5.3)**.**
(a) The functions w and w' are continuous; w' is absolutely continuous; w'' is bounded on the closed interval $[r_1, r_2]$;
(b) $w \ge g$; $w(r_i) = g(r_i)$, $i = 1, 2$; $Lw + f \le 0$; $Lw + f = 0$ on the set $[r_1, r_2] \cap (w > g)$;

Proof. **1.** The continuity of w and w' follows from the lemma. Further, it follows from lemma (b) that on any connected sub-interval where $w > g$, we have $\tilde{v}_n > g$ for n large enough. Then on any compact subset of any interval where $w > g$ we get for large values of n

$$(g - \tilde{v}_n)_+ = 0,$$

and, hence, the equation (d) of the lemma reduces on such subsets and for such n to the following:

$$[(\sigma^2/2)\tilde{v}_n'' + b\tilde{v}_n' + f] = 0,$$

or, equivalently, to

$$[\tilde{v}_n'' + (2b/\sigma^2)\tilde{v}_n' + (f/\sigma^2)] = 0.$$

In the integral version by the First Theorem of the Calculus (aka Newton – Leibniz formula)

$$\tilde{v}_n'(x_1) - \tilde{v}_n'(x_2) = \int_{x_1}^{x_2} \tilde{v}_n''(s)ds$$

$$= -\int_{x_1}^{x_2} [b\tilde{v}_n'(s) + f(s) + n(g - \tilde{v}_n)_+(s)]ds$$

Using Arzela – Ascoli theorem we obtain in the limit over some subsequence that $|w'(x) - w'(y)| \le C|x - y|$ due to the a priori bound on $\tilde{v}_n'$. So, w' is absolutely continuous and, moreover, $\exists\, w''$ almost everywhere, & $|w''| \le N$.

2. More than that, where $w > g$, we have $\tilde{v}_n > g$ for n large enough; so, $(g - \tilde{v}_n)_+ = 0$ on this set, and the equation actually reads

$$\tilde{v}_n'(x_1) - \tilde{v}_n'(x_2) = -\int_{x_1}^{x_2} [b\tilde{v}_n'(s) + f(s)]ds.$$

For such x_1, x_2 in the limit we get

$$w'(x_1) - w'(x_2) = -\int_{x_1}^{x_2} [bw'(s) + f(s)]ds.$$

which is equivalent (a.e.) to the Poisson equation

$$\frac{\sigma^2}{2} w'' + bw' + f = 0.$$

3. On the complement to the set $w > g$ we have $w = g$ and everywhere we have

$$L\tilde{v}_n + f \leq 0$$

(see lemma (d)). In the integral form it means for any $r_1 \leq x_1 < x_2 \leq r_2$ that

$$\tilde{v}_n'(x_2) - \tilde{v}_n'(x_1) - \int_{x_1}^{x_2} [b\tilde{v}_n'(s) + f(s)]ds \leq 0.$$

In the limit over some subsequence it implies

$$w_n'(x_2) - w_n'(x_1) - \int_{x_1}^{x_2} [bw_n'(s) + f(s)]ds \leq 0.$$

or, equivalently,

$$\frac{\sigma^2}{2} w'' + bw' + f \leq 0.$$

Recall that $\frac{\sigma^2}{2} w'' + bw' + f \leq 0$ on any interval where $w > g$. QED

12.5 Appendix: Bellman's integral principle rigorously, $d = 1$

Let $d = 1$. Here the Bellman integral principle is rigorously proved for the controlled diffusion on the interval $[r_1, r_2]$. In this setting the coefficients of the SDE as well as the function f may depend on the control: $\sigma(a, x)$, $b(a, x)$, $c(a, x)$, and $f(a, x)$ with $a \in U$.

Theorem 60. *Let σ, b, c, f be Lipschitz in both variables (a, x) and bounded along with σ^{-1}. Let $\tau = \inf(t \ge 0 : X_t \notin [r_1, r_2])$, and let $\gamma \le \tau$ be a stopping time. Then for any $u \in C^2$ such that*

$$F[u] = \sup_a [\frac{\sigma^2}{2} u'' + bu' - cu + f](x) = 0, \; x \in [r_1, r_2],$$

the following holds true,

$$u(x) = \sup_{a(\cdot)\in \mathcal{A}_M} E_x^a [\int_0^\gamma e^{-\varphi_t} f^a(X_t) dt + e^{-\varphi_\gamma} u(X_\gamma)]. \qquad (189)$$

Proof. Denote the rhs of (189) by $\tilde u^\gamma(x)$. Applying Ito's formula with any strategy, using $[\frac{\sigma^2}{2}u'' + bu' - cu + f](x) \le 0$, that is, $[\frac{\sigma^2}{2}u'' + bu' - cu] \le -f(x)$ and integrating from zero to $\gamma \wedge t$, we get

$$u(x) \ge \tilde u^{\gamma, a(\cdot)}(x).$$

So,

$$u(x) \ge \sup_{a(\cdot)} \tilde u^{\gamma, a(\cdot)}(x).$$

Now taking the strategy $\bar a(\cdot)$ such that $F[u] = [\frac{\sigma^2}{2}u'' + bu' - cu + f]|_{a(\cdot)}(x) = 0$, we obtain the equality

$$u(x) = \tilde u^{\gamma, \bar a(\cdot)}(x)$$

with a given γ. Therefore,

$$u(x) = \sup_{a(\cdot)\in \mathcal{A}_M} \tilde u^{\gamma, a(\cdot)}(x).$$ QED

13 Lecture 13, 03.12.2021: Backward Pardoux – Peng SDEs (BSDEs): existence and uniqueness

Let $T > 0$, and let W be a standard WP, and let $\mathcal{F}^W_t, t \ge 0$ be its filtration. Although, in fact, the history started a little earlier in the paper [J.M. Bismut, Conjugate convex functions in optimal stochastic control, J. Math. Anal. Appl. 44, 384-404 (1973)]; however, the real beginning of the BSDE theory relates to the paper

Список литературы

[1] É. Pardoux & S.G. Peng, Adapted solution of a backward stochastic differential equation, Systems & Control Letters 14 (1990) 55-61. (Denoted by [PP90] in what follows)

13.1 BSDEs: introduction

The setting may look a little crazy at the first sight: suppose there is a r.v. $\xi \in \mathcal{F}_T^W$, a function $f(s,y)$ – which will be generalised a bit later – and we wonder whether there is a solution of the SDE (where Y_0 should be also found)

$$Y_t = Y_0 - \int_0^t f(s, Y_s)ds - \int_0^t Z_s dW_s, \quad 0 \le t \le T, \tag{190}$$

with a whatever (adapted) diffusion process Z_s such that

$$Y_T = \xi.$$

Why minus signes? Of course, they may be replaced by plus signes, but the convenience of this form will be clear on the next page. In some or even in many papers the signes are chosen differently, though.

Let us rewrite the equation in the equivalent form:

$$Y_t = \xi + \int_t^T f(s, Y_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T, \tag{191}$$

and Y_t must be $\mathcal{F}_t^W$-adapted. Indeed, if we subtract

$$\xi = Y_T = Y_0 - \int_0^T f(s, Y_s)ds - \int_0^T Z_s dW_s, \quad 0 \le t \le T,$$

to the equation (191), then we will get back to (190). The equation (191) is a standard form of a BSDE (up to the $\pm$ signes). Even a bit more general equations will be discussed in what follows and in [PP90] (and in many other papers): let f also depend on Z; we will be also studying the equations in the form

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T, \tag{192}$$

i.e., where f may depend on Z. A solution is an **adapted pair** (Y, Z), with a compulsory requirement $Y_T = \xi$. The assumptions are

- $\xi \in \mathcal{F}_T^W$ & $\mathrm{E}\xi^2 < \infty$;
- $|f(t,z)| \le C \quad \forall\, t, z$;
- $|f(t, y_1, z_1) - f(t, y_2, z_2)| \le C(|y_1 - y_2| + |z_1 - z_2|)$.

Theorem 61 (Pardoux – Peng)**.** *Under the conditions stated above, the equation (192) has a unique solution* (Y, Z)*, such that* $\mathrm{E}\int_0^T Z_s^2 ds < \infty$.

13.2 The "simplest" case: Clark's theorem

Let $f \equiv 0$. This is called Clark's representation. In this case the equation (191) reduces to

$$Y_t = \xi + \int_t^T Z_s dW_s, \quad 0 \le t \le T, \tag{193}$$

with $(Y_t, Z_t) \in \mathcal{F}_t^W$, or, equivalently, to

$$\xi = Y_t - \int_t^T Z_s dW_s, \quad 0 \le t \le T. \tag{194}$$

In particular, for $t = 0$ we write

$$\xi = Y_0 - \int_0^T Z_s dW_s, \quad 0 \le t \le T. \tag{195}$$

Note that here Y_0 must be **a constant**.

Theorem 62 (J.M.C. Clark (1970), H.Kunita & C. Watanabe (1967), A.D. Wentsell (1961))**.** *For any $\xi \in L_2(\Omega, \mathcal{F}_T^W, \mathbb{P})$ there exists a unique adapted process $Z_s \in \mathcal{F}_s$ such that*

$$\xi = \mathrm{E}\xi - \int_0^T Z_s dW_s. \tag{196}$$

Traditionally, this is called Clark's theorem, or Clark's representation, although, in fact, this is a simplification of what was done by professor Martin Clark, and despite the fact that this simplified version was established earlier by other investigators[133]*.*

Corollary 16. *If $\xi \in L_2(\Omega, \mathcal{F}_T^W, \mathbb{P})$, then there exists a solution of the BSDE problem (193) (i.e., (191) with $f \equiv 0$) given by the formula*

$$Y_t := \mathrm{E}(\xi|\mathcal{F}_t^W).$$

Proof of corollary. Naturally, we have $\mathrm{E}\int_0^T Z_s^2 ds = \mathrm{var}(\xi) < \infty$. Then, since $\xi = \mathrm{E}\xi - \int_0^T Z_s dW_s$, the $\mathcal{F}^W$-adapted martingale $Y_t := \mathrm{E}(\xi|\mathcal{F}_t^W)$ may be presented as

[133] Hence, perhaps, it would be more fair to call it Itô – Wentsell's, or, Itô – Wentsell – Clark's representation.

follows,

$$Y_t = \mathrm{E}(\mathrm{E}\xi - \int_0^T Z_s dW_s | \mathcal{F}_t^W) = \mathrm{E}\xi - \int_0^t Z_s dW_s \stackrel{*}{=}$$

and replacing back $\mathrm{E}\xi$ via $\mathrm{E}\xi = \xi + \int_0^T Z_s dW_s$, we continue

$$\stackrel{*}{=} \xi + \int_0^T Z_s dW_s - \int_0^t Z_s dW_s = \xi + \int_t^T Z_s dW_s.$$

That is,

$$\xi = Y_t - \int_t^T Z_s dW_s$$

with adapted Y_t and Z_t as required. QED

Proof of Clark's theorem. Note that any element of the space $L_2(\Omega, \mathcal{F}_T^W, P)$ can be approximated[134] in L_2 by finite sums of the form

$$\prod_{i=1}^n f_{t_i}(W_{t_i}), \quad \text{with } f_{t_i} \in C_b^\infty.$$

1. Take just one multiplier $\xi = f_{t_i}(W_{t_i})$: it has a natural representation due to Ito's formula

$$f_{t_i}(W_{t_i}) = f_{t_i}(0) + \int_0^{t_i} f'_{t_i}(W_s) dW_s + \frac{1}{2}\int_0^{t_i} f''_{t_i}(W_s) ds.$$

However, this is not the unique possible representation for $f_{t_i}(W_{t_i})$ via stochastic integrals and not the most useful. There exists a better one which does not involve any Lebesgue integral. Consider a Cauchy problem

$$u_t(t,x) + \frac{1}{2} u_{xx}(t,x) = 0, \ 0 \le t \le t_i, \ u(t_i, x) = f_{t_i}(x).$$

Since $f_{t_i} \in C^\infty$, this problem has a unique bounded classical solution $u(t,x)$; more than that, $u(t,\cdot) \in C_b^\infty$ for any $t_i > 0$. Let us apply Ito's formula to $u(t, W_t)$: due to the equation,

$$f_{t_i}(W_{t_i}) - u(0,0) = \int_0^{t_i} u_x(s, W_s) dW_s.$$

Denote

$$u(s,x) =: (T_{s,t_i} f_{t_i})(x).$$

[134] *(Homework!)*

Then we get a required representation

$$f_{t_i}(W_{t_i}) = T_{0,t_i} f_{t_i}(0) + \int_0^{t_i} \partial_x T_{s,t_i} f_{t_i}(W_s) dW_s, \tag{197}$$

where, in particular, $\mathrm{E} f_{t_i}(W_{t_i}) = T_{0,t_i} f_{t_i}(0) = u(0,0)$.

2. Let us tackle the product $\prod_{i=1}^2 f_{t_i}(W_{t_i})$. Let $0 \le t_1 < t_2$, and let us show how to tackle the product

$$\prod_{i=1}^2 f_{t_i}(W_{t_i}), \quad \forall\, f_{t_i} \in C_b^\infty.$$

Firstly, we write as in (197),

$$f_{t_2}(W_{t_2}) = T_{t_1,t_2} f_{t_2}(W_{t_1}) + \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s.$$

Now we should represent the expression

$$\xi = f_{t_1}(W_{t_1}) T_{t_1,t_2} f_{t_2}(W_{t_1}) + f_{t_1}(W_{t_1}) \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s.$$

Let $f_{t_1}(W_{t_1}) T_{t_1,t_2} f_{t_2}(W_{t_1}) =: \tilde f_{t_1}(W_{t_1})$, and note that $\tilde f_{t_1} \in C_b^\infty$. For the first term we have as earlier,

$$\tilde f_{t_1}(W_{t_1}) = T_{0,t_1} \tilde f_{t_1}(0) + \int_0^{t_1} \partial_x T_{s,t_1} \tilde f_{t_1}(W_s) dW_s.$$

For the second term we have,

$$f_{t_1}(W_{t_1}) \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s = T_{0,t_1} f_{t_1}(0) \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s$$

$$+ \int_0^{t_1} \partial_x T_{s,t_1} f_{t_1}(W_s) dW_s \times \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s$$

The product

$$\int_0^{t_1} \partial_x T_{s,t_1} f_{t_1}(W_s) dW_s \times \int_{t_1}^{t_2} \partial_x T_{s,t_2} f_{t_2}(W_s) dW_s$$

equals a.s.[135]

$$\int_{t_1}^{t_2}\left(\int_0^{t_1}\partial_x T_{r,t_1} f_{t_1}(W_r)dW_r\right)\partial_x T_{s,t_2} f_{t_2}(W_s)dW_s.$$

Denote $F_{t_1} := \int_0^{t_1}\partial_x T_{r,t_1} f_{t_1}(W_r)dW_r$. Then overall obtain,

$$\xi = \prod_{i=1}^{2} f_{t_i}(W_{t_i}) = T_{0,t_1}\tilde f_{t_1}(0) + \int_0^{t_1}\partial_x T_{s,t_1}\tilde f_{t_1}(W_s)dW_s$$

$$+\int_{t_1}^{t_2}(F_{t_1} + T_{0,t_1} f_{t_1}(0))\partial_x T_{s,t_2} f_{t_2}(W_s)dW_s$$

$$= T_{0,t_1}\tilde f_{t_1}(0) + \int_0^{t_2}\Big(1(s\le t_1)\partial_x T_{s,t_1}\tilde f_{t_1}(W_s)$$

$$+1(t_1 < s \le t_2)(F_{t_1} + T_{0,t_1} f_{t_1}(0))\partial_x T_{s,t_2} f_{t_2}(W_s)\Big)dW_s$$

as required.

3. For a general product $\prod_{i=1}^n f_{t_i}(W_{t_i})$ we obtain a similar representation by induction.

4. Now we have a sequence $\xi^n \stackrel{L_2}{\to} \xi$ and

$$\xi^n = \mathrm{E}\xi^n + \int_0^T Z^n_s dW_s.$$

Here clearly $\mathrm{E}\xi^n \to \mathrm{E}\xi$, so we also have

$$\int_0^T Z^n_s dW_s \stackrel{L_2}{\to} \xi - \mathrm{E}\xi.$$

Let us choose a subsequence $n' \to \infty$ and re-number it again by (n) so that

$$\mathrm{E}\left(\int_0^T Z^n_s dW_s - (\xi - \mathrm{E}\xi)\right)^2 \le 2^{-n}.$$

[135] *(Homework!)*

Then (Z^n) is a Cauchy sequence in L_2 and we can choose again a converging subsequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$ to some adapted process Z:

$$\mathrm{E}\int_0^T (Z_s^n - Z_s)^2 ds \to 0.$$

Then in the limit we obtain

$$\xi = \mathrm{E}\xi + \int_0^T Z_s dW_s.$$

as required.

5. Let us show uniqueness. Assume there exists another solution Z' such that also

$$\xi = \mathrm{E}\xi - \int_0^T Z'_s dW_s.$$

Then we have[136],

$$\int_0^T (Z_s - Z'_s) dW_s = 0 \quad (a.s.)$$

So, by Ito's isometry

$$0 = \mathrm{E}(\int_0^T (Z_s - Z'_s) dW_s)^2 = \mathrm{E}\int_0^T (Z_s - Z'_s)^2 ds.$$

It follows that $\mathbb{P}(Z'_t = Z_t,\, 0 \le t \le T) = 1$, as required. QED

13.3 Simplified Pardoux – Peng theorem with random $f(s)$

As a second step[137] in the proof of the Pardoux – Peng theorem 61, we consider a special case, the equations in the form

$$Y_t = \xi + \int_t^T f(s) ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T, \tag{198}$$

where f depends neither on Y nor on Z, but may be random (and, of course, adapted). Its solution is by definition an adapted pair (Y, Z), with the terminal condition $Y_T = \xi$. The assumptions are

[136] *Note here the difference between stochastic and ordinary calculi: for the nonrandom functions and Lebesgue integral the equation* $\int_0^T (z(s) - z'(s)) ds = 0$ *does not imply that* $z(s) - z'(s) = 0$ *a.e.*

[137] The first step was Clark's representation.

- $\xi \in \mathcal{F}_T^W$ & $\mathrm{E}\xi^2 < \infty$;
- $|f(t)| \leq C$

Proposition 2. *Under the above assumptions the equation (198) has a unique solution* (Y, Z)*.*

Proof. Equivalently (subtracting $Y_0 = \xi + \int_0^T f(s)ds + \int_0^T Z_s dW_s$), we are to solve the equation

$$Y_t = Y_0 - \int_0^t f(s)ds - \int_0^t Z_s dW_s,\ 0 \leq t \leq T, \quad Y_T = \xi.$$

Denote

$$\hat{Y}_t := \mathrm{E}\left(\xi + \int_t^T f(s)ds|\mathcal{F}_t^W\right) \in \mathcal{F}_t^W$$

By virtue of the calculus in the proof of Clark's theorem $\exists\, Z_t \in \mathcal{F}_t$, such that

$$\mathrm{E}\left(\xi + \int_0^T f(s)ds|\mathcal{F}_t^W\right) = \hat{Y}_0 - \int_0^t Z_s dW_s$$

Then $(\hat{Y}_t, Z_t)$ is the solution of (198).
Indeed,

$$\hat{Y}_t := \mathrm{E}\left(\xi + \int_t^T f(s)ds|\mathcal{F}_t^W\right)$$

together with

$$\mathrm{E}\left(\xi + \int_0^T f(s)ds|\mathcal{F}_t^W\right) = \hat{Y}_0 - \int_0^t Z_s dW_s$$

imply that

$$\begin{aligned}\hat{Y}_t = \mathrm{E}\left(\xi + \int_t^T f(s)ds|\mathcal{F}_t^W\right) &= \mathrm{E}\left(\xi + \int_0^T f(s)ds|\mathcal{F}_t^W\right) - \int_0^t f(s)ds \\ &= \hat{Y}_0 - \int_0^t f(s)ds - \int_0^t Z_s dW_s. \qquad \text{QED}\end{aligned}$$

Homework: Explain why solution is unique[138] *and why* $\int_0^T \mathrm{E} Z_s^2 ds < \infty$*.*

[138] *The uniqueness will be proved in the next subsections for the general case; however, for the special case it is a bit easier and is worth to be tried independently.*

Hint: Use that

$$\hat{Y}_t := \mathrm{E}\left(\xi + \int_t^T f(s)ds | \mathcal{F}_t^W\right),$$

$$\mathrm{E}\left(\xi + \int_0^T f(s)ds | \mathcal{F}_t^W\right) = \hat{Y}_0 - \int_0^t Z_s dW_s,\ 0 \le t \le T,$$

so that, in particular,

$$\hat{Y}_0 := E\left(\xi + \int_0^T f(s)ds\right) \qquad \textit{is bounded,}$$

$$\textit{and that} \quad \int_0^T Z_s dW_s = -\xi - \int_0^T f(s)ds + \hat{Y}_0.$$

13.4 Simplified Pardoux – Peng theorem with $f = f(s, z)$

In this subsection the function $f = f(s, z)$ is deterministic. We will prove that under the conditions of the theorem (including the Lipschitz condition $|f(t, y_1) - f(t, y_2)| \le C(|y_1 - y_2|)$) the statement of Pardoux – Peng's theorem holds true. Only existence will be shown here because it is convenient to prove uniqueness straight for the most general case. Let us use successive approximations for this "intermediate" equation

$$Y_t = \xi + \int_t^T f(s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T. \tag{199}$$

Take $(Y^0, Z^0) \equiv (0, 0)$ (this pair does not solve anything reasonable, but never mind) and by induction define $(Y^n, Z^n), n \ge 1$ so that

$$Y_t^n = \xi + \int_t^T f(s, Z_s^{n-1})ds + \int_t^T Z_s^n dW_s, \quad 0 \le t \le T.$$

The pair (Y^n, Z^n) exists (and is unique) for each $n \ge 1$ due to the previous subsection. Let us show its convergence. Note that $\mathrm{E} \int_0^T (Z^n)_s^2 ds \le K < \infty$ uniformly[139] in n.

[139] *(Homework!)*

We have,

$$Y_t^{n+1} = \xi + \int_t^T f(s, Z_s^n)ds + \int_t^T Z_s^{n+1}dW_s$$

$$= Y_0^{n+1} - \int_0^t f(s, Z_s^n)ds - \int_0^t Z_s^{n+1}dW_s;$$

$$Y_t^n = \xi + \int_t^T f(s, Z_s^{n-1})ds + \int_t^T Z_s^n dW_s$$

$$= Y_0^n - \int_0^t f(s, Z_s^{n-1})ds - \int_0^t Z_s^n dW_s.$$

By Ito's formula,

$$d|Y_t^{n+1} - Y_t^n|^2 = d(Y_t^{n+1} - Y_t^n)^2$$

$$= 2(Y_t^{n+1} - Y_t^n)d(Y_t^{n+1} - Y_t^n) + (Z_t^{n+1} - Z_t^n)^2 dt$$

$$= -2(Y_t^{n+1} - Y_t^n)(f(t, Z_t^n) - f(t, Z_t^{n-1}))dt$$

$$-2(Y_t^{n+1} - Y_t^n)(Z_t^{n+1} - Z_t^n)dW_t + (Z_t^{n+1} - Z_t^n)^2 dt.$$

So, integrating from t to T and using that $Y_T^{n+1} = Y_T^n = \xi$, we get

$$0 - |Y_t^{n+1} - Y_t^n|^2 = -2\int_t^T (Y_s^{n+1} - Y_s^n)(f(s, Z_s^n) - f(s, Z_s^{n-1})ds \tag{200}$$

$$-2\int_t^T (Y_s^{n+1} - Y_s^n)(Z_s^{n+1} - Z_s^n)dW_s + \int_t^T (Z_s^{n+1} - Z_s^n)^2 ds.$$

Note by the way that the equality[140]

$$-\mathrm{E}|Y_t^{n+1}-Y_t^n|^2 = -2\int_t^T \mathrm{E}(Y_s^{n+1}-Y_s^n)(f(s,Z_s^n)-f(s,Z_s^{n-1})ds \qquad (201)$$

$$+\int_t^T \mathrm{E}(Z_s^{n+1}-Z_s^n)^2 ds$$

implies the differentiability of the lhs $\mathrm{E}|Y_t^{n+1}-Y_t^n|^2$ and the equality of the derivative to the respective expression. In turn, the equality

$$-|Y_t^{n+1}-Y_t^n|^2 = -2\int_t^T (Y_s^{n+1}-Y_s^n)(f(s,Z_s^n)-f(s,Z_s^{n-1})ds$$

$$-2\int_t^T (Y_s^{n+1}-Y_s^n)(Z_s^{n+1}-Z_s^n)dW_s + \int_t^T (Z_s^{n+1}-Z_s^n)^2 ds$$

implies that

$$|Y_t^{n+1}-Y_t^n|^2 + \int_t^T (Z_s^{n+1}-Z_s^n)^2 ds$$

$$=2\int_t^T (Y_s^{n+1}-Y_s^n)(f(s,Z_s^n)-f(s,Z_s^{n-1})ds-2\int_t^T (Y_s^{n+1}-Y_s^n)(Z_s^{n+1}-Z_s^n)dW_s.$$

Taking expectations we obtain by the CBS inequality with $\epsilon = 1/(2L_f)$ (where L_f is

[140] *Justify that the stochastic integral has a zero expectation here!*

the Lipschitz constant of the function f) and a corresponding $C = C_\epsilon$

$$\begin{aligned} &\mathrm{E}|Y_t^{n+1} - Y_t^n|^2 + \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds \\ &= 2\int_t^T \mathrm{E}(Y_s^{n+1} - Y_s^n)(f(s, Z_s^n) - f(s, Z_s^{n-1})ds \\ &\leq C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \epsilon \mathrm{E}\int_t^T (f(s, Z_s^n) - f(s, Z_s^{n-1})^2 ds \\ &\leq C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^n - Z_s^{n-1})^2 ds \end{aligned} \tag{202}$$

Thus,

$$\begin{aligned} &\mathrm{E}|Y_t^{n+1} - Y_t^n|^2 + \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds \\ &\leq C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^n - Z_s^{n-1})^2 ds. \end{aligned} \tag{203}$$

Recall that $\mathrm{E}\int_0^T (Z^n)_s^2 ds \leq K < \infty$ uniformly in n, and denote

$$u_t^{n+1} := \mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds, \qquad v_t^{n+1} := \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds.$$

Then

$$-\dot{u}_t^{n+1} + v_t^{n+1} \leq C u_t^{n+1} + \frac{1}{2} v_t^n. \tag{204}$$

~~Of course, this is not a result of the differentiation of the inequality (203): we cannot differentiate inequalities, in general! However, if we differentiate the *equality* (202) and then apply CBS inequality, we get exactly (204).~~
No differentiation: just (203) is **equivalent** to (204).

Further, denoting $U^n_t = u^n_{T-t}$ and $V^n_t = v^n_{T-t}$ (because in this way it is easier to use Grönwall's inequality), we obtain a similar bound

$$\dot U^{n+1}_t + V^{n+1}_t \le C U^{n+1}_t + \frac{1}{2} V^n_t. \tag{205}$$

Note that $\dot v^{n+1}_t \le 0$, so, $\dot V^{n+1}_t \ge 0$, and $u^{n+1}_T = v^{n+1}_T = 0$, i.e., $U^n_0 = V^n_0 = 0$. Let us multiply each term in (205) by $\exp(C(T-t)) =: e_t$. Then it follows

$$(e_t U^{n+1}_t)' + e_t V^{n+1}_t \le \frac{1}{2} e_t V^n_t.$$

Integrate from 0 to t:

$$e_t U^{n+1}_t + \int_0^t e_s V^{n+1}_s ds \le \frac{1}{2}\int_0^t e_s V^n_s ds.$$

For the function u^n and v^n this means

$$e_{T-t} u^{n+1}_{T-t} + \int_t^T e_s v^{n+1}_s ds \le \frac{1}{2}\int_t^T e_s v^n_s ds.$$

Recall that by definition all u^n and v^n are non-negative. So, as a corollary and by iterations

$$\int_t^T e_s v^{n+1}_s ds \le \frac{1}{2}\int_t^T e_s v^n_s ds \le \ldots \le \frac{1}{2^n}\int_t^T e_s v^1_s ds = \frac{C}{2^n}.$$

Thus, we get for u^n

$$u^{n+1}_{T-t} \le C 2^{-n},$$

and for v^n

$$\int_t^T v^{n+1}_s ds \le C 2^{-n}$$

with some $C > 0$. Also iterating (204) ($-\dot u^{n+1}_t + v^{n+1}_t \le C u^{n+1}_t + \frac{1}{2} v^n_t$) and using that $-\dot u^{n+1}_t \ge 0$, we get[141]

$$v^{n+1}_t \le C(1+n)2^{-n}. \tag{206}$$

It follows that both (Y^n) and (Z^n) are Cauchy sequences in $L_2((0,T)\times\Omega); \Lambda \times P)$. So, they have their limits as $n \to \infty$ (possibly over some subsequence), which are both adapted:

$$(Y^n, Z^n) \overset{L_2((0,T)\times\Omega,\Lambda\times P)}{\to} (Y, Z).$$

[141] *An easy exercise.*

Therefore, we can pass to the limit in the equation

$$Y_t^n = \xi + \int_t^T f(s, Z_s^{n-1})ds + \int_t^T Z_s^n dW_s, \quad 0 \le t \le T,$$

obtaining for the limiting processes the desired equation

$$Y_t = \xi + \int_t^T f(s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T.$$

NB: The existence of a limit for Y_t^n *not just in* L_2 *but for any particular* t follows from the existence of the limit for the right hadn side. Existence in the special case of $f(s, z)$ is established. QED.

Homework: Explain why the limit Y_t *is a.s. continuous wrt* t.

13.5 General case $f(s, y, z)$: uniqueness

Consider the general case of the equation

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T,$$

and show **uniqueness of solution**. Assume that there are two solutions (Y', Z') and (Y'', Z''). We have,

$$\begin{aligned} Y_t^{'} &= \xi + \int_t^T f(s, Y_s^{'}, Z_s^{'})ds + \int_t^T Z_s^{'} dW_s \\ &= Y_0^{'} - \int_0^t f(s, Y_s^{'}, Z_s^{'})ds - \int_0^t Z_s^{'} dW_s; \\ Y_t^{''} &= \xi + \int_t^T f(s, Y_s^{''}, Z_s^{''})ds + \int_t^T Z_s^{''} dW_s \\ &= Y_0^{''} - \int_0^t f(s, Y_s^{''}, Z_s^{''})ds - \int_0^t Z_s^{''} dW_s. \end{aligned}$$

By Ito's formula,

$$d|Y_t^{'} - Y_t^{''}|^2 = d(Y_t^{'} - Y_t^{''})^2$$

$$= 2(Y_t^{'} - Y_t^{''})d(Y_t^{'} - Y_t^{''}) + (Z_t^{'} - Z_t^{''})^2 dt$$

$$= -2(Y_t^{'} - Y_t^{''})(f(t, Y_t^{'}, Z_t^{'}) - f(t, Y_t^{''}, Z_t^{''})dt$$

$$-2(Y_t^{'} - Y_t^{''})(Z_t^{'} - Z_t^{''})dW_t + (Z_t^{'} - Z_t^{''})^2 dt.$$

So, integrating from t to T we get

$$\begin{aligned} & 0 - |Y_t^{'} - Y_t^{''}|^2 \\ &= -2\int_t^T (Y_s^{'} - Y_s^{''})(f(s, Y_s^{'}, Z_s^{'}) - f(s, Y_s^{''}, Z_s^{''}))ds \\ & -2\int_t^T (Y_s^{'} - Y_s^{''})(Z_s^{'} - Z_s^{''})dW_s + \int_t^T (Z_s^{'} - Z_s^{''})^2 ds. \end{aligned} \tag{207}$$

Note by the way, that the equality

$$\begin{aligned} & -\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 \\ &= -2\int_t^T \mathrm{E}(Y_s^{'} - Y_s^{''})(f(s, Y_s^{'}, Z_s^{'}) - f(s, Y_s^{''}, Z_s^{''})ds + \int_t^T \mathrm{E}(Z_s^{'} - Z_s^{''})^2 ds \end{aligned} \tag{208}$$

implies the differentiability of the lhs $\mathrm{E}|Y_t^{'} - Y_t^{''}|^2$ wrt t and the equality of the derivative to the respective expression in the rhs. Using CBS inequality, we obtain

$$\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 + \mathrm{E}\int_t^T (Z_s^{'} - Z_s^{'})^2 ds$$

$$= 2\int_t^T E(Y_s^{'} - Y_s^{''})(f(s, Y_t^{'}, Z_s^{'}) - f(s, Y_t^{''}, Z_s^{''})ds$$

$$\leq C\mathrm{E}\int_t^T (Y_s^{'} - Y_s^{''})^2 ds + \epsilon E\int_t^T (f(s, Y_t^{'}, Z_s^{'}) - f(s, Y_t^{''}, Z_s^{''})^2 ds$$

$$\leq C\mathrm{E}\int_t^T (Y_s^{'} - Y_s^{''})^2 ds + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^{'} - Z_s^{''})^2 ds.$$

Thus,

$$\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 + \mathrm{E}\int_t^T (Z_s^{'} - Z_s^{''})^2 ds$$

$$\leq C\mathrm{E}\int_t^T (Y_s^{'} - Y_s^{''})^2 ds + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^{'} - Z_s^{''})^2 ds.$$

Hence, we obtain

$$\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^{'} - Z_s^{'})^2 ds \leq C\mathrm{E}\int_t^T (Y_s^{'} - Y_s^{''})^2 ds.$$

By virtue of Gröwnall's inequality[142], this implies

$$\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 = 0,$$

and

$$\mathrm{E}|Y_t^{'} - Y_t^{''}|^2 + \frac{1}{2}\mathrm{E}\int_t^T (Z_s^{'} - Z_s^{''})^2 ds = 0.$$

This means that $(Y', Z') = (Y'', Z'')$, as required. QED

[142] Exercise!

13.6 General case: convergence & existence

Using successive approximations we write (let $(Y^0, Z^0) = (0, 0)$)

$$Y_t^{n+1} = \xi + \int_t^T f(s, Y_s^n, Z_s^n)ds + \int_t^T Z_s^{n+1} dW_s$$
$$= Y_0^{n+1} - \int_0^t f(s, Y_s^n, Z_s^n)ds - \int_0^t Z_s^{n+1} dW_s;$$
$$Y_t^n = \xi + \int_t^T f(s, Y_s^{n-1}, Z_s^{n-1})ds + \int_t^T Z_s^n dW_s$$
$$= Y_0^n - \int_0^t f(s, Y_s^{n-1}, Z_s^{n-1})ds - \int_0^t Z_s^n dW_s.$$

By Ito's formula,

$$d|Y_t^{n+1} - Y_t^n|^2 = d(Y_t^{n+1} - Y_t^n)^2$$

$$= 2(Y_t^{n+1} - Y_t^n)d(Y_t^{n+1} - Y_t^n) + (Z_t^{n+1} - Z_t^n)^2 dt$$

$$= -2(Y_t^{n+1} - Y_t^n)(f(t, Y_t^n, Z_t^n) - f(t, Y_s^{n-1}, Z_t^{n-1})dt$$

$$-2(Y_t^{n+1} - Y_t^n)(Z_t^{n+1} - Z_t^n)dW_t + (Z_t^{n+1} - Z_t^n)^2 dt$$

So, integrating from t to T we get

$$0 - |Y_t^{n+1} - Y_t^n|^2 \tag{209}$$

$$= -2\int_t^T (Y_s^{n+1} - Y_s^n)(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1})ds$$

$$-2\int_t^T (Y_s^{n+1} - Y_s^n)(Z_s^{n+1} - Z_s^n)dW_s + \int_t^T (Z_s^{n+1} - Z_s^n)^2 ds$$

Note by the way that the equality

$$-\mathrm{E}|Y_t^{n+1} - Y_t^n|^2 \tag{210}$$

$$= -2\int_t^T \mathrm{E}(Y_s^{n+1} - Y_s^n)(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1})ds + \int_t^T \mathrm{E}(Z_s^{n+1} - Z_s^n)^2 ds$$

implies the differentiability of the lhs $\mathrm{E}|Y_t^{n+1} - Y_t^n|^2$ and the equality of the derivative to the respective expression in the rhs.

Returning to the equation without expectation, we rewrite it as follows,

$$
\begin{aligned}
|Y_t^{n+1} - Y_t^n|^2 + \int_t^T (Z_s^{n+1} - Z_s^n)^2 ds \\
= 2\int_t^T (Y_s^{n+1} - Y_s^n)(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1})ds \\
+2\int_t^T (Y_s^{n+1} - Y_s^n)(Z_s^{n+1} - Z_s^n)dW_s
\end{aligned}
$$

Taking expectations and applying CBS inequality we obtain

$$
\begin{aligned}
\mathrm{E}|Y_t^{n+1} - Y_t^n|^2 + \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds \\
= 2\int_t^T E(Y_s^{n+1} - Y_s^n)(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1}))ds \\
\le 2\int_t^T E|(Y_s^{n+1} - Y_s^n)|\,|(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1})|ds \\
\le C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \epsilon \mathrm{E}\int_t^T \underbrace{(f(s, Y_s^n, Z_s^n) - f(s, Y_s^{n-1}, Z_s^{n-1})^2}_{\le C(|Y_s^n - Y_s^{n-1}|^2 + |Z_s^n - Z_s^{n-1}|^2)} ds \\
\le C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \delta \mathrm{E}\int_t^T (Y_s^n - Y_s^{n-1})^2 ds + \delta \mathrm{E}\int_t^T (Z_s^n - Z_s^{n-1})^2 ds,
\end{aligned}
$$

assuming that ϵ is small enough. Here, of course, $C = C(\delta)$. We used the inequality

$$2ab \le a^2 + b^2 \quad \& \quad 2ab = 2(a\epsilon)(b/\epsilon) \le \epsilon^2 a^2 + b^2/\epsilon^2.$$

The value δ will be also chosen small enough a bit later. Thus,

$$\mathrm{E}|Y_t^{n+1} - Y_t^n|^2 + \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds \tag{211}$$
$$\le C\mathrm{E}\int_t^T (Y_s^{n+1} - Y_s^n)^2 ds + \delta\mathrm{E}\int_t^T (Y_s^n - Y_s^{n-1})^2 ds + \delta\mathrm{E}\int_t^T (Z_s^n - Z_s^{n-1})^2 ds.$$

Recall that $\mathrm{E}\int_0^T (Z^n)_s^2 ds \le K < \infty$ uniformly in n and denote

$$u_t^{n+1} := \mathrm{E}\int_t^T |Y_s^{n+1} - Y_s^n|^2 ds, \quad v_t^{n+1} := \mathrm{E}\int_t^T (Z_s^{n+1} - Z_s^n)^2 ds.$$

Then similarly to the previous subsections, from (211) we obtain[143]

$$-\dot u_t^{n+1} + v_t^{n+1} \le Cu_t^{n+1} + \delta u_t^n + \delta v_t^n. \tag{212}$$

Denoting $U_t^n = u_{T-t}^n$ and $V_t^n = v_{T-t}^n$, we rewrite it in the form

$$\dot U_t^{n+1} + V_t^{n+1} \le CU_t^{n+1} + \delta U_t^n + \delta V_t^n. \tag{213}$$

Note that $\dot v_t^{n+1} \le 0$, so, $\dot V_t^{n+1} \ge 0$, and $u_T^{n+1} = v_T^{n+1} = 0$, i.e., $U_0^n = V_0^n = 0$. Let us multiply each term in (213) by $\exp(C(T-t)) =: e_t$. Then it follows that

$$(e_t U_t^{n+1})' + e_t V_t^{n+1} \le \delta e_t V_t^n + e_t \delta\, U_t^n.$$

Let us integrate the latter inequality from 0 to $t \ge 0$:

$$e_t U_t^{n+1} + \int_0^t e_s V_s^{n+1} ds \le \delta\int_0^t e_s V_s^n ds + \delta\int_0^t e_s U_s^n ds,$$

or, equivalently (with $r \ge 0$),

$$e_r U_r^{n+1} + \int_0^r e_s V_s^{n+1} ds \le \delta\int_0^r e_s V_s^n ds + \delta\int_0^r e_s U_s^n ds. \tag{214}$$

In any case, dropping here a nonnegative term $e_r U_r^{n+1}$ we have

$$\int_0^r e_s V_s^{n+1} ds \le \delta\int_0^r e_s V_s^n ds + \delta\int_0^r e_s U_s^n ds. \tag{215}$$

[143] Simply, (211) is equivalent to (212).

By integration we get from (214) (for $t \geq 0$)

$$\int_0^t e_r U_r^{n+1} dr + \int_0^t dr \int_0^r e_s V_s^{n+1} ds \leq \delta \int_0^t dr \int_0^r e_s V_s^n ds + \delta \int_0^t dr \int_0^r e_s U_s^n ds.$$

After the change of order in the integration we obtain (recall that $t \leq T$)

$$\int_0^t e_r U_r^{n+1} dr + \int_0^t dr \int_0^r e_s V_s^{n+1} ds \leq \delta \int_0^t e_s V_s^n ds \int_s^t dr + \delta \int_0^t e_s U_s^n ds \int_s^t dr$$

$$= \delta \int_0^t e_s V_s^n (t-s) ds + \delta \int_0^t e_s U_s^n (t-s) ds \leq \delta T \int_0^t e_s V_s^n ds + \delta T \int_0^t e_s U_s^n ds.$$

At this stage let us choose $\delta > 0$ so small that $\delta \wedge (\delta T) \leq 1/4$. Then the bound reads

$$\int_0^t e_r U_r^{n+1} dr + \int_0^t dr \int_0^r e_s V_s^{n+1} ds \leq \frac{1}{4} \int_0^t e_s V_s^n ds + \frac{1}{4} \int_0^t e_s U_s^n ds.$$

In any case,

$$\int_0^t e_r U_r^{n+1} dr \leq \frac{1}{4} \int_0^t e_s V_s^n ds + \frac{1}{4} \int_0^t e_s U_s^n ds. \tag{216}$$

By adding up (215) with $\delta \leq 1/4$ and (216) we obtain

$$\int_0^t e_s U_s^{n+1} ds + \int_0^t e_s V_s^{n+1} ds \leq \frac{1}{2} \int_0^t e_s V_s^n ds + \frac{1}{2} \int_0^t e_s U_s^n ds. \tag{217}$$

Iterating this bound we get by induction

$$\int_0^t e_s U_s^{n+1} ds + \int_0^t e_s V_s^{n+1} ds \leq \ldots \leq \frac{1}{2^n} \int_0^t e_s V_s^1 ds + \frac{1}{2^n} \int_0^t e_s U_s^1 ds = \frac{C}{2^n}. \tag{218}$$

Thus, it follows from (214) that for all $0 \leq t \leq T$

$$e_t U_t^{n+1} \leq \frac{C}{2^n} \quad \Longleftrightarrow \quad e_t u_t^{n+1} \leq \frac{C}{2^n} \quad \Longleftrightarrow \quad \boxed{u_t^{n+1} \overset{\text{(another } C)}{\leq} \frac{C}{2^n}}. \tag{219}$$

Now, returning to the bound (212) and dropping the nonnegative term $-\dot{u}_t^{n+1}$ there, i.e., using only

$$v_t^{n+1} \leq C u_t^{n+1} + \delta u_t^n + \delta v_t^n, \tag{220}$$

and iterating it we have by induction (and using the already established bound (219))

$$v_t^{n+1} \le \underbrace{C u_t^{n+1} + \delta u_t^n}_{\le \frac{C}{2^n} + \frac{C\delta}{2^{n-1}}} + \delta v_t^n \le \underbrace{\frac{C}{2^n} \overset{(\text{recall } \delta \le 1/4)}{+} \frac{C\delta}{2^{n-1}}}_{\le \frac{C}{2^n}} + \delta \left(\frac{C}{2^{n-1}} \right) + \delta v_t^{n-1} \le \ldots \le \frac{C(n+1)}{2^n}. \tag{221}$$

It follows from (219) and (221) that both (Y^n) and (Z^n) are Cauchy sequences in $L_2((0,T)\times\Omega); \Lambda\times\mathbb{P})$. So, they have their limits as $n \to \infty$, which are both adapted:

$$(Y^n, Z^n) \overset{L_2((0,T)\times\Omega,\Lambda\times P)}{\rightarrow} (Y, Z)$$

Therefore[144], we can pass to the limit in

$$Y_t^n = \xi + \int_t^T f(s, Y_s^{n-1}, Z_s^{n-1}) ds + \int_t^T Z_s^n dW_s, \quad 0 \le t \le T,$$

obtaining for the limiting processes the desired equation

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s) ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T.$$

NB: Similarly to the special case, an existence of a limit for Y_t^n for *any particular* t follows from the existence of the limit for the rhs. The theorem is proved. QED

[144] *Exercise!*

14 Lecture 14, 10.12.2021: FBSDEs, quasi-linear PDEs, and optimal control

One of the possibilities for a functional of a Wiener process is, of course, a strong solution of an ordinary Ito's SDE. Let $\xi = X_T$, or, more generally, $\xi = g(X_T)$, where X_t is a strong solution of the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x,$$

or, in the integral form,

$$X_t = x + \int_0^t b(s, X_s)dt + \int_0^t \sigma(s, X_s)dW_s, \quad t \geq 0, \tag{222}$$

Then we are looking for the process Y_t which must be a solution to the BSDE

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \tag{223}$$

or, a bit more generally,

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \tag{224}$$

This setting is called **FBSDE** (Forward - backward SDE). In the next subsection a slightly more general and complicated SDE/BSDE system will be tackled.

14.1 FBSDE

Let $\xi = g(X_T)$, where

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s, Y_s, Z_s)dt + \int_0^t \sigma(s, X_s, Y_s, Z_s)dW_s, \\ Y_t &= \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \end{aligned} \tag{225}$$

A further complication/generalisation made here relates to the coefficients b and σ which *may* depend on Y and Z. Such a system is usually called *a system with full dependence.*

Theorem 63. *If the vector of coefficients (b, σ, f) is Lipschitz and bounded (or, of no more than a linear growth), then there exists a unique (strong) solution (X, Y, Z) of the equation (225).*

Without proof.

14.2 FBSDEs and quasi-linear PDEs

Let $\xi = g(X_T)$, where X_t is again strong solution of an SDE (independent of Y, Z)

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad t \geq 0,$$

and consider the Cauchy problem for a quasi-linear PDE

$$u_t(t,x) + Lu(t,x) = -f(t,x,u,\sigma^*\nabla_x u), \quad u(T,x) = g(x), \tag{226}$$

with $a = \sigma^2/2$ and $L = \langle a, \nabla^2\rangle + \langle b, \nabla\rangle - c$. A typical application to the stochstic control is the optimal control problem for a controlled SDE with σ independent on control:

$$dX_t^a = b(t, a(t, X_t^a), X_t^a)dt + \sigma(t, X_t)dW_t, \quad X_0 = x,$$

with the value function

$$v(s,x) = \sup_{a(\cdot)\in\mathcal{A}_M} E_{s,x}^{a(\cdot)}[\int_s^T e^{-\int_s^t c(r,X_r)dr} h(t, X_t)dt + e^{-\int_s^T c(r,X_r)dr} g(X_T)]. \tag{227}$$

Here we should use *(note the difference between two objects, $f(t,x,u(t,x), -\sigma^*(t,x)\nabla_x u(t,x))$ and $f^a(t,x)$)*

$$f(t,x,u(t,x), -\sigma^*(t,x)\nabla_x u(t,x))$$

$$= \sup_{a\in U}[b(t,a,x)\nabla_x u(t,x)) - c(t,a,x)u(t,x)) + f^a(t,x)].$$

Theorem 64. *If u is a solution of the equation (226) in $C^{1,2}$ (or, with Sobolev derivatives integrable in L_{d+1} in the case if $\sigma\sigma^*$ is nondegenerate), and $\xi = g(X_T)$, then*

$$u(0,x) = Y_0, \quad u(t, X_t) = Y_t, \quad \sigma^*(t, X_t)\nabla u(t, X_t) = -Z_t, \tag{228}$$

where the pair (Y, Z) is a unique solution of the BSDE

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \leq t \leq T, \tag{229}$$

In other words, there exists another way for computing the value function in the optimal control problem (227): instead of solving Bellman's equation, there is a possibility to solve the FBSDE (229), exactly or approximately, which is some (if not in all) situations may be easier.

Proof. For simplicity we assume in the calculus that $c \equiv 0$. Let us apply Ito's formula to $u(t, X_t)$:

$$du(t, X_t) = (u_t + Lu)(t, X_t)dt + \sigma^* \nabla u(t, X_t)dW_t$$

$$= -f(t, X_t, u(t, X_t), -\sigma^* \nabla u(t, X_t))dt + \sigma^* \nabla u(t, X_t)dW_t$$

In the integral form,

$$u(t, X_t) - u(0, x) = \int_0^t (u_s + Lu)(s, X_s)ds + \int_0^t \sigma^* \nabla u(s, X_s)dW_s$$

$$= -\int_0^t f(s, X_s, u(s, X_s), -\sigma^* \nabla u(s, X_s))ds + \int_0^t \sigma^* \nabla u(s, X_s)dW_s.$$

Substitute here $t = T$:

$$g(X_T) - u(0, x)$$

$$= -\int_0^T f(s, X_s, u(s, X_s), -\sigma^* \nabla u(s, X_s))ds + \int_0^T \sigma^* \nabla u(s, X_s)dW_s.$$

Subtract from the previous equation:

$$u(t, X_t) - g(X_T)$$

$$= \int_t^T f(s, X_s, u(s, X_s), -\sigma^* \nabla u(s, X_s))ds - \int_t^T \sigma^* \nabla u(s, X_s)dW_s.$$

Comparing the equation

$$u(t, X_t) = g(X_T)$$

$$+ \int_t^T f(s, X_s, u(s, X_s), -\sigma^* \nabla u(s, X_s))ds - \int_t^T \sigma^* \nabla u(s, X_s)dW_s.$$

with BSDE written on $\xi = g(X_T)$,

$$Y_t = \xi + \int_t^T f(s, X_s, Y_s, Z_s)ds + \int_t^T Z_s dW_s, \quad 0 \le t \le T,$$

we observe that solution $u(t, X_t)$ coincides with Y_t due to the uniqueness of the BSDE solution, and the term $-Z_t$ equals $\sigma^* \nabla u(t, X_t)$, which shows (228). QED

Exercise 29. *Perform a similar calculus with a variable potential c.*

14.3 FBSDE and control

For example, we can treat in this way the nonlinearity of the form

$$f(s,x,z) = \sup_{v\in U}[\sigma^{-1}b^v(s,x)z + h^v(t,x)].$$

Hence, we may represent solutions of Bellman's equation

$$u_t + \frac{\sigma\sigma^*}{2}u_{xx} + \sup_v[b^v(s,x)\nabla u(t,x) + h^v(t,x)] = 0,\ u(T,x) = g(x),$$

via FBSDEs. This may be applied to the stochastic control problem of the type

$$E^{a(\cdot)}_{t,x}[\int_t^T h(s,X_s)ds + g(X_T)] \to \sup_{a(\cdot)\in\mathcal{A}_M} =: u(t,x).$$

Bellman's equation for this problem has a form,

$$u_t + \frac{\sigma^2}{2}u''(t,x) + \sup_a[b^a(s,x)\nabla u(t,x) + h^a(t,x)] = 0, \quad u(T,x) = g(x).$$

Exercise 30. *Write down the answer, i.e., represent solution $u(t,x)$ via the solution of a corresponding BSDE (FBSDE). Make sure that you have specified the controlled SDE to which the above Bellman equation corresponds.*

14.4 Appendix: Wiener - Ito chaos expansion

Proposition 3. *For any $\xi \in L_2(\Omega, \mathcal{F}^W_T, \mathbb{P})$ there exist deterministic Borel measurable functions of variables of increasing dimension $h_n(s_1,\dots,s_n), n\ge 1$, such that a.s.*

$$\xi = \mathrm{E}\xi + \sum_{n\ge 1}\int_0^T\int_0^{s_1}\dots\int_0^{s_{n-1}} h_n(s_1,\dots,s_n)dW_{s_n}\dots dW_{s_1}, \tag{230}$$

where the series in the rhs converges in $L_2(\mathbb{P})$. All summands on this series are orthogonal in $L_2(\mathbb{P})$. The equality holds

$$\mathrm{E}\xi^2 = (\mathrm{E}\xi)^2 + \sum_{n\ge 1}\int_0^T\int_0^{s_1}\dots\int_0^{s_{n-1}} h_n^2(s_1,\dots,s_n)ds_n\dots ds_1. \tag{231}$$

In fact, this could be an alternative way to the approach based on Clark's theorem; but we go the reverse way.

Denote the nth term in the sum by $I_n(h_n)$. The orthogonality follows from the remark that by Ito's isometry $\mathrm{E}I_1(f)I_1(g) = \int_0^T \mathrm{E}fg ds$: so, e.g.,

$$\mathrm{E}I_2I_1 = \int_0^T \mathrm{E}\left(h_1(s_1)\int_0^{s_1} h_2(s_1,s_2)dW_{s_2}\right)ds = 0,$$

and likewise for all other pairs of non-equal dimensions.

Assume that ξ is orthogonal to all $I_n(h_n)$, $n \geq 0$. (Here $I_0 = \mathrm{E}\xi$.) Then it is orthogonal to any stochastic exponential

$$\rho_T = \exp(\int_0^T g(s)dW_s - \frac{1}{2}\int_0^T g^2(s)ds),$$

at least, if the function g is bounded. Indeed, by Ito's formula we have

$$\rho_T - 1 = \int_0^T g(s)\rho_s dW_s. \tag{232}$$

The latter equality may be iterated:

$$\rho_T - 1 = \int_0^T g(s)\rho_s dW_s$$

$$= \int_0^T g(s)\left(1 + \int_0^s g(s_2)\rho_{s_2}dW_{s_2}\right)dW_s \stackrel{\text{induction}}{=} \dots$$

$$= \int_0^T g(s)dW_s + \int_0^T \left(\int_0^s g(s)g(s_2)dW_{s_2}\right)dW_s + \dots,$$

where, if this "…" is a finite sum with a **remainder** then this **remainder** converges to zero in $L_2(\mathbb{P})$, as its order $n \to \infty$ due to the proposition. This is the result by K. Ito [J. Math. Soc. Japan 3(1): 157-169 (May, 1951). DOI: 10.2969/jmsj/00310157] provided here without proof.

Exercise 31. *Derive from (232) that for any $t \le T$*

$$\rho_t - 1 = \int_0^t g(s)\rho_s dW_s$$

with the same[145] *function g.*

15 Lecture 15, 17.12.2021: Second order (F)BSDEs/2BSDEs, optimal control

A natural question arise: if FBSDEs are suitable for a representations of the solutions of quasi-linear Bellman's PDEs with the diffusion coefficient not depending on control, which means that the nonlinearity $f(t, x, y, z)$ in the FBSDE has a form

$$f(t, x, u(t, x), \sigma^*(t, x)\nabla_x u(t, x)),$$

involving only the gradient of the PDE solution, for example, $f(s, x, z) = \sup_{v\in U}[\sigma^{-1}b^v(s, x)z + h^v(t, x)]$, yet, is there any other way to include control into the diffusion coefficient at all (except for the Bellman equation)? Such a way exists and is called **2BSDE**, or, a **BSDE of the second order**. Here is a brief introduction into this topic. Consider an SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \ge s, \quad X_s = x, \tag{233}$$

(which solution is denoted by $X_t^{s,x}$) along with the following extended BSDE-type system:

$$dY_t = f(t, X_t, Y_t, Z_t, \Gamma_t)dt + Z_t \circ dX_t, \quad s \le t \le T, \tag{234}$$

$$dZ_t = \underbrace{A_t\, dt}_{\text{“A” for “any”}} + \Gamma_t dX_t, \quad s \le t \le T, \tag{235}$$

$$Y_T = g(X_T), \tag{236}$$

where $Z_t \circ dX_t$ is a so called *Stratonovich stochastic differential*[146], namely[147],

$$Z_t \circ dX_t = Z_t dX_t + \frac{1}{2}\operatorname{Tr}[\Gamma_t\sigma\sigma^*(X_t)]dt. \tag{237}$$

[145] This was crucial in the above calculus.

[146] Usually the term Stratonovich stochastic **integral** is used, of course, for the integral version.

[147] This is just a particular case of Stratonovich differential where the integrand is “linear”.

Solution of the system (234)– (236) is by definition an adapted tuple (A, Y, Z, Γ) such that the integral version of this system is satisfied with probability one. Let[148]

$$\mathcal{L} = \frac{\partial}{\partial t} + \frac{1}{2}(\sigma\sigma^*(t,x)\nabla_x, \nabla_x).$$

Theorem 65 (P.Cheridito, H.M.Soner, N.Touzi, N.Victoir (2006); without proof)**.** *Let g be continuous, let X_t be a strong solution of the equation (233), let $f(t,x,y,z,\gamma)$ be continuous and bounded,[149] Lipschitz in y, let g be continuous and bounded.[150] Then the 2BSDE system (234)– (236) has a unique solution.*

The link of 2BSDE to optimal control problems is given by the following result.[151]

Theorem 66. *Suppose the function $v(t,x) \in C^{1,2}([0,T] \times \mathbb{R}^d)$ is a solution of the nonlinear PDE*

$$v_t(t,x) + f(t, v(t,x), \nabla_x v(t,x), \nabla_x^2 v(t,x)) = 0, \quad 0 \le t \le T, \quad v(T,x) = g(x), \tag{238}$$

and such that all the derivatives in the expression $\mathcal{L}\nabla_x v$ exist and are continuous. Then for each (s,x), $0 \le s < T$, the solution of the 2BSDE (234)– (236) is given by the formulae

$$Y_t = v(t, X_t^{s,x}),\ Z_t = \nabla_x v(t, X_t^{s,x}),\ \Gamma_t = \nabla_x^2 v(t, X_t^{s,x}),\ A_t = \mathcal{L}\nabla_x v(t, X_t^{s,x}). \tag{239}$$

Vice versa, the solution $v(s,x)$ is given by the first part[152] in (239):

$$v(s,x) = Y_s.$$

Proof of theorem 66. Let us apply **Ito's formula**[153] to $v(t, X_t)$:

$$dv(t, X_t) = \mathcal{L}v(t, X_t)dt + v_x(t, X_t)dX_t,$$

[148] No drift with the first order derivative, unlike what could be expected; it is correct.
[149] Some polynomial growth may be allowed.
[150] Some polynomial growth may be allowed.
[151] It is "provisional" because the existence of $C^{1,2}$-solution of the PDE (238) below is rarely guaranteed. The paper [CSTV-2006] (https://doi.org/10.1002/cpa.20168), actually, deals with viscosity solutions under weaker assumptions.
[152] Recall that Y_s here is a constant, not random.
[153] Explain that this is exactly Ito's formula.

or, in the integral form,

$$v(t, X_t) = v(s, x) + \int_s^t \mathcal{L}v(r, X_r)dr + \int_s^t \underbrace{v_x(r, X_r)}_{Z_r} dX_r$$

$$= v(s, x) - \int_s^t f(r, v(r, X_r), \nabla_x v(r, X_r), \nabla_x^2 v(r, X_r)))dr$$

$$+ \int_s^t \frac{1}{2}\sigma\sigma^*(r, X_r)\nabla_x^2 v(r, X_r)dr + \int_s^t v_x(r, X_r)dX_r.$$

It is clear that $v(T, x) = g(x)$, and it is straightforwardly checked that

$$dZ_t = dv_x(t, X_t) = A_t dt + \Gamma_t dX_t,$$

as required. QED

This is the end of the course.

16 Lecture 16, 24.12.2021: Revision

17 Topics for the exam/Примерные темы к экзамену

1. Винеровский процесс (ВП); свойства рекуррентности и транзиентности.
2. Квадратическая вариация ВП; стохастический интеграл Ито (СИ), его свойства
3. Мартингальные неравенства Дуба и Дуба – Колмогорова, в т.ч. для СИ.
4. Формула Ито
5. Теорема Ито о существовании сильного решения СДУ
6. Теорема о непрерывной зависимости от начального условия.

7. Марковское свойство решений СДУ
8. Стохастические экспоненты; теорема Гирсанова
9. Экспоненциальные мартингальные неравенства; условия типа Новикова.
10. Слабые решения СДУ на основе теоремы Гирсанова.
11. Сильные решения, принцип Ямада – Ватанабе
12. Сильные решения, теоремы Ямада – Ватанабе
13. Сильные решения, упрощенная теорема Накао-Звонкина
14. Сильные решения, теоремы сравнения.
15. Оценки Крылова, формула Ито-Крылова для функций с соболевскими производными.
16. Слабые решения по Скороходу.
17. Слабые решения по Крылову.
18. Характеризация Леви винеровского процесса (многомерный случай).
19. Случайная замена времени.
20. СДУ с отражением на полупрямой и в (многомерной) области; задача Скорохода.
21. Связь винеровского процесса с оператором и уравнением Лапласа.
22. Связь винеровского процесса с уравнением теплопроводности.
23. Связь решения общего СДУ с общим эллиптическим оператором.
24. Связь решения общего СДУ с общими параболическим оператором.
25. Оценка среднего момента выхода невырожденной диффузии[154] из шара.
26. Производные решений СДУ.
27. Процесс Орнштейна – Уленбека.

[154]= Решения невырожденного СДУ.

28. Эргодические решения СДУ.
29. СДУ на торе; инвариантная мера; скорость сходимости; каплинг (метод склейки).
30. Уравнение Пуассона «во всем пространстве»
31. Обратные и прямые уравнения Фоккера – Планка – Колмогорова (ФПК); генераторы (квази)диффузионного процесса.
32. Стохастическое управление на конечном интервале времени; уравнение Беллмана
33. Алгоритм улучшения цены (функции выигрыша) Ховарда – Беллмана; понятие о вязких решениях уравнения Беллмана.
34. Эргодическое управление «в среднем»; эргодическое уравнение Беллмана
35. Эргодическое управление «в среднем»; алгоритм Ховарда.
36. Задача об оптимальной остановке.
37. Задача об оптимальной остановке, подход на основе интенсивности убивания.
38. Обратные СДУ Парду – Пенга (BSDEs); теорема о представлении Кларка.
39. Обратные СДУ Парду – Пенга (BSDEs); существование.
40. Обратные СДУ Парду – Пенга (BSDEs); единственность.
41. Прямые-обратные СДУ Парду – Пенга (FBSDEs) и их связь с нелинейными уравнениями в частных производных и с задачами управления.
42. 2BSDEs (обратные уравнения Парду – Пенга второго порядка) и их связь с нелинейными уравнениями в частных производных и с задачами управления.

18 Further studies around the area

The following directions may be regarded as further studies on the edge of the modern stochastics and applications. This is, of course, a personal view of your lecturer, and clearly the list is not complete.

- SDEs with integrable and generalised drift; highly degenerate SDEs
- SPDEs: stochastic partial differential equations
- SDEs in Hilbert spaces
- Filtering equations[155] for hidden Markov models
- McKean – Vlasov (aka mean-field) stochastic equations
- Viscosity solutions including for integro-differential equations (aka Bellman equations ofr jump-diffusions)
- Stochastic averaging
- Reaction–diffusion equations
- Diffusion approximations
- Large deviations
- Numerical approximations for all of the above
- Statistical issues (parameter estimations et al.) for the models listed above
- Games; mean-field games

[155] The items 2–4 are fairly close one to each other