PROBABILITY

FOR FINANCE AND INSURANCE

Bohdan Maslowski

1 Introduction

In principle, the present Lecture Notes comprehend the contents of the course NMFM408 "Probability for Finance and Insurance" at the Faculty of Mathematics and Physics of the Charles University in Prague, which serves as a brief introduction to continuous time martingale theory and stochastic analysis for students of Finance and Insurance Mathematics.

The aim of the course (and of the notes) is to provide, in a concise form, a relatively rigorous mathematical treatment of basic facts which are needed in continuous time modelling in finance and insurance mathematics. The material collected here is by no means original. It is (in author's opinion) a carefully selected minimal set of mathematical tools concerning stochastic continuous time modelling that students of this branch of study should possess in their portfolio.

More comprehensive treatments in this field are numerous, for example the monographs [KS], [O], [SV], [Fr], cf. also the lecture notes [Se] (in Czech, which may be recommended to all interested readers. The present lecture notes were built upon a selection from these books, notably from [KS] and [O].

There is a nonempty symmetric difference between the contents of actual course and exam requirements on one side (which may vary from year to year) and these Notes on the other one. Only very few proofs and just selection of examples can be found here. On the other hand, some facts are presented here only for better overview (in particular, Section 6 is usually not a part of the course).

Section 2 contains Preliminaries; some basic definitions and facts from probability theory are recalled and some notation is introduced. Section 3 is a crash course to fundamentals of stochastic processes. Section 3 contains motivation, definition and basic properties of the Wiener process (Brownian motion). In Section 4 continuous-time martingales are dealt with. Section 5 provides and introduction to stochastic analysis, definition and some properties of stochastic integral, the concept of stochastic differential, Ito formula and some applications. In Section 6, a simple existence and uniqueness result for nonlinear stochastic differential equation is given.

2 Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measure space; that is, a triple consisting of a set Ω, a σ-algebra $\mathcal{F} \subset \exp \Omega$ and a measure $\mathbb{P}$ on $\mathcal{F}$, $\mathbb{P}(\Omega) = 1$. The space $\Omega = (\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*, the sets in $\mathcal{F}$ are *events* and $\mathbb{P}$ is interpreted as *probability* by which the events from $\mathcal{F}$ are measured. The points in Ω (often denoted by ω) are called *sample points* or *elementary events*. A property that is true except for an event of probability zero is said to hold *almost surely* (abbreviated "a.s."). Let $(E, \mathcal{B})$ be a measurable space ($\mathcal{B}$ is a σ-algebra consisting of sets from $\exp E$); a measurable mapping $X : \Omega \to E$ is called *an E-valued random variable*. In the sequel, E (often called a *state space*) will be always a separable Banach space $E = (E, \|\cdot\|)$ equipped with the σ-algebra $\mathcal{B}$ of Borel sets (in fact, in most cases $E = \mathbb{R}^n$).

Let X be an E-valued random variable. Then

$$\hat{\mathcal{F}}^X := \{X^{-1}(B); B \in \mathcal{B}\}$$

is called the *σ-algebra generated by X*; the measure μ_X on $\mathcal{B}$ defined as

$$\mu_X(A) := \mathbb{P}(X^{-1}(A)),\ A \in \mathcal{B}$$

is called the *probability distribution* or the *probability law* of the random variable X (we also use the notation $\mu_X = \mathrm{Law}(X)$). Thus μ_X is the image of the measure $\mathbb{P}$ under X and for $A \in \mathcal{B}$, $\mu_X(A)$ is interpreted as the probability that X evaluates in A; another usual notation is

$$\mu_X(A) = \mathbb{P}(\{\omega \in \Omega;\ X(\omega) \in A\}) =: \mathbb{P}[X \in A].$$

The *mathematical expectation* (or the *mean value*) of X is defined as the abstract Lebesgue (or Bochner) integral

$$\mathbb{E}X := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_E x\mu_X(dx)$$

whenever it makes sense. The *variance* is defined as

$$\operatorname{Var} X = \mathbb{E}\|X - \mathbb{E}X\|^2 = \|X - \mathbb{E}X\|^2_{L_2(\Omega;E)}.$$

Definition 2.1. (i) Two events $A, B \in \mathcal{F}$ are said to be stochastically independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

(ii) Two σ-algebra $\mathcal{G}_1 \subset \mathcal{F}$ and $\mathcal{G}_2 \subset \mathcal{F}$ are stochastically independent if each two events $A \in \mathcal{G}_1$, $B \in \mathcal{G}_2$, are stochastically independent.

(iii) Two random variables X, Y defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (not necessarily taking values in the same state space) are said to be stochastically independent if the respective σ-algebras $\hat{\mathcal{F}}^X$, $\hat{\mathcal{F}}^Y$ are stochastically independent.

Next we give a definition of conditional expectation (and conditional probability), which again is one of the most important concepts in probability theory.

Note that if $\mathcal{G}$ is a σ-algebra, $\mathcal{G} \subset \mathcal{F}$, then $L^2(\Omega, \mathcal{G}; E)$ is a closed subspace of $L^2(\Omega, \mathcal{F}; E) = L^2(\Omega; E)$.

Definition 2.2. Let $X \in L^2(\Omega, \mathcal{F}; E)$ and $\mathcal{G} \subset \mathcal{F}$ be given, where $\mathcal{G}$ is a σ-algebra. Then

(i) the conditional expectation of X given $\mathcal{G}$ (denoted by $\mathbb{E}[X|\mathcal{G}]$) is the orthogonal projection of X on the subspace $L^2(\Omega, \mathcal{G}; E)$.

(ii) If $\mathcal{G} = \mathcal{F}^Y$ where Y is a random variable defined on Ω, then $\mathbb{E}[X|\mathcal{F}^Y] =: \mathbb{E}[X|Y]$ is called the conditional expectation of X given Y.

Remark 2.3. (i) By Definition 2.2 the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ is a random variable that is $\mathcal{G}$-measurable and is the best approximation of X in $L^2(\Omega, \mathcal{G}; E)$ with respect to the $L^2(\Omega, \mathcal{F}; E)$-norm.

(ii) In general, the conditional expectation may be defined for $X \in L^1(\Omega, \mathcal{F}; E)$ (as a $\mathcal{G}$-measurable r.v. such that $\int_A X\, d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P}$ for each $A \in \mathcal{G}$).

(iii) It is easy to check that

- $\mathbb{E}X = \mathbb{E}[X|\mathcal{G}]$ if $\mathcal{G} = \{\emptyset, \Omega\}$ is trivial,
- $\mathbb{E}(\mathbb{E}[X|\mathcal{G}]) = \mathbb{E}X$
- $\mathbb{E}[X|\mathcal{G}] = X$ iff X is $\mathcal{G}$-measurable
- $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$ if X and $\mathcal{G}$ are stochastically independent.

The conditional expectation may be interpreted as follows: The σ-algebra $\mathcal{G}$ contains the "information" available to us, on basis of which we construct an estimate of the random variable X.

3 Stochastic Processes - General Concepts

Our next aim is to define an E-valued stochastic process and introduce some related concepts. In general, by E-valued *stochastic (or random) process* one can understand arbitrary family $(X_t)_{t\in\Gamma}$ of E-valued random variables where the index set Γ (usually interpreted as time) may be $\mathbb{N}$, $\mathbb{N}\cup\{0\}$, $\mathbb{Z}$ (discrete time), $[0,T]$, $\mathbb{R}_t = [0,+\infty)$ or $\mathbb{R}$ (continuous time). We are only interested in the continuous time case when $\Gamma = [0,T]$ or $\Gamma = \mathbb{R}_+$ (if not stated otherwise) . By a *sample path* (or *trajectory*) of the process we understand the function $t \mapsto X_t(\omega)$ for each given $\omega \in \Omega$. Clearly, the stochastic process may be viewed as a mapping $X : [0,T]\times\Omega \to E$ and we are free to use the notation $X = (X(t,\omega)) = (X_t)$. The process is called measurable if X is measurable w.r.t. the product σ-algebra $\mathcal{B}([0,T])\times\mathcal{F}$. Clearly, the process (X_t) may be also viewed as a random variable taking valued in a suitable path space. For example, if $X \in L^2((0,T)\times\Omega; \mathcal{B}([0,T]\times\mathcal{F};\ dt\times\mathbb{P}; E)$ then X may be viewed as $L^2(0,T;E)$-valued random variable.

Each stochastic process defines a family of measures, so called *finite-dimensional distributions*,

$$\mu^X_{t_1,\dots,t_n}(A_1\times\cdots\times A_n) = \mathbb{P}[X_{t_1}\in A_1,\dots,X_{t_n}\in A_n], \qquad A_j\in\mathcal{B},\ t_j\in[0,T],\ j=1,\dots,n, \tag{3.1}$$

for $n\in\mathbb{N}$.

From the definition (3.1) it follows that

$$\mu^X_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(A_1\times\cdots\times A_n) = \mu_{t_1,\dots,t_n}(A_{\sigma^{-1}(t_1)}\times\cdots\times A_{\sigma^{-1}(t_n)}) \tag{3.2}$$

for each permutation σ on the set $1, 2, ..., n$ and

$$\mu^X_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \mu^X_{t_1,\dots,t_n,t_{n+1},\dots,t_{n+k}}(A_1 \times \cdots \times A_n \times E_{n+1} \times \cdots \times E_{n+k})$$

for $k \in \mathbb{N}$, where $E_j = E$, $j = n+1, \dots, n+k$. Conversely, we have the following result:

Theorem 3.1. (Daniell-Kolmogorov extension theorem). *Let a family* $\{\mu^X_{t_1,\dots,t_n},\ t_j \in [0,T];\ j = 1, 2, \dots, n;\ n \in \mathbb{N}\}$ *of normalized measures be given that satisfies consistency conditions* (3.1) *and* (3.2). *Then there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a family of* E*-valued random variables* $(X_t)_{t\in[0,T]}$ *on* Ω *such that* (3.1) *holds.*

The above Theorem (and its proof) makes it possible to construct a stochastic process from the family of its finite-dimensional distributions (that must satisfy natural consistency conditions). It also explains a way to "construct" a probability space for a process that is usually "given" just by its finite-dimensional distributions. Namely, one can set

$$\Omega = E^{[0,T]},\ \mathcal{F} = \sigma\{C_{t_1,\dots,t_n}(A_1, \dots, A_n),\ t_j \in [0,T],\ A_j \in \mathcal{B}\}$$

where C denotes the cylindrical set

$$C_{t_1,\dots,t_n}(A_1, \dots, A_n) = \{\omega \in E^{[0,T]};\ \omega(t_1) \in A_1, \dots, \omega(t_n) \in A_n\}, \tag{3.3}$$

$$\mathbb{P}[C_{t_1,\dots,t_n}(A_1, \dots, A_n)] := \mu^X_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) \tag{3.4}$$

($\mathbb{P}$ may be extended to a measure on $\mathcal{F}$) and for $\omega \in \Omega = E^{[0,T]}$,

$$X(t, \omega) := \omega(t)$$

(X is often called a "canonical" process).

The space $E^{[0,T]}$ is very large and sometimes it may be convenient to choose a smaller "space of possible sample paths" for Ω, when we have a good reason to believe that the paths really do live in such a space.

Now we should try to formalize the general feeling that changing the process on $\mathbb{P}$-measure zero does not mean so much. As we have uncountable many "times", there is a reason to be careful.

Definition 3.2. Let (X_t), (Y_t) be stochastic processes defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) The process (Y_t) is a version (modification) of a process (X_t) if

$$\forall\ t \in [0,T] \quad \mathbb{P}[X_t \neq Y_t] = 0.$$

(ii) The processes X and Y are indistinguishable if

$$\mathbb{P}[X_t = Y_t \quad \forall\, t \in [0,T]] = 1.$$

We have the following simple result:

Proposition 3.3. *Let X be a modification of Y and suppose that both X, Y have $\mathbb{P}$-a.s. right- or left-continuous paths. Then X and Y are indistinguishable.*

The finite-dimensional distributions in principle cannot provide enough information to verify such properties like path continuity. So we need to find a relatively simple criterion that would make it possible to verify that the process has a continuous modification. Such a tool is provided in the following theorem that is often called a Kolmogorov-Chentsov theorem (or the Kolmogorov continuity test):

Theorem 3.4. *Let there exist $K > 0$, $m > 0$, $\alpha > 0$ such that*

$$\mathbb{E}\|X_t - X_s\|^m \leq K|t-s|^{1+\alpha}, \quad t, s \in [0,T]. \tag{3.5}$$

Then the process (X_t) has a version with β-Hölder continuous paths for $\beta < \frac{\alpha}{m}$ (in particular, it has a version with continuous paths).

Note that the left side of (3.5) is determined just by finite-dimensional distributions of the process (X_t) (in particular, by the measure $\mu^X_{t,s}$ on $\mathcal{B}\times\mathcal{B}$).

4 Brownian motion

In the present Section the standard Wiener process (or Brownian motion) is defined and some of its basic properties are listed. The Wiener process is a special stochastic process that is, in a sense, one of the central notions in Stochastic Analysis. At first some motivations and historical remarks are given.

4.1 Historical remarks and heuristics

With a little bit of exaggeration we may say that two main areas of application of stochastic analysis are physics (and more generally, science), where it serves as a useful model of diffusive behaviour and of various types of random perturbations, and finance mathematics, where it models "market" in a broad sense, for instance stock prices. It is quite interesting that stochastic analysis was standing on these two legs from the very beginning.

The process that is nowadays called " Wiener process" or "Brownian motion" was used for the first time in 1900 by L. Bachelier (who obtained it as a limit of discrete random walks) to describe fluctuations in stock prices. Just a few years (1905) later it has been rediscovered by A. Einstein and Smoluchowski ([Ei], [ES]) in an attempt to study Brownian phenomena. Let us explain their model in more details using the terminology introduced in the previous section (rather than their original language). R. Brown in 1826–27 observed an irregular motion of pollen particles suspended in water. He noted that the path of each particle is very irregular, having a tangent at no point, and the motions of two distinct particles appear to be independent. Because of the radial symmetry the "state space" of the particles is considered to be one-dimensional (we may think of a projection of the movement on a given straight line). Let ω denote the particle and $W(t,\omega)$ its position at time t. The average displacement of a large numbers of particles at a given time interval $[s,t]$ is zero, i.e.

$$\mathbb{E}(W(t,\cdot) - W(s,\cdot)) = 0 \tag{4.1}$$

and as was observed, the "dispersion" of the moving particles is directly proportional to the square root of elapsed time, that is

$$\mathbb{E}(W(t,\cdot) - W(s,\cdot))^2 = c(t-s),\ s<t. \tag{4.2}$$

Under the assumption that "higher moments" are negligible on small time intervals, i.e.

$$\mathbb{E}(W(t,\cdot) - W(s,\cdot))^n = o(t-s),\ n>2, \tag{4.3}$$

Einstein derived an equation for the density $p(t,x,y)$ (we may interpret it as a density of the probability distribution of the position at time t of a random particle that starts from x at time zero), namely

$$\frac{\partial p}{\partial t}(t,x,y) = \frac{1}{2}c\frac{\partial^2 p}{\partial y^2}(t,x,y), \quad (t,x,y)\in(0,T)\times\mathbb{R}\times\mathbb{R}, \tag{4.4}$$

with the initial condition

$$p(0,x,y) = \delta_x(y) \tag{4.5}$$

where δ_x is the Dirac function at x. PDE's experts recognize p as the fundamental solution to the parabolic equation (4.4) that has the form

$$p(t,x,y) = \frac{1}{\sqrt{2\pi ct}} \exp\Big\{ \frac{-(x-y)^2}{2ct} \Big\}, \tag{4.6}$$

and all probabilists know that (4.6) is a formula for the Gaussian density. This relation between PDE's and probabilistic structures will be elaborated in the next sections. To finish the story, Einstein computed that

$$c = \frac{RT}{Nf}$$

where R is the gas constant, T is the absolute temperature, f is the friction coefficient and N the Avogadro number. This equation and observed properties of Brownian motion allowed J. Perrin to compute $N \sim 6 \cdot 10^{23}$ (the number of molecules in a mole) and in fact, it helped to confirm the atomic theory of matter (see e.g. [Cs] for more details on physical background of the theory). Now we will try to approach the definition of the Wiener process from another side, through a random perturbation of a deterministic system. This will also motivate the form and definition of stochastic differential equation (and stochastic integral) that are studied in subsequent sections. We will start from a discrete time deterministic model

$$\begin{aligned} x(t_{i+1}) - x(t_i) &= f(t_i, x(t_i))\Delta t_i, \quad i = 0,1,2,\ldots,n, \\ x(0) &= x_0, \end{aligned} \tag{4.7}$$

where $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ and $x_0 \in \mathbb{R}$ are given, (t_i) is a partition of the interval $[0,T]$, $t_0 = 0 < t_1 < t_2 < \cdots < t_n = T$ and $\Delta t_i = t_{i+1} - t_i$. Such a system may be often derived to model some specific phenomena that are of interest. It is trivial to "solve" the system (4.7), one has

$$x(t_k) = x_0 + \sum_{i=0}^{k-1} f(t_i, x(t_i))\Delta t_i, \ k = 1,2,\ldots,n. \tag{4.8}$$

When $\max|\Delta t_i|$ becomes very small, by a limiting passage in (4.7) (divided by Δt_i) we obtain a continuous-time analogue of (4.7) that is an ordinary differential equation

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)), \quad t \in (0, T) \\ x(0) &= x_0, \end{aligned} \tag{4.9}$$

the solution of which is a continuous-time analogue of (4.8),

$$x(t) = x_0 + \int_0^t f(s, x(s))ds, \ t \in [0, T]. \tag{4.10}$$

Obviously, unlike the explicit recursive formula (4.8), (4.10) is again an equation that is to be solved. As is now generally accepted, we will interpret the integral in (4.10) as the Lebesgue integral. Let us consider the "canonical" case of random perturbations of these systems. In the discrete-time case it should be

$$\begin{aligned} x(t_{i+1}) - x(t_i) = f(t_i, x(t_i))\Delta t_i + \sigma(t_i, x(t_i))N(t_i, \omega)\Delta t_i, \\ x(0) = x_0, \quad i = 0, 1, 2, \ldots, n \end{aligned} \tag{4.11}$$

where $\sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $N(t_i, \omega) = N_{t_i}$ is a discrete-time random process (or a sequence of random variables) with the following properties:

$$\mathbb{E}N_{t_i} = 0 \tag{4.12}$$

(otherwise we would "add" the expectation to the first term on the r.h.s. of (4.11)),

$$\text{Law}(N_{t_i}) \equiv \mu \text{ does not depend on } i, \ \mathbb{E}N_{t_i}^2 = 1 \tag{4.13}$$

(i.e., the process is stationary, the probabilistic characteristics does not change in time; the variance is finite and may be normalized) and

$$N_{t_i} \text{ and } N_{t_j} \text{ are independent for } i \neq j. \tag{4.14}$$

(i.e., the process has "no memory").

The process with properties (4.12)–(4.14) is often called *white noise* (discrete-time) and one can solve (4.11) recursively as in (4.8). However, what is the continuous counterpart of (4.11)? Obviously, this should be the differential equation

$$\dot{x}(t) = f(t, x(t)) + \sigma(t, x(t))N(t, \omega), \quad t \in (0, T) \tag{4.15}$$

$$x(0) = x_0 \tag{4.16}$$

where $N(t, \omega) = N_t(\omega)$ is now a stochastic process satisfying the conditions

(i) $\mathbb{E}N_t = 0$

(ii) N_t is stationary, $\mathbb{E}N_t^2 = 1$

(iii) N_t and N_s are independent for any $t \neq s$.

The process satisfying (i)–(iii) (often used in applications) and is called a (continuous-time) *white noise*. It has however one special feature: It does not exist (as a measurable process). Indeed, we have

Proposition 4.1. *Let (N_t), $t \in [0,T]$, be a stochastic process satisfying* (i)–(iii). *Then the mapping $N : [0,T] \times \Omega \to \mathbb{R}$ is not measurable w.r.t. the product σ-algebra $\mathcal{B}([0,T]) \times \mathcal{F}$.*

For the proof cf. [Kal].

It is apparent that a non-measurable stochastic process is of little use – it cannot be handled by usual mathematical tools. So something must be done here: One may introduce a class of generalized random processes where white noise may be treated. This has been successfully done (see e.g. [Hi]; the story resembles the Dirac delta function and theory of distributions). However, this theory is closely related to some advanced parts of functional analysis and for our purpose we may try something simpler. Assume that the random perturbation is written as an **increment** of a process, i.e. instead of (4.11) we consider

$$\begin{aligned} x(t_{i+1}) - x(t_i) &= f(t_i, x(t_i))\Delta t_i + \sigma(t_i, x(t_i))(W(t_{i+1}, \omega) - W(t_i, \omega)), \\ x(0) &= x_0, \ i = 1, 2, \ldots, n \end{aligned} \tag{4.17}$$

where $W(t, \omega) = W_t$ is a process on $[0,T]$ (to simplify the presentation we may consider directly a continuous time process). The solution to (4.17) is given recursively,

$$\begin{aligned} x(t_k) = x_0 &+ \sum_{i=0}^{k-1} f(t_i, x(t_i))\Delta t_i \\ &+ \sum_{i=0}^{k-1} \sigma(t_i, x(t_i))(W(t_{i+1}, \omega) - W(t_i, \omega)), \ k = 1, 2, \ldots, n. \end{aligned} \tag{4.18}$$

of course, we try to model the same situation as above. Therefore, roughly speaking, "the increments of W" should analogous properties as "the values of N". More specifically, we assume

(W1) $W_0 = 0$, $\mathbb{E}(W_t - W_s) = 0$, $0 \le s \le t < \infty$,

(W2) *the process* $t \mapsto W_{t+s} - W_t$ *is stationary for each* $s \ge 0$ *and*

$$\mathbb{E}(W_t - W_s)^2 = t - s, \quad 0 \le s \le t < \infty,$$

(W3) *for each* $0 \le t_1 \le t_2 \le t_3 \le t_4$ *the random variables* $W_{t_4} - W_{t_3}$ *and* $W_{t_2} - W_{t_1}$ *are stochastically independent.*

Having made a bad experience with the paths of the white noise, we would like to ensure that the paths of W are reasonable. We might wish

(W4) *The paths* $t \mapsto W(t, \omega)$ *are continuous for* $\mathbb{P}$*-almost all* $w \in \Omega$.

The following remarkable result holds true:

Proposition 4.2. *The following two sets of conditions are equivalent:* (W1)–(W4) $\Leftrightarrow$ (W1), (W2′), (W3), (W4) *where*

(W2′) *For* $0 \le s < t \le T$ *the random variable* $W_t - W_s$ *is Gaussian with the mean zero and variance* $(t - s)$*, that is, the density of its probability law is given*

$$\frac{d\,\mathrm{Law}(W_t - W_s)}{dx}(x) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\Big\{\frac{-x^2}{2(t-s)}\Big\}.$$

The proof of $\Leftarrow$ is obvious but the converse implication is not trivial at all (cf. [KS] for the proof). It may be surprising that conditions (W1)–(W4) are rather general, with no specific requirements on density of the law of (W_t), but in (W2′) we already got a formula for the density!

Definition 4.3. A stochastic process (W_t) satisfying (W1)–(W4) (or, equivalently, (W1), (W2′), (W3), (W4)) is called a (standard, one-dimensional) Wiener process (or Brownian motion).

Having defined an appropriate process (W_t) we may try to guess what the continuous-time counterparts of (4.17) and (4.18) should be. In place of (4.17) we can easily imagine a formal differential

$$dx(t) = f(t, x(t))dt + \sigma(t, x(t))dW(t, \omega), \quad t > 0$$

$$x(0) = x_0,$$

which is the common form of a stochastic differential equation. Its solution may be defined via a continuous-time counterpart of (4.18),

$$x(t) = x_0 + \int_0^t f(s.x(s))ds + \int_0^t \sigma(s, x(s))dW(s, \omega), \quad t \geq 0,$$

where the first integral on the r.h.s. is the Lebesgue integral (with $\omega \in \Omega$ as a parameter) and the second one should be obtained as a suitable limit of partial integral sums of Stieltjes type! This is a motivation for the definition of stochastic integral of the Ito type studied below.

4.2 Basic properties

In this subsection some basic properties of the Wiener process are listed. Its first (and very important) feature is that it does exist, which is a nontrivial fact (because Wiener process is a nontrivial object).

Theorem 4.4. *The Wiener process does exist.*

Proof. It is possible to construct a Wiener process as an infinite series of much simpler, "explicitly given" processes as was done by Ciesielski [Ci]. However, we may also use the results of the previous section. The finite-dimensional distributions of (W_t) may be guessed from (W2′) and (W3), namely

$$\mu^W_{t_1,\ldots,t_n}\Big(\prod_{j=1}^n (a_j, b_j)\Big) = \mathbb{P}[W_{t_1} \in (a_1, b_1), \ldots, W_{t_n} \in (a_n, b_n)] = \tag{4.19}$$
$$= \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} p(t_1, x_1) \cdot p(t_2 - t_1, x_2 - x_1)$$
$$\ldots p(t_n - t_{n-1}, x_n - x_{n-1}) dx_n \ldots dx_1$$

where

$$p(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \tag{4.20}$$

and $a_j \leq b_j$, $j = 1, 2, \ldots, n$. It may be checked that the set of measures $(\mu^W_{t_1,\ldots,t_n})$ satisfies the consistency conditions (3.1), (3.2) of the (Daniell-Kolmogorov) Theorem 3.1, therefore there exists a process $(\widetilde{W}_t)$ satisfying (W1), (W2′) and (W3). We still need the paths continuity (W4) and to this

end we may apply the Kolmogorov continuity test (Theorem 3.4). Indeed, using the density (4.20) it is easy to compute

$$\mathbb{E}|W_t - W_s|^{2m} = \int_{-\infty}^{\infty} |x|^{2m} p(t-s,x)dx = C_m|t-s|^m, \ m \in \mathbb{N}, \qquad (4.21)$$

where C_m is a constant depending only on m. Therefore, taking arbitrary $m \geq 2$ we obtain (3.5) and hence there exist a version (W_t) of $(\widetilde{W}_t)$ with continuous sample paths, which is the process we are looking for. In fact, Theorem 3.4 yields even Hölder continuity of paths with exponent $\beta < \frac{m-1}{2m}$. Since m is arbitrarily large, we arrive at □

Corollary 4.5. *The paths of the Wiener process are β-Hölder continuous for $\beta < \frac{1}{2}$.*

It may be worth to note that $\beta < \frac{1}{2}$ is optimal: The paths of Wiener process are **not** $\frac{1}{2}$-Hölder continuous at any point in $[0,T]$. In fact, it is possible to find exactly the modulus of continuity of paths of the Wiener process, which is $cg(\delta)$ almost surely for $c > 1$, where

$$g(\delta) = (2\delta \log(1/\delta))^{1/2}, \quad \delta > 0,$$

while for $c < 1$ it is is a modulus of continuity for almost no Wiener path (cf. [KS]). This remarkable result has been proved by P.Lévy in 1937 and its celebrated consequence is

Theorem 4.6. *For $\mathbb{P}$-almost every $\omega \in \Omega$, the sample path $t \mapsto W(t,\omega)$ is nowhere differentiable and is of infinite variation on each subinterval of $[0,T]$.*

What concerns the large and small time behaviour of the Wiener process, the precise description is given in the following Theorem.

Theorem 4.7. (The Law of Iterated Logarithm, Khinchine 1933) *Let (W_t) be a Wiener process. Then*

$$\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1, \quad \liminf_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = -1,$$
$$\limsup_{t\to 0+} \frac{W_t}{\sqrt{2t\log\log \frac{1}{t}}} = 1, \quad \liminf_{t\to 0+} \frac{W_t}{\sqrt{2t\log\log \frac{1}{t}}} = -1.$$

As an immediate consequence we obtain the following statement.

Corollary 4.8. (The Strong Law of Large Numbers for Wiener process) *We have that*

$$\lim_{k\to\infty} \frac{W_t}{t} = 0, \quad \mathbb{P}\text{-}a.s.$$

Wiener process is also a particular case of processes called martingales which are studied in the next section.

5 Martingales

5.1 Filtration and measurability

Definition 5.1. (i) *A filtration on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is an arbitrary system* $(\mathcal{F}_t)$, $t \geq 0$, *of* σ*-algebras of* Ω, *such that* $\mathcal{F}_t \subset \mathcal{F}$ *and* $\mathcal{F}_s \subset \mathcal{F}_t$ *for each* $0 \leq s \leq t < \infty$.

(ii) *Given a stochastic process* (X_t) *on* $(\Omega, \mathcal{F}, \mathbb{P})$, *the filtration generated by* (X_t) *is defined by*

$$\hat{\mathcal{F}}_t^X := \sigma\{X_s \in A;\ 0 \leq s \leq t,\ A \in \mathcal{B}\}$$

(iii) *Let* (W_t) *and* $(\mathcal{F}_t)$ *be a Wiener process and a filtration, respectively, on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *Then* (W_t) *is said to be a Wiener process w.r.t.* $(\mathcal{F}_t)$ *is* W_s *is* $\mathcal{F}_s$*-measurable and* $W_t - W_s$ *and* $\mathcal{F}_s$ *are stochastically independent for each* $0 \leq s < t < \infty$.

Obviously, each Wiener process (W_t) is a W.p. with respect to its own filtration $(\hat{\mathcal{F}}_t^W)$.

Given a filtration $(\mathcal{F}_t)$, the quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is called the stochastic basis. Furthermore, for $t \geq 0$ set $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$. Clearly, $(\mathcal{F}_{t+})$ is a filtration and if $\mathcal{F}_{t+} = \mathcal{F}_t$, $t \geq 0$, the filtration $(\mathcal{F}_t)$ is said to be right-continuous. Analogously, define $\mathcal{F}_{t-} := \sigma a\{\bigcup_{s<t} \mathcal{F}_s\}$, $t > 0$ and the left-continuity of the filtration $(\mathcal{F}_t)$ by the demand $\mathcal{F}_{t-} = \mathcal{F}_t$, $t > 0$.

We say that a fitration $(\mathcal{F}_t)$ satisfies the usual conditions (UC) if it is right-continuous and $\mathcal{F}_0$ (hence $\mathcal{F}_t$ for each $t \geq 0$) contains the system $\mathcal{N} = \{N \in \mathcal{F};\ \mathbb{P}(N) = 0\}$ of all sets of probability zero.

Proposition 5.2. *Given a filtration $(\mathcal{F}_t)$ there exists the smallest filtration $(\mathcal{G}_t)$ satisfy (UC) and containing $(\mathcal{F}_t)$. It holds that*

$$\mathcal{G}_t = \bigcap_{s>t} \sigma(\mathcal{F}_s \cup \mathcal{N}),\ t \geq 0.$$

The filtration $(\mathcal{G}_t)$ is called augmentation of $(\mathcal{F}_t)$.

For a stochastic process (X_t) we denote by $(\mathcal{F}_t^X)$ the augmentation of $(\widehat{\mathcal{F}}_t^X)$.

Note that $(\widehat{\mathcal{F}}_t^W)$ does not satisfy (UC), fortunately (W_t) is also Wiener process with respect to the augmentation $(\mathcal{F}_t^W)$.

The importance of (UC) will be seen in subsequent sections where we work with random times.

In what follows, we introduce several concepts of measurability of a stochastic process $(X_t)_{t\geq 0}$ defined on a stochastic basis. To avoid ambiguities we say that a mapping $f : Y \to Z$ is $(\mathcal{A}/\mathcal{B})$-measurable where $(Y, \mathcal{A})$ and $(Z, \mathcal{B})$ are measurable spaces, if $f^{-1}(\mathcal{B}) \subset \mathcal{A}$.

Recall that $(X_t)_{t\geq 0}$ may be understood as a map $X : \mathbb{R}_t \times \Omega \to \mathbb{R}^m$.

The process $(X_t)_{t\geq 0}$ is said to be

– measurable if $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}^m$ is $(\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F})/\mathcal{B}(\mathbb{R}^m)$-measurable

– adapted (w.r.t. $(\mathcal{F}_t)$) if $X_t : \Omega \to \mathbb{R}^m$ is $\mathcal{F}_t/\mathcal{B}(\mathbb{R}^m)$-measurable for each $t \geq 0$

– progressively measurable (or progressive) if for each $t > 0$ the mapping $X : [0, t] \times \Omega \to \mathbb{R}^m$ is $(\mathcal{B}([0, t]) \otimes \mathcal{F}_t)/\mathcal{B}(\mathbb{R}^m)$-measurable.

Remark 5.3. Each progressively measurable process is both measurable and adapted. Each adapted, right- or left-continuous process is progressively measurable (prove!). Each measurable and adapted process has a progressively measurable version (the proof of this is difficult).

If the (UC) are satisfied, each modification of an adapted process is again adapted. This allows to consider ALL paths of an (almost surely) continuous process to be continuous, without loosing adaptiveness.

5.2 Martingales-Fundamentals

In this subsection we assume that a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t,)\mathbb{P})$ is given.

Definition 5.4. Let (X_t) and $(\mathcal{F}_t)$ be an $\mathbb{R}^n$-valued stochastic process and a filtration, respectively, such that (X_t) is $(\mathcal{F}_t)$-adapted, $\mathbb{E}|X_t| < \infty$ for each $t \geq 0$ and

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s \quad \mathbb{P}\text{-a.s.}, \ 0 \leq s \leq t < \infty. \tag{5.1}$$

Then (X_t) is called a martingale with respect to $(\mathcal{F}_t)$ Analogously, for $n = 1$, (X_t) is called submartingale and supermartingale for $\geq$, $\leq$ in (5.1), respectively. If moreover, $\mathbb{E}|X_t|^2 < \infty$, (X_t) is called an L^2-martingale or square integrable (sub-, super-) martingale.

Martingales have a lot of nice properties that can be found in most monographs or textbooks devoted to stochastic analysis (e.g. [KS]). Here we list some of them.

Proposition 5.5. *If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}(\varphi(X_t))^+ < \infty$ for each $t \geq 0$ then*

(i) *If (X_t) is a martingale then $(\varphi(X_t))$ is a submartingale.*

(ii) *If (X_t) is a submartingale an φ is nondecreasing then $(\varphi(X_t))$ is a submartingale.*

Proof. By a "conditional" version of Jensen inequality. □

Contemplation: Compare the definition of martingale with the one of Markov property:

$$\mathbb{E}[X_t \mid \widehat{\mathcal{F}}_s^X] = \mathbb{E}[X_t \mid X_s] := \mathbb{E}[X_t \mid \hat{\mathcal{F}}^{X_s}] \ \mathbb{P}\text{-a.s.}, \ t > s.$$

The filtration $(\mathcal{F}_t)$ may be augmented so that the "bigger" filtration satisfies (UC) and the process (X_t) is still a martingale w.r.t. this filtration.

Theorem 5.6. *Assume that (X_t) is a submartingale w.r.t. a filtration $(\mathcal{F}_t)$ satisfying (UC) and the function $t \mapsto \mathbb{E}X_t$ is right-continuous (this is trivially satifies if (X_t) is a martingale). Then (X_t) has a right-continuous modification.*

Remark. More precisely, (X_t) then has càdlàg paths (right-continuous and such that there exist finite limits from the left); càdlàg = abbrev. of "continue à droite, limite à gauche".

Examples of martingales: Processes with independent increments, compensated power of a Wiener process, stochastic exponential, compensated Poisson process.

Definition 5.7. The random variable $\tau : \Omega \to \mathbb{R}_+ \cup \{+\infty\}$ is called a stopping time (w.r.t. $(\mathcal{F}_t)$) if

$$\{\omega \in \Omega;\ \tau(\omega) \leq t\} = [\tau \leq t] \in \mathcal{F}_t,$$

(or, equivalently, $[\tau > t] \in \mathcal{F}_t$) for each $t \geq 0$. Similarly, τ is called an optional time if $[\tau < t] \in \mathcal{F}_t$, $t \geq 0$.

Exercise. Each stopping time is optional. If $(\mathcal{F}_t)$ is right-continuous, the converse is true as well.

Exercise. If τ is a stopping time (ST), $\sigma = 2\tau$ is a stopping time, while $\kappa = \frac{1}{2}\tau$ (in general) is not.

Proposition 5.8. *Let $(\mathcal{F}_t)$ be right-continuous. Then*
(i) *τ, σ are ST $\Rightarrow$ $\tau + \sigma$ is ST (not $\tau - \sigma$), $\tau \wedge \sigma$ $\tau \vee \sigma$ are ST.*
(ii) *(τ_n) is a sequence of ST $\Rightarrow$ $\sup_n \tau_n$, $\inf_n \tau_n$, $\limsup \tau_n$, $\liminf \tau_n$ are ST.*

Theorem 5.9. *Assume that (UC) are satisfied and let $B \subset \mathbb{R}^n$ be either open or closed set, (X_t) an adapted continuous process in $\mathbb{R}^n$. Let*

$$\tau_B := \inf\{t \geq 0,\ X_\tau \in B\}$$

define the hitting time of B (= exit time of $\mathbb{R}^n \setminus B$) with the convention $\inf\{\emptyset\} = +\infty$. Then τ_B is a stopping times.

Definition 5.10. Let τ be a stopping time w.r.t. $(\mathcal{F}_t)_{t\geq 0}$. The σ-algebra $\mathcal{F}_\tau$ of events taking place before time τ consists of those events $A \in \mathcal{F}$ for which

$$A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \text{for each } t \geq 0.$$

Theorem 5.11. (random stopping) *Set $X_\tau(\omega) := X_{\tau(\omega)}$ (assuming $\tau < \infty$ $\mathbb{P}$-a.s.). If (X_t) is progressively measurable and the "stopped process" (Y_t), $Y_t := X_{\tau\wedge t}$ is progressively measurable.*

Proof. A nice exercise on measurability of composition maps. In general, the statement is false if "progressive measurability" is replaced by "adaptivity". □

Theorem 5.12. (optional sampling theorem) *Let (X_t). be a right-continuous submartingale w.r.t. filtration $(\mathcal{F}_t)$ satisfying (UC), τ and σ are bounded stopping times such that $\sigma \leq \tau$ $\mathbb{P}$-a.s. Then*

(i)
$$\mathbb{E}[X_\tau \mid \mathcal{F}_\sigma] \geq X_\sigma \quad \mathbb{P}\text{-}a.s.$$
If (X_t) is a martingale, (i) holds with equality.

(ii) *A stopped (sub)martingale (Y_t), $Y_t := X_{\tau\wedge t}$ is a (sub)martingale.*

One of the most important properties of martingales is the possibility of estimating maxima of paths by their values at the endpoints of intervals:

Theorem 5.13. (Doob maximal inequalities) *Let (X_t) be a martingale or (for $n = 1$) a nonnegative submartingale. Then*

(i) $\forall\ p \geq 1,\ \lambda > 0,\ T > 0$
$$\lambda^p \mathbb{P}[\sup_{0\leq t\leq T} |X_t| \geq \lambda] \leq \sup_{t\in[0,T]} \mathbb{E}|X_t|^p = \mathbb{E}|X_T|^p$$

(ii) $\forall\ p > 1,\ T > 0$
$$\mathbb{E}|X_T|^p \leq \mathbb{E}\sup_{t\in[0,T]} |X_t|^p \leq \left(\frac{p}{p-1}\right)^p \sup_{t\in[0,T]} \mathbb{E}|X_t|^p = \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_T|^p.$$

5.3 Doob-Meyer Decomposition

In the present subsection, $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ denotes a stochastic basis satisfying (UC).

Definition 5.14. An adapted (1-dimensional) process (A_t) is called increasing if $A_0 = 0$, $t \mapsto A_t$ are $\mathbb{P}$-a.s. nondecreasing continuous functions and $\mathbb{E}A_T < \infty$ for each $T > 0$. An increasing process is called natural if
$$\mathbb{E}\int_{[0,t]} M_s dA_s = \mathbb{E}\int_{[0,t]} M_{s-} dA_s,\ t \geq 0,$$
holds for each right-continuous bounded martingale (M_t).

Remark 5.15. Let $\mathcal{P}$ denote the smallest σ-algebra on the product space $\mathbb{R}_+ \times \Omega$ such that all left-continuous, adapted processes $Y : \mathbb{R}_+ \times \Omega \to \mathbb{R}^n$ are $\mathcal{P}$-measurable (the sigma-algebra of predictable sets). A predictable process is defined as a mapping $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}^n$ which is $\mathcal{P}$-measurable. It may be shown that each predictable process is progressive (In the discrete-time case, an adapted random sequence (X_n) is predictable if X_n is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$). It may be shown that an increasing process is natural if and only if it is predictable.

Definition 5.16. A right-continuous, adapted process (X_t) is said to be of class (DL) is for each $a > 0$ the set $(X_\tau)_{\tau \in \mathcal{S}_a}$ is uniformly integrable where

$$\mathcal{S}_a := \{\tau : \Omega \to [0, a],\ \tau \text{ is a stopping time}\}.$$

Remark 5.17. If (X_t) is right-continuous one-dimensional submartingale then

(i) $(X_t \geq 0\ \mathbb{P}\text{-a.s.} \quad \forall\ t \geq 0) \Rightarrow (X_t)$ is of class (DL).

(ii) For any increasing process (A_t) and a martingale (M_t), $(X_t = M_t + A_t \quad \forall\ t \geq 0) \Rightarrow (X_t)$ is of class (DL).

The converse is the celebrated Doob-Meyer theorem:

Theorem 5.18. (Doob-Mayer decomposition) *If (X_t) is a right-continuous one-dimensional submartingale of class (DL) then*

$$X_t = M_t + A_t,\ t \geq 0,$$

where (M_t) is a (right-continuous) martingale and (A_t) is an increasing process. The process (A_t) may be chosen natural and with this condition is the above decomposition unique.

Proof. Quite instructive in the discrete-time case. In the rest of this section the processes are supposed to be one-dimensional.

Denote by $\mathcal{M}_2$ the linear space of L^2-martingales, right-continue and such that $X_0 = 0$. By $\mathcal{M}_2^c$ we denote the subspace of $\mathcal{M}_2$ consisting of continuous martingales. □

Proposition 5.19. *For $(X_t) \in \mathcal{M}_2$ there is a unique decomposition*

$$X_t^2 = M_t + A_t,$$

where (M_t) is a in $\mathcal{M}_2$ and (A_t) is a natural (predictable) increasing process. If $(X_t) \in \mathcal{M}_2^c$ then (M_t), (A_t) are continuous.

Proof. Follows directly from Theorem 5.18 and Remark 5.16 a since (X_t^2) is a nonnegative submartingale. □

Definition 5.20. For $(X_t) \in \mathcal{M}_2$, the quadratic variation process (denoted by $(\langle X \rangle_t)$) is the increasing process (A_t) from Proposition 5.19. Thus it is the unique natural increasing process such that $(X_t^2 - \langle M \rangle_t)$ is a martingale.

Exercises. Compute the quadratic variation of the Wiener and comparsated Poisson processes.

Consider an arbitrary random process (X_t).

Given a partition $\mathbb{T} = \{t_0, t_1, \dots, t_m\}$ of the interval $[0, t]$, $0 = t_0 < t_1 < \dots < t_m = t$, and $p > 0$, set

$$V_t^{(p)}(\mathbb{T}) := \sum_{k=1}^{m} |X_{t_n} - X_{t_{k-1}}|^p, \ \|\mathbb{T}\| := \max_k |t_{k-1} - t_n|.$$

Theorem 5.21. *Let $(X_t) = \mathcal{M}_2^c$. Then*

$$\lim_{\|\mathbb{T}\|\to 0} V_t^{(2)}(\mathbb{T}) = \langle X\rangle_t \quad \textit{in } \mathbb{P}$$

for each $t \geq 0$, that is, for each $\varepsilon > 0$ there exists $\vartheta > 0$ such that for each partition $\mathbb{T}$, $\|\mathbb{T}\| < \vartheta$, we have

$$\mathbb{P}[|V_t^{(2)}(\mathbb{T}) - \langle X\rangle_t| > \varepsilon] < \varepsilon.$$

Remarks. 1) The above statement may be false for discontinuous processes.

2) In general if

$$\lim_{\|\mathbb{T}\|\to 0} V_t^{(p)}(\mathbb{T}) = L_t \quad \text{in } \mathbb{P},$$

and $L_t \in (0, +\infty)$ $\mathbb{P}$ a.s., $t \geq 0$, then

$$\lim_{\|\mathbb{T}\|\to 0} V_t^{(q)}(\mathbb{T}) = 0 \quad \text{in } \mathbb{P},$$
$$\lim_{\|\mathbb{T}\|\to 0} V_t^{(q')}(\mathbb{T}) = +\infty \quad \text{in } \mathbb{P}$$

for each $q > p$, $q' < p$. In particular, the paths of processes in $\mathcal{M}_2^c$ typically do not have bounded variation and may not be used as integrators in Lebesgue-Stieltjes integral.

3) If $X_t = W_t$ is a Wiener process, the statement of Theorem 5.22 may be modified: If $(\mathbb{T}_n)$ is a sequence of partitions of the interval $[0, t]$; $\|\mathbb{T}_n\| \to 0$, then

$$\mathbb{E}|V_t^{(2)}(\mathbb{T}_n) - t|^2 \to 0, \ n \to \infty.$$

If, moreover, either $\sum_{n=1}^{\infty} \|\mathbb{T}_n\| < \infty$ or $(\mathbb{T}_n)$ are nested ($\mathbb{T}_{n+1} \supset \mathbb{T}_n$ for each $n \in \mathbb{N}$) then

$$V_t^{(2)}(\mathbb{T}_n) \to \langle W \rangle_t = t \quad \mathbb{P}\text{-a.s.}$$

We conclude the present paragraph by the following very useful result:

Theorem 5.22. (Lévy) *Let (X_t) be a continuous adapted process in $\mathbb{R}$ such that the process (M_t) defined by $M_t := X_t - X_0$ is in $\mathcal{M}_2^c$ and $\langle M \rangle_t = t$, $t \geq 0$. Then (M_t) is the Wiener process.*

6 Stochastic integration

In this section we will define a stochastic integral that is usually written in the form

$$\int_0^T \varphi(s) dW_s \tag{6.1}$$

where (W_t) is a Wiener process and φ is another stochastic process. The form (6.1) resembles the Stieltjes integral and indeed, it will be defined as a suitable limit of Stieltjes-type partial sums. A motivation for this definition has been given in Section 2.1. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a stochastic basis and (W_t) a standard, one-dimensional Wiener process defined on it. At first we define a class of stochastic processes on Ω that will turn out to be "admissible integrands". Set

$$M_w^2(0,T) := \{(X_t) \in L^2((0,T) \times \Omega); (X_t) \text{ is } (\mathcal{F}_t)\text{-progressive}\} \tag{6.2}$$

and

$$\begin{aligned} \mathcal{E} = \mathcal{E}_w(0,T) :=& \{(Y_t) \in M_w^2(0,T);\ \exists\, t_0 = 0 < t_1 < \cdots < t_n = T; \\ & Y_t \equiv Y^{(i)} \text{ for } t \in [t_i, t_{i+1})\}. \end{aligned} \tag{6.3}$$

Alternatively,

$$Y_t = \sum_{i=0}^{n-1} Y^{(i)} 1_{[t_i, t_{i+1})}(t),\ t \in [0,T], \tag{6.4}$$

where (t_i) is a partition of the interval $[0,T]$ as in (6.3) and $Y^{(i)}$ are random variables on Ω, $Y^{(i)} \in L^2(\Omega)$ is $\mathcal{F}_{t_i}$-measurable for each i. The elements of $\mathcal{E}$ are called *step functions*. We have the following

Lemma 6.1. (i) $M_w^2(0,T)$ *is a closed subspace of* $L^2((0,T)\times\Omega)$.

(ii) $\mathcal{E}_w(0,T)$ *is dense in* $M_w^2(0,T)$ *with respect to the norm of* $L^2((0,T)\times\Omega)$.

While the proof of (i) is obvious, (ii) requires some technical work. The basis idea follows the lines of standard analytical methods, using e.g. either mollifiers or "averages over small intervals". However, it must be made sure that the chosen technique will produce a sequence of step functions that are really in $\mathcal{E}$, i.e., that are adapted (cf. [KS] or [O] for details).

Definition 6.2. Let $Y\in\mathcal{E}$ take the form (6.4). Then the random variable

$$I(Y)=\int_0^T Y_s dW_s := \sum_{i=0}^{n-1} Y^{(i)}(W_{t_{i+1}}-W_{t_i}) \tag{6.5}$$

is called the stochastic (Itô) integral of (Y_t) with the driving process (integrator) (W_t).

The following Proposition will help us to extend the integral (as an operator) from step functions to the whole space $M_w^2(0,T)$.

Proposition 6.3. (i) $I:\mathcal{E}_w(0,T)\to L^2(\Omega)$ *is a linear operator.*
(ii) *For each* $Y\in\mathcal{E}$,

$$\mathbb{E}\int_0^T Y_s dW_s = \mathbb{E}I(Y)=0. \tag{6.6}$$

(iii) *For each* $Y\in\mathcal{E}$,

$$\mathbb{E}|I(Y)|^2=\mathbb{E}\left|\int_0^T Y_s dW_s\right|^2=\mathbb{E}\int_0^T |Y_s|^2 ds=\int_0^T \mathbb{E}|Y_s|^2 ds, \tag{6.7}$$

that is, the operator $I:\mathcal{E}\to L^2(\Omega)$ *is an isometry (called the Itô isometry) when* $\mathcal{E}$ *is equipped with* $L^2((0,T)\times\Omega)$*-norm.*

Proof. The proof of (i) is obvious. In order to prove (ii) we may note that for each i the random variables $Y^{(i)}$ and $W_{t_{i+1}}-W_{t_i}$ are stochastically independent, therefore

$$\mathbb{E}\int_0^T Y_s dW_s=\mathbb{E}\sum_{i=1}^{n-1} Y^{(i)}(W_{t_{i+1}}-W_{t_i})=\sum_{i=1}^{n-1}\mathbb{E}Y^{(i)}\cdot\mathbb{E}(W_{t_{i+1}}-W_{t_i})=0$$

since the process (W_t) is centered. Also, we have

$$\mathbb{E}\Big(\sum_{i=1}^{n-1} Y^{(i)}(W_{t_{i+1}} - W_{t_i})\Big)^2 = \mathbb{E}\Big[\sum_i (Y^{(i)})^2 (W_{t_{i+1}} - W_{t_i})^2 + 2\sum_{i<j} Y^{(i)} Y^{(j)} (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})\Big]. \tag{6.8}$$

By the property (W2) of the Wiener process and using again adaptiveness of (Y_t) we have

$$\mathbb{E}\sum_i (Y^{(i)})^2 (W_{t_{i+1}} - W_{t_i})^2 = \sum_i \mathbb{E}(Y^{(i)})^2 \cdot \mathbb{E}(W_{t+1} - W_{t_i})^2 = \sum_i \mathbb{E}(Y^{(i)})^2 (t_{i+1} - t_i) = \int_0^T \mathbb{E} Y_t^2 dt. \tag{6.9}$$

Furthermore, since $t_{i+1} \le t_j$, the product $Y^{(i)} Y^{(j)} (W_{t_{i+1}} - W_{t_i})$ is $\mathcal{F}_{t_j}$-measurable and clearly, $W_{t_{j+1}} - W_{t_j}$ is independent of $\mathcal{F}_{t_j}$, which yields

$$\mathbb{E}\sum_{i<j} Y^{(i)} Y^{(j)} (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) = \sum_{i<j} \mathbb{E} Y^{(i)} Y^{(j)} (W_{t_{i+1}} - W_{t_i}) \cdot \mathbb{E}(W_{t_{j+1}} - W_{t_j}) = 0 \tag{6.10}$$

as the last expectation on the r.h.s. of (6.10) is zero. Now, substituting (6.9) and (6.10) into (6.8) we get

$$\mathbb{E}\left| \int_0^T Y_s dW_s \right|^2 = \mathbb{E}\Big(\sum_i Y^{(i)} (W_{t_{i+1}} - W_{t_i}) \Big)^2 = \int_0^T \mathbb{E}|Y_s|^2 ds,$$

which completes the proof of (iii). □

Now we are in position to complete the definition of the stochastic integral. The linear operator $I : (\mathcal{E}, \|\cdot\|_{L^2((0,T)\times\Omega)}) \to L^2(\Omega)$ is bounded (in fact, isometric) by (iii) of the previous Proposition and $\mathcal{E}$ is dense in $M_w^2(0,T)$ by Lemma 6.1. Thus there exists a unique extension of I (denoted again by I) to the space $M_w^2(0,T)$ that is a bounded linear operator $I : M_w^2(0,T) \to L^2(\Omega)$.

Definition 6.4. For $Y \in M_w^2(0,T)$, the random variable $I(Y)$ is also denoted by $\int_0^T Y_s dW_s$ and called the stochastic (Itô) integral.

The above definition may be reformulated in the following way: Given $Y \in M_w^2(0,T)$, there exists a sequence of step functions $Y_n \in \mathcal{E}$, $Y_n \to Y$ in $L^2((0,T)\times\Omega)$ (by Lemma 6.1). Obviously, (Y_n) is a Cauchy sequence in $L^2((0,T)\times\Omega)$ and therefore by Proposition 6.3 (iii) $I(Y_n) = \int_0^T Y_n(s)dW_s$ is a Cauchy sequence in $L^2(\Omega)$, hence convergent. The limit $I(Y) = \lim I(Y_n)$ is the *stochastic integral* $\int_0^T Y_s dW_s$.

As an immediate consequence of the above construction and Proposition 6.3 we obtain the following

Corollary 6.5. *The operator $I : M_w^2(0,T) \to L^2(\Omega)$ is linear and for $Y \in M_w^2(0,T)$,* (ii) *and* (iii) *of Proposition* 6.3 *hold true (i.e., I is an isometry).*

In most cases (in the SDE's theory) we deal with integrands that have continuous sample paths. In this case we have a statement that more lucidly justifies the definition of stochastic integrals in models described in the previous Section.

Remark 6.6. The concept of stochastic integral may be significantly extended in many directions. The most usual extension is to consider the space $\mathcal{S}_w(0,T)$ of all $(\mathcal{F}_t)$-progressive integrands (Y_t) satisfying

$$\int_0^T |Y_t|^2 dt < \infty \quad \mathbb{P}\text{-a.s.} \tag{6.11}$$

(i.e. without the mathematical expectation). The stochastic integral $\int_0^T Y_t dW_t$ is then obtained as a limit in probability of integrals of approximate step functions. Such an integral, however, need not satisfy (ii) and (iii) of Proposition 6.3 (in general, it is not an element of $L^2(\Omega)$ nor even integrable on Ω.

Proposition 6.7. *Let $D_n = \{t_i^n\}$ be a sequence of partitions of the interval $[0,T]$ and assume that $Y \in \mathcal{S}_w(0,T)$ has continuous sample paths. Then*

$$\sum_i Y_{t_i^n}(W_{t_{i+1}^n} - W_{t_i^n}) \to \int_0^T Y_t dW_t \text{ as } |D_n| \to 0 \quad \text{in } \mathbb{P}. \tag{6.12}$$

Proof. It may be easily checked that the sequence of step functions $Y^n(t) = Y_{t_i^n}$, $t \in [t_i^n, t_{i+1}^n)$, satisfies the conditions required in the definition of stochastic integral. □

Next we list some useful properties of stochastic integrals.

Theorem 6.8. *Let $Y \in \mathcal{S}_w(0,T)$ set $I_t(Y) := \int_0^t Y_s dW_s$, $t \in [0,T]$*
(i) *(I_t) is a $(\mathcal{F}_t)$-progressive process.*
(ii) *for each $0 \le u \le T$ we have*

$$\int_0^T Y_s dW_s = \int_0^u Y_s dW_s + \int_u^T Y_s dW_s, \quad \mathbb{P}\text{-a.e.}$$

(iii) *There is a modification of (I_t) which has continuous paths.*
(iv) *If $Y \in M_w^2(0,T)$ then (I_t) is a martingale w.r.t $(\mathcal{F}_t)$ and*

$$< I >_t = \int_0^t |Y_s|^2 ds.$$

(v) *If $Y \in M_w^2(0,T)$ we have that*

$$\mathbb{E}\left[\max_{0\le t\le T}\left|\int_0^t Y_s dW_s\right|^2\right] \le 4\mathbb{E}\int_0^T |Y_s|^2 ds. \tag{6.13}$$

Proof. The statements (i), (ii) and (iv) follow easily from the definition of the stochastic integral. The part (iii) may be proved in a natural way by a limit passage of integrals of simple functions (note that if (Y_t) is deterministic it also easily follows by the Kolmogorov continuity test if we use the Itô isometry and Gaussianity of the integral). The inequality (6.13) is a simple consequence of (iii), (iv), and the Doob maximal inequality. □

Remark 6.9. (i) In the sequel, when speaking about a stochastic integral as a process $t \mapsto I_t(Y)$, we always have in mind the continuous version provided by Theorem 6.7 (iv).
(ii) Note that if only $Y \in \mathcal{S}_w(0,T)$ (not necessarily $Y \in M_w^2(0,T)$, cf. Remark 6.6) the stochastic integral does not have to be a martingale (the statements of Theorem 6.8 (iv) and (v) no longer hold) because of lack of integrability on Ω. However, it is so-called local martingale (see e.g. [KS] for details).

Example 6.10. Compute $\int_0^T W_t dW_t$! Obviously, the integral does exist since $(W_t) \in M_w^2(0,T)$. The integrand has $\mathbb{P}$-a.s. continuous paths, so by Proposition 6.7

$$\int_0^T W_t dW_t = \lim_{|D_n|\to 0} \sum_{i=0}^{n-1} W_{t_i^n}(W_{t_i^{n+1}} - W_{t_i^n}) \tag{6.14}$$

over a sequence of partitions $D_n = \{t_i^n\}$ of the interval $[0, T]$. For clarity, we suppress n in the notation $t_i^n = t_i$. Then

$$\begin{aligned}\sum_i W_{t_i}(W_{t_{i+1}} - W_{t_i}) &= \sum_i \Big(-\frac{1}{2}W_{t_{i+1}}^2 + W_{t_i}W_{t_{i+1}} - \frac{1}{2}W_{t_i}^2\Big) \\ &+ \sum_i \Big(\frac{1}{2}W_{t_{i+1}}^2 - \frac{1}{2}W_{t_i}^2\Big) \\ &= -\frac{1}{2}\sum (W_{t_{i+1}} - W_{t_i})^2 + \frac{1}{2}W_T^2\end{aligned} \tag{6.15}$$

and computing the $L^2(\Omega)$-limit by Theorem **??** we obtain

$$\int_0^T W_t dW_t = -\frac{1}{2}T + \frac{1}{2}W_T^2. \tag{6.16}$$

It is interesting to compare this result with his analogue for the Stieltjes integral in case when the integrand (and integrator) φ is smooth, $\varphi(0) = 0$. Then we have by integration by parts formula

$$\int_0^T \varphi(s)d\varphi(s) = \int_0^T \varphi(s)\varphi'(s)ds = \varphi^2(T) - \int_0^T \varphi'(s)\varphi(s)ds$$

which yields

$$\int_0^T \varphi(s)d\varphi(s) = \frac{1}{2}\varphi^2(T). \tag{6.17}$$

Comparing (6.16) to (6.17) we can see that an extra term $-\frac{1}{2}T$ has been obtained that is missing in the "smooth" case (6.17). This term has arisen as a quadratic variation of the Wiener process, which is nonzero (unlike in the case of a smooth function φ). A similar phenomenon will later appear in a more general situation in the so-called Itô formula.

Remark 6.11. In the standard Stieltjes integration theory, the integral is found as a limit of partial sums of the form

$$\sum_{i=1}^{n-1} \varphi(\tau_i)(W_{t_{i+1}} - W_{t_i}) \approx \int_0^T \varphi(t)dW_t \tag{6.18}$$

where τ_i, roughly speaking, may vary in the interval $[t_i, t_{i+1}]$. It is easy to demonstrate that such an approach would be fruitless in our case. Let

$\bar{t}_i = \frac{1}{2}(t_i + t_{i+1})$ be the midpoints of the interval $[t_i, t_{i+1}]$. Proceeding similarly as in (6.15) it is easy to see that

$$\lim_{|D_n|\to 0} \sum_{i=0}^{n-1} W_{\bar{t}_i}(W_{t_{i+1}} - W_{t_i}) = \frac{1}{2} W_T^2 \text{ in } L^2(\Omega) \tag{6.19}$$

which is different from the right-hand side of (6.16) and seems to resemble the "smooth" case (6.17). However, it may be easily seen that the similarity with (6.17) is rather incidental; the influence of quadratic variation of the process does not "disappear" (as in the smooth case) but is just "balanced", so at the end of the day it is cancelled out. In this way we might consider a whole scale of stochastic integrals, depending on the position of τ_i inside $[t_i, t_{i+1}]$.

The choice $\tau_i = \bar{t}_i$ is quite popular and leads to the concept of so called *Stratonovich* (or *symmetric*) *integral*, denoted usually $\int_0^T Y_s \circ dW_s$, and defined for a class of continuous integrands (Y_t) as a limit of partial sums of the form

$$\sum_{i=1}^{n-1} Y_{\bar{t}_i}(W_{t_{i+1}} - W_{t_i}) \approx \int_0^T Y_s \circ dW_s.$$

Both Itô a Stratonovich integrals are commonly used in stochastic analysis. There is a lengthy discussion which one is better in each particular instance in applications. A major advantage of the Itô approach was illustrated in the intuitive derivation of SDE carried out in Section 2.1 - it leads to the Itô-type integral. An analogous procedure leading to Stratonovich integral would mean that the noise is "looking to the future". On the other hand, if the paths of the Wiener process in the SDE are approximated by continuously differentiable processes (or polygons) then the solutions of such approximative equations (interpreted as ODE's depending on a parameter $\omega \in \Omega$) converge to a solution of a SDE in which, however, the stochastic integral is understood in the Stratonovich sense. From the technical viewpoint, the Stratonovich integrals are not martingales (unlike the Itô integrals), but on the other hand there are no second order terms in the transformation formula (like in the Itô formula below), which makes them suitable especially when dealing with SDE's on manifolds.

From pure mathematical viewpoint, the Itô a Stratonovich integrals are basically mutually "convertible", for example, in the framework of the so-called Malliavin calculus (se e.g. [Nu]). Also the SDE's consider in either Itô

or Stratonovich sense may be (in reasonable case) mutually converted (after the change of the type of integral, they will have the same "diffusion" part but different "drift" part).

Our next aim is to cope with the situation, when the upper bound in the stochastic integral is random, i.e.,

$$\int_0^\tau Y_s dW_s,\ \tau : \Omega \to \mathbb{R}_+.$$

It turns out that the above integral may be defined and reasonably handled for τ which is a stopping time.
We have the following natural definition:

Definition 6.12. Let $\tau \leq T$ be a stopping time w.r.t. $(\mathcal{F}_t)$, W_t a Wiener process w.r.t. $(\mathcal{F}_t)$ and $(Y_t) \in \mathcal{S}_w(0,T)$. Then

$$I_\tau(Y) = \int_0^\tau Y_t dW_t := \int_0^T 1_{[\tau > t]} Y_t dW_t.$$

Obviously the process $t \mapsto 1_{[\tau>t]}Y_t$ is in $\mathcal{S}_w(0,T)$, so the definition is correct. We immediately obtain

Proposition 6.13. *For a stopping time $\tau \leq T$, $Y \in M_w^2(0,T)$ we have*

$$\mathbb{E} \int_0^\tau Y_t dW_t = 0$$

So far we have been dealing with one-dimensional Wiener process and stochastic integral. However, extension of the above definitions and results to a multidimensional case is quite straightforward. By an n-dimensional, standard Wiener process we understand the $\mathbb{R}^n$-valued process $W_t = (W_t^1, \ldots, W_t^n)'$ the components (W_t^i) of which are standard one-dimensional Wiener processes that are stochastically independent. For a Hilbert space H, set $M_w^2(0,T;H) = \{Y \in L^2((0,T)\times\Omega; H),\ Y$ is $(\mathcal{F}_t)$-progressive$\}$ where $(\mathcal{F}_t)$ is a filtration such that (W_t) is a Wiener process w.r.t. $(\mathcal{F}_t)$ (this concept remains the same as in one-dimensional case). The definition of $\mathcal{S}_w(0,T;H)$ is analogous. Let $Y \in \mathcal{S}_w(0,T;\mathbb{R}^{m\times n})$, $Y = (Y^{ij})$; then

$$I(Y) = \int_0^T Y_t dW_t := \left(\sum_{j=1}^n \int_0^T Y_t^{1j} dW_t^j, \ldots, \sum_{j=1}^m \int_0^T Y_t^{mj} dW_t^j \right)'.$$

It is easy to modify all statements of this Section (in particular, Proposition 6.3, Corollary 6.5 [the Itô isometry!], and Theorem 6.8) for the multidimensional case.

Definition 6.14. An $\mathbb{R}^m$-valued process (X_t) is said to have a stochastic differential of the form

$$dX_t = a_t dt + b_t dW_t,\ t \in (0,T] \tag{6.20}$$

where $(a(t)) \in \mathcal{S}_w(0,T;\mathbb{R}^m)$, $(b_t) \in \mathcal{S}_w(0,T;\mathbb{R}^{m\times n})$, (W_t) being an n-dimensional, standard Wiener process, if

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dW_s,\ t \in [0,T],\ \mathbb{P}\text{-a.s.} \tag{6.21}$$

where the first integral on the r.h.s. of (6.21) is the Lebesgue integral almost surely and the second integral is the stochastic (Itô) integral.

Next, we formulate the celebrated *Itô Lemma*, a chain rule for stochastic differentials. By $\mathcal{C}^{1,2}([0,T]\times\mathbb{R}^m)$ we denote the space of functions whose first time derivative and second space derivatives are continuous on $[0,T]\times\mathbb{R}^m$.

Theorem 6.15. (the Itô formula). *Let $V \in \mathcal{C}^{1,2}([0,T)\times\mathbb{R}^m)$ and assume that the process (X_t) is given by the stochastic differential* (6.20)*, where (a_t) and (b_t) satisfy the conditions from Definition* 6.14*. Set $\sigma_t = (\sigma_t^{ij}) = b_t b_t^T$. Then the process $Y_t = V(t,X_t)$ has a stochastic differential,*

$$\begin{aligned} dY_t =& \left[\frac{\partial V}{\partial t}(t,X_t) + \langle \nabla_x V(t,X_t), a_t\rangle + \frac{1}{2}\sum_{i,j=1}^m \sigma_t^{ij}\frac{\partial^2 V}{\partial x_i \partial x_i}(t,X_t)\right]dt \\ &+ \sum_{i=1}^m \sum_{j=1}^n b_t^{ij}\frac{\partial V}{\partial x_i}(t,X_t)dW_t^j,\ t\in(0,T). \end{aligned} \tag{6.22}$$

In the one-dimensional case ($n=m=1$) the formula (6.22) *takes the form*

$$\begin{aligned} dY_t =& \Big(\frac{\partial V}{\partial t}(t,X_t) + a_t\frac{\partial V}{\partial x}(t,X_t) + \frac{1}{2}b_t^2\frac{\partial^2 V}{\partial x^2}(t,X_t)\Big)dt \\ &+ b_t\frac{\partial V}{\partial X}(t,X_t)dW_t. \end{aligned} \tag{6.23}$$

The proof of Theorem 6.15 is relatively lengthy and technical, it may be found practically in all monographs and textbooks on stochastic calculus (e.g., [KS] or [O]).

Here we give a short "heuristic" proof (not a mathematically rigorous proof!) in the one-dimensional case.

Let dt and dW_t be the "infinitesimal" increments of time and the Wiener process, respectively. We postulate that

$$(dt)^k = 0,\ (dW_t)^2 = dt,\ (dt)^l(dW_t)^r = 0 \tag{6.24}$$

for $k \geq 2$, $l \geq 1$, $r \geq 1$ (trace what is behind these postulates!). We may write

$$X_{t+dt} - X_t = dX_t = a_t dt + b_t dW_t,\ (dX_t)^2 = a_t^2(dt)^2 + 2a_t b_t (dt)dW_t + b_t^2 (dW_t)^2, \tag{6.25}$$

etc. Using the Taylor expansion we get for $Y_t = V(t, X_t)$ that

$$\begin{aligned} dY_t =& V(t+dt, X_{t+dt}) - V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t)dt + \frac{\partial V}{\partial x}(t, X_t)(X_{t+dt} - X_t) \\ &+ \frac{1}{2}\left(\frac{\partial^2 V}{\partial t^2}(t, X_t)(dt)^2 + \frac{\partial^2 V}{\partial x^2}(t, X_t)(a_t dt + b_t dW_t)^2 + 2\frac{\partial^2 V}{\partial t \partial x}(t, X_t)(dX_t)(dt) \right) \end{aligned}$$

(the higher order terms vanish, cf. (6.24). Substituting (6.25) and taking into account (6.24) we obtain (6.23).

Example (stochastic bilinear equation). Given $x_0 \in \mathbb{R}$ and $f, g \in \mathcal{C}(0, T)$ find a process (Y_t) satisfying

$$dY_t = f(t)Y_t dt + g(t)Y_t dW_t,\ Y_0 = x_0.$$

Hint: Use the Itô formula with

$$\begin{aligned} V(t, x) &= x_0 \exp\left\{ \int_0^t (f(s) - \frac{1}{2}g^2(s))ds + x \right\}, \\ dX_t &= g(t)dW_t,\ X_0 = 0, \end{aligned}$$

to obtain

$$Y_t = x_0 \exp\left\{ \int_0^t (f(s) - \frac{1}{2}g^2(s))ds + \int_0^t g(s)dW_s \right\},\ t \in [0, T].$$

Example (stochastic linear equation). Given $x_0 \in \mathbb{R}$, $a \in L^1(0,T)$, be $L^2(0,T)$, show that the process

$$Y_t := \mathrm{e}^{\int_0^t a(s)ds} x_0 + \int_0^t \mathrm{e}^{\int_r^t a(s)ds} b(r) dW_r, \ t \in [0,T]$$

satisfies

$$dY_t = a(t)Y_t dt + b(t)dW_t, \ Y_0 = x_0, \ t \in [0,T].$$

Hint: Use the Itô formula for the product $Y_t\Phi_t$, where

$$\dot{\Phi}(t) = a(t)\Phi(t), \ \Phi(t) = 1, \text{ and } dX_t = \mathrm{e}^{-\int_0^t a(\lambda)d\lambda}\sigma(t)dW_t, \ X_0 = x_0.$$

The Itô formula (6.22) is a basic tool in stochastic calculus. It is used in most computations and in many theoretical arguments involving stochastic differentials and SDE's.

Example 6.16. (= Example 6.10 revisited). Using the Itô formula, we get a quick and elegant method to compute the integral $\int_0^T W_t dW_t$, where (W_t) is a 1-dimensional Wiener process. Setting $X_t = W_t$, $V(x) = \frac{1}{2}x^2$, we may use formula (6.23) where $Y_t = \frac{1}{2}W_t^2$, $a_t \equiv 0$, $b_t \equiv 1$, $V'(x) = x$, $V''(x) = 1$, which yields

$$d\Big(\frac{1}{2}W_t^2\Big) = \frac{1}{2}dt + W_t dW_t,$$

thus integrating over $[0,T]$,

$$\frac{1}{2}W_T^2 = \frac{1}{2}T + \int_0^T W_t dW_t,$$

hence

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T.$$

7 Stochastic differential equations: Basic results

The concept of stochastic differential equation has been motivated and heuristically introduced in Section 3.1. The aim of the present Section is to define some notions of solution to SDE's rigorously and to present a basic result on existence and uniqueness of solutions.

Let $b : [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0,T] \times \mathbb{R}^m \to \mathbb{R}^{m \times n}$ be measurable mappings and $(W_t) = (W_t^{(j)})$, $j = 1, 2, \ldots, n$, a standard n-dimensional Wiener process. The *stochastic differential equation* is a stochastic differential, involving an unknown process (X_t), of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \ t \in (0,T) \tag{7.1}$$

amended with an initial condition

$$X_0 = x \in \mathbb{R}^m \tag{7.2}$$

that may be, in general, random.

Definition 7.1. A (strong) solution to the equation (7.1)–(7.2) is an $\mathbb{R}^m$-valued stochastic process (X_t) satisfying

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad t \in [0,T], \tag{7.3}$$

where the first integral on the r.h.s. is understood in the Lebesgue sense $\mathbb{P}$-a.s. while the second one is an Itô stochastic integral (and both integrals are well defined). Alternatively, (X_t) is a strong solution of (7.1)–(7.2) if $X_0 = x$ and (X_t) has the stochastic differential of the form (7.1).

We will prove the following classical theorem on existence and uniqueness of strong solutions, which is in principle due to K. Itô (1942):

Theorem 7.2. *For $T > 0$, let $b : [0,T] \times \mathbb{R}^m \to \mathbb{R}^m$, $\sigma : [0,T] \times \mathbb{R}^m \to \mathbb{R}^{m \times n}$ be measurable functions satisfying*

$$|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|), \ x \in \mathbb{R}^m, \ t \in [0,T], \tag{7.4}$$

and

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq C|x - y|, \ x, y \in \mathbb{R}^m, \ t \in [0,T], \tag{7.5}$$

for a constant C. Then the equation (7.1)–(7.2), where the initial value $X_0 = x$ is $\mathcal{F}_0$-measurable (or independent of (W_t)) and satisfying

$$\mathbb{E}|X_0|^2 < \infty \tag{7.6}$$

has a continuous solution $(X_t) \in M_w^2(0,T)$. The solution is unique (if $(\widehat{X}_t)$ is another such solution, then (X_t) and $(\widehat{X}_t)$ are indistinguishable).

Proof. The uniqueness follows easily from the isometry (4.8) and the Lipschitz property (7.5):

Let $X_t^x(\omega)$ and $\widehat{X}_t^y(\omega)$ be solutions with initial values x, y respectively.

Put $a(s,\omega) = b(s, X_s^x) - b(s, \widehat{X}_s^y)$ and $\gamma(s,\omega) = \sigma(s, X_s^x) - \sigma(s, \widehat{X}_s^y)$. Then

$$\begin{aligned}
\mathbb{E}[|X_t^x - \widehat{X}_t^y|^2] &= \mathbb{E}\left[\left(x - y + \int_0^t a ds + \int_0^t \gamma dW_s\right)^2\right] \\
&\leq 3|x-y|^2 + 3 \cdot \mathbb{E}\left[\left(\int_0^t a ds\right)^2\right] + 3\mathbb{E}\left[\left(\int_0^t \gamma dW_s\right)^2\right] \\
&\leq 3|x-y|^2 + 3t \cdot \mathbb{E}\left[\int_0^t a^2 ds\right] + 3\mathbb{E}\left[\int_0^t \gamma^2 ds\right] \\
&\leq 3|x-y|^2 + 3(1+t)C^2 \cdot \int_0^t \mathbb{E}[|X_s^x - X_s^y|^2] ds.
\end{aligned}$$

So the function

$$v(t) = \mathbb{E}[|X_t^x - \widehat{X}_t^y|^2];\ 0 \leq t \leq T$$

satisfies

$$v(t) \leq 3|x-y|^2 + A \cdot \int_0^t v(s) ds, \text{ where } A = 3(1+T)C^2.$$

Let $w(t) = \int_0^t v(s) ds$. Then $w'(t) \leq 3|x-y|^2 + A \cdot w(t)$, so since $w(0) = 0$, $w(t) \leq 3|x-y|^2 t \cdot \exp(At)$. (Consider $f(t) = w(t)\exp(-At)$.) Therefore

$$v(t) \leq 3|x-y|^2(1 + At\exp(At)).$$

Now assume that $x = y$. Then $v(t) = 0$ for all $t \geq 0$. Hence, writing $X_t = X_t^x$, $\widehat{X}_t = \widehat{X}_t^x$ we have

$$\mathbb{P}[|X_t - \widehat{X}_t| = 0 \text{ for all } t \in \mathbb{Q} \cap [0,T]] = 1,$$

where $\mathbb{Q}$ denotes the rational numbers. By continuity of $t \to |X_t - \widehat{X}_t|$ it follows that

$$\mathbb{P}[|X_1(t,\omega) - X_2(t,\omega)| = 0 \text{ for all } t \in [0,T]] = 1,$$

and the uniqueness is proved.

The proof of the existence is similar to the familiar existence proof for ordinary differential equations: Define $Y^{(0)} = X_0$ and $Y^{(k)} = Y_t^{(k)}(\omega)$ inductively as follows

$$Y^{(k+1)} = X_0 + \int_0^t b(s, Y^{(k)})ds + \int_0^t \sigma(s, Y^{(k)})dW_s. \tag{7.7}$$

Then, similar computation as for the uniqueness above gives

$$\mathbb{E}[|Y^{(k+1)} - Y^{(k)}|^2] \leq (1+T)3C^2 \cdot \int_0^t \mathbb{E}[|Y^{(k)} - Y^{(k-1)}|^2]ds,$$

for $k \geq 1$ and

$$\mathbb{E}[|Y^{(1)} - Y^{(0)}|^2] \leq 2C^2t^2(1 + \mathbb{E}(|X_0|^2]) + 2C^2t(1 + E[|X_0|^2]) \leq A_1 \cdot t$$

where the constant A_1 only depends on C, T and $\mathbb{E}[|X_0|^2]$. So by induction on k we obtain

$$\mathbb{E}[|Y^{(k+1)} - Y^{(k)}|^2] \leq \frac{A_2^{k+1}t^{k+1}}{(k+1)!};\ k \geq 0,\ t \in [0,T]$$

for some suitable constant A_2 depending on C, T and $E[|X_0|^2]$ and it follows that $Y^{(k)}$ is a Cauchy (hence convergent) sequence in $C([0,T], L^2(\Omega, \mathbb{R}^m))$. A standard limit passage in (7.7) together with the Itô isometry imply that the limit $X = \lim Y^{(k)}$ solves (7.3). By the continuous dependence on the upper bound (Theorem 6.8 (iv)) and continuity of the Lebesgue integral we obtain $\mathbb{P}$-a.s. continuity of (X_t), which completes the proof. □

Remark 7.3. (i) If the equation is autonomous, i.e. the coefficients b and σ do not depend on t, then (7.5) obviously implies (7.4).
(ii) Conditions (7.4) and (7.5) may be substantially weakened in various manners. The most common case is that the global Lipschitz condition (7.5) is turned to a local one: Then the statement of the Theorem remains unchanged (just the proof of slightly more complicated). If the linear growth condition (7.4) is removed and the Lipschitz condition is only local, we still may obtain existence and uniqueness of a local solution which however may explode at a finite random time. The "linear growth" condition (7.4) basically prevents such explosions, and it also may be relaxed (a more general condition for nonexplosion called "the Khasminskii test" may be given in

terms of Lyapunov functions, cf. [Ch]). As examples of "almost explicitly solved" SDE's we may consider the linear and bilinear stochastic equations studied in the preceding Section. The solution to the linear equation is usually called *Ornstein-Uhlenbeck process*. In applications, it corresponds to a stochastic perturbation of a linear deterministic system, which is independent of the solution ("the additive noise"). The solution (X_t) is a Gaussian process, which easily follows from the fact that the stochastic integral of a deterministic function is Gaussian.

The solution to bilinear equation is provided as an "explicit" solution in terms of the paths of the Wiener process (note however that usually these paths are not observable). The solution in case of constant coefficients is also is called *geometric Brownian motion* and is often used in applications (e.g. in finance mathematics, as a model of stock prices).

References

[CS] M. Csörgö, Random walking around financial mathematics, Random Walks (Budapest 1998), 59–111, Bolyai Soc. Math. Stud. 9, Janos Bolyai Math. Soc., Budapest, 1999.

[Ci] Z. Ciesielski, Hoelder conditions for realizations of Gaussian processes, Trans. Amer. Math. Soc. 99 (1961), 403-413.

[Ei] A. Einstein, On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory, Ann. Physik 17 (1905), 549–560.

[Fr] A. Friedman, Stochastic Differential Equations and Applications, Vol. 2, Academic Press, New York, 1976.

[Hi] T. Hida, H.-H. Kuo, J. Potthoff and L. Streit, White Noise, An Infinite-Dimensional Calculus, Kluwer, Dodrecht, 1993.

[Kal] G. Kallianpur, Stochastic Filtering Theory, Springer-Verlag, 1980.

[Ch] R. Z. Khas'minskiĭ, Stability of systems of differential equations under random perturbations of their parametes, Nauka, Moscow, 1969 (in Russian); English translation: Stochastic stability of differential equations, Sijthoff & Noordhoff, Alphen aan den Rijn 1980.

[KS] Y. Karatzas and S. E. Shrieve, Brownian Motion and Stochastic Calculus, Springer-Verlag, New York, 1988.

[Nu] D. Nualart, The Malliavin Calculus and Related Topics, Springer-Verlag, New York, 1995.

[O] B. Oksendal, Stochastic Differential Equations, An Introduction with Applications, Springer-Verlag, 1995.

[Se] J. Seidler, Vybrané kapitoly ze stochastické analysy, Matfyzpress, Praha 2011.

[Sm] M. Smoluchowski, Zur Kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, Ann. Physik 21 (1906), 756.

[SV] D.W. Strook and S.R.S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, 1979.