Stochastic Differential Equations

Jianing Yao
Department of MSIS-RUTCOR
Rutgers University, the State University of New Jersey
Piscataway, NJ 08854 USA

April 2, 2015

1 Multi-dimensional Stochastic Integration and It$\bar{o}$'s formula

1.1 Multi-dimensional Brownian motion

Let W_1, W_2, ..., W_d be independent Brownian motion set,

$$W(t) = (W_1(t), W_2(t), ..., W_d(t)), \ t \geq 0 \tag{1}$$

W is called a *standard multi-dimensional Brownian motion*. A filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is said to be filtration for W if $W(t) - W(s) \perp \mathcal{F}_s$ for $0 \leq s \leq t$ and $W(t)$ is $\mathcal{F}_t$–measurable $\forall t$. In other words, $W_i(t) - W_i(s)$ is always independent of $\mathcal{F}_s$ whenever $0 \leq s \leq t$, for each individual Brownian motion W_i and, of course, $W_i(t)$ is $\mathcal{F}_t$–measurable for all i and t. Thus, $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration for W_i, for each $1 \leq i \leq d$.

Let's give an intuition about what is going on here: $\mathcal{F}_t$ is the σ-algebra generated by $W_1(s)$, $W_2(s)$, ..., $W_d(s)$ for all $s \leq t$. Observing $\mathcal{F}_t$ is equivalent to observing $W_1(s)$, $W_2(s)$, ..., $W_d(s)$ whenever $0 \leq s \leq t$, because $\mathcal{F}_s$ contains only information about $W_i(r)$ for $r \leq s$, which is independent fo $W_i(t) - W_i(s)$, and information about $W_j(r)$, $r \leq s$, $j \neq i$, which is independent of $W_i(t) - W_i(s)$ by virtue of the assumption that W_i is independent of W_j when $i \neq j$.

Assume that $\{\mathcal{F}_t\}_{t\ geq0}$ is a filtration for $W = (W_1, ..., W_d)$. let $\{\alpha_1(t)\}_{t\geq 0}$, ..., $\{\alpha_d(t)\}_{t\geq 0}$ be processes which are each adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, each of which satisfy:

$$\mathbb{P}\left(\int_0^t \alpha_i^2(s)ds < +\infty\right) = 1 \tag{2}$$

Since $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration for each W_i,

$$\int_0^t \alpha_i(s)dW_i(s), \ t \geq 0 \tag{3}$$

is well-defined for each i. In the meanwhile, let $\{\beta(t)\}_{t\geq 0}$ also be adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ and assume

$$\mathbb{P}\left(\int_0^t |\beta(s)|ds < +\infty\right) = 1 \tag{4}$$

Let $X(0)$ be $\mathcal{F}_0$–measurable. Then, define

$$X(t) = X(0) + \int_0^t \beta(s)ds + \int_0^t \alpha_1 dW_1(s) + \cdots + \int_0^t \alpha_d(s)dW_d(s) \tag{5}$$

(Now, I realize d is not a good notation for dimension, sorry!) In differential notation:

$$dX(t) = \beta(t)dt + \sum_{i=1}^{d} \alpha_i(t)dW_i(t) \tag{6}$$

We call X an Itō process, as well.

1.2 Little Fact

Let $\Pi = \Pi[t_0, ..., t_n]$ of $[0, t]$, define:

$$Y_\Pi := [W_i, W_j]^\Pi(t) = \sum_{k=0}^{n-1} \left[W_i(t_{k+1}) - W_i(t_k)\right]\left[W_j(t_{k+1}) - W_j(t_k)\right] \tag{7}$$

Theorem 1.1 If W_i and W_j are independent Brownian motion, then $y_\Pi = 0$.

Remember the identity is in the sense that as partition goes finer and finer, it will converge in L^2-norm to 0., i.e.,

$$\lim_{||\Pi||\to 0} \mathbb{E}[y_\Pi^2] = 0 \tag{8}$$

Proof. Observe

$$\begin{aligned} y_\Pi^2 = &\sum_{k=0}^{n-1}[W_i(t_{k+1}) - W_i(t_k)]^2[W_j(t_{k+1}) - W_j(t_k)]^2 \\ &+ 2\sum_{l<k}^{n-1}[W_i(t_{l+1}) - W_i(t_l)][W_j(t_{l+1}) - W_j(t_l)] \cdot [W_i(t_{k+1}) - W_i(t_k)][W_j(t_{k+1}) - W_j(t_k)] \end{aligned}$$

All increments appearing in the sum of cross-term are independent of one another and all have mean zero. Therefore,

$$\mathbb{E}[y_\Pi^2] = \mathbb{E}\sum_{k=0}^{n-1}[W_i(t_{k+1}) - W_i(t_k)]^2[W_j(t_{k+1}) - W_j(t_k)]^2 \tag{9}$$

But $[W_i(t_{k+1}) - W_i(t_k)]^2$ and $[W_j(t_{k+1}) - W_j(t_k)]^2$ are independent of one another, and each has expectation $t_{k+1} - t_k$. It follows that

$$\mathbb{E}[y_\Pi^2] = \sum_{k=0}^{n-1}(t_{k+1} - t_k)^2 \leq ||\Pi|| \cdot \sum_{k=0}^{n-1}(t_{k+1} - t_k) = ||\Pi|| \cdot T \tag{10}$$

As $||\Pi|| \to 0$, we have $\mathbb{E}[y_\Pi^n] \to 0$, the assertion follows. □

Remark 1.2 In differential notation, we have

$$dW_i(t)dW_j(t) = 0 \tag{11}$$

if $i \neq j$ and $W_i \perp W_j$.

1.3 It$\bar{o}$'s Formula

Rather than write a very general formula, we do an example computation at the formal level. Let X and Y be It$\bar{o}$ process defined by:

$$dX(t) = \beta(t)dt + \sum_{i=1}^{d} \alpha_i(t)dW_i(t) \tag{12}$$

$$dY(t) = \gamma(t)dt + \sum_{i=1}^{d} \delta_i(t)dW_i(t) \tag{13}$$

By using $dW_i(t)dW_j(t) = 0$, $(dt^2) = 0$, $dtd_W(t) = 0$ and $(dW_i(t))^2 = dt$,

$$\begin{aligned}
(dX(t))^2 &= \left(\beta(t)dt + \sum_{i=1}^{d} \alpha_i(t)dW_i(t)\right)^2 \\
&= \beta^2(t)(dt)^2 + 2\sum_{i=1}^{d}\ beta(t)\alpha_i(t)dtdW_i(t) + \sum_{i=1}^{d}\sum_{j=1}^{d} \alpha_i(t)\alpha_j(t)dW_i(t)dW_j(t) \\
&= \left(\sum_{i=1}^{d} \alpha_i^2(t)\right) dt
\end{aligned}$$

Similarly,

$$(dY(t))^2 = (\sum_{i=1}^{d} \delta_i^2(t))dt$$

and (Trust me),

$$dX(t)dY(t) = \left(\sum_{i=1}^{d} \alpha_i(t)\delta_i(t)\right) dt \tag{14}$$

Then, computing formally,

$$\begin{aligned}
& df(X(t),Y(t)) \\
=& f_x dX(t) + f_y dY(t) + f_{xy} dX(t) d(Y(t)) + \frac{1}{2}\left(f_{xx}(dX(t))^2 + f_{yy}(dY(t))^2\right) \\
=& \left[f_x\beta(t) + f_y\gamma(t) + \frac{1}{2} f_{xx}\left(\sum_{i=1}^{d}\alpha_i^2\right) + \frac{1}{2} f_{yy}\left(\sum_{i=1}^{d}\delta_i^2\right) + f_{xy}\left(\sum_{i=1}^{d}\alpha_i(t)\delta_i(t)\right)\right] dt \\
& + \sum_{i=1}^{d}\left(f_x\alpha_i(t) + f_y\delta_i(t)\right) dW_i(t)
\end{aligned} \tag{15}$$

(In this calculation, the dependence on X, Y of f is omitted just for the simplicity of the notation.) This calculation can be stated formally as a theorem, with appropriate assumptions:

Theorem 1.3 *(Two-Dimensional Itō-Doeblin's formula)* Let $f(t,x,y)$ be a function whose partial derivative f_x, f_y, f_{xx}, f_{yy}, f_{xy} and f_{yx} are defined and are continuous. Let $X(t)$ and $Y(t)$ be Itō processes as discussed above. The two-dimensional Itō-Doeblin formula in differential form is (15).

Remark 1.4 Obviously, we can generalize the formula to n-dimensional, i.e., $f(X_,1, X_2, ..., X_n)$ with a little bit effort. We will not write down the formula since it is extremely messy, but rather, we write down the multi-dimensional Itō process:

$$\begin{aligned}
dX_1(t) &= \beta_1(t)dt + \sum_{i=1}^{d}\alpha_{1i}(t)dW_i(t), \\
dX_2(t) &= \beta_2(t)dt + \sum_{i=1}^{d}\alpha_{2i}(t)dW_i(t), \\
&\cdots\cdots\cdots \\
dX_n(t) &= \beta_n(t)dt + \sum_{i=1}^{d}\alpha_{ni}(t)dW_i(t)
\end{aligned} \tag{16}$$

If we define $\beta(t) = [\beta_1(t)\ \beta_2(t)\ \cdots\ \beta_n(t)]^\top$ and

$$\alpha(t) := \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1d} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nd} \end{pmatrix}(t)$$

then, (16) can be re-written as:

$$dX(t) = \beta(t)dt + \alpha(t)dW(t) \tag{17}$$

where $X(t), W(t)$ are column vectors.

Remark 1.5 As a final remark, we are always interested in the case where β, α have explicit dependence on $X(\cdot)$, i.e.,

$$dX(t) = \beta(t), X(t))dt + \alpha(t, X(t))dW(t) \tag{18}$$

Our discussion in the following sections are based on (18)

2 Existence and Uniqueness of Solution to SDE

2.1 Definition, Basics

In last recitation, we verified an original equation is the solution to SDE and also we directly solved SDE to get the same solution. The question arises whether a solution of SDE is always guaranteed, furthermore, is the solution unique? Let's address these issues now.

If you have undergraduate study in math, when treating ordinary differential equation (ODE), we need ensure the well-posedness and uniqueness. There, we impose the *Lipschitz condition* on the parameter and take advantage of *Picard iteration* to obtain the desired results. It turns out that this technique works almost identically in the stochastic setting. Let's work out the details. Recall the definition of *Lipschitz continuity*:

Definition 2.1 A function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is called *Lipschitz continuous* if there exists a constant $L < \infty$ such that

$$||f(x) - f(y)|| \leq K||x - y|| \tag{19}$$

for all $x, y \in \mathbb{R}^n$. A function $g : S \times \mathbb{R}^n \mapsto \mathbb{R}^m$ is *Lipschitz uniformly in S* if

$$||g(s, x) - g(s, y)|| \leq K||x - y|| \tag{20}$$

for any constant $K < +\infty$ which does not depend on s.

Remark 2.1 $|| \cdot ||$ is Euclidean norm.

Loosely speaking, a Lipschitz continuous function is restricted in how dramatical it can behave: the constant K is an upper bound for the slope of any secant line. In this sense, the family of Lipschitz function is large, as long as the function does not behave wildly, we will have Lipschitz continuity.

Next, let's have a short and informal discussion on *fixed point iteration.* Suppose we want to solve $x = f(x)$ (then x is called a *fixed point of* f), where f is a real-valued function. One approach is to find analytical solution of x, however, sometimes, it is very difficult to solve it and even not worthy doing so. The fixed point iteration method offers a numerical way to approximate the solution. The algorithm is quite simple, we choose any x_0 in the domain of f and apply the recursive formula:

$$x_{n+1} = f(x_n) \tag{21}$$

which give rise to the sequence $x_1, x_2, \ldots..$ Our hope is that $\{x_n\}_{n\geq 0}$ converges to a point x.

It can be realized if we assume that f satisfies Lipschitz continuity with constant $K < 1$. Under this circumstance, f has precisely one fixed point, regardless which initial point x_0 one choose. Let's sketch the proof:

Proof. Since f is Lipschitz continuous with $K < 1$, then

$$|x_n - x_{n-1}| = |f(x_{n-1}) - f(x_{n-2})| \leq K|x_{n-1} - x_{n-2}| \tag{22}$$

This is valid for $n \geq 2$. Thus,

$$|x_n - x_{n-1}| \leq K^{n-1}|x_1 - x_0| \tag{23}$$

Since $K < 1$, $K^{n-1} \to 0$ as $n \to \infty$. Therefore, $\{x_n\}_{n\geq 0}$ forms a *Cauchy sequence*. Since $\mathbb{R}$ is complete, $\lim_{n\to\infty} x_n = x^*$. For the iteration $x_n = f(x_{n-1})$, pass limit on both sides, by the continuity of f, we obtain:

$$x^* = f(x^*) \tag{24}$$

This shows that x^* is the fixed point for f. □

Remark 2.2 We use $\mathbb{R}$ to illustrate the idea of fixed point, but if we can only have it on real line, it won't be so powerful. Fortunately, *Stephan Banach* has the following result:

Theorem 2.3 *(Banach Fixed Point Theorem)* Let (X, d) be a non-empty complete metric space with a contraction mapping $T : X \mapsto X$. Then T admits a unique fixed-point x^* in X. x^* can be found by the recursive formula:

$$x_n = T(x_{n-1}) \tag{25}$$

It is a straightforward generalization. Notice in the proof of elementary fixed point method, the implicit, but most important, assumption is that real line is complete, i.e., every *Cauchy sequence* converges to a point in the space. Obviously, we only need the completeness of the vector space. A real valued function f can be generalized to a map T and T is a contraction mapping if T is Lipschitz continuous with constant less than 1. *Banach fixed point theorem* is ubiquitous, especially for the purpose of showing existence and uniqueness, and we will use it shortly.

2.2 Proof for SDE*

This is beyond the scope of the course, readers who has strong interest in math can appreciate the beauty of the proof. Maybe, I will mention the general idea in the recitation. When you are reviewing for the final, DO NOT waste your time on this part. It will never appear there !!! I just noticed that in this section, I use subscript to denote the time instead of brackets, i.e., $X(t)$ is the same as X_t, sorry for this confusion.

Here is the setup: we have finite time horizon $[0, T]$, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{0\leq t\leq T}$ generated by d-dimensional Brownian motion $\{W(t)\}_{0\leq t\leq T}$. We choose X_0 to be an $\mathcal{F}_0$-measurable n-dimensional random variable (we often consider the X_0 to be deterministic, but it is not necessary), and we seek a solution to the following SDE:

$$X_t = X_0 + \int_0^t \beta(s, X_s)ds + \int_0^t \alpha(s, X_s)dW_s,\ 0 \leq t \leq T \tag{26}$$

Here, $\beta : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}^n$ and $\alpha : [0, T] \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times d}$ are at least *Borel measurable*.

Theorem 2.4 *(Existence)* Suppose that ($X_0 \in L^2$)[1]; (2) α, β are Lipschitz continuous uniformly on $[0,T]$; and (3) $||\alpha(t,0)||$ and $||\sigma(t,0)||$ are bounded on $t \in [0,T]$. Then there exists a solution $\{X_t\}_{0\le t\le T}$ to the associated SDE, and moreover for this solution $\{X_t\}_{0\le t\le T}$ is in $(\mathcal{L}^2(T))^2$.

Proof. For any $\mathcal{F}_t$–adapted $Y. \in \mathcal{L}^2(T)$, we introduce the following nonlinear map $\mathscr{D}$:

$$(\mathscr{D}(Y.))_t := X_0 + \int_0^t \beta(s, Y_s)ds + \int_0^t \alpha(s, Y_s)dW_s \tag{27}$$

We claim that under the condition which we have imposed, $\mathscr{D}(Y.)$ is again an $\mathcal{F}_t$–adapted process in $\mathcal{L}^2(T)$. Our goal is to find a *fixed point* of the nonlinear operator $\mathscr{D}$, i.e., we wish to find an $\mathcal{F}_t$–adapted process $X. \in \mathcal{L}^2(T)$ such that $\beta(X.). = X..$ As a result, $X.$ is a solution of SDE.

We begin by showing that $\mathscr{D}$ does indeed map to an $\mathcal{F}_t$–adapted process in $\mathcal{L}^2(T)$. Note that

$$||\beta(t,x)|| \le ||\beta(t,x) - \beta(t,0)|| + ||\beta(t,0)|| \le K||x|| + K' \le C(1 + ||x||), \tag{28}$$

where K, K', $C < +\infty$ are constants that do not depend on t. We say that β satisfies a *linear growth condition.* Clearly the same argument holds for α, i.e.,

$$||\alpha(t,x)|| \le C(1 + ||x||) \tag{29}$$

We can now estimate:

$$||\mathscr{D}(Y.)||_{\mathcal{L}^2(T)} \le ||X_0||_{\mathcal{L}^2(T)} + ||\int_0^{\cdot} \beta(s, Y_s)ds||_{\mathcal{L}^2(T)} + ||\int_0^{\cdot} \alpha(s, Y_s)dW_s||_{\mathcal{L}^2(T)} \tag{30}$$

The first term gives $||X_0||_{\mathcal{L}^2(T)} = \sqrt{T}||X_0||_{\mathcal{L}^2(T)} < +\infty$ by assumption. Next,

$$||\int_0^{\cdot} \beta(s, Y_s)ds||_{\mathcal{L}^2(T)}||^2 \le T^2||\beta(\cdot, Y.)||^2_{\mathcal{L}^2(T)} \le T^2C^2||(1 + ||Y.||)||^2_{\mathcal{L}^2(T)} < +\infty \tag{31}$$

where we have used $(\int_0^t g(s)ds)^2 \le \frac{1}{t}\int_0^t g(s)^2 ds$ (*Jensen's inequality*), the linear growth condition and $Y. \in \mathcal{L}^2(T)$. Finally, let us estimate the stochastic integral term:

$$||\int_0^{\cdot} \alpha(s, Y_s)dW_s||^2_{\mathcal{L}^2(T)} \le T||\alpha(\cdot, Y.)||^2_{\mathcal{L}^2(T)} \le TC^2||(1 + ||Y.||)||^2_{\mathcal{L}^2(T)} < +\infty \tag{32}$$

where we used *Itō isometry.* Hence, $||\mathscr{D}(Y.)||_{\mathcal{L}^2(T)} < +\infty$ for $\mathcal{F}_t$–adapted $Y. \in \mathcal{L}^2(T)$, and clearly, $\mathscr{D}(Y.)$ is $\mathcal{F}_t$–adapted, so the claim is established.

Our next claim is that $\mathscr{D}$ is a continuous map, that is, $||Y.^n - Y.||_{\mathcal{L}^2(T)} \to 9$, then $||\mathscr{D}(Y.^n) - \mathscr{D}(Y.)||_{\mathcal{L}^2(T)} \to 0$ as well. But proceeding exactly as before, we find that:

$$||\mathscr{D}(Y.^n) - \mathscr{D}(Y.)||_{\mathcal{L}^2(T)} \le T||\beta(\cdot, Y.^n) - \beta(\cdot, Y.)||_{\mathcal{L}^2(T)} + \sqrt{T}||\alpha(\cdot, Y^n.) - \alpha(\cdot, Y.)||_{\mathcal{L}^2(T)} \tag{33}$$

---

[1]Suppose Z is a random variable, $Z \in L^2$ means that it has finite mean square, i.e., $\mathbb{E}[|Z|^2] < +\infty$

[2]Suppose $\gamma(\cdot)$ is an $\mathcal{F}_t$–adapted process, $\gamma(\cdot) \in \mathcal{L}^2(T)$ means $\mathbb{E}\big[\int_0^T \gamma(s)^2 ds\big] < +\infty$

In particular, using the Lipschitz condition, we find that

$$||\mathscr{D}(Y.^n) - \mathscr{D}(Y.)||_{\mathcal{L}^2(T)} \leq K\sqrt{T}(\sqrt{T}+1)||Y.^n - Y.||_{\mathcal{L}^2(T)} \tag{34}$$

where K is a Lipschitz constant for both β and α. This establish the claim.

Now, lets carry out the heart of the proof. Starting from an arbitrary $\mathcal{F}_t$–adapted process $Y^0. \in \mathcal{L}^2(T)$, consider the sequence $Y^1. = \mathscr{D}(Y^0.)$, $Y^2. = \mathscr{D}(Y^1.) = \mathscr{D}^2(Y.^0)$, e.t.c.. We will show below that $Y.^n$ is a *Cauchy sequence* in $\mathcal{L}^2(T)$; hence it converges to some $\mathcal{F}_t$–adapted process $Y_t \in \mathcal{L}^2(T)$. But then $Y.$ is necessarily a fixed point of $\mathscr{D}$; after all, $\mathscr{D}(Y.^n) \to \mathscr{D}(Y.)$ by the continuity of $\mathscr{D}$. Thus, $\mathscr{D}(Y.) = Y.$.

It only remains to show that $Y.^n$ is a Cauchy sequence in $\mathcal{L}^2(T)$. This follows from a slightly refined version of the argument that we used to prove continuity of $\mathscr{D}$. Note that

$$||(\mathscr{D}(Z.))_t - (\mathscr{D}(Y.))_t||_{\mathcal{L}^2(T)} \leq \sqrt{t}||\beta(\cdot, Z.) - b(\cdot, Y.)||_{\mathcal{L}^2(T)} + ||\alpha(\cdot, Z.) - \sigma(\cdot, Y.)||_{\mathcal{L}^2(T)} \tag{35}$$

which follows exactly as above. In particular, using Lipschitz property, we find

$$||(\mathscr{D}(Z.))_t - (\mathscr{D}(Y.))_t||_{\mathcal{L}^2(T)}|| \leq K(\sqrt{T}+1)||Z. - Y.||_{\mathcal{L}^2(T)} \tag{36}$$

Set $L = K(\sqrt{T}+1)$. Iterating this bound, we obtain:

$$\begin{aligned}
||(\mathscr{D}^n(Z.))_t - (\mathscr{D}^n(Y.))_t||^2_{\mathcal{L}^2(T)} &= \int_0^T ||(\mathscr{D}^n(Z.))_t - (\mathscr{D}^n(Y.))_t||^2_{\mathcal{L}^2(T)} dt \\
&\leq L^2 \int_0^T ||(\mathscr{D}^{n-1}(Z.))_t - (\mathscr{D}^{n-1}(Y.))_t||^2_{\mathcal{L}^2(t_1)} dt_1 \\
&\leq \cdots \leq L^{2n} \int_0^T \int_0^{t_1} \cdots \int_0^{t_{n-1}} ||Z. - Y.||^2_{\mathcal{L}^2(t_n)} dt_n \cdots dt_1 \\
&\leq \frac{L^{2n}T^n}{n!} ||Z. - Y.||^2_{\mathcal{L}^2(T)}
\end{aligned}$$

In particular, this implies that:

$$||\mathscr{D}^{n+1}(Y^0.) - \mathscr{D}^n Y^0.||_{\mathcal{L}^2(T)} \leq ||\mathscr{D}(Y^0.) - Y.^0||_{\mathcal{L}^2(T)} \sqrt{\frac{L^{2n}T^n}{n!}} \tag{37}$$

Thus,

$$\sum_{n=0}^{\infty} ||\mathscr{D}^{n+1}(Y^0.) - \mathscr{D}^n Y^0.||_{\mathcal{L}^2(T)} \leq ||\mathscr{D}(Y^0.) - Y.^0||_{\mathcal{L}^2(T)} \sum_{n=0}^{\infty} \sqrt{\frac{L^{2n}T^n}{n!}} < +\infty \tag{38}$$

which establishes $\mathscr{D}^n(Y^0.)$ is a Cauchy sequence in $\mathcal{L}^2(T)$. We are done □

Theorem 2.5 *(Uniqueness)* The solution of *Theorem* 2.4 is unique almost surely.

Proof. Let $X.$ be the solution of *Theorem* 2.4 and let $Y.$ be any other solution. It suffices to show that $X = Y$ almost surely; after all, both X_t and Y_t must have continuous sample paths, so $X. = Y.$ implies that they are indistinguishable.

Let us first suppose that $Y. \in \mathcal{L}^2(T)$ as well; then $\mathscr{D}^n(Y.) = Y.$ and $\mathscr{D}^n(X.) = X.$. Using the estimate in the proof above, we find that

$$||Y. - X.||^2_{\mathcal{L}^2(T)} = ||\mathscr{D}^n(Y.) - \mathscr{D}^n(X.)||^2_{\mathcal{L}^2(T)} \leq \frac{L^{2n}T6n}{n!}||Y. - X.||^2_{\mathcal{L}^2(T)} \tag{39}$$

Letting $n \to \infty$, we find that $||Y. - X.||_{\mathcal{L}^2(T)} = 0$, so $X. = Y.$ almost surely.

We now claim that any solution Y_t with $Y_0 = X_0 \in L^2$ must necessarily be an element of $\mathcal{L}^2(T)$; once this is established, the proof is complete. Let us write, using Itō's rule,

$$||Y_t||^2 = ||X_0||^2 + \int_0^t (2(Y_s)\beta(s, Y_s)) + ||\alpha(s, Y_s)||^2)ds + \int_0^t 2(Y_s)\alpha(s, Y_s)dW_s \tag{40}$$

Now, let $\tau_n = \inf\{t : ||Y_t|| \geq n\}$, and note that this sequence of stopping times is a localizing sequence for the stochastic integral; in particular,

$$\mathbb{E}(||Y_{t\wedge\tau_n}||^2) = \mathbb{E}(||X_0||^2) + \mathbb{E}\big[||\alpha(s, Y_s)||^2)ds\big] \tag{41}$$

Using *Fatou's lemma* on the left and *monotone convergence* on the right to let $n \to \infty$, applying *Tonelli's theorem*, and using the simple estimates $(a + b)^2 \leq 2(a^2 + b^2)$, we obtain

$$\mathbb{E}(1 + ||Y_t||^2) \leq \mathbb{E}[1 + ||X_0||^2] + 2C(2 + C)\int_0^t \mathbb{E}[1 + ||Y_s||^2]ds \tag{42}$$

But then we find that $\mathbb{E}[1 + ||Y_t||^2] \leq \mathbb{E}[1 + ||X_0||^2]e^{2C(2+C)t}$ using *Gronwall's lemma*, from which the claim follows easily. Hence the proof is complete. □

2.3 Examples – Solving Linear SDE

Let's assume that $\beta(t, x) = A(t)x + a(t)$ which is linear in x and a general α, i.e.,

$$dX(t) = (AX(t) + a(t))dt + \alpha(t)dW(t), \ X(0) = X_0 \tag{43}$$

To solve above SDE, let's first remove the *Wiener part* and solve the following ODE:

$$\dot{x}(t) = Ax(t) + a(t), \ t \geq 0 \tag{44}$$

with initial condition x_0. The solution is of the form:

$$x(t) = \Phi(t)(x_0 + \int_0^t \Phi^{-1}(s)a(s)ds), \ t \geq 0 \tag{45}$$

where Φ is so called *fundamental solution*. This means that Φ solves the matrix equation:

$$\dot{\Phi}(t) = A(t)\Phi(t), \text{ with } \Phi(0) = I \tag{46}$$

where I is the Identity matrix. The fundamental solution is given by:

$$\Phi(t) = e^{At} := \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} \tag{47}$$

The strong solution X of equation (43) with initial condition X_0 is given by

$$X(t) = \Phi(t)(X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW(s)),\ t \geq 0 \tag{48}$$

This can be verified by simply applying Itō's formula. Since the solution is unique, we actually solve the linear SDE. As a special case,

$$dX(t) = AX(t)dt + BdW(t),\ X(0) = x \tag{49}$$

has solution

$$X(t) = e^{At}x + \int_0^t e^{A(t-s)}BdW(s) \tag{50}$$

To see the uniqueness, let $Y(t)$ be another solution with same initial condition. Then

$$X(t) - Y(t) = \int_0^t A(X(s) - Y(s))ds \tag{51}$$

which implies

$$\frac{d}{dt}(X(t) - Y(t)) = A(X(t) - Y(t)),\ X(0) - Y(0) = 0 \tag{52}$$

and it is a standard fact that the unique solution of this equation is $X(t) - Y(t) = 0$.

Remark 2.6 In the class, you already saw several solved SDE's, in the homework, you will also get practice. The general idea of solving SDE is to guess the proper function $f(s, x)$, then apply Itō's formula to $df(s, X(s))$, finally, we want to match the SDE.

3 Simulation of SDE*

This part is also not covered in class, but I want to make this supplementation to let you know the usefulness of SDE. It is not just a theoretical model, rather Quants took advantage of simulating SDE to price various financial derivatives. Furthermore, maybe you will be doing some projects for other courses or for your own interests, simulation of SDE may give you some fresh ideas.

Stochastic differential equations, like their non-random counterparts, rarely admit analytical solution. For this reason, it is important to have numerical methods to simulate such equations on a computer. In the SDE case, we are seeking a numerical method that can simulate sample paths of the SDE with the correct distribution. We will discuss the simplest of these methods, which is nevertheless one of the most widely used in practice – the *Euler-Maruyama method*.

The method is in fact very close to the classical Euler method for discretization of ODEs. Consider our usual SDE, and let's discretize the interval $[0, T]$ into time steps of length T/p, i.e., we introduce the discrete grid $t_k = kT/p$, $k = 0, ..., p$. Then

$$X_{t_n} = X_{t_{n-1}} + \int_{t_{n-1}}^{t_n} \beta(s, X_s)ds + \int_{t_{n-1}}^{t_n} \alpha(s, X_s)dW_s \tag{53}$$

This expression can not be used as a numerical method, as X_{t_n} depends not only on $X_{t_{n-1}}$ but on all X_s in the interval $s \in [t_{n-1}, t_n]$. As X_s has continuous sample paths, however, it seems plausible that $X_s \approx X_{t_{n-1}}$ for $s \in [t_{n-1}, t_n]$, provided that p is sufficiently large. Then we try to approximate

$$X_{t_n} \approx X_{t_{n-1}} + \beta(t_{n-1}, X_{t_{n-1}})(t_n - t_{n-1}) + \alpha(t_{n-1}, X_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \tag{54}$$

The simple recursion is easily implemented on a computer, where we can obtain a suitable sequence of $W_{t_n} - W_{t_{n-1}}$ by generating i.i.d. d-dimensional Gaussian random variables with zero mean and covariance $(T/p)I$ using a (pseudo-)random number generator. Specifically, if we can generate $Z \sim N(0,1)$, then we shall simulate

$$X_{t_n} \approx X_{t_{n-1}} + \beta(t_{n-1}, X_{t_{n-1}})(t_n - t_{n-1}) + \alpha(t_{n-1}, X_{t_{n-1}})\sqrt{T/p}Z \tag{55}$$

And there is a theorem guaranteeing that this algorithm really does approximation to the solution of SDE when p is sufficiently large.

Suppose in Black-Scholes model, we can simulate

$$S_{t_n} = S_{t_{n-1}} + b(t_n - t_{n-1}) + \sigma\sqrt{T/p}Z, \text{ with inital condition } S_0 = s_0 \tag{56}$$

for N times, then we basically get N sample paths of stock price, i.e., $S^1, S^2, ..., S^N$. For example, we want to price the European call, the final pay-off is defined as: $V(S(T)) = (S(T) - K)^+$. Then we can calculate $V(S(T))$ for N times, by strong law of large number,

$$V(0, s_0) = e^{-rT}\mathbb{E}[(S(T) - K)^+] \approx \frac{1}{N}\sum_{i=1}^{N}(S^i(T) - K)^+ \tag{57}$$

for N sufficiently large. Isn't that interesting !

Remark 3.1 We will talk about some details about simulation on Friday's recitation if time permits. Also if you are interested in this topic, you can discuss with me or read the monograph by *Prof. Paul Glasserman, "Monte Carlo Methods in Financial Engineering"*.

4 Reference

1. *Jean Jacod, Philip Protter, "Probability Esentials", Springer, 2004;*
2. *E.Cinlar, "Probability and Stochastics", Springer, 2011*
3. *Daniel, Ocone, "Notes in mathematical finance I 2011"*
4. *Alison Etheridge, "A course in financial calculus", Cambridge, 2002*