Stochastic Differential Equations: Two Methods of Solution Construction and Wong-Zakai Approximations

Muhammad Alif Aqsha

Advisor: Dr Pierre Portal

September 25, 2022

Contents

1 Introduction

Let $F = (F_t)_{t\geq 0}$ be a stochastic process adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$. Let a and b some appropriate function mapping $\Omega \times \mathbb{R}$ to $\mathbb{R}$. A one-dimensional stochastic differential equation takes the form of

$$dX_t = a(\cdot, X_t)\,dt + b(\cdot, X_t)\,dF_t \tag{1}$$

which is just the differential representation of the following integral

$$X_t(\omega) = X_0(\omega) + \int_0^t a(\omega, X_t(\omega))\,dt + \int_0^t b(\omega, X_t(\omega))\,dF_t(\omega) \tag{2}$$

The differential form of the equation in (1) may be interpreted as follows: the change in X is driven by a deterministic source due to time and a random change due to some (or many) unexplainable sources. These form of equations is particularly useful to model real-world behaviors where there is perceived randomness influencing them e.g. stock price movements, particle trajectory on a fluid, or the spread of a disease.

The adapted stochastic process $X = (X_t)_{t \geq 0}$ satisfying (2) almost-surely is called the solution to (1) with initial condition X_0. If F has a finite variation a.s., then the second integral in (2) may be interpreted as the Lebesgue-Stiltjes type. If that is the case, the solution may be constructed using the theory of deterministic ordinary differential equations.

Unfortunately in general, F may have infinite variation. Indeed, the Brownian motion, which is one of the most well-known adapted process, has infinite variation and undifferentiable everywhere. Consequently, the theory of stochastic integration such as Itô and Stratonovich calculus were developed to tackle this problem.

The existence or construction of SDE solutions may be obtained using analytical (or limiting) arguments such as Picard iteration. Unfortunately, these methods may not be practical to implement computationally as they may rely on being able to observe the source process F at any time (no information gaps). Instead, an approximation of F is used where the process are sampled on some finite points on the time interval and interpolated on the gaps. Then the solution may be approximated by

$$dX_t^* = a(\cdot, X_t^*)\, dt + b(\cdot, X_t^*)\, dF_t^* + \text{correction term} \tag{3}$$

where F^* is the approximating (sampled) process. This type of approximations is called the Wong-Zakai approximation.

In this paper, we first discuss the theory of conditional expectations, martingales, Itô's integral and Itô's rule. We do not discuss the proof of these preliminaries; we focus on the motivations instead (and perhaps the rough outline of the proof). Next, we present the construction of an SDE solution using Picard's iteration. Furthermore, we describe two methods of solving two special case of SDE involving turning these into a problem of solving deterministic ODEs. We then prove the convergence of Wong-Zakai approximations for non-regime-switching SDEs. Lastly, we present a result developed by (Nguyen & Peralta, 2021) on the convergence rate of Wong-Zakai approximations for the regime-switching SDEs.

2 Preliminaries

In this section we will describe preliminary theorems needed for stochastic differential equations theory but we will not prove them. Instead, we will present some motivations on the "basic" definitions and theorems of conditional expectation, martingales, and Itô's calculus.

2.1 Conditional Expectation

Let us imagine a sequence of prices (which is of course random) from a stock, each element in the sequence came from different times in the past. Of course, if we know the price at, let's say, time t, we automatically know the previous prices, but not necessarily the future prices. We might say that all information from the past is included in the next information. Using concepts from measure theory, we may mathematically define this characteristics of increasing information like this.

Definition 2.1 (Filtration)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple. A collection of sub-sigma algebras $(\mathcal{F}_t)_{t \in I}$ with respect to sigma-algebra $\mathcal{F}$ and a totally-ordered time-index set I is a *filtration* if $\mathcal{F}_s \subseteq \mathcal{F}$ and $\mathcal{F}_s \subseteq \mathcal{F}_t$ for every $s, t \in I$ and $s < t$.

Of course, knowing the stock price at a time means knowing the previous prices, but not necessarily the future values. Using concept of measurability, we can define such collection of random variables as below.

Definition 2.2 (One-dimensional Adapted Stochastic Process)**.** Let $(\mathcal{F}_t)_{t\in I}$ be a filtration with respect to the sigma-algebra $\mathcal{F}$ and the totally-ordered time-index set I. An *adapted stochastic process* X is a collection of random variables $(X_t)_{t\in I}$ indexed by I such that for every $t \in I$, X_t is $\mathcal{F}_t$ measurable, that is,

$$X_t^{-1}(B) = \{\omega \in \Omega : X_t(\omega) \in B\} \in \mathcal{F}_t$$

for every Borel set B in $\mathbb{R}$.

Definition 2.3 (One-dimensional Progressively Measurable Stochastic Process)**.** Let $(\mathcal{F}_t)_{t\in I}$ be a filtration with respect to the sigma-algebra $\mathcal{F}$ and the time-index set I. Let $\mathcal{A}_t = \sigma(B([0,t]) \times \mathcal{F}_t)$ (the smallest sigma-algebras containing the product of the Borel σ-algebra of $[0,t]$ and $\mathcal{F}_t$). A stochastic process $X = (X_t)_t$ is *progressively measurable* if the map $[0,t] \times \Omega \to X_t(\omega)$ defined by $(s,\omega) \mapsto X_s(\omega)$ is $\mathcal{A}_t$ measurable, that is

$$X^{-1}(B) = \{(s,\omega) : X_s(\omega) \in B\} \in \mathcal{A}_t$$

for every Borel set B in $\mathbb{R}$.

Now, we are interested to determine the expected price in the future, based on the information we have gathered consisting of historical data of the prices. Let us say we are currently at time s and we have observed the initial price X_t. We would like to know the expected value of X_t at time t in the future based on all information available at the present. Thus we may collapse all possible information about the past and only consider the ones matching our observations. This kind of "collapsing possible informations" is akin to a concept in linear algebra called *orthogonal projection*. With orthogonal projection, we only care about the smaller sub-vector space and ignore anything orthogonal to it (in the case of random variables, orthogonality is interpreted as uncorrelation).

To make sense of this "orthogonal projection", we need a Hilbert space for random variables. Of course, the natural space of functions (or random variables) with Hilbert characteristics is the square-integrable one, that is, all random variable X to the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ with finite 2-moment.

$$\mathbb{E}\left[|X|^2\right] = \int_\Omega |X|^2 \, d\mathbb{P} < \infty$$

Let us denote such space as $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Now, we may collapse the collection of all possible events $\mathcal{F}$ into a smaller sub-sigma algebras $\mathcal{G}$ after taking all present information into account. From functional analysis, the space $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a sub-Hilbert space of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, which guarantees that the orthogonal projection from $L^2(\Omega, \mathcal{F}, P)$ to $L^2(\Omega, \mathcal{G}, \mathbb{P})$ exists and is unique (this is from Riesz representation theorem). Let us denote such projection as T.

The inner product in Hilbert function (or random variable) space is the integral of the multiplication

$$\langle X, Y \rangle = \int_\Omega XY \, d\mathbb{P}$$

so from the properties of orthogonal projection, we need to have

$$0 = \langle T(X) - X, Y \rangle \quad \text{for all } X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$$

Of course for all $G \in \mathcal{G}$, the indicator function $\mathbf{1}_G$ is $\mathcal{G}$-measurable and has finite second-moment. Also, from functional analysis, the space of linear combinations of such indicator functions is dense in $L^2(\Omega, \mathcal{G}, \mathbb{P})$, so we may restrict Y to be such indicator function and still obtain the same orthogonal projection.

Theorem 2.1. *If T is a continuous linear operator from $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to $L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$0 = \langle T(X) - X, \boldsymbol{1}_G \rangle \quad \text{for all } G \in \mathcal{G}$$

then T is the orthogonal projection from $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

To make things clear, from now on the inner product will be written in its integral form. Thus we may write

$$\int_\Omega (T(X) - X)\mathbf{1}_G \, d\mathbb{P} = 0 \quad \text{for all } X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), G \in \mathcal{G} \tag{4}$$

$$\Leftrightarrow \int_G T(X) \, d\mathbb{P} = \int_G X \, d\mathbb{P} \tag{5}$$

From this, we may define conditional expectation as follows.

Definition 2.4 (Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and $\mathcal{G} \subseteq \mathcal{F}$ is a sub-sigma algebra.

Then the *conditional expectation* with respect to $\mathcal{G}$ is the unique continuous linear operator $T : L^2(\Omega, \mathcal{F}, \mathbb{P}) \to L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\int_G T(X) \, d\mathbb{P} = \int_G X \, d\mathbb{P} \quad \forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), G \in \mathcal{G} \tag{6}$$

We denote $T(X)$ as $\mathbb{E}\left[X|\mathcal{G}\right]$, which reads as the conditional expectation of X w.r.t. $\mathcal{G}$.

Conditional expectations have several properties as presented in the following theorem.

Theorem 2.2 (Properties of Conditional Expectation)**.** *Let $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_0$ is a sub-sigma algebra of $\mathcal{F}$. Then*

(i) (Linearity) $\mathbb{E}\left[aX + bY|\mathcal{F}_0\right] = a\mathbb{E}\left[X|\mathcal{F}_0\right] + b\mathbb{E}\left[Y|\mathcal{F}_0\right]$

(ii) (Monotonicity) If $X \leq Y$ a.s., then $\mathbb{E}\left[X|\mathcal{F}_0\right] \leq \mathbb{E}\left[Y|\mathcal{F}_0\right]$

(iii) (The Tower Property) If $\mathcal{F}_1$ is a subsigma algebra of $\mathcal{F}_0$, then $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}_0\right]|\mathcal{F}_1\right] = \mathbb{E}\left[X|\mathcal{F}_1\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}_1\right]|\mathcal{F}_0\right]$

(iv) (Law of Total Expectation) $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}_0\right]\right] = \mathbb{E}\left[X\right]$

(v) (Jensen's Inequality) If ϕ is a convex function, then $\phi(\mathbb{E}\left[X|\mathcal{F}_0\right]) \leq \mathbb{E}\left[\phi(X)|\mathcal{F}_0\right]$

2.2 Martingales

Let us consider an easier process, for example a gambling process. Suppose one gains 1 dollar if a coin toss resulted in the number part faced upward, loses 1 dollar otherwise. It is reasonable to suppose that the probability of getting the number part of the coin equals the picture part, 1/2 each. Suppose also that every coin toss is independent of each other. Then of course at any time, the expected additional gains we will get at any future time is zero because the expected gains of each coin toss is 0 (and the coin tosses are independent). Because it is expected that we gain no additional coins, the expected number of coins we have at any future time is exactly equal the number of coins we currently have.

This type of process, where the expected value of a process at any future time equals the current value is called a *martingale*. In fact, this term *martingale* came from gambling.

Definition 2.5 (Martingale)**.** Suppose that $(X_t)_{t\in I}$ is a process adapted to filtration $(\mathcal{F}_t)_{t\in I}$ on the probability space $(\Omega, \mathcal{F}, P)$. This process is called a martingale if

$$\mathbb{E}\left[X_s|\mathcal{F}_t\right] = X_t \quad \forall s < t \text{ in } I$$

The most famous example of martingale is the one we've described earlier on gambling, which is also called a *random walk*. Of course it is a discrete-time martingale. For the continuous-time martingale, the most well-known example is the *Wiener process*, or more commonly known as the *Brownian motion*.

Definition 2.6 (One-dimensional Brownian motion)**.** A continuous-time process $(X_t)t \geq 0$ is called a Brownian motion if it satisfies the following properties:

1. $X_0 = 0 \quad a.s.$
2. For every $n \in \mathbb{N}$ and $0 < t_1 < t_2 < ... < t_n < \infty$, the random variables $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}$, ..., $X_{t_n} - X_{t_{n-1}}$ are mutually independent.
3. For every $0 \leq s < t < \infty$, the increment $X_t - X(s)$ is normally distributed with mean 0 and variance $t - s$.
4. The process $(X_t)t \geq 0$ is continuous almost surely, that is

$$P(t \mapsto X_t \text{ is continuous}) = 1$$

Several constructions have been proposed to ensure the existence of such Brownian motion, most of which restrict the time interval on $[0, 1]$. One of them is using the Haar wavelets as described in (Steele, 2001). The other one is described in (McKnight, 2009) by defining $X_0 = 0$ and X_1 a $N(0, 1)$-distributed random variable, then successively interpolating on each midpoint of all dyadic intervals $[j2^{-k}, (j + 1)2^{-k}]$ for all j, k nonnegative integers and adding independent normally-distributed noises of mean 0 and scaled variances, then define the process as the limit of each dyadic points.

Now let us consider back our gambling process. Perhaps we devise a strategy where we will stop whenever we have accumulated certain number of coins, or that we stop whenever we have lost certain number of coins. In both cases, our strategy of stopping at a certain turn no more than, say the n-th turn, requires information we've gained from the n-th turn and beyond previously. This can be defined mathematically as a *stopping time*.

Definition 2.7 (Stopping time)**.** Suppose that $(\mathcal{F}_t)_{t\in\mathbb{N}}$ is a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and τ a random variable which takes value on the nonnegative integers. The random variable τ is called a *stopping time* if

$$\{\tau \leq n\} \in \mathcal{F}_n$$

It turns out that no matter what stopping strategy we are using, as long as our original gambling process is a martingale, we will still be expected to gain no additional coins, that is, our stopped process is also martingale. First let us consider the discrete-time ones.

Theorem 2.3. *If $(X_n)_{n=0}^{\infty}$ is a martingale and τ is a stopping-time, both w.r.t filtration $(\mathcal{F}_n)_{n=0}^{\infty}$, then $(X_{n\wedge\tau})_{n=0}^{\infty}$ is also a martingale.*

PROOF. This can be seen on (Steele, 2001). □

Now, a stochastic process may not be a martingale, but the stopped process might be one. Then we may be interested to approximate this process by some sort of the stopped version. If we can find a sequence of stopping time that diverges to ∞ such that if we stop our original process with any stopping time in the sequence we obtain a martingale, then we call this process a *local martingale*.

Definition 2.8. Let $X = (X_t)_{t \in I}$ be a process adapted to the filtration $(\mathcal{F}_t)_{t \in I}$. If there exists a nondecreasing sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ w.r.t filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ such that $\mathbb{P}\left(\lim_{n \to \infty} \tau_n\right) = 1$ and the stopped process $(X_{t \wedge \tau_n})_{t \in I}$ is a martingale for every n, then X is a *local martingale*.
Such nondecreasing sequence of stopping times is called a *localizing sequence*.

Now let us return for a moment to the Brownian martingale $(B_t)_{t \geq 0}$. From Jensen's inequality for conditional expectation, we have for $0 \leq s < t < \infty$,

$$X_s^2 = \mathbb{E}\left[X_t | \mathcal{F}_s\right]^2 \leq \mathbb{E}\left[X_t^2 | \mathcal{F}_s\right] \tag{7}$$

Then $(X_t^2)_{t \geq 0}$ is called a *submartingale*, that is, the conditional expectation is greater than the present value. We also have that the process is nonnegative, continuous, and $\mathbb{E}\left[X_s^2\right] = s < \infty$ (square-integrable). Consequently, we have the unique *Doob-Meyer* decomposition

$$X_t^2 = M_t + A_t \tag{8}$$

where M is a right-continuous martingale and A is a nondecreasing process with initial value $A_0 = 0$. It turns out $M_t = X_t^2 - t$ and $A_t = t$ as $(X_t^2 - t)_{t \geq 0}$ is a continuous martingale.

It turns out that this decomposition generalizes to other right-continuous and square-integrable submartingale (with different M and A of course).

Theorem 2.4. *If a martingale $(X_t)_{t \geq 0}$ is right-continuous and square-integrable, there exists a unique Doob-Meyer decomposition*

$$X_t^2 = M_t + A_t \tag{9}$$

where $(M_t)_{t \geq 0}$ is a right-continuous martingale and $(A_t)_{t \geq 0}$ is a nondecreasing process where $A_0 = 0$.

PROOF. The proof is quite long. It can be seen on (Karatzas & Shreve, 1991). □

Then we can define a *quadratic-variation* of $(X_t)_{t \geq 0}$, which is the "limit" of the discrete quadratic-variation $\sum_{t=0}^{n} (X_{t_{i+1}} - X_{t_i})^2$ where $0 = t_0 < t_1 < ... < t_n < t_{n+1} = t$ is a partition of $[0, t]$.

Definition 2.9. Let $(X_t)_{t \geq 0}$ be a right-continuous, square integrable martingale. For fixed $0 \leq t < \infty$, let $0 = t_0 < t_1 < ... < t_n < t_{n+1} = t$ be a partition of $[0, t]$ and $\Pi = \max_{0 \leq i \leq n} t_{i+1} - t_i$ be the mesh of the partition. The *quadratic variation* of X at time t, $\langle X \rangle_t$, is defined as the limit in probability to $\sum_{t=0}^{n} (X_{t_{i+1}} - X_{t_i})^2$ as $\Pi \to 0$, that is

$$\sum_{t=0}^{n} (X_{t_{i+1}} - X_{t_i})^2 \xrightarrow{\mathbb{P}} \langle X \rangle_t \text{ as } \Pi \to 0 \tag{10}$$

It turns out that this quadratic variation equals to the A part in the Doob-Meyer decomposition in (9).

$$\langle X \rangle_t = A_t \tag{11}$$

This is due to the fact that M is a martingale, so the increments over two non-overlapping intervals are uncorrelated (using the conditional expectation properties), so

$$\mathbb{E}\left[(X_t - X_0)^2\right] = \mathbb{E}\left[\left(\sum_{i=0}^{n} X_{t_{i+1}} - X_{t_i}\right)^2\right] = \mathbb{E}\left[\sum_{t=0}^{n}(X_{t_{i+1}} - X_{t_i})^2\right] \tag{12}$$

But $\mathbb{E}\left[X_t X_0\right] = \mathbb{E}\left[\mathbb{E}\left[X_t X_0 | \mathcal{F}_0\right]\right] = \mathbb{E}\left[X_0^2\right]$, so

$$\begin{aligned}\mathbb{E}\left[(X_t - X_0)^2\right] &= \mathbb{E}\left[X_t^2 - X_0^2\right] \\ &= \mathbb{E}\left[M_t + A_t - M_0 - A_0\right] \\ &= \mathbb{E}\left[A_t\right]\end{aligned} \tag{13}$$

Now, using stopping time to bound the process X in $[0, t]$, we obtain that the variance of the discrete quadratic-variation converges to 0 (Karatzas & Shreve, 1991). Thus, using Chebyshev's inequality, we obtain the convergence in probability.

2.3 Itô's Integral

Now, let $X = (X_t)_{t \geq 0}$ be a right-continuous, square-integrable martingale. We may interpret this X as the price of certain asset (e.g. stock). Let $0 = t_0 < t_1 ... < t_n < t_{n+1} = t$ be a partition of $[0, t]$. Suppose we may determine the number of that asset we hold at each time-interval $[t_i, t_{i+1})$. Naturally, we may only determine it using past informations, not future ones. We may write the number of asset as

$$A(s) = \sum_{i=0}^{n} A_{t_i} \mathbf{1}_{[t_i, t_{i+1}) \cap [0,s]} \tag{14}$$

where A_{t_i} is $\mathcal{F}_{t_i}$-measurable and square-integrable.

As an example, we hold A_0 assets at time $[0, t_1)$. Then at time t_1, we update our holdings to A_{t_1} using past informations and keep it until time t_2. If $A_{t_1} > A_0$, we buy $A_{t_1} - A_0$ assets at time t_1, and if $A_{t_1} < A_0$, we sell $A_0 - A_{t_1}$ assets. Then our gains equal to

$$(A_{t_1} - A_0)(X_{t_2} - X_{t_1}) + A_0(X_{t_2} - X_0) = A_0(X_{t_1} - X_0) + A_{t_1}(X_{t_2} - X_{t_1}) \tag{15}$$

We may generalize this gain at any time s as

$$I(A)(s) = \sum_{i=0}^{k} A_{t_i}(X_{t_{i+1}} - X_{t_i}) + A_{t_k}(X_s - X_{t_k}) \tag{16}$$

where k is the unique integer satisfying $t_k \leq s < t_{k+1}$.

The equation (16) is called the *Ito integral* of *simple process* (14). It is also called a *martingale transform* of X because the process $(A'_{t_i})_{i=1}^n = (A_{t_{i-1}})_{i=1}^n$ is *predictable* w.r.t filtration $(\mathcal{F}_{t_{i-1}})_{i=1}^n$ i.e. A'_{t_i} is $\mathcal{F}_{t_{i-1}}$-measurable. From (Steele, 2001), if $\|A'_{t_i}\|_\infty < \infty$, then the martingale transform is also a martingale & right-continuous with respect to time. By approximating A'_{t_i}s by an L_∞ random variables using stopping times, we also obtain the Itô integral in (16) is also a right-continuous martingale.

Now, we denote the collection of such adapted simple processes (or Borel-measurable real-valued functions with domain in sample space Ω and time $[0,T]$) as in (14) as $H^2_{0,X}$. It turns out that this $H^2_{0,X}$ space is dense in H^2_X, which is the space of all functions (or processes) f satisfying

$$\mathbb{E}\left[\int_0^T |f(\cdot,s)|^2 \, d\langle X\rangle_s\right] = \int_{[0,T]\times\Omega} |f(\omega,s)|^2 \, d\langle X\rangle_s \, d\mathbb{P} < \infty \tag{17}$$

equipped with the same norm as the equation above. We may also denote $H^2_X = L^2([0,T]\times\Omega, \langle M\rangle \times \mathbb{P})$ which is the space of square-integrable function in $[0,T]\times\Omega$ equipped with measure $\langle M\rangle \times \mathbb{P}$.

Thus we may define the Itô integral of $f \in H^2_{0,X}$ from $t=0$ to T as the limit of the Itô integral of the approximating simple processes.

$$I(f)(T) = \lim_{\substack{n\to\infty \\ L^2(\Omega)}} I(f_n)(T), \quad f_n \xrightarrow{H^2_X} f \tag{18}$$

To expand this Itô integral on fixed interval $[0,T]$ to any interval $[0,t] \subseteq [0,T]$, we use the following theorem.

Theorem 2.5. *Let $X = (X_t)_{0\le t\le T}$ be a square-integrable martingale and $f \in H^2_X$. Then there exists a $(\mathcal{F}_t)_{0\le t\le T}$-adapted continuous martingale $Y = (Y_t)_{0\le t\le T}$ such that*

$$\mathbb{P}\left(Y_t = I(f)(t)\right) = 1 \quad \forall t \in [0,T] \tag{19}$$

Let us denote $I(f)(t) = I_t(f)$ for $f \in H^2_X$. Then this Itô integral satisfies these properties:

(i) $I_0(f) = 0$ a.s.

(ii) (Martingale property) $\mathbb{E}\left[I_t(f)|\mathcal{F}_s\right] = I_s(f)$ for $0 \le s < t < \infty$ a.s.

(iii) (Itô isometry) $\mathbb{E}\left[I_t(f)^2\right] = \mathbb{E}\left[\int_0^t f(\cdot,u)^2 \, d\langle X\rangle_u\right]$

(iv) (Conditional Itô isometry) $\mathbb{E}\left[(I_t(f) - I_s(f))^2|\mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t f(\cdot,u)^2 \, d\langle X\rangle_u |\mathcal{F}_s\right]$ for $0 \le s < t < \infty$ a.s.

(v) (Linearity) $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for $f, g \in H^2_X$

2.4 Itô's Formula

From ordinary calculus, if we have a differentiable path y_t and a differentiable function F, then we have the Fundamental Theorem of Calculus

$$F(y_t) - F(0) = \int_0^t F'(y_s) \, dy_s \tag{20}$$

where the last integral may be interpreted as the Riemann-Stiltjes (or Lebesgue-Stiltjes) type. Alternatively, we may just substitute $dy_s = y'_s\, ds$.

Unfortunately, a stochastic process $X = (X_t)_{o \leq t \leq T}$ might not be differentiable for a.s. every realization. Thus, we might need to look at this integral as the Itô type.

Consider a process $X = (X_t)_{t \geq 0}$ of the form

$$X_t = X_0 + M_t + A_t \tag{21}$$

where $(M_t)_{t \geq 0}$ is a local continuous martingale with $M_0 = 0$ and $A = (A_t)_{t \geq 0}$ is a process with bounded variation.

As we can see, this looks like a generalization of the Doob-Meyer decomposition. This type of process is called a *semimartingale*.

Definition 2.10 (Continuous Semimartingale)**.** A continuous semimartingale $X = (X_t)_t$ is an adapted process which has decomposition as in (21) where $(M_t)_{t \geq 0}$ is a local continuous martingale with $M_0 = 0$ and $A = (A_t)_{t \geq 0}$ is a process with bounded variation or equivalently a difference of two continuous, nondecreasing processes

$$A = A^+ - A^-, \quad A_0^\pm = 0$$

We may also see that $\langle X \rangle_t = \langle M \rangle_t$ as the quadratic variation of a monotone process is always 0.

We may define a generalised Itô integral for the semimartingale (21) as

$$\int_0^t f(\cdot, s)\, dX_s = \int_0^t f(\cdot, s)\, dM_s + \int_0^t f(\cdot, s)\, dA_s \tag{22}$$

where the first integral on the righthand-side is the Itô integral (the local martingale is handled by a localizing sequence to turn it into a martingale), while the second integral is the Lebesgue-Stiltjes type (as the process A has finite variation a.s).

Theorem 2.6. *Suppose a process Y is constructed as*

$$Y(\omega, t) = \int_0^t a(\omega, s)\, dM_s + \int_0^t b(\omega, s)\, dA_s \tag{23}$$

where M and A is obtained from the decomposition of continuous semimartingale (21)*, and*

$$\mathbb{P}\left(\int_0^t |a(\cdot, s)|^2\, d\langle M \rangle_s < \infty \right) = 1 \quad \forall\, t \in [0, T] \tag{24}$$

$$\mathbb{P}\left(\int_0^t |b(\cdot, s)|\, d\langle A \rangle_s < \infty \right) = 1 \quad \forall\, t \in [0, T] \tag{25}$$

Then

$$\langle Y \rangle_t = \int_0^t |a(\cdot, s)|^2\, d\langle M \rangle_s \quad \forall 0 \leq t \leq T \tag{26}$$

Theorem 2.6 basically tells us that the quadratic variation of a process only comes from the stochastic (Itô) integral part. The Lebesgue-Stiltjes integral has quadratic variation 0 which follows directly from triangle inequality and the requirement (25).

To calculate the quadratic variation of the Itô integral part, we may compute the expectation

$$d_j = \left(\int_{t_{j-1}}^{t_j} a(\cdot, s)\, dM_s \right)^2 - \int_{t_{j-1}}^{t_j} |a(\cdot, s)|^2\, d\langle M \rangle_s \tag{27}$$

using conditional Itô's isometry and the law of total expectations, which will gives us 0. Adding them all for all time-partitions, we get the expectation of the discrete quadratic-variation equals the expectation of the right-hand side of (26).

To calculate the variance of $\left(\sum_{j=1}^{n} d_j \right)$, again we use the tower and total expectation property to obtain $\mathbb{E}[d_j d_k] = 0$ for $j \neq k$. Using AM-QM and Cauchy-Schwarz inequality, we can make the variance arbitrarily small by taking a partition with small enough mesh.

Now we may describe the Itô's formula.

Theorem 2.7 (Itô's formula)**.** *Suppose* $X = (X_t)_{0 \le t \le T}$ *is a continuous semimartingale and* $F : \mathbb{R} \to \mathbb{R}$ *is twice continuously-differentiable. Then for all* $t \in [0, T]$,

$$F(X_t) - F(0) = \int_0^t F'(X_s)\, dX_s + \frac{1}{2} \int_0^t F''(X_s)\, d\langle X \rangle_s \tag{28}$$

The proof of the Itô's formula usually starts by assuming F has a compact support, so it has bounded F, F', and F''. The F difference on each time partition is then written in its 2nd order Taylor series plus the error term

$$|r(X_{t_{i-1}}, X_{t_i})| \le \frac{1}{2} |X_{t_{i-1}} - X_{t_i}|^2 \sup_{u \in [X_{t_{i-1}} \wedge X_{t_i}, X_{t_{i-1}} \vee X_{t_i}]} |F''(u) - F''(X_{t_{i-1}})| \tag{29}$$

Then by decomposing X as in (21),

$$F(X_t) - F(X_0) = \sum_{i=1}^{n} F'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) + \frac{1}{2} \sum_{i=1}^{n} F''(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}})^2 + \sum_{i=1}^{n} r(X_{t_{i-1}}, X_{t_i}) \tag{30}$$

$$\begin{aligned} &= \sum_{i=1}^{n} F'(X_{t_{i-1}})(M_{t_i} - M_{t_{i-1}}) + \sum_{i=1}^{n} F'(X_{t_{i-1}})(A_{t_i} - A_{t_{i-1}}) \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} F''(X_{t_{i-1}})(M_{t_i} - M_{t_{i-1}})^2 + \sum_{i=1}^{n} F''(X_{t_{i-1}})(M_{t_i} - M_{t_{i-1}})(A_{t_i} - A_{t_{i-1}}) \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} F''(X_{t_{i-1}})(A_{t_i} - A_{t_{i-1}})^2 + \sum_{i=1}^{n} r(X_{t_{i-1}}, X_{t_i}) \end{aligned} \tag{31}$$

The first sum is the Itô's integral of the approximating simple process, and by bounded convergence theorem, it converges to the Itô integral of F'. The second one converges to the Lebesgue-Stiltjes integral as A has finite variation. The fourth one converges to 0 because F'' is bounded, A has finite variation, and M is continuous on the closed interval $[0, T]$ that we may find a small enough mesh such that $M_{t_i} - M_{t_{i-1}}$ is arbitrarily small. The fifth one also converges to 0 using similar argument as the fourth one. The sixth sum also converges to 0 because F'' is uniformly continuous, so we may choose a space-mesh δ_1 such that $|F(x) - F(y)| < \varepsilon$ for $|x - y| < \delta_1$, time-mesh $\delta_2 := \delta_2(\omega, \delta_1)$ such that $|X_s - X_t| < \delta_1$ for $s, t \in [0, T]$,

$|s-t|<\delta_2$, and choosing our time-partition with mesh $<\delta_2$. Then this sum is bounded by ε times the discrete-quadratic variation, converging to 0 for arbitrarily small ε.

For the third sum,

$$\begin{aligned}\sum_{i=1}^{n} F''(X_{t_{i-1}})(M_{t_i}-M_{t_{i-1}})^2 = &\sum_{i=1}^{n} F''(X_{t_{i-1}})\Big[(M_{t_i}-M_{t_{i-1}})^2 - \Big(\langle M\rangle_{t_i} - \langle M\rangle_{t_{i-1}}\Big)\Big] \\ &+ \sum_{i=1}^{n} F''(X_{t_{i-1}})\Big(\langle M\rangle_{t_i} - \langle M\rangle_{t_{i-1}}\Big)\end{aligned} \tag{32}$$

Let us assume first M is a martingale. Then the second integral on (32) converges to the Lebesgue-Stiltjes integral w.r.t $\langle M\rangle$ as $\langle M\rangle$ is a non-decreasing process with $\langle M\rangle_0 = 0$

Observe that $M^2 - \langle M\rangle$ is also a martingale. Using the fact that $\mathbb{E}\left[M_{t_{i-1}}M_{t_i}|\mathcal{F}_{t_{i-1}}\right] = 0$, the conditional expectation (w.r.t $\mathcal{F}_{t_{i-1}}$) of each $F''(X_{t_{i-1}})\Big[(M_{t_i}-M_{t_{i-1}})^2 - \Big(\langle M\rangle_{t_i} - \langle M\rangle_{t_{i-1}}\Big)\Big]$ equals to $2F''(X_{t_{i-1}})\langle M\rangle_{t_{i-1}}$. Consequently, the unconditioned expectation is $2\langle M\rangle_0 = 0$, and the expectation of the first sum (32) equals 0.

Furthermore, using the martingale property such as the tower property, we may see that the variance of the first sum on (32) converges to 0, so the first sum converges to 0 in probability.

For more general twice continuously-differentiable F, we multiply F by some twice continuously-differentiable bell function b_M such that $b_M(x) = 1$ for $x \in [-M, M]$ and $0 \in [-M-1, M+1]$. Itô's formula then applies for Fb_m. Using stopping time argument, it should also applies to F.

From Itô's formula, we may obtain the following Itô's rule.

Theorem 2.8 (Itô's rule)**.** *Suppose $X = (X_t)_{0\le t\le T}$ is a continuous semimartingale and $F : [0,T]\times\mathbb{R}\to\mathbb{R}$ is a function that is continuously-differentiable in the first (time) entry and twice continuously-differentiable in the second (space) entry. Then for all $t\in[0,T]$,*

$$F(t,X_t) - F(0,0) = \int_0^t \partial_s F(s,X_s)\,ds + \int_0^t \partial_x F(s,X_s)\,dX_s + \frac{1}{2}\int_0^t \partial_{xx}F(s,X_s)\,d\langle X\rangle_s \tag{33}$$

3 Stochastic Differential Equations

Now, consider a Brownian motion $(B_t)_{t\ge 0}$ and a process $(X_t)_{t\ge 0}$ which satisfies

$$X_t = X_0 + \int_0^t \mu(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s, \quad 0\le t\le T \tag{34}$$

for all $t\ge 0$, with μ and σ having certain nice properties. To shorten the notation, we may write equation (34) as

$$dX_t = \mu(s,X_t)dt + \sigma(t,X_t)dB_t, \quad 0\le t\le T \tag{35}$$

The equation (35) really is an integral equation (34), but it gives us a nice interpretation: the change in X_t is driven by the deterministic change $\mu(s,X_t)dt$ with randomness $\sigma(t,X_t)dB_t$. The deterministic change μ (without the dt part) is called the *drift* term, and the random change σ (without dB_t) is called the *diffusion* term.

Now we need both integrals on (34) to make sense outcome-wise (ω), so we need μ to be integrable in the ordinary Lebesgue sense and σ to be Itô-integrable, so we restrict ourselves to standard processes as discussed previously.

In the theory of ordinary differential equations (ODEs), we only need μ to be locally-Lipschitz and bounded by linear growth to ensure the existence and uniqueness of the (local) solution. For the stochastic ones, we can prove uniqueness using the same requirements. But to prove existence for the stochastic ones, we need a stronger requirements for μ and σ, that is, in addition to be bounded linear growth, they are globally-Lipschitz. This is because the Brownian motion can progress arbitrarily large depending on the outcome ω, and thus we cannot obtain a constant bound across all ω.

Theorem 3.1 (Uniqueness)**.** *Let μ and σ both be $\mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ functions with the first entry is the time and the second entry is the space. Assume that μ and σ satisfy the local Lipschitz entry, that is*

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K_{T,M}|x-y| \quad \forall t \in [0,T], x, y \in [-M, M] \tag{36}$$

Assume there is a process $(X_t)_{0\leq t\leq T}$ which satisfies (34) with initial condition $X_0 \in L^2(\Omega)$. If there is another $(Y_t)_{0\leq t\leq T}$ also satisfying (34) with the same initial condition, then X_t and Y_t is indistinguishable, that is

$$P(X_t = Y_t \text{ for all } t \in [0,T]) = 1 \tag{37}$$

PROOF. Let us assume there are 2 solutions, X_t and Y_t, that satisfy SDE (34).
Define $\tau_n = \inf\{t \geq 0 : |X_t| \geq n \text{ or } |Y_t| \geq n \text{ or } t \geq T\}$. Clearly $(\tau_n)_{n=0}^{\infty}$ is a localizing sequence for both X_t and Y_t. Also, $t \wedge \tau_n \to t$ a.s because X and Y are continuous a.s, thus are bounded on the time interval $[0,T]$ a.s. Then, by triangle inequality for the 2-moment norm, Itô isometry, Cauchy-Schwarz inequality, and the local Lipschitz property consecutively,

$$\begin{aligned}
\left(\mathbb{E}\left[|X_{t\wedge\tau_n} - Y_{t\wedge\tau_n}|^2\right]\right)^{\frac{1}{2}} &\leq \left(\mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} \mu(s,X_s) - \mu(s,Y_s)ds\right)^2\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} \sigma(s,X_s) - \sigma(s,Y_s)dB_s\right)^2\right]\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left[\left(\int_0^{t\wedge\tau_n} \mu(s,X_s) - \mu(s,Y_s)ds\right)^2\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\int_0^{t\wedge\tau_n} |\sigma(s,X_s) - \sigma(s,Y_s)|^2 ds\right]\right)^{\frac{1}{2}} \\
&\leq t^{\frac{1}{2}}\left(\mathbb{E}\left[\int_0^{t\wedge\tau_n} |\mu(s,X_s) - \mu(s,Y_s)|^2 ds\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\int_0^{t\wedge\tau_n} |\sigma(s,X_s) - \sigma(s,Y_s)|^2 ds\right]\right)^{\frac{1}{2}} \\
&\leq T^{\frac{1}{2}}K_{T,n}\left(\mathbb{E}\left[\int_0^{t} |X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}|^2 ds\right]\right)^{\frac{1}{2}} + K_{T,n}\left(\mathbb{E}\left[\int_0^{t} |X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}|^2 ds\right]\right)^{\frac{1}{2}} \\
&= (T^{\frac{1}{2}}+1)K_{T,n}\left(\int_0^t \mathbb{E}\left[|X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}|^2\right]\right)^{\frac{1}{2}} \\
&= 0 + (T^{\frac{1}{2}}+1)K_{T,n}\left(\int_0^t \mathbb{E}\left[|X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}|^2\right]\right)^{\frac{1}{2}}
\end{aligned} \tag{38}$$

Using Gronwall inequality,

$$\left(\mathbb{E}\left[|X_{t\wedge\tau_n} - Y_{t\wedge\tau_n}|^2\right]\right)^{\frac{1}{2}} \leq (T^{\frac{1}{2}}+1)K_{T,n}\int_0^t 0\cdot\exp\left((T^{\frac{1}{2}}+1)^2K_{T,n}^2(t-s)\right)ds$$
$$= 0$$

Thus, $\left(\mathbb{E}\left[|X_{t\wedge\tau_n} - Y_{t\wedge\tau_n}|^2\right]\right)^{\frac{1}{2}} = 0$, implying for each $t \in [0,T]$, $X_{t\wedge\tau_n} = Y_{t\wedge\tau_n}$ a.s. Then $X_{t\wedge\tau_n} = Y_{t\wedge\tau_n}$ for all $t \in \mathbb{Q}\cap[0,T]$ a.s. By continuity, $X_{t\wedge\tau_n}$ and $Y_{t\wedge\tau_n}$ are indistinguishable. Taking $n \to \infty$, X_t and Y_t are also indistinguishable. □

To ensure existence of solution, we are assuming additional properties satisfied by the drift and diffusion term, that is, they are now both globally Lipschitz and bounded by linear growth. We may the drift locally Lipschitz, however we have to add additional properties to the diffusion as will be described next on Wong-Zakai approximations using Lamperti transformation.

We may adopt the use of Picard's iteration to construct the solution of (34). Let $X_t^{(0)} = X_0 = \xi$ and define an iteration

$$X_t^{(n)} = \xi + \int_0^t \mu(s, X_s^{(n-1)})ds + \int_0^t \sigma(s, X_s^{(n-1)})dB_s \tag{39}$$

It turns out we may bound the expectation of these iterations as follows.

Lemma 3.2. *Let us fix a Brownian motion $(B_t)_{t\geq 0}$. Let μ and σ both be $\mathbb{R}_{\geq 0}\times\mathbb{R}\to\mathbb{R}$ functions with the first entry is the time and the second entry is the space. Assume that μ and σ satisfy these properties:*

1. $|\mu(t,x)-\mu(t,y)|+|\sigma(t,x)-\sigma(t,y)| \leq K|x-y| \quad \forall t\in[0,T], x,y\in\mathbb{R}$ *(globally Lipschitz on the space entry)*
2. $|\mu(t,x)|^2+|\sigma(t,x)|^2 \leq K(1+|x|^2) \quad \forall t\in[0,T], x\in\mathbb{R}$ *(linear growth boundedness)*

Assume that the initial condition $X_0 = \xi$ is of finite 2-moment.Then there exists a constant C depending only on K and T such that the sequence of processes obtained from Picard's iteration in (39) *can be bounded as*

$$\mathbb{E}\left[|X_t^{(n)}|^2\right] \leq C(1+\mathbb{E}\left[|\xi|^2\right])\exp(Ct) \quad \forall t\in[0,T], n\in\mathbb{N} \tag{40}$$

PROOF. We may compute, using triangle inequality for the 2-norm, Cauchy-Schwarz inequality, Itô isometry, the linear growth boundedness, and AM-QM inequality consecutively,

$$\begin{aligned}
\left(\mathbb{E}\left[|X_t^{(n+1)}|^2\right]\right)^{\frac{1}{2}} &\leq \left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\left|\int_0^t \mu(s, X_s^{(n)})\, ds\right|^2\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\left|\int_0^t \sigma(s, X_s^{(n)})\, dB_s\right|^2\right]\right)^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + t^{\frac{1}{2}}\left(\mathbb{E}\left[\int_0^t |\mu(s, X_s^{(n)})|^2\, ds\right]\right)^{\frac{1}{2}} + \left(\mathbb{E}\left[\int_0^t |\sigma(s, X_s^{(n)})|^2\, ds\right]\right)^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + Kt^{\frac{1}{2}}\left\{t + \left(\mathbb{E}\left[\int_0^t |X_s^{(n)}|^2\, ds\right]\right)\right\}^{\frac{1}{2}} + K\left\{t + \left(\mathbb{E}\left[\int_0^t |X_s^{(n)}|^2\, ds\right]\right)\right\}^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + \left(2K^2t\left\{t + \left(\mathbb{E}\left[\int_0^t |X_s^{(n)}|^2\, ds\right]\right)\right\} + 2K^2\left\{t + \left(\mathbb{E}\left[\int_0^t |X_s^{(n)}|^2\, ds\right]\right)\right\}\right)^{\frac{1}{2}} \\
&= \left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + \left(2\left(K^2t + K^2t^2\right) + 2\left(K^2t + K^2\right)\int_0^t \mathbb{E}\left[|X_s^{(n)}|^2\right]\, ds\right)^{\frac{1}{2}} \\
&\leq C_1^{\frac{1}{2}}\left(\mathbb{E}\left[|\xi|^2\right]\right)^{\frac{1}{2}} + C_1^{\frac{1}{2}}\left(1 + \int_0^t \mathbb{E}\left[|X_s^{(n)}|^2\right]\, ds\right)^{\frac{1}{2}}
\end{aligned}$$

where $C_1 = \max\left\{2\left(K^2T + K^2T^2\right), 2\left(K^2T + K^2\right), 1\right\}$. We apply QM-AM inequality once again, getting

$$\left(\mathbb{E}\left[|X_t^{(n+1)}|^2\right]\right)^{\frac{1}{2}} \leq \left(2C_1\mathbb{E}\left[|\xi|^2\right] + 2C_1\left(1 + \int_0^t \mathbb{E}\left[|X_s^{(n)}|^2\right]\, ds_1\right)\right)^{\frac{1}{2}} \tag{41}$$

We now have an iterable inequality. Iterating this inequality backward to $\mathbb{E}\left[|X_s^{(0)}|^2\right] = \mathbb{E}\left[|\xi|^2\right]$, we may obtain

$$\begin{aligned}
\left(\mathbb{E}\left[|X_t^{(n+1)}|^2\right]\right)^{\frac{1}{2}} &\leq \left[\left(2C_1\mathbb{E}\left[|\xi|^2\right] + 2C_1\right)\left(\sum_{j=0}^{n} \frac{2^j C_1^j t^j}{j!}\right) + 2^{n+1}C_1^{n+1}\int_0^t\int_0^{s_1} \dots \int_0^{s_n} \mathbb{E}\left[|\xi|^2\right] ds_{n+1}\dots ds_2 ds_1\right]^{\frac{1}{2}} \\
&\leq \left[\left(2C_1\mathbb{E}\left[|\xi|^2\right] + 2C_1\right)\exp(2C_1t)\right]^{\frac{1}{2}} \quad (\text{because } 2C_1 \geq 2) \\
&= \left[C_2(1 + \mathbb{E}\left[|\xi|^2\right])\exp(C_2t)\right]^{\frac{1}{2}}
\end{aligned}$$

where $C_2 = 2C_1$. □

Now we may begin proving the solution existence.

Theorem 3.3 (Existence)**.** *Let μ and σ both be $\mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ functions with the first entry is the time and the second entry is the space. Assume that μ and σ satisfy these properties:*

1. $|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y| \quad \forall t \in [0, T], x, y \in \mathbb{R}$ *(globally Lipschitz on the space entry)*
2. $|\mu(t, x)|^2 + |\sigma(t, x)|^2 \leq K(1 + |x|^2) \quad \forall t \in [0, T], x \in \mathbb{R}$ *(linear growth boundedness)*

Let us fix a Brownian motion $(B_t)_{t\geq 0}$. Then there is a unique solution $(X_t)_{t\geq 0}$ to the SDE (34) *with initial condition $X_0 = \xi$ of finite 2-moment. Furthermore, X_t has finite 2-moment for any $t \in [0, T]$ and that there exists a constant C, depending only on K and T such that*

$$\mathbb{E}\left[|X_t|^2\right] \leq C(1 + \mathbb{E}\left[|\xi|^2\right])\exp(Ct) \quad \forall t \in [0, T] \tag{42}$$

PROOF. Let us follow a Picard's iteration as we have described before. Fix $t \in [0, t]$.

Now we may denote

$$X_t^{(n+1)} - X_t^{(n)} = D_t^{(n)} + M_t^{(n)} \tag{43}$$

where

$$\begin{aligned} D_t^{(n)} &= \int_0^t \mu(s, X_s^{(n)}) - \mu(s, X_s^{(n-1)})\, ds \\ M_t^{(n)} &= \int_0^t \sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})\, dB_s \end{aligned}$$

From AM-QM inequality,

$$|X_s^{(n+1)} - X_s^{(n)}|^2 \leq 2({D_s^{(n)}}^2 + {M_s^{(n)}}^2) \tag{44}$$

Observe that using the bound on lemma 3.2 and the global Lipschitz condition on μ, we can obtain

$$\begin{aligned} \mathbb{E}\left[\int_0^T |\sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})|^2\, ds\right] &= \int_0^T \mathbb{E}\left[|\sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})|^2\right]\, ds \\ &\leq \int_0^T K\mathbb{E}\left[|X_s^{(n)} - X_s^{(n-1)}|^2\right]\, ds \\ &\leq \int_0^T 4KC_2(1 + \mathbb{E}\left[|\xi|^2\right]) \exp(C_2 t)\, ds \\ &< \infty \end{aligned}$$

Then $\sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})$ is an $\mathcal{H}^2$ process, implying that $(M_s^{(n)})_{0 \leq s \leq T}$ is a martingale. Thus we may use Doob's L^p inequality and Itô isometry consecutively as follows.

$$\begin{aligned} \mathbb{E}\left[\sup_{0 \leq s \leq t} |M_s^{(n)}|^2\right] &\leq 4\mathbb{E}\left[|M_t^{(n)}|^2\right] \\ &= 4\int_0^t \mathbb{E}\left[\left|\sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})\right|^2\right]\, ds \\ &\leq 4K^2 \int_0^t \mathbb{E}\left[|X_s^{(n)} - X_s^{(n-1)}|^2\right]\, ds \end{aligned} \tag{45}$$

Using Cauchy-Schwarz inequality, we may obtain

$$\begin{aligned} |D_s^{(n)}|^2 &\leq K^2 s \int_0^s \left|X_u^{(n)} - X_u^{(n-1)}\right|^2\, du \\ \Rightarrow \sup_{0 \leq s \leq t} |D_s^{(n)}|^2 &\leq K^2 T \int_0^t \left|X_s^{(n)} - X_s^{(n-1)}\right|^2\, ds \end{aligned} \tag{46}$$

Using equation (43), triangle inequality and the inequalities (44), (45), (46),

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le s\le t}\left|X_s^{(n+1)}-X_s^{(n)}\right|^2\right] &\le (16K^2+4K^2T)\int_0^t \mathbb{E}\left[\left|X_s^{(n)}-X_s^{(n-1)}\right|^2\right] ds \\ &\le L\int_0^t \mathbb{E}\left[\sup_{0\le u\le s}\left|X_u^{(n)}-X_u^{(n-1)}\right|^2\right] ds \end{aligned} \tag{47}$$

where $L = 16K^2 + 4K^2T$. Iterating this inequality, we obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le s\le t}\left|X_s^{(n+1)}-X_s^{(n)}\right|^2\right] &\le L\int_0^t \frac{L^{n-1}s^{n-1}}{(n-1)!}\mathbb{E}\left[\left|X_s^{(1)}-X_s^{(0)}\right|^2\right] ds \\ &\le L\int_0^t \frac{L^{n-1}s^{n-1}}{(n-1)!}\sup_{0\le s\le T}\mathbb{E}\left[\left|X_s^{(1)}-X_s^{(0)}\right|^2\right] ds \\ &= \frac{L^n t^n}{n!}\sup_{0\le s\le T}\mathbb{E}\left[\left|X_s^{(1)}-X_s^{(0)}\right|^2\right] \\ &= C^*\frac{L^n t^n}{n!}\infty \end{aligned} \tag{48}$$

where $C^* = \sup_{0\le s\le T}\mathbb{E}\left[\left|X_s^{(1)}-X_s^{(0)}\right|^2\right] < \infty$ because of lemma 3.2.

Using Chebyshev inequality,

$$\mathbb{P}\left(\sup_{0\le s\le T}\left|X_s^{(n+1)}-X_s^{(n)}\right| > \frac{1}{2^{n+1}}\right) \le 4C^*\frac{4^n L^n t^n}{n!} \tag{49}$$

If we sum the right-hand part of inequality (49) across all n, the summation is absolutely convergent. Thus, using Borel-Cantelli lemma, for each ω in a set $\Omega^* \subseteq \Omega$ of probability 1, there exists $N(\omega) \in \mathbb{N}$ such that

$$\begin{aligned}&\sup_{0\le s\le T}\left|X_s^{(n+1)}(\omega)-X_s^{(n)}(\omega)\right| \le \frac{1}{2^{n+1}} \quad \forall n \ge N(\omega) \\ \Rightarrow &\sup_{0\le s\le T}\left|X_s^{(n+k)}(\omega)-X_s^{(n)}(\omega)\right| \le \frac{1}{2^{n}} \quad \forall n \ge N(\omega)\end{aligned} \tag{50}$$

Then $X_s^{(n)}$ converges uniformly for each $s \in [0,T]$ almost surely. Denote the limit as X_s. Because of uniform convergence and the almost-sure continuity of each $X_s^{(n)}$, it should follow that X_s is also continuous almost-surely.

From lemma 3.2 and Fatou's lemma,

$$\begin{aligned}\mathbb{E}\left[|X_t|^2\right] &\le \lim_{n\to\infty}\mathbb{E}\left[|X_t^{(n)}|^2\right] \\ &\le C(1+\mathbb{E}\left[|\xi|^2\right])\exp(Ct)\end{aligned} \tag{51}$$

Then the process $(X_t)_{t\geq 0}$ is also an $\mathcal{H}^2$ process. Thus, by Itô isometry

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t \sigma(s, X_s^{(n)}) - \sigma(s, X_s)\, dB_s\right|^2\right] &\leq K^2 \mathbb{E}\left[\left|\int_0^t |X_s - X_s^{(n)}|\, dB_s\right|^2\right] \\ &= K^2 \int_0^t \mathbb{E}\left[|X_s - X_s^{(n)}|^2\right]\, ds \end{aligned} \tag{52}$$

Because $X_s^{(n)}$ converges uniformly to X_s on $[0,T]$ almost surely and the bound on $\mathbb{E}\left[|X_s^{(n)}|^2\right]$ and $\mathbb{E}\left[|X_s|^2\right]$ in lemma 3.2 and inequality (51), by dominated convergence theorem, the right-hand side of (52) converges to 0.

Thus,

$$\int_0^t \sigma(s, X_s^{(n)})\, dB_s \xrightarrow{L^2(\Omega)} \int_0^t \sigma(s, X_s) dB_s \tag{53}$$

Similarly,

$$\int_0^t \mu(s, X_s^{(n)})\, ds \xrightarrow{L^2(\Omega)} \int_0^t \mu(s, X_s)\, ds \tag{54}$$

Because of this $L^2(\Omega)$ convergence, there is a subsequence $(X_t^{(n_k)})_{k=1}^{\infty}$ such that the convergence in (53) and (54) becomes almost-sure convergence. But because $X_t^{(n)}$ converges almost surely to X_t, it should follow that

$$X_t = X_0 + \int_0^t \mu(s, X_s^{(n)})\, ds + \int_0^t \sigma(s, X_s^{(n)})\, dB_s \quad a.s. \tag{55}$$

Because $\mathbb{Q}$ is countable, there exists a set $\Omega_1 \in \mathcal{F}$ of probability 1 such that

$$X_t = X_0 + \int_0^t \mu(s, X_s^{(n)})\, ds + \int_0^t \sigma(s, X_s^{(n)})\, dB_s \quad \forall t \in [0,T] \cap \mathbb{Q}\, a.s. \tag{56}$$

Because integration (ordinary Lebesgue or Itô) is continuous, equation (56) should also be satisfied for all $t \in [0,T]$ almost surely.

□

4 Some Methods of Solving SDEs/Solution Construction

4.1 Doss-Sussman Method

Now let us consider a class of one-dimensional stochastic differential equations of the form

$$X_t = \xi + \int_0^t \left(\mu(X_s) + \frac{1}{2}\sigma(X_s)\sigma'(X_s)\right)\, ds + \int_0^t \sigma(X_s)\, dB_s \tag{57}$$

Now, what is so special about this form of SDE? It turns out that we may "turn" this SDE into a deterministic ordinary differential equations. By turning SDE into ODE, we may obtain the solution X_t in a form of deterministic function with random inputs (note that generally speaking, the solution X_t is a *random function*, not just a deterministic function with random inputs). From Mean Value Theorem,

Computationally, when we are executing a (computational) methods of solving for X_t, we need to be sure that the method works consistently for *all* outcome ω (almost surely at least), not just the outcome we observed. But if X_t is in a form of deterministic function with random inputs, we do not need to simulate all possible outcome ω. We can be sure that indeed our computations are consistent.

Now, we assume some nice properties for the initial condition ξ, the diffusion σ, and μ.

Assumption 4.1. (i) The initial condition ξ is square-integrable i.e. a $L^2(\Omega, \mathcal{F}, \mathbb{P})$ random variable.

(ii) μ is globally Lipschitz-continuous, i.e. there exists a constant L such that

$$|\mu(x) - \mu(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R} \tag{58}$$

(iii) σ has bounded first and second derivatives, i.e. there exists a constant A such that

$$|\sigma'(x)| \leq A \text{ and } |\sigma''(x)| \leq a \quad \forall x \in \mathbb{R} \tag{59}$$

Because σ has bounded first derivatives, it is also globally Lipschitz-continuous, so from the theory of ODEs, there exists (uniquely and globally) a function $u(x, y)$ which solves the ODE with initial condition

$$\frac{\partial u}{\partial x} = \sigma(u), \quad u(0, y) = y \tag{60}$$

Then we have

$$\frac{\partial^2 u}{\partial x^2} = \sigma(u)\sigma'(u), \quad \frac{\partial^2 u}{\partial x \partial y} = \sigma'(u)\frac{\partial u}{\partial y}, \quad \frac{\partial}{\partial y}u(0, y) = 1 \tag{61}$$

Using standard calculus, we may obtain

$$\frac{\partial}{\partial y}u(x, y) = \exp\left(\int_0^x \sigma'(u(z, y))dz\right) := \frac{1}{\rho(x, y)} \tag{62}$$

Now, from (58), we may obtain $\exp(-A|x|) \leq \rho(x, y) \leq \exp(A|x|)$ for all x and y, implying $\frac{\partial}{\partial y}u(x, y)$ is bounded by the same lower and upper bound.

$$|u(x, y_1) - u(x, y_2)| = \left|\frac{\partial}{\partial y}u(x, y^*)\right| |y_1 - y_2| \leq \exp(A|x|)||y_1 - y_2| \tag{63}$$

Thus, for fixed x, we have $\mu(u(x, y))$ is Lipschitz-continuous in y because

$$|\mu(u(x, y_1)) - \mu(u(x, y_2)) \leq L\exp(A|x|)||y_1 - y_2| \tag{64}$$

Furthermore, from (64). $|\mu(u(x, y))| \leq L\exp(A|x|)||y| + |\mu(u(x, 0))|$, implying $\mu(u(x, y))$ has linear growth in y for fixed x.

Then, using the inequality $|\exp(z_1) - \exp(z_2)| \leq \max\{\exp(z_1), \exp(z_2)\}\,|z_1 - z_2|$, we have

$$
\begin{aligned}
|\rho(x, y_1) - \rho(x, y_2)| &\leq \max\{\rho(x, y_1), \rho(x, y_2)\} \int_0^{|x|} \left|\sigma'(u(z, y_1)) - \sigma'(u(z, y_2))\right| dz \\
&\leq \exp(A|x|) \int_0^{|x|} A\,|u(z, y_1) - u(z, y_2)|\, dz \\
&\leq \max\{\rho(x, y_1), \rho(x, y_2)\} \int_0^{|x|} A \exp(A|z|)||y_1 - y_2|\, dz \\
&\leq exp(A|x|) \int_0^{|x|} A \exp(A|x|)|y_1 - y_2|\, dz \\
&= A|x| \exp(2A|x|)|y_1 - y_2| \qquad (65)
\end{aligned}
$$

implying that for fixed x, $\rho(x, y)$ is Lipschitz-continuous in y.

Define $f(x, y) := \rho(x, y)\mu(u(x, y))$. Then for $x, y_1, y_2 \in [-K, K]$, using the linear growth of μ,

$$
\begin{aligned}
|f(x, y_1) - f(x, y_2)| &= |\rho(x, y_1)\mu(u(x, y_1)) - \rho(x, y_2)\mu(u(x, y_2))| \\
&\leq |\rho(x, y_1)\mu(u(x, y_1)) - \rho(x, y_1)\mu(u(x, y_2))| + |\rho(x, y_1)\mu(u(x, y_2)) - \rho(x, y_2)\mu(u(x, y_2))| \\
&\leq \exp(A|x|)\,|\mu(u(x, y_1)) - \mu(u(x, y_2))| + \Big[L \exp(A|x|)||y_2| + |\mu(u(x, 0))|\Big]|\rho(x, y_1) - \rho(x, y_2)| \qquad (66)
\end{aligned}
$$

Because u has continuous partial derivatives with respect to both x and y, $u(x, 0)$ is bounded for $x \in [-K, K]$, implying that $\mu(u(x, y))$ is also bounded on $[-K, K]$. Then using (64) and (65), we may find a constant L_K such that

$$
|f(x, y_1) - f(x, y_2)| \leq L_k|y_1 - y_2| \quad \forall x, y_1, y_2 \in [-K, K] \qquad (67)
$$

Furthermore, from the linear growth of $\mu(u(x, y))$ and the boundedness of $\rho(x, y)$,

$$
\begin{aligned}
|f(x, y)| &\leq \exp(A|x|)\Big[L \exp(A|x|)||y| + |\mu(u(x, 0))|\Big] \\
&\leq K_{1k} + K_{2k}|y| \quad \forall |x| \leq k, y \in \mathbb{R} \qquad (68)
\end{aligned}
$$

Because of (67) and (68), for every $y_0 \in \mathbb{R}$ and continuous function $x : \mathbb{R}_{\geq 0} \to \mathbb{R}$, there exists a unique solution $Y^*(x(t), y_0, t)$ to the ordinary differential equation (in the integral form) such that

$$
Y^*(x(t), y_0, t) = y_0 + \int_0^t f(x(s), Y^*(x(s), y_0, s))\, ds \qquad (69)
$$

For every $\omega \in \Omega$, fix $Y_t(\omega) = Y^*(B_t(\omega), \xi(\omega), t)$. Because B_t is continuous almost surely, then

$$
Y_t(\omega) = \xi(\omega) + \int_0^t f(B_s(\omega), Y_s(\omega))\, ds \qquad (70)
$$

Because f and Y_t is continuous (the latter almost surely), it follows that for any $T > 0$,

$$\int_0^T |f(B_s(\omega), Y_s(\omega))| \, ds < \infty \quad \text{a.s.}$$

Thus, the process $(Y_t)_{t\geq 0}$ has 0 quadratic variation or $\langle Y \rangle_T = 0$ for any $T > 0$. Furthermore,

$$B_t \pm Y_t = B_0 \pm \xi + \int_0^t dB_s \pm \int_0^t f(B_s(\omega), Y_s(\omega)) \, ds \tag{71}$$

so $\langle B \pm Y \rangle = \langle B \rangle_T = T$. This implies the quadratic covariaton of W and Y is also 0 because

$$\begin{aligned} \langle B, Y \rangle_T &= \frac{1}{4}\Big[\langle B + Y \rangle_T - \langle B - Y \rangle_T \Big] \\ &= \frac{1}{4}\Big[T - T \Big] \\ &= 0 \end{aligned} \tag{72}$$

Define $X_t(\omega) = u(B_t(\omega), Y_t(\omega))$. Obviously $X_0(\omega) = u(0, \xi(\omega)) = \xi(\omega)$. Then, from Itô's rule, for every $t > 0$,

$$\begin{aligned} X_t(\omega) &= u(B_0(\omega), Y_0(\omega)) + \int_0^t u_x(B_s(\omega), Y_s(\omega)) \, dB_s + \int_0^t u_y(B_s(\omega), Y_s(\omega)) \, dY_s \\ &\quad + \frac{1}{2}\int_0^t u_{xx}(B_s(\omega), Y_s(\omega)) \, d\langle B \rangle_s + \int_0^t u_{xy}(B_s(\omega), Y_s(\omega)) \, d\langle W, Y \rangle_s \\ &\quad + \frac{1}{2}\int_0^t u_{yy}(B_s(\omega), Y_s(\omega)) \, d\langle Y \rangle_s \\ &= \xi(\omega) + \int_0^t \sigma\Big(u(B_s(\omega), Y_s(\omega))\Big) \, dB_s + \int_0^t \frac{1}{\rho\Big(B_s(\omega), Y_s(\omega)\Big)} f(B_s(\omega), Y_s(\omega)) \, ds \\ &\quad + \frac{1}{2}\int_0^t \sigma\Big(u(B_s(\omega), Y_s(\omega))\Big)\sigma'\Big(u(B_s(\omega), Y_s(\omega))\Big) \, ds + 0 \\ &\quad + 0 \\ &= \xi(\omega) + \int_0^t \sigma(X_s(\omega)) \, dB_s + \int_0^t \mu(X_s) \, ds + \frac{1}{2}\int_0^t \sigma(X_s(\omega))\sigma'(X_s(\omega)) \, ds \quad \text{a.s.} \end{aligned} \tag{73}$$

Because of the countability and density of $\mathbb{Q}$, in addition to the fact that integration is a continuous operator, the equation (73) holds for any $t > 0$ almost surely. The process X which satisfies this equation is unique due to the uniqueness of ODE solution in . If there is another process Y satisfying (73), then X and Y is indistinguishable.

This result, first developed by Doss and Sussman independently, can be stated as follows.

Theorem 4.1 (Doss-Sussman)**.** *Let ξ, μ, and σ satisfy the assumption 4.1. Then the one-dimensional stochastic differential equation*

$$X_t = \xi + \int_0^t \mu(X_s) + \frac{1}{2}\sigma(X_s)\sigma'(X_s) \, ds + \int_0^t \sigma(X_s(\omega)) \, dB_s \tag{74}$$

has a unique solution X_t. Furthermore, X_t can be written as

$$X_t(\omega) = u(B_t(\omega), Y_t(\omega)) \tag{75}$$

for a suitable continuous deterministic function $u : \mathbb{R}^2 \to \mathbb{R}$ and a process Y which solves an ordinary differential equation, for all ω almost surely.

Example 4.2. Let us consider the SDE

$$dX_t = \left(\sqrt{1+X_t^2} + \frac{1}{2}X_t\right) dt + \sqrt{1+X_t^2}\, dB_t \tag{76}$$

Now we may fit this equation into the Doss-Sussman type with $\mu(x) = \sigma(x) = \sqrt{1+x^2}$. After some calculations, we may see these functions satisfy the assumption 4.1 and that

$$\sigma'(x)\sigma(x) = x$$

Now, the unique solution to the ODE

$$\frac{\partial u}{\partial x} = \sqrt{1+u^2} \quad u(0,y) = y \tag{77}$$

is $u(x,y) = \sinh(x + \sinh^{-1} y)$. Thus $\sigma'(u(z,y)) = \frac{u(z,y)}{\sqrt{1+u^2(z,y)}} = \frac{\sinh(z+\sinh^{-1} y)}{\cosh(z+\sinh^{-1} y)}$ and

$$\begin{aligned} \rho(x,y) &= \exp\left(-\int_0^x \frac{\sinh(z+\sinh^{-1} y)}{\cosh(z+\sinh^{-1} y)}\, dz\right) \\ &= \exp\left(-\log\left(\cosh(z+\sinh^{-1} y)\right)\Big|_0^x\right) \\ &= \frac{\cosh(\sinh^{-1} y)}{\cosh(x+\sinh^{-1} y)} \\ &= \frac{\sqrt{1+y^2}}{\cosh(x+\sinh^{-1} y)} \end{aligned} \tag{78}$$

Next, we obtain f

$$\begin{aligned} f(x,y) &= \rho(x,y)\mu(u(x,y)) \\ &= \frac{\sqrt{1+y^2}}{\cosh(x+\sinh^{-1} y)}\sqrt{1+\sinh^2(x+\sinh^{-1} y)} \\ &= \sqrt{1+y^2} \end{aligned} \tag{79}$$

Then, we solve for Y^* as the unique solution to the following ODE.

$$\begin{aligned} &\frac{d}{dt}Y^*(x,y_0,t) = \sqrt{1+(Y^*)^2(x,y_0,t)}, \quad Y^*(x,y_0,0) = y_0 \\ &\Rightarrow Y^*(x,y_0,t) = \sinh\left(t+\sinh^{-1} y_0\right) \end{aligned} \tag{80}$$

Finally, we fix $Y_t(\omega) = \sinh\left(t + \sinh^{-1} X_0(\omega)\right)$ and

$$\begin{aligned} X_t(\omega) &= \sinh\left(B_t(\omega) + \sinh^{-1}(Y_t(\omega))\right) \\ &= \sinh\left(B_t(\omega) + \sinh^{-1}(\sinh\left(t+\sinh^{-1} X_0(\omega)\right))\right) \\ &= \sinh\left(B_t(\omega) + t + \sinh^{-1} X_0(\omega))\right) \end{aligned} \tag{81}$$

Using Itô's rule, we may check that this is indeed the solution to the equation (76).

4.2 Lamperti Transformation Method for Regime-Switching SDE

The Doss-Sussman method we described may only apply to SDE of the form (57), so we are interested to see other method to solve other types of SDE.

Other form of SDE is the *regime-switching* type where the drift and diffusion may change according to some event. Let $(J_t)_{t\geq 0}$ a right-continuous jump process where all realization of each J_t take values from a common finite set $\mathcal{E}$. Furthermore, we also assume that the number of jumps on any compact interval is finite almost surely.

Now, consider a one-dimensional equation

$$dX_t = \mu_{J_t}(t, X_t)\, dt + \sigma_{J_t}(t, X_t)\, dB_t, \quad X_0 = x_0 \in \mathbb{R} \tag{82}$$

First, we need several assumptions to guarantee the existence of unique solution to (82).

Assumption 4.3.

(i) For each $\imath \in \mathcal{E}$, μ_i

  (a) is locally-Lipschitz continuous with respect to both entries

  (b) is bounded by global linear growth with respect to both entries

(ii) For each $\imath \in \mathcal{E}$, σ_i

  (a) is continuously-differentiable with respect to both entries

  (b) is locally-Lipschitz continuous with respect to both entries

  (c) is globally-Lipschitz continuous with respect to the second (or space) entry

  (d) has locally-Lipschitz continuous (w.r.t both entries) partial derivatives

  (e) is positive, bounded, and does not vanish i.e. the infimum is not 0.

  (f) has a partial derivative (w.r.t first/time entry) bounded by quadratic growth i.e. there is a constant K such that

$$\partial_1[\sigma_i]/\sigma_i^2 \leq K$$

Before constructiong the solution, (Nguyen & Peralta, 2021) first assume there exists a unique solution to (82). From this, the SDE (82) is transformed, using Lamperti transformation, into an SDE with only unit diffusion and show that this new SDE also has a unique solution. By the injective nature of Lamperti transform, the reversed direction is obtained if the new SDE also has a unique solution.

Theorem 4.2. *Suppose that μ_i and σ_i satisfy assumption 4.3. Assume there exists a unique solution to SDE* (82)*. Define the Lamperti transformation of σ_i as*

$$h_i(t, x) = \int_{x_0}^{x} \frac{1}{\sigma_i(t, y)}\, dy \tag{83}$$

Let $L_t = h_{J_t}(t, X_t)$. Then the process $(L_t)_{t\geq 0}$ satisfies the regime-switching SDE with jumps and unit diffusion

$$dL_t = \mu^*_{J_t}(t, L_t)\, dt + dB_t + \sum_{s\leq t;\, J_s \neq J_{s^-}} \left(h_{J_s}\left(s, h^{-1}_{J_{s^-}}(s, L_{s^-})\right) - L_{s^-} \right) \tag{84}$$

where h_i^{-1} is the inverse of $h_i : \mathbb{R} \to \mathbb{R}$ and

$$\mu_i^*(t,\ell) = \partial_1\big[h_i\big(t, h_i^{-1}(t,\ell)\big)\big] + \frac{\mu_i\big(t, h_i^{-1}(t,\ell)\big)}{\sigma_i\big(t, h_i^{-1}(t,\ell)\big)} - \frac{1}{2}\partial_2\big[h_i\big(t, h_i^{-1}(t,\ell)\big)\big] \tag{85}$$

Note: ∂_i denotes the partial derivatives with respect to the i-th entry.

PROOF. The proof is quite straightforward. By assumption 4.3, there exists positive constants v and V such that $\frac{1}{V} \leq \frac{1}{\sigma_i(t,y)} \leq \frac{1}{v}$, making sure that $\frac{1}{\sigma_i(t,y)}$ is also continuously-differentiable with respect to both entries. Thus, the Lamperti transformation is strictly monotonically increasing, once continuously-differentiable in the first (time) entry, and twice continuously-differentiable in the second (space) entry.

We may compute that

$$\begin{aligned} \partial_2\big[h_{J_t}(t,X_t)\big]\,dX_t &= \frac{1}{\sigma_{J_t}(t,X_t)}\Big(\mu_{J_t}(t,X_t)\,dt + \sigma_{J_t}(t,X_t)\,dB_t\Big) \\ &= \frac{\mu_{J_t}(t,X_t)}{\sigma_{J_t}(t,X_t)}\,dt + dB_t \end{aligned} \tag{86}$$

Applying Itô's rule for regime-switching processes, we obtain $(L_t)_{t\geq 0}$ satisfies (84). The diffusion coefficient is transformed into unity because of (86). □

For the other way around, if we find a unique solution to the transformed equation in (84), we may guarantee the unique solution to the original equation (82) is $X_t = h_{J_t}^{-1}(t, L_t)$.

Now we turn to construct the solution to (84) using, once again, the theory of ODEs. First, let us consider the following lemma.

Lemma 4.3. *Under assumption 4.3, the constructed function μ_i^* in (85) is locally-Lipschitz continuous and bounded by global linear growth, with respect to both entries.*

PROOF. The proof may be found in (Nguyen & Peralta, 2021). □

Now, let us construct

$$Y_{i,b,r}(t) = b + \int_0^t \mu_i^*(r+u, Y_{i,b,r}(u) + B_{r+u} - B_r)\,du \tag{87}$$

For any fixed $\omega \in \Omega$, $i \in \mathcal{E}$, $b \in \mathbb{R}$, and $r \geq 0$, there exists a unique solution $Y_{i,b,r}(t)$ satisfying (87) because μ_i^* satisfies lemma 4.3, and the equation (87) is really an ODE.

Next, for each ω, let t_1, t_2, ..., t_n be the time-jumps on $[0,t]$, and $t_0 = 0$. Then we define the process $(S_{t'})_{0\leq t'\leq t}$ recursively as follows:

$$S_0 = x_0 \tag{88}$$

$$S_{t'} = Y_{J_{t_k}, S_{t_k}, t_k}(t' - t_k) + B_{t'} - B_{t_k} \quad \text{for } t' \in (t_k, t_{k+1}) \tag{89}$$

$$S_{t_k} = h_{J_{t_k}}\Big(t_k, h_{J_{t_{k-1}}}^{-1}(t_k, S_{t_k^{-1}})\Big) \quad \text{for } k = 1, 2, ..., n \tag{90}$$

This process is continuous except on t_1,t_2,...,t_n. Furthermore, from (89), the process is right-continuous on t_1,t_2,...,t_n because the Brownian is continuous. Thus, $(S_{t'})_{0\leq t'\leq t}$ is right-continuous.

Moreover, for $t' \in (t_k, t_{k+1})$,

$$\begin{aligned} S_{t'} - S_{t_n} &= Y_{J_{t_k}, S_{t_k}, t_k}(t' - t_k) + B_{t'} - B_{t_k} - S_{t_n} \\ &= \int_0^{t'-t_k} \mu^*_{J_{t_k}}(t_k + u, Y_{J_{t_k}, S_{t_k}, t_k}(u) + B_{t_k+u} - B_{t_k})\, du + B_{t'} - B_{t_k} \\ &= \int_{t_k}^{t'} \mu^*_{J_{t_k}}(u, Y_{J_{t_k}, S_{t_k}, t_k}(u - t_k) + B_u - B_{t_k})\, du + +B_{t'} - B_{t_k} \\ &= \int_{t_k}^{t'} \mu^*_{J_{t_k}}(u, S_u)\, du + +B_{t'} - B_{t_k} \end{aligned} \tag{91}$$

On the other hand,

$$S_{t_k} - S_{t_k^{-1}} = h_{J_{t_k}}\left(t_k, h^{-1}_{J_{t_{k-1}}}(t_k, S_{t_k^{-1}})\right) - S_{t_k^{-1}} \tag{92}$$

and we may compute $S_{t_k^{-1}} - S_{t_{k-1}}$ using (91). Thus, we may calculate $S_{t'} - S_0$ recursively backwards and obtain

$$S_{t'} = S_0 + \int_0^{t'} \mu^*_{J_u}(u, S_u)\, du + B_{t'} + \sum_{u\leq t'; J_u \neq J_{u^-}} \left(h_{J_u}\big(u, h^{-1}_{J_{u^-}}(u, S_{u^-})\big) - S_{u^-}\right) \tag{93}$$

Thus we have shown that $(S_{t'})_{0\leq t'\leq t}$ is the solution to (84). This solution is also unique due to the fact that the constructed $Y_{J_{t_k}, S_{t_k}, t_k}$ is unique and that $S_{t_{k^-}}$ is completely determined by $Y_{J_{t_k}, S_{t_k}, t_k}$ and the jump process $(J_{t'})_{0\leq t'\leq t}$. Because t is an arbitrary positive real number, then we may extend this result to construct $(S_t)_{t\geq 0}$ and also conclude that this process is also the unique solution to (84).

Now we have this theorem.

Theorem 4.4. *The process $(S_t)_{t\geq 0}$ is the unique solution to the regime-switching SDE with jumps given by* (84)*.*

Note that we may also use this method to solve the solution of non-regime-switching SDE as long as the drift and diffusion term satisfy assumption 4.3.

5 Wong-Zakai Approximations of SDEs solution

5.1 Wong-Zakai Approximations of Non-Regime-Switching SDEs

Now, let us assume that we want to model a trajectory/movement in real life (e.g. stock prices, disease spread, birds flight pattern) where it is influenced by deterministic component plus randomness i.e. stochastic component.

$$dX^*_t = \mu(t, X^*_t)\, d_t + \sigma(t, X^*_t)\, dF^*_t \tag{94}$$

Assume that the drift and the diffusion term are suitable enough. We may never know the exact nature of the random, stochastic component F_t^*, but from observation, we perceive that it may well be assume as Brownian. Thus, we may model

$$dX_t = \mu(X_t)\, d_t + \sigma(X_t)\, dB_t \tag{95}$$

We may be able to solve this equation analytically and hope that our solution is close enough to the actual real-life observation. Unfortunately, even though we perceive the original randomness as Brownian, real world random paths are rarely true Brownian motion. Furthermore, the random path is likely of finite-variation, which is not a characteristic of Brownian path.

On the other hand, we may wish to go the other way around. Suppose that we are trying to solve an SDE where the randomness source is Brownian but it is impossible to solve analytically. Then we may employ a numerical method. Unfortunately, there is no way to exactly simulate a truly Brownian motion on any machine. The feasible thing to do is to approximate the Brownian by some finite-variation process e.g. sampling finite points from the Brownian, then linearly interpolating the in-betweens. Of course as in any other numerical methods, we wish the computations do indeed approximate the actual solution well enough.

It turns out that a better model includes a correction term seen in Doss-Sussman result.

$$dX_t = \mu(X_t)\, dt + \sigma(X_t(\omega))\, dB_t + +\frac{1}{2}\sigma(X_t)\sigma'(X_t)\, dt \tag{96}$$

This correction term (the last integral) is also known as the Wong-Zakai approximation term. .

But before we proceed, we first turn to SDE where the source of randomness is of finite-variation. We have to be sure that such SDE has a unique solution.

Lemma 5.1. *Let $(V_t)_{t\geq 0}$ is a continuous and finite-variation process on every interval $[0,T], T < \infty$ $(\ell[0,T]\times\mathbb{P})$-a.e. ($\ell[0,T]$ is the Lebesgue measure on $[0,T]$) and $V_0 = 0$. Assume that the initial condition ξ is of finite 2-moment. If μ, σ are Lipschitz continuous, then there exists a unique solution to the SDE*

$$X_t^* = \xi + \int_0^t \mu(X_s^*)\, ds + \int_0^t \sigma(X_s^*)\, dV_s \tag{97}$$

where the last integral is interpreted as the Stiltjes type.

PROOF. Following Picard iteration again, define

$$X_t^{(n+1)*} = \xi + \int_0^t \mu(X_s^{(n)*})\, ds + \int_0^t \sigma(X_s^{(n)*})\, dV_s \tag{98}$$

Define also $D_t^{(n+1)*} := \sup\limits_{0\leq s\leq t} |X_s^{(n+1)*} - X_s^{(n)*}|$. Then we have

$$\begin{aligned} D_t^{(n+1)*} &= \sup_{0\leq u\leq t} \left| \int_0^u \mu(X_s^{(n)*}) - \mu(X_s^{(n-1)*})\, ds + \int_0^u \sigma(X_s^{(n)*}) - \sigma(X_s^{(n-1)*})\, dV_s \right| \\ &\leq L \left[\int_0^t D_s^{(n)*}\, ds + \int_0^t D_s^{(n)*}\, dV_s \right] \\ &=\leq L \int_0^t D_s^{(n)*}\, (ds + dV_s) \end{aligned} \tag{99}$$

where L is the Lipschitz constant for μ and σ. Iterating this inequality, we can obtain

$$D_t^{(n+1)*} \leq D_t^{(1)*} L^n \frac{(t+V_t)^n}{n!} \tag{100}$$

For fixed $t \in [0, \infty)$, the right-hand side of (100) is summable across n. If we bound s such that $s \in [0, t]$, $s + V_s$ attains a maximum on the closed interval because of the continuity of $(V_s)_{s \geq 0}$. Because of this and the way $D_t^{(n+1)*}$ is defined, the process $X^{(n)*}$ converges uniformly on the interval $[0, t]$ to a continuous process X^*. Letting $n \to \infty$ in (98) and using the uniform convergence of $X^{(n)*}$ on $[0, t]$ and the fact that μ and σ are Lipschitz, we conclude that X^* satisfies (97). If Y^* is another solution to (97), then by defining $D_t^* = \sup_{0 \leq s \leq t} |X_t^* - Y_t^*|$, we may compute by iteration

$$D_t^* \leq D_t^* \frac{L^n (t+V_t)^n}{n!} \quad \forall n \in \mathbb{N} \tag{101}$$

implying that $D_t^* = 0$. □

Now that we have established the existence and uniqueness to the SDE where the source of randomness/diffusion term is from a continuous process with finite variation. We will now see how the solution approximates the solution to (96) when the continuous & finite-variation process converges to a Brownian motion, in the appropriate strong sense.

Theorem 5.2. *Suppose that ξ, μ, and σ satisfy assumption 4.1. Let $\left((V_t^{(n)})_{t \geq 0}\right)_{n=1}^{\infty}$ be sequence of almost surely continuous process with finite variation and $(B_t^{(n)})_{t \geq 0}$ be a Brownian motion where $V_t^{(n)}$ converges strongly (uniformly and almost surely) to $B_t^{(n)}$ on every compact time-interval, that is*

$$\lim_{n \to \infty} \sup_{0 \leq s \leq T} |V_s^{(n)} - B_s| = 0 \quad a.\ s. \quad \forall\, 0 \leq T < \infty \tag{102}$$

Then the solutions to (97) where $V_s := V_s^{(n)}$ also converge strongly (uniformly and almost surely) to the solution of (96) on every compact time-interval.

PROOF. Let u and f be as in the construction of Doss-Sussman method and

$$Y_t^{(n)}(\omega) = Y^*(V_t^{(n)}(\omega), \xi(\omega), t) \tag{103}$$

$$X_t^{(n)}(\omega) = u(V_t^{(n)}, Y_t^{(n)}) \tag{104}$$

Using similar argument as to the construction of Doss-Sussman method, $\left(X_t^{(n)}\right)_{0 \leq t < \infty}$ is the unique solution to (97) where $V_s := V_s^{(n)}$ (in this case, there is no correction term involving $\frac{1}{2}\sigma'(X_s^{(n)})\sigma(X_s^{(n)})$ because $\langle V \rangle_s = 0$).
Fix $\omega in \Omega$, $0 \leq t < \infty$ and a positive integer k. With L_k as the Lipschitz constant to f with respect to the second-entry as in (67). Now choose ε small enough such that $\varepsilon < e^{-L_k t} \wedge 1$.
Define the stopping times

$$\tau_k(\omega) = t \wedge \inf\{0 \leq s \leq t :\ |Y_s(\omega)| \geq k-1 \text{ or } |B_s(\omega)| \geq k-1\} \tag{105}$$

$$\tau_k^{(n)}(\omega) = t \wedge \inf\left\{0 \leq s \leq t :\ |Y_s^{(n)}(\omega)| \geq k\right\} \tag{106}$$

Because $\left((V_t^{(n)})_{t \geq 0}\right)_{n=1}^{\infty}$ converge strongly to $(B_t)_{0 \leq t < \infty}$ on every compact time-interval, we may choose n sufficiently large (depending on ω) so that $V_s^{(n)}(\omega)$ is near enough to $B_s(\omega)$ so that $|f(V_s^{(n)}, Y_s(\omega)) -$

$f(B_s, Y_s(\omega))| \leq \varepsilon^2$ and $|V_s^{(n)}| \leq k$ hold for every $s \in [0, \tau_k(\omega) \wedge \tau_k^{(n)}(\omega)]$. Then

$$\begin{aligned}\left|\frac{d}{ds}(Y_s^{(n)}(\omega) - Y_s(\omega))\right| &\leq |f(V_s^{(n)}(\omega), Y_s^{(n)}(\omega)) - f(V_s^{(n)}(\omega), Y_s(\omega))| \\ &\quad + |f(V_s^{(n)}(\omega), Y_s(\omega)) - f(V_s(\omega), Y_s(\omega))| \\ &\leq L_k|Y_s^{(n)}(\omega) - Y_s(\omega)| + \varepsilon^2 \end{aligned} \tag{107}$$

Using Gronwall's inequality, we obtain

$$|Y_s^{(n)}(\omega) - Y_s(\omega)| \leq \varepsilon^2 e^{L_k t} < \varepsilon \quad \forall\, s \in [0, \tau_k(\omega) \wedge \tau_k^{(n)}(\omega)] \tag{108}$$

If $\tau_k^{(n)}(\omega) < \tau_k(\omega)$, we have an open interval in $[0, t]$ such that $|Y_s(\omega)| < k - 1$ but $|Y_s^{(n)}(\omega)| \geq k$. This means that $|Y_s^{(n)}(\omega) - Y_s(\omega)| \geq 1$ for $s \in (\tau_k^{(n)}(\omega), \tau_k(\omega))$. But we have $|Y_s^{(n)}(\omega) - Y_s(\omega)| < \varepsilon$ for $s \in [0, (\tau_k^{(n)}(\omega)]$. It is impossible for the difference to suddenly jump from less than ε at $\tau_k^{(n)}(\omega)$ to something greater than 1 immediately because both processes Y and $Y^{(n)}$ are continuous. Thus, it should be certain that $\tau_k(\omega) \leq \tau_k^{(n)}(\omega)$. We have

$$\lim_{n\to\infty} \sup_{0\leq s\leq \tau_k(\omega)} |Y_s^{(n)}(\omega) - Y_s(\omega)| = 0 \tag{109}$$

We may choose k large enough such that $\tau_k(\omega) = t$. Then

$$\lim_{n\to\infty} \sup_{0\leq s\leq t} |Y_s^{(n)}(\omega) - Y_s(\omega)| = 0 \tag{110}$$

The above equation holds for any $\omega \in \Omega$ a.s. □

5.2 Wong-Zakai Approximations of Regime-Switching SDEs

We have discussed a method using Lamperti transformation to solve regime-switching SDEs under certain regularities in assumption 4.3. Next, as we have described in approximations of non-regime switching SDE in (82), we might be interested to approximate this solution and obtain the rate of convergence i.e. how good the approximation is.

Consider a collection of finite-variation processes $\big((V_t^\lambda)\big)_{\lambda\geq 0}$ which converges to a Brownian motion $(B_t)_{t\geq 0}$ in some strong enough sense. In the non-regime-switching case, the limit of the approximations include a correction term. We may first correct for this term in the approximation so that the limit does not include it anymore as follows:

$$dX_t^\lambda = \left[\mu_{J_t}(t, X_t^\lambda) - \frac{1}{2}\sigma'_{J_t}(t, X_t^\lambda)\sigma_{J_t}(t, X_t^\lambda)\right] dt + \sigma_{J_t}(t, X_t^\lambda)\, dV_t^\lambda, \quad X_0 = x_0 \in \mathbb{R} \tag{111}$$

$$dX_t = \mu_{J_t}(t, X_t)\, dt + \sigma_{J_t}(t, X_t)\, dB_t, \quad X_0 = x_0 \tag{112}$$

We have constructed the solution to (112) by basically "sewing" the individual solutions on the interval between two jumps. We may also adopt this method to solve the approximation (111). This might lead a not-small-enough errors at exactly the time-jumps. These not-small-enough errors might accumulate over and over.

(Nguyen & Peralta, 2021) treated this problem by making additional restrictions or assumptions as follows.

Assumption 5.1. For each $i \in \mathcal{E}$, the function μ_i^* is Lipschitz continuous in the first (time) entry.

By this assumption, they showed that the errors increase geometrically each time a jump occurs in in $(J_t)_{t\geq 0}$. The problem is, even though it is assumed that there are only finite number of jumps on every compact time-interval, the number might be arbitrarily large. (Nguyen & Peralta, 2021) treated this by assuming for every fixed compact time-interval, the number of jumps has "light" enough tail. Without loss of generality, we may restrict ourselves to the time-interval $[0,1)$. Let N be the number of jumps on $[0,1)$ i.e.

$$N = \#\left\{s \in [0,1) : \, J_{s^-} \neq J_s\right\}$$

Assumption 5.2. There exists a constant $\gamma_0 >$ such that $\mathbb{P}\left(N > n\right) = o\Big(e^{-n(\log n - \gamma_0)}\Big)$.

The assumption above is a slight-generalization of saying "the number of jumps N is more-or-less a Poisson process". (Nguyen & Peralta, 2021) proved if N is stochastically-dominated by a Poisson process (i.e. there exists a Poisson process N^* such that $\mathbb{P}\left(N \geq n\right) \leq \mathbb{P}\left(N^* \geq n\right)$), then it satisfies assumption 5.2. This assumption also generalizes to a case where the jump intensity is, although bounded, might not be "uniform" across the interval.

Under assumption 4.3 and the additional assumptions 5.1 and 5.2, Nguyen & Peralta, 2021 obtained a theorem which basically said that if $\big((V_t^\lambda)\big)_{\lambda\geq 0}$ converges to $(B_t)_{t\geq 0}$ strongly at some rate $\delta(\lambda)$, then the approximations $\big((X_t^\lambda)\big)_{\lambda\geq 0}$ also converges strongly to $(X_t)_{t\geq 0}$ at *almost* the same rate $\delta(\lambda)\lambda^\varepsilon$ for any $\varepsilon > 0$.

Theorem 5.3 (Nguyen & Peralta, 2021)**.** *Let $\big((V_t^\lambda)\big)_{\lambda\geq 0}$ be a collection of finite-variation processes such that for some decay function $\delta : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{\lambda\to\infty} \delta(\lambda) = 0$, we have that for all $q > 0$,*

$$\mathbb{P}\left(\sup_{t\in[0,T]} |V_t^\lambda - B_t| \geq \alpha\delta(\lambda)\right) = o(\lambda^{-q}) \tag{113}$$

where $\alpha = \alpha(q,T)$ only depends on q and T. For $\lambda > 0$, let $(X_t^\lambda)_{t\in[0,T]}$ be the unique solution to the SDE (111)*.*

Then

1. *The family of processes $\big((X_t)_{t\in[0,T]}\big)_{\delta\geq 0}$ converges strongly (uniformly and almost-surely) to the unique solution to the SDE* (112)*.*
2. *For all $q, T > 0$, there exists a constant $\gamma = \gamma(q,T) > 0$ such that*

$$\mathbb{P}\left(\sup_{t\in[0,T]} |X_t^\lambda - X_t| \geq \gamma\delta(\lambda)\lambda^\varepsilon\right) = o(\lambda^{-q}) \tag{114}$$

6 References

Karatzas, Ioannis, Steven E. Shreve (1991). *Brownian Motion and Stochastic Calculus*. 2nd ed. Springer.

McKnight, Aaron (2009). *Some Basic Properties of Brownian Motion*. URL: `http://www.math.uchicago.edu/~may/VIGRE/VIGRE2009/REUPapers/McKnight.pdf`.

Nguyen, Giang T., Oscar Peralta (2021). *Wong–Zakai approximations with convergence rate for stochastic differential equations with regime switching*. DOI: `10.48550/ARXIV.2101.03250`. URL: `https://arxiv.org/abs/2101.03250`.

Portal, Pierre (n.d.). *MATH3015/MATH6115: Stochastic Analysis and its Financial Applications Lecture Notes*.

Steele, J. Michael (2001). *Stochastic Calculus and Financial Applications*. Springer.

Wong, Eugene, Moshe Zakai (1965). "On the Convergence of Ordinary Integrals to Stochastic Integrals". In: *The Annals of Mathematical Statistics* 36.5, pp. 1560–1564. DOI: 10.1214/aoms/1177699916. URL: https://doi.org/10.1214/aoms/1177699916.