Notes on PSTAT 223

Haosheng Zhou

Sept, 2022

Mid-term: Nov. 2, take-home

Final: Dec. 7, 8-11 am in class

BM

Approximation by Random Walk

Set $X_1, .., X_n, ...$ *i.i.d.* with probability $\frac{1}{2}$ taking value $\sqrt{\varepsilon}$ and probability $\frac{1}{2}$ taking value $-\sqrt{\varepsilon}$ with $\lfloor \frac{t}{\varepsilon} \rfloor = n$ and set $S_t = \sum_{i=1}^n X_i$, then by CLT

$$\frac{S_t}{\sqrt{t}} \xrightarrow{d} N(0,1) \tag{1}$$

which gives the informal approximation to the BM (no path regularity mentioned).

BM existence is ensured by Kolmogorov's extension theorem and the Kolmogorov's lemma (the first use marginal distributions to construct the continuous-time stochastic process with the same finite dimensional distributions and the second ensures the regularity of path so that it's continuous).

Property of BM

- BM not differentiable. If differentiable on $[0, T]$ then total variation is finite. Note that with $T = n \cdot \Delta t$

$$\sum_i |B_{t_i} - B_{t_{i-1}}| \sim n\mathbb{E}|B_{\Delta t}| \sim \frac{T}{\Delta t}\sqrt{\Delta t} \to \infty \ (n \to \infty, \Delta t \to 0) \tag{2}$$

- Quadratic variation of BM on $[0, T]$ is just T. Partition $\Delta : 0 = t_0 < t_1 < ... < t_n = T$ with $||\Delta|| = \sup_i |t_i - t_{i-1}|$, then

$$\sum_{i=1}^n |B_{t_i} - B_{t_{i-1}}|^2 \xrightarrow{L_2} T \ (n \to \infty) \tag{3}$$

  actually such limit can be lifted to *a.s.* sense.

- Levy's characterization of BM: $B_0 = 0, B_t$ is continuous $a.s., \mathbb{E}(e^{iu(B_t - B_s)}|\mathscr{F}_s) = e^{-\frac{1}{2}u^2(t-s)}$

- Markov Property

- Martingale

Refer to HW 1 for more properties, or GTM 274

Week 2

Ito's Integral

Differential form of SDE:

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \tag{4}$$

with initial value $X_0 = x_0$. But it's actually not rigorous since dB_t is not well defined. Instead, use the integral form:

$$X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s \tag{5}$$

for well definition. Here b is the **drift** coefficient (mean return, controls the speed of evolution), σ is the **diffusion** coefficient (volatility, controls the size of the noise).

Define for $0 \le S \le T$, the space of all measurable (actually progressive), adapted and L^2 process.

$$V(S,T) = \{f : [0, +\infty) \times \Omega \to \mathbb{R}\} \tag{6}$$

such that $(t, \omega) \to f_t(\omega)$ is measurable w.r.t. $\mathscr{B}_{\mathbb{R}_+} \times \mathscr{F}$, $f_t \in \mathscr{F}_t$ and $\mathbb{E}\left(\int_S^T f_t^2\,dt\right) < \infty$ (say $f \in L^2([S,T] \times \Omega)$ since $\langle f_t, g_t \rangle = \mathbb{E}\left(\int_S^T f_t g_t\,dt\right)$ is the inner product on such space under $a.s.$ sense). Our goal is to **define the stochastic integral $I(f) = \int_S^T f_t\,dB_t$ for a general process f in such Hilbert space**.

We follow several steps, first consider defining this for a "simple" process and then extend it to general processes. The main thought is to find a **dense subset** of the Hilbert space and define the stochastic integral on such dense subset to prove that it's actually an **isometry**. After that, **extend** it as the isometry on the whole Hilbert space.

Step 0: Consider **elementary process** defined by $\varphi_t(\omega) = \sum_{j=1}^{\infty} e_j(\omega)\mathbb{I}_{[t_j, t_{j+1})}(t)$ **where** $e_j \in \mathscr{F}_{t_j}, e_j \in L^2(\Omega)$ and n, t_i are all fixed. t_j is the j-th **dyadic number** within $[S,T]$, i.e. it is $\frac{j}{2^n}$ if such value is in $[S,T]$. If such value is less than S, $t_j = S$. If such value is greater than T, $t_j = T$. (In simple words, only care about the dyadic partition within $[S,T]$).

Remark. *Note that the elementary process is a generalization of the step function, replacing the fixed constant with a random variable. The important point here is that this random variable here is* ***measurable w.r.t. the left endpoint*** *t_j. Changing the measurability here as the one w.r.t. the midpoint $e_j \in \mathscr{F}_{\frac{t_j + t_{j+1}}{2}}$ results in a different integration scheme.*

Naturally, the stochastic integral for elementary process is defined as

$$\int_S^T \varphi_t\,dB_t = \sum_{j=1}^{\infty} e_j(B_{t_{j+1}} - B_{t_j}) \tag{7}$$

Remark. *Let's show a counterexample here why $e_j \in \mathscr{F}_{t_j}$ can't be removed.*

For $\varphi_t^{(1)}$, take $e_j = B_{t_j}$. By definition, the integral should be $\sum_{j=1}^{\infty} B_{t_j}(B_{t_{j+1}} - B_{t_j})$, its expectation is

$$\mathbb{E}\sum_{j=1}^{\infty} B_{t_j}(B_{t_{j+1}} - B_{t_j}) = \sum_{j=1}^{\infty} \mathbb{E}B_{t_j}(B_{t_{j+1}} - B_{t_j}) \tag{8}$$

$$= \sum_{j=1}^{\infty} \mathbb{E}B_{t_j} \cdot \mathbb{E}(B_{t_{j+1}} - B_{t_j}) = 0 \tag{9}$$

For $\varphi_t^{(2)}$, take $e_j = B_{t_{j+1}}$. By definition, the integral should be $\sum_{j=1}^{\infty} B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j})$, its expectation is

$$\mathbb{E}\sum_{j=1}^{\infty} B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j}) = \sum_{j=1}^{\infty} \mathbb{E}B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j}) \tag{10}$$

$$= \sum_{j=1}^{\infty} \mathbb{E}(B_{t_{j+1}} - B_{t_j})^2 + \mathbb{E}B_{t_j}(B_{t_{j+1}} - B_{t_j}) \tag{11}$$

$$= \sum_{j=1}^{\infty} \mathbb{E}(B_{t_{j+1}} - B_{t_j})^2 = T - S \tag{12}$$

As we can see, a slight change in measurability at the endpoints results in the large change of the integral. This is due to the infinite total variation of BM, and this tells us from another perspective that the Lebesgue-Stieljes integration does not work for BM any longer!

The following lemma shows that such definition is actually an isometry between Hilbert spaces.

Lemma 1. *If φ_t is a bounded elementary process, then $\mathbb{E}\left(\int_S^T \varphi_t\, dB_t\right)^2 = \mathbb{E}\left(\int_S^T \varphi_t^2\, dt\right)$. This means that $||\int_S^T \varphi_t\, dB_t||_{L^2(\Omega)} = ||\varphi_t||_{L^2([S,T]\times\Omega)}$, the* ***Ito's isometry for elementary process****.*

Proof.

$$\mathbb{E}\left(\int_S^T \varphi_t\, dB_t\right)^2 = \mathbb{E}\left(\int_S^T \sum_{j=1}^{\infty} e_j \mathbb{I}_{[t_j,t_{j+1})}\, dB_t\right)^2 \tag{13}$$

$$= \mathbb{E}\left(\sum_{j=1}^{\infty} e_j(B_{t_{j+1}} - B_{t_j})\right)^2 \tag{14}$$

$$= \mathbb{E}\left(\sum_{i,j=1}^{\infty} e_i e_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right) \tag{15}$$

$$= \sum_{i,j=1}^{\infty} \mathbb{E}\left(e_i e_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right) \tag{16}$$

$$= \sum_{i=1}^{\infty} \mathbb{E}\left(e_i^2(B_{t_{i+1}} - B_{t_i})^2\right) + 2\sum_{i<j}^{\infty} \mathbb{E}\left(e_i e_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right) \tag{17}$$

Note that the interchange of expectation and the infinite sum is the consequence of Fubini's theorem with the fact that φ_t is $a.s.$ bounded, so $\forall i, |e_i| \leq M$ $a.s.$ are uniformly bounded.

$$\sum_{i,j=1}^{\infty} \mathbb{E}(|e_ie_j| \cdot |B_{t_{i+1}} - B_{t_i}| \cdot |B_{t_{j+1}} - B_{t_j}|) \leq M^2 \sum_{i,j=1}^{\infty} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}| \cdot |B_{t_{j+1}} - B_{t_j}|) \tag{18}$$

$$= M^2 \sum_{i=1}^{\infty} \mathbb{E}(B_{t_{i+1}} - B_{t_i})^2 + 2M^2 \sum_{i<j} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|) \cdot \mathbb{E}(|B_{t_{j+1}} - B_{t_j}|) \tag{19}$$

$$= M^2 \sum_{i=1}^{\infty} (t_{i+1} - t_i) + 2M^2 \sum_{i<j} \frac{2}{\pi} \sqrt{(t_{i+1} - t_i)(t_{j+1} - t_j)} \tag{20}$$

$$\leq M^2(T-S) + \frac{4M^2}{\pi}(T-S)^2 2^{n-1} < \infty \tag{21}$$

Also note that for $i < j$

$$\mathbb{E}\left(e_ie_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right) = \mathbb{E}\left[\mathbb{E}\left(e_ie_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})|\mathscr{F}_{t_j}\right)\right] \tag{22}$$

$$= \mathbb{E}\left[e_ie_j(B_{t_{i+1}} - B_{t_i})\mathbb{E}\left(B_{t_{j+1}} - B_{t_j}|\mathscr{F}_{t_j}\right)\right] = 0 \tag{23}$$

and use the independency of e_i and $B_{t_{i+1}} - B_{t_i}$

$$\mathbb{E}\left(\int_S^T \varphi_t \, dB_t\right)^2 = \sum_{i=1}^{\infty} \mathbb{E}\left(e_i^2(B_{t_{i+1}} - B_{t_i})^2\right) \tag{24}$$

$$= \sum_{i=1}^{\infty} \mathbb{E}e_i^2 \cdot (t_{i+1} - t_i) \tag{25}$$

$$= \mathbb{E}\sum_{i=1}^{\infty} e_i^2 \cdot (t_{i+1} - t_i) \tag{26}$$

$$= \mathbb{E}\left(\int_S^T \varphi_t^2 \, dt\right) \tag{27}$$

the interchange of the expectation and the infinite sum is due to the non-negativity and the Fubini theorem.

□

Remark. *The Ito's isometry provides a link between stochastic integral and Lebesgue-Stieljes integral in that the L^2 norm of the stochastic integral $\int_S^T \varphi_t \, dB_t$ is equal to the expectation of a Lebesgue-Stieljes integral $\int_S^T \varphi_t^2 \, dt$ that only integrates w.r.t. the time.*

To extend the definition of Ito's integral and the Ito's isometry property onto the whole process space $V(S,T)$, the core is to prove that elementary processes are actually a dense subset of $V(S,T)$.

Step 1: Consider any **bounded continuous process** $g_t \in V(S,T)$, it's quite natural to notice that there exist bounded elementary process $\varphi_t^{(n)}$ such that $||g_t - \varphi_t^{(n)}||_{L^2([S,T]\times\Omega)} \to 0$ $(n \to \infty)$. This is done by construction

$$\varphi_t^{(n)} = \sum_{j=1}^{\infty} g_{t_j} \mathbb{I}_{[t_j, t_{j+1})}(t) \tag{28}$$

where t_i are truncated dyadic numbers in the interval $[S,T]$, i.e. $\frac{i}{2^n}$ if it's in the interval and the endpoint if not. The L^2 convergence is ensured by

$$\mathbb{E}\int_S^T \left(g_t - \varphi_t^{(n)}\right)^2 dt = \mathbb{E}\int_S^T \left(\sum_{j=1}^{\infty}(g_t - g_{t_j})\mathbb{I}_{[t_j,t_{j+1})}(t)\right)^2 dt \tag{29}$$

$$= \mathbb{E}\int_S^T \sum_{j=1}^{\infty}(g_t - g_{t_j})^2\mathbb{I}_{[t_j,t_{j+1})}(t)\, dt \tag{30}$$

$$= \mathbb{E}\sum_{j=1}^{\infty}\int_{t_j}^{t_{j+1}} (g_t - g_{t_j})^2\, dt \tag{31}$$

Since g_t is continuous $a.s.$ for $t \in [S,T]$, it's uniformly continuous, so $\forall \varepsilon > 0, \exists \delta > 0, \forall x, y \in [S,T]$ $s.t.$ $|x-y| < \delta, |g_x - g_y| < \varepsilon$. Now for large enough n, exists ε, δ such that $\delta > \frac{1}{2^n}$, then

$$\mathbb{E}\sum_{j=1}^{\infty}\int_{t_j}^{t_{j+1}} (g_t - g_{t_j})^2\, dt \le \mathbb{E}\sum_{j=1}^{\infty}\varepsilon^2(t_{j+1} - t_j) \tag{32}$$

$$= \varepsilon^2(T-S) \tag{33}$$

$$\mathbb{E}\int_S^T \left(g_t - \varphi_t^{(n)}\right)^2 dt \to 0 \ (n \to \infty) \tag{34}$$

Remark. *The construction of $\varphi_t^{(n)}$ is just similar to that in the deterministic case (use step functions to approximate continuous bounded functions well enough), where we replace the fixed constant as a $\mathscr{F}_{t_j}$ measurable random variable g_{t_j}.*

Step 2: Extend the definition for all **bounded process** $h_t \in V(S,T)$. We hope to prove that there always exists a bounded continuous process $g_t^{(n)} \in V(S,T)$ such that $||h_t - g_t^{(n)}||_{L^2([S,T]\times\Omega)} \to 0$ $(n \to \infty)$.

The construction uses **the convolution w.r.t. the mollifier**. To be specific, let $\psi_n \ge 0$ be a continuous mollifier on $\mathbb{R}$ such that $\forall x \ge 0, \forall x \le -\frac{1}{n}, \psi_n(x) = 0$ and $\int_{\mathbb{R}} \psi_n(x)\, dx = 1$. Consider

$$g_t^{(n)} = \int_0^t \psi_n(s-t) \cdot h_s\, ds \tag{35}$$

It's then not hard to see that if ω is fixed,

$$|g_t^{(n)}| \le \sup_s |h_s| \cdot \int_0^t \psi_n(s-t)\, ds \le \sup_s |h_s| \tag{36}$$

so $g_t^{(n)}$ is uniformly bounded by the same bound of h_t.

The continuity follows from the fact that

$$\begin{aligned}
\left|g_{t+\Delta t}^{(n)} - g_t^{(n)}\right| &= \left|\int_0^{t+\Delta t} \psi_n(s-t-\Delta t)\cdot h_s\,ds - \int_0^t \psi_n(s-t)\cdot h_s\,ds\right| &(37)\\
&= \left|\int_0^t [\psi_n(s-t-\Delta t) - \psi_n(s-t)]\cdot h_s\,ds\right| + \left|\int_t^{t+\Delta t} \psi_n(s-t-\Delta t)\cdot h_s\,ds\right| &(38)\\
&\le \sup_s |h_s|\cdot\left(\left|\int_0^t [\psi_n(s-t-\Delta t) - \psi_n(s-t)]\,ds\right| + \left|\int_t^{t+\Delta t} \psi_n(s-t-\Delta t)\,ds\right|\right) &(39)\\
&= \sup_s |h_s|\cdot\left(\left|\int_{-t}^0 [\psi_n(u-\Delta t) - \psi_n(u)]\,du\right| + \left|\int_{-\Delta t}^0 \psi_n(u)\,du\right|\right) &(40)
\end{aligned}$$

Note that $\left|\int_{-\Delta t}^0 \psi_n(u)\,du\right| \to 0$ as $\Delta t \to 0$ since the continuity of ψ_n ensures its boundedness on $[-\Delta t, 0]$ and the integration is on a small enough range. For the other term, notice that ψ_n is uniformly continuous on $[-t-\Delta t, 0]$, so $\forall \varepsilon > 0, \exists \delta > 0, \forall u_1, u_2 \in [-t-\Delta t, 0],\ if\ |u_1 - u_2| < \delta, |\psi_n(u_1) - \psi_n(u_2)| < \varepsilon$.

Since we hope to investigate this term as $\Delta t \to 0$, there exists δ such that $\Delta t < \delta$, so

$$\left|\int_{-t}^0 [\psi_n(u-\Delta t) - \psi_n(u)]\,du\right| < \varepsilon t \tag{41}$$

is also small enough.

As a result, we have shown that the convoluted process $g_t^{(n)}$ is continuous and bounded. Now we only have to show the convergence in L^2.

$$\begin{aligned}
||h_t - g_t^{(n)}||_{L^2([S,T]\times\Omega)} &= \mathbb{E}\int_S^T \left(h_t - g_t^{(n)}\right)^2 dt &(42)\\
&= \mathbb{E}\int_S^T \left(h_t - \int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot h_s\,ds\right)^2 dt &(43)\\
&= \mathbb{E}\int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot (h_t - h_s)\,ds\right)^2 dt &(44)
\end{aligned}$$

note that if we can prove the property that

$$\int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot(h_t - h_s)\,ds\right)^2 dt \to 0\ (n\to\infty) \tag{45}$$

then the L^2 convergence is proved by the bounded convergence theorem since

$$\left|\int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot(h_t-h_s)\,ds\right)^2 dt\right| \le \left(2\sup_s |h_s|\right)^2 \cdot \int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\,ds\right)^2 dt \tag{46}$$

$$= \left(2\sup_s |h_s|\right)^2 \cdot (T-S) < \infty \tag{47}$$

Let's now try to prove the fact that

$$\int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot(h_t-h_s)\,ds\right)^2 dt \to 0 \ (n\to\infty) \tag{48}$$

actually holds. The **Minkowski integral inequality** gives

$$\left[\int_S^T \left(\int_{t-\frac{1}{n}}^t \psi_n(s-t)\cdot(h_t-h_s)\,ds\right)^2 dt\right]^{\frac{1}{2}} \le \int_{t-\frac{1}{n}}^t \left(\int_S^T \psi_n^2(s-t)\cdot(h_t-h_s)^2\,dt\right)^{\frac{1}{2}} ds \tag{49}$$

$$= \int_{-\frac{1}{n}}^0 \psi_n(u)\left(\int_S^T (h_t-h_{u+t})^2\,dt\right)^{\frac{1}{2}} du \tag{50}$$

$$\tag{51}$$

with the fact that the **translation in time h_t to h_{u+t} is continuous**, so $\forall\varepsilon>0, \exists\delta>0$, if $|u|<\delta$, then $|h_t-h_{t+u}|<\varepsilon$. Since $\int_{\mathbb{R}}\psi_n = 1$, $\exists n_0$ such that for any $n>n_0$, $\int_{u\le-\delta}\psi_n(u)\,du<\varepsilon$. Split the integral above into two parts:

$$\int_{u>-\delta,u\in[-\frac{1}{n},0]} \psi_n(u)\left(\int_S^T (h_t-h_{u+t})^2\,dt\right)^{\frac{1}{2}} du + \int_{u\le-\delta,u\in[-\frac{1}{n},0]} \psi_n(u)\left(\int_S^T (h_t-h_{u+t})^2\,dt\right)^{\frac{1}{2}} du \tag{52}$$

$$\le \varepsilon\sqrt{T-S}\int_{u>-\delta,u\in[-\frac{1}{n},0]} \psi_n(u)\,du + (2\sup_s|h_s|)\cdot\sqrt{T-S}\int_{u\le-\delta,u\in[-\frac{1}{n},0]} \psi_n(u)\,du \tag{53}$$

$$\le \varepsilon\sqrt{T-S} + (2\sup_s|h_s|)\cdot\sqrt{T-S}\cdot\varepsilon \tag{54}$$

so the conclusion gets proved. This property is exactly the **approximation identity of the mollifier**, which tells us that when the support of the mollifier goes to 0, the convolution converges to the true mollified function in L^p sense.

Remark. *This is a classic technique to use in analysis. First write the difference between the function and the convolution as an integral form, then use inequalities to change the order of the integral and at last tear the integral into two parts. The first part is* ***near the singularity of the mollifier****, where the* ***continuity of translation*** *is used. The second part is* ***far away from the singularity of the mollifier****, where* ***the support can be shrunk*** *such that the integral of the mollifier is always small enough.*

Note that there might be issue proving the measurability of $g_t^{(n)}$ (not verified here)

Step 3: For **general process** $f_t \in V(S,T)$, always exists bounded process $h_t^{(n)} \in V(S,T)$ such that $||f_t - h_t^{(n)}||_{L^2([S,T]\times\Omega)} \to 0$ $(n \to \infty)$. The construction is given by simple truncation of function value that

$$h_t^{(n)} = f_t \wedge n \vee (-n) \tag{55}$$

It's then quite obvious to see that by using the Fubini theorem, the monotone convergence theorem for $(f_t + n)^2 \mathbb{I}_{f_t<-n} \searrow 0$ $(n \to \infty)$ and the monotone convergence theorem for $\mathbb{E}[(f_t+n)^2\mathbb{I}_{f_t<-n}] \searrow 0$ $(n \to \infty)$:

$$\mathbb{E}\int_S^T \left(f_t - h_t^{(n)}\right)^2 dt = \mathbb{E}\int_S^T (f_t+n)^2\mathbb{I}_{t:f_t<-n} + (f_t-n)^2\mathbb{I}_{t:f_t>n}\, dt \tag{56}$$

$$= \mathbb{E}\int_{t\in[S,T],f_t<-n} (f_t+n)^2\, dt + \mathbb{E}\int_{t\in[S,T],f_t>n} (f_t-n)^2\, dt \tag{57}$$

$$= \int_S^T \mathbb{E}[(f_t+n)^2\mathbb{I}_{f_t<-n}]\, dt + \int_S^T \mathbb{E}[(f_t-n)^2\mathbb{I}_{f_t>n}]\, dt \to 0\ (n \to \infty) \tag{58}$$

As a result, we have proved that the set of all elementary process is a **dense** subset of $V(S,T)$. For any $f_t \in V(S,T)$, its stochastic integral is defined as the L^2 limit of the stochastic integral of the approximation process $\varphi_t^{(n)}$, i.e.

$$\mathbb{E}\int_S^T \left(f_t - \varphi_t^{(n)}\right)^2 dt \to 0\ (n \to \infty) \tag{59}$$

$$\int_S^T \varphi_t^{(n)}\, dB_t \overset{L^2(\Omega)}{\to} \int_S^T f_t\, dB_t\ (n \to \infty) \tag{60}$$

Remark. *Note that here* $\left\{\int_S^T \varphi_t^{(n)}\, dB_t\right\}_{n=1}^{\infty}$ *has to converge in* L^2 *sense since the set of elementary functions is a dense subset in the Hilbert space and that there's already an isometry on this dense subset. That is to say, if there are Hilbert spaces* H_1, H_2 *with norms* $||\cdot||_1, ||\cdot||_2$*, and* $D \subset H_1$ *is dense with* $f: D \to H_2$ *an isometry, then there exists an extension of* f *denoted* $g: H_1 \to H_2$ *to be an isometry.*

The construction is intuitive

$$\forall x \in H_1, \exists d_n \in D, ||d_n - x||_1 \to 0\ (n \to \infty) \tag{61}$$

$$g(x) \overset{def}{=} \lim_{n\to\infty} g(d_n) \tag{62}$$

to notice that the completeness of Hilbert space ensures that g *is well-defined (such limit exists).*

The Ito's isometry holds for general process in $V(S,T)$. Assume that f_t is a bounded process in $V(S,T)$, then exists a series of bounded elementary processes $\varphi_t^{(n)}$ such that $\varphi_t^{(n)} \overset{L^2([S,T]\times\Omega)}{\to} f_t$ $(n \to \infty)$.

$$\mathbb{E}\left(\int_S^T f_t\, dB_t\right)^2 = \lim_{n\to\infty} \mathbb{E}\left(\int_S^T \varphi_t^{(n)}\, dB_t\right)^2 = \lim_{n\to\infty} \mathbb{E}\left(\int_S^T \left(\varphi_t^{(n)}\right)^2 dt\right) = \mathbb{E}\left(\int_S^T f_t^2\, dt\right) \tag{63}$$

where the first and third equations are the consequences of L^2 convergence in space $L^2(\Omega)$ (space of r.v.) and $L^2([S,T]\times\Omega)$ (space of process) since L^2 convergence implies the convergence of L^2 norms. The second equation comes from the Ito's isometry for bounded elementary process.

Theorem 1. *$\forall f_t \in L^2([S,T]\times\Omega)$ (which means that $\mathbb{E}\int_S^T f_t^2\,dt < \infty$), $f_t \in \mathscr{F}_t$ and f_t progressive (which means that $(t,\omega)\to f_t(\omega)$ is measurable w.r.t. $\mathscr{B}_{[S,T]}\times\mathscr{F}$), then the definition of stochastic integral above gives **the Ito's isometry for general process** $\mathbb{E}(\int_S^T f_t\,dB_t)^2 = \mathbb{E}\int_S^T f_t^2\,dt < \infty$ a.s..*

Proof. We proved above that any bounded process $f_t \in V(S,T)$ satisfies the Ito's isometry. Now for general $f_t \in V(S,T)$, find a series of bounded process to approximate

$$\exists h_t^{(n)} \in V(S,T), \forall t, |h_t| \le M \tag{64}$$

$$h_t^{(n)} \overset{L^2([S,T]\times\Omega)}{\to} f_t \ (n\to\infty) \tag{65}$$

It's natural that

$$\mathbb{E}\left(\int_S^T f_t\,dB_t\right)^2 = \lim_{n\to\infty}\mathbb{E}\left(\int_S^T h_t^{(n)}\,dB_t\right)^2 = \lim_{n\to\infty}\mathbb{E}\left(\int_S^T \left(h_t^{(n)}\right)^2\,dt\right) = \mathbb{E}\left(\int_S^T f_t^2\,dt\right) \tag{66}$$

□

Example

Compute $\int_0^t B_s dB_s$.

First check if B_s is in the Hilbert space

$$\mathbb{E}\int_0^t B_s^2\,ds = \int_0^t \mathbb{E}B_s^2\,ds = \frac{t^2}{2} < \infty \tag{67}$$

so $B_s|_{s\in[0,t]} \in V(0,t)$.

Now we can first try to follow the definition of stochastic integral. Find elementary process

$$\varphi_s^{(n)} = \sum_{j=1}^{\infty} B_{s_j}\mathbb{I}_{[s_j,s_{j+1})}(s) \tag{68}$$

to approximate BM such that

$$\begin{aligned}
\mathbb{E}\int_0^t \left(\varphi_s^{(n)} - B_s\right)^2 ds &= \mathbb{E}\int_0^t \sum_{j=1}^{\infty}(B_{s_j} - B_s)^2 \mathbb{I}_{[s_j,s_{j+1})}(s)\, ds && (69)\\
&= \mathbb{E}\sum_{j=1}^{\infty}\int_{s_j}^{s_{j+1}} (B_{s_j} - B_s)^2\, ds && (70)\\
&= \sum_{j=1}^{\infty}\int_{s_j}^{s_{j+1}} \mathbb{E}(B_{s_j} - B_s)^2\, ds && (71)\\
&= \sum_{j=1}^{\infty}\int_{s_j}^{s_{j+1}} (s - s_j)\, ds && (72)\\
&= \sum_{j=1}^{\infty}\frac{(s_{j+1} - s_j)^2}{2} \to 0 \ (n \to \infty) && (73)
\end{aligned}$$

That's why the stochastic integral is formulated as

$$\sum_{j=1}^{\infty} B_{s_j}(B_{s_{j+1}} - B_{s_j}) \overset{L^2(\Omega)}{\to} \int_0^t B_s\, dB_s \ (n \to \infty) \tag{74}$$

A transformation gives

$$\begin{aligned}
\sum_{j=1}^{\infty} B_{s_j}(B_{s_{j+1}} - B_{s_j}) &= \sum_{j=1}^{\infty}\frac{B_{s_{j+1}} + B_{s_j}}{2}(B_{s_{j+1}} - B_{s_j}) - \sum_{j=1}^{\infty}\frac{B_{s_{j+1}} - B_{s_j}}{2}(B_{s_{j+1}} - B_{s_j}) && (75)\\
&\to \frac{B_t^2}{2} - \frac{t}{2} \ (n \to \infty) && (76)
\end{aligned}$$

note that the first limit is a telescoping and the second limit comes from the quadratic variation of BM. So we get

$$\int_0^t B_s\, dB_s = \frac{B_t^2}{2} - \frac{t}{2} \tag{77}$$

note that there's another **Ito's correction term** as the quadratic variation so **the chain rule does not hold!**

Wiener Integral

Just to mention, the wiener integral is defined as a special situation for stochastic integral, where the integrand is a deterministic function in time instead of a process. For function $f(s) \in C^1, \int_0^t f^2(s)\, ds < \infty$, define by the integration by parts:

$$\int_0^t f(s)\,dB_s = f\cdot B\Big|_0^t - \int_0^t B_s\,df(s) \tag{78}$$

$$= f(t)\cdot B_t - \int_0^t f'(s)\cdot B_s\,ds \tag{79}$$

This turns the stochastic integral into a Lebesgue-Stieljes integral. Now we want to show that this definition is consistent with that for the stochastic integral set up above. The continuity of f ensures that it's bounded and continuous on $[0,t]$, so it can be easily approximated by elementary process (which, in deterministic case, is just the step function). By the definition above,

$$\int_0^t f(s)\,dB_s = \lim_{n\to\infty}\sum_j f(t_j)(B_{t_{j+1}} - B_{t_j}) \tag{80}$$

for t_j as truncated dyadic numbers $\frac{j}{2^n}$ in $[0,t]$ and the limit is in the $L^2(\Omega)$ sense.

Let's assume that the truly effective dyadic numbers are $0 = t_0^n < t_1^n < ... < t_{p_n}^n \le t < t_{p_n+1}^n$, so the sum for fixed n is actually

$$\sum_{j=0}^{p_n-1} f(t_j^n)\left(B_{t_{j+1}^n} - B_{t_j^n}\right) + f(t_{p_n}^n)\left(B_t - B_{t_{p_n}^n}\right) \tag{81}$$

$$= f(t_{p_n-1}^n)B_{t_{p_n}^n} - f(t_0^n)B_{t_0^n} - \sum_{j=1}^{p_n-1} B_{t_j^n}(f(t_j^n) - f(t_{j-1}^n)) + f(t_{p_n}^n)\left(B_t - B_{t_{p_n}^n}\right) \tag{82}$$

$$= f(t_{p_n}^n)B_t - \sum_{j=1}^{p_n} B_{t_j^n}(f(t_j^n) - f(t_{j-1}^n)) \tag{83}$$

here we use the Abel's lemma (summation by parts), and set $n\to\infty$ to find that

$$f(t_{p_n}^n)B_t \to f(t)B_t \tag{84}$$

$$\sum_{j=1}^{p_n} B_{t_j^n}(f(t_j^n) - f(t_{j-1}^n)) = \sum_{j=1}^{p_n} B_{t_j^n}\Delta t\frac{f(t_j^n) - f(t_{j-1}^n)}{\Delta t} \tag{85}$$

$$\xrightarrow{L^2} \int_0^t B_s f'(s)\,ds \tag{86}$$

The L^2 convergence comes from the fact that

$$\mathbb{E}\left[\sum_{j=1}^{p_n} B_{t_j^n} f'(t_j^n)\Delta t - \sum_{j=1}^{p_n} B_{t_j^n}\Delta t \frac{f(t_j^n) - f(t_{j-1}^n)}{\Delta t}\right]^2 \tag{87}$$

$$= \mathbb{E}\left[\sum_{j=1}^{p_n} B_{t_j^n}\Delta t \left(f'(t_j^n) - \frac{f(t_j^n) - f(t_{j-1}^n)}{\Delta t}\right)\right]^2 \tag{88}$$

$$\leq \mathbb{E}\left[\sum_{j=1}^{p_n} B_{t_j^n}\Delta t \cdot \varepsilon\right]^2 \tag{89}$$

$$\leq \varepsilon^2 t \tag{90}$$

note that the existence of the uniform ε for all t_j^n is ensured by the continuity of f', and the last equation can be derived by expanding the square.

As a result, we have proved that the Wiener integral is consistent with the definition of general stochastic integral (uniqueness of L^2 limit). After the work, we shall also notice from

$$\sum_j f(t_j)(B_{t_{j+1}} - B_{t_j}) \xrightarrow{L^2} \int_0^t f(s)\,dB_s \ (n \to \infty) \tag{91}$$

that the Wiener integral is actually the L^2 limit of a linear combination of independent Gaussian random variables (independent increment), so the Wiener integral is the L^2 limit of Gaussian. Note that L^2 limit of Gaussian must be Gaussian and the expectation and variance of the limit are just the limit of the expectation series and the variance series (this is from the pointwise limit of the characteristic function), so Wiener integral must be Gaussian.

$$B_{t_{j+1}} - B_{t_j} \sim N(0, \Delta t) \tag{92}$$

$$f(t_j)(B_{t_{j+1}} - B_{t_j}) \sim N(0, f^2(t_j)\Delta t) \tag{93}$$

$$\mathbb{E}\left(\int_0^t f(s)\,dB_s\right) = 0 \tag{94}$$

$$Var\left(\int_0^t f(s)\,dB_s\right) = \sum_j f^2(t_j)\Delta t = \int_0^t f^2(s)\,ds \tag{95}$$

or we can conclude from the Ito's isometry that $\mathbb{E}\left(\int_0^t f(s)\,dB_s\right)^2 = \mathbb{E}\left(\int_0^t f^2(s)\,ds\right) = \int_0^t f^2(s)\,ds$. As a result, we get the distribution of the Wiener integral

$$\int_0^t f(s)\,dB_s \sim N\left(0, \int_0^t f^2(s)\,ds\right) \tag{96}$$

Property of Stochastic Integral

The first one is **linearity**.

$$\forall c,d\in\mathbb{R},\forall f_t,g_t\in V(S,T),\int_S^T(cf_s+dg_s)\,dB_s=c\int_S^T f_s\,dB_s+d\int_S^T g_s\,dB_s \tag{97}$$

Let's prove this with the definition using elementary process to approximate general process.

$$\exists\varphi_t^n,\varphi_t^n\overset{L^2([S,T]\times\Omega)}{\to}f_t,\exists\psi_t^n,\psi_t^n\overset{L^2([S,T]\times\Omega)}{\to}g_t\ (n\to\infty) \tag{98}$$

$$c\varphi_t^n+d\psi_t^n\overset{L^2([S,T]\times\Omega)}{\to}cf_t+dg_t\ (n\to\infty) \tag{99}$$

That's why by the definition of the stochastic integral,

$$\sum_j\Big(ce_{t_j}^n(\varphi)+de_{t_j}^n(\psi)\Big)\cdot\big(B_{t_{j+1}}-B_{t_j}\big)\overset{L^2}{\to}\int_S^T(cf_s+dg_s)\,dB_s\ (n\to\infty) \tag{100}$$

$$\sum_j ce_{t_j}^n(\varphi)\cdot\big(B_{t_{j+1}}-B_{t_j}\big)\overset{L^2}{\to}c\int_S^T f_s\,dB_s\ (n\to\infty) \tag{101}$$

$$\sum_j de_{t_j}^n(\psi)\cdot\big(B_{t_{j+1}}-B_{t_j}\big)\overset{L^2}{\to}d\int_S^T g_s\,dB_s\ (n\to\infty) \tag{102}$$

and linearity is proved by the uniqueness of L^2 limit. Here $e_{t_j}^n(\varphi)$ is the $\mathscr{F}_{t_j}$ measurable random variable used in the construction of φ_t^n such that $\varphi_t^n=\sum_j e_{t_j}^n(\varphi)\mathbb{I}_{[t_j,t_{j+1})}(t)$

The second one is the **partition of integration area**.

$$\forall S\le U\le T,\forall f_t\in V(S,T),\int_S^T f_s\,dB_s=\int_S^U f_s\,dB_s+\int_U^T f_s\,dB_s \tag{103}$$

which is also natural from the definition of stochastic integral and the approximation of elementary process.

The third property is that **the process of Ito integral $M_t=\int_0^t f_s\,dB_s$ is an L^2 martingale adapted to the filtration generated by BM**. These are all observations directly from the definition using elementary process that

$$\exists\varphi_t^n,\varphi_t^n\overset{L^2([S,T]\times\Omega)}{\to}f_t\ (n\to\infty) \tag{104}$$

$$\sum_j e_{t_j}^n(\varphi)(B_{t_{j+1}}-B_{t_j})\overset{L^2}{\to}M_t\ (n\to\infty) \tag{105}$$

since $e_{t_j}^n(\varphi)\in\mathscr{F}_{t_j}$, so $\varphi_{t_j}^n(B_{t_{j+1}}-B_{t_j})\in\mathscr{F}_{t_{j+1}}\subset\mathscr{F}_t$, so the L^2 limit $M_t\in\mathscr{F}_t$ is adapted. $M_t\in L^2$ is also obvious from the Ito's isometry. Now let's prove the martingale property for f_r:

$$\exists \varphi_r^n, \varphi_r^n \overset{L^2([s,t]\times\Omega)}{\to} f_r \ (n \to \infty) \tag{106}$$

$$\sum_j e_{r_j}^n(\varphi)(B_{r_{j+1}} - B_{r_j}) \overset{L^2}{\to} \int_s^t f_r \, dB_r \ (n \to \infty) \tag{107}$$

$$\forall s \le t, \mathbb{E}(M_t - M_s|\mathscr{F}_s) = \mathbb{E}\left(\int_s^t f_r \, dB_r \middle| \mathscr{F}_s\right) \tag{108}$$

$$= \lim_{n\to\infty} \sum_j \mathbb{E}\left(e_{r_j}^n(\varphi)(B_{r_{j+1}} - B_{r_j}) \middle| \mathscr{F}_s\right) \tag{109}$$

$$= \lim_{n\to\infty} \sum_j \mathbb{E}\left[\mathbb{E}\left(e_{r_j}^n(\varphi)(B_{r_{j+1}} - B_{r_j}) \middle| \mathscr{F}_{r_j}\right) \middle| \mathscr{F}_s\right] \tag{110}$$

$$= \lim_{n\to\infty} \sum_j \mathbb{E}\left[\mathbb{E}\left(e_{r_j}^n(\varphi)(B_{r_{j+1}} - B_{r_j}) \middle| \mathscr{F}_{r_j}\right) \middle| \mathscr{F}_s\right] \tag{111}$$

$$= \lim_{n\to\infty} \sum_j \mathbb{E}\left[e_{r_j}^n(\varphi) \cdot \mathbb{E}\left(B_{r_{j+1}} - B_{r_j} \middle| \mathscr{F}_{r_j}\right) \middle| \mathscr{F}_s\right] \tag{112}$$

$$= 0 \tag{113}$$

note that the appearance of the limit is due to the L^2 convergence to the stochastic integral and the martingale property follows directly from the tower property of C.E. and the fact that BM is itself a martingale. (r_j^n are the truncated dyadic numbers within $[s,t]$ with the grid of partition to be $\frac{1}{2^n}$, for fixed n, the sum w.r.t. j is actually a finite sum)

Path Regularity for Ito Integral

Now that M_t is a martingale, we know that under some special conditions (filtration to be right-continuous and complete, and $t \to \mathbb{E}M_t$ to be right-continuous), a martingale has a modification with **Cadlag sample paths** (refer to GTM 274 P57). To verify those conditions, the completeness of filtration is trivial and the right-continuity of filtration is also satisfied (Blumenthal's 0-1 law of BM), actually this filtration is called **the canonical filtration** of BM. Moreover, $\mathbb{E}M_t = \mathbb{E}M_0 = 0$ so it's continuous (MG property). As a result, the Ito integral process has a Cadlag modification and the one-sided continuity **enables the application of MG inequalities**. (Doob's maximal, Doob's L^p etc.). However, due to the special structure of Ito's integral (the continuity of BM), we can actually show that this process has a modification with **continuous sample paths**.

Theorem 2. *There exists a unique continuous modification of M_t.*

Proof. The uniqueness under indistinguishability directly follows from the continuity of sample path, so only need to prove existence.

To apply the definition of stochastic integral, there exists elementary process $\varphi_s^n \overset{L^2([0,t]\times\Omega)}{\to} f_s$ $(n \to \infty)$ with

$$\varphi_s^n = \sum_j e_j^n \mathbb{I}_{[t_j^n, t_{j+1}^n)}(s), e_j^n \in \mathscr{F}_{t_j^n} \tag{114}$$

where t_j^n is the truncated dyadic number in $[0,t]$ with grid size $\frac{1}{2^n}$ and consider

$$I_t^n = \int_0^t \varphi_s^n \, dB_s = \sum_j e_j^n \cdot (B_{t_{j+1}^n} - B_{t_j}^n) \tag{115}$$

which is obviously continuous in t for each fixed n. This is due to the uniform continuity of BM on closed intervals. Since I_t^n is itself an Ito integral, it's also an adapted martingale.

Notice that $I_t^n \overset{L^2}{\to} M_t$ $(n \to \infty)$, to prove the path continuity of M_t, it suffices to prove that I_t^n **converges uniformly on any compact set** $[0,T]$.

Let's take $\forall T \geq 0$ and consider the convergence on $t \in [0,T]$. Our goal is to prove that

$$\forall \varepsilon > 0, \exists N, \forall m, n > N, \sup_{t \leq T} |I_t^n - I_t^m| < \varepsilon \tag{116}$$

so we recall the Borel-Cantelli lemma and hope to prove that $\sup_{t\leq T} |I_t^n - I_t^m| < \varepsilon$ holds eventually. However, the difficulty is that here we have both m and n going to infinity and we also have to deal with ε, so we hope that there is a way for us to turn these three things into the dependency on a same variable going to infinity, a natural thought is to **take a good enough subsequence**.

To do this, note that $I_t^m - I_t^n$ is always a martingale with continuous sample path, so Doob's maximal inequality gives

$$\forall \varepsilon > 0, \mathbb{P}\left(\sup_{t \leq T} |I_t^m - I_t^n| \geq \varepsilon\right) \leq \frac{\mathbb{E}(I_T^m - I_T^n)^2}{\varepsilon^2} = \frac{\mathbb{E}\left(\int_0^T (\varphi_s^m - \varphi_s^n) \, dB_s\right)^2}{\varepsilon^2} \tag{117}$$

$$= \frac{\mathbb{E}\left(\int_0^T (\varphi_s^m - \varphi_s^n)^2 \, ds\right)}{\varepsilon^2} \to 0 \ (n \to \infty) \tag{118}$$

by the Cauchy principle of the $L^2([0,T]\times\Omega)$ convergence of φ_s^n. As a result, there exists a subsequence $n_k \to \infty$ such that (a simple construction)

$$\forall k \geq 1, \mathbb{P}\left(\sup_{t\leq T} |I_t^{n_k} - I_t^{n_{k+1}}| \geq 2^{-k}\right) \leq \frac{1}{2^k} \tag{119}$$

Now it's easy to use Borel-Cantelli:

$$\sum_{k\geq 1}\mathbb{P}\left(\sup_{t\leq T}|I_t^{n_k}-I_t^{n_{k+1}}|\geq 2^{-k}\right)<\infty \tag{120}$$

$$\mathbb{P}\left(\sup_{t\leq T}|I_t^{n_k}-I_t^{n_{k+1}}|\geq 2^{-k}\ i.o.\right)=0 \tag{121}$$

so $a.s.$ $\sup_{t\leq T}|I_t^{n_k}-I_t^{n_{k+1}}|\leq 2^{-k}$ eventually for large enough k, which means that $a.s.$ $I_t^{n_k}$ converges uniformly on $[0,T]$ as $k\to\infty$. Since the L^2 limit is unique, it's easy to see that the limit of this subsequence has to be equal to M_t almost surely, so this limit is just a modification of M_t and uniform convergence ensures the continuity of sample path.

□

Remark. *Doob's maximal inequality bounds the tail probability of the tail supreme of the approximation, leading to the existence of a good enough subsequence and the uniform convergence on any compact set. This is a frequently used criterion for* ***proving path continuity: bound the tail probability of the tail supreme, take a good subsequence and show uniform convergence with Borel-Cantelli****.*

Remark. *It's quite obvious that the upcrossing inequality is the key to the Cadlag modification of martingales. However, in order to get continuous modification, the continuity of BM is the key, i.e. if we are integrating w.r.t. a general semi-martingale, the continuity won't necessarily hold.*

From now on, we always assume that the **Ito's process w.r.t. process** f_t

$$M_t=\int_0^t f_s\,dB_s \tag{122}$$

is a continuous martingale.

Week 3

Extension of Ito Integral

There are two main extensions for Ito integral. One is that the filtration can be slightly enlarged. We can choose **the filtration $\mathscr{F}_t$ such that B_t is a $\mathscr{F}_t$-BM and $f_t \in \mathscr{F}_t$**, instead of the one directly generated by the BM, which is $\mathscr{F}_t = \sigma(X_s, 0 \leq s \leq t)$. The main motivation for this is that when considering the stochastic integral for multi-dimensional BM $B_t = (B_t^1, ..., B_t^n)$, an object of interest would be $\int_S^T B_t^2 \, dB_t^1$. If we consider this integral under the old settings, then the filtration $\mathscr{F}_t$ should be the one generated by B_t^1. However, B_t^2 is not necessarily adapted to this filtration.

However, by setting $\mathscr{F}_t = \sigma((B_s^1, ..., B_s^n), 0 \leq s \leq t)$ (the jointly generated sigma field), $B_t^2 \in \mathscr{F}_t$ is adapted now and let's verify that this enlargement of sigma field still guarantees that B_t^1 is $\mathscr{F}_t$-adapted BM. This is quite obvious since multi-dimensional BM has independent coordinates, so B_t^1 conditioning on $\sigma((B_s^1, ..., B_s^n), 0 \leq s \leq t)$ is the same as B_t^1 conditioning on $\sigma(B_s^1, 0 \leq s \leq t)$.

The other extension is made for the Hilbert space $L^2([S,T] \times \Omega)$, in which the processes f_t should satisfy $\mathbb{E} \int_S^T f_t^2 \, dt < \infty$. Now we would like to weaken the finiteness of expectation into almost surely finite

$$\mathbb{P}\left(\int_S^T f_t^2 \, dt < \infty \right) = 1 \tag{123}$$

As a result, in the following context, the **Ito integral $\int_S^T f_t \, dB_t$ is actually defined for processes f_t and filtration $\mathscr{F}_t$ such that**

$$\begin{cases} B_t \ is \ \mathscr{F}_t - BM \\ (t,\omega) \to f_t(\omega) \in \mathscr{B}_{\mathbb{R}_+} \times \mathscr{F} \\ \forall t, f_t \in \mathscr{F}_t \\ \mathbb{P}\left(\int_S^T f_t^2 \, dt < \infty \right) = 1 \end{cases} \tag{124}$$

So how does this generalization work? Actually still by approximation using elementary process but the convergence is expected in a weaker sense (converge in probability). Our task is to find a sequence of elementary process φ_t^n such that

$$\int_S^T |f_t - \varphi_t^n|^2 \, dt \xrightarrow{p} 0 \ (n \to \infty) \tag{125}$$

and the Ito integral is formed as the limit in probability

$$\int_S^T \varphi_t^n \, dB_t \xrightarrow{p} \int_S^T f_t \, dB_t \ (n \to \infty) \tag{126}$$

The approximation is basically the same as what we've done above. For continuous bounded process $f_t, \forall t, |f_t| \leq$

M, use the endpoint to construct a sequence of elementary processes at the truncated dyadic numbers t_j of $[S,T]$

$$\varphi_t^n = \sum_j f_{t_j}\mathbb{I}_{[t_j,t_{j+1})}(t) \tag{127}$$

and the approximation is well enough by the uniform continuity of f_t on $[S,T]$, for which $|f_t - f_{t_j}|$ is controlled by $\forall\varepsilon' > 0$

$$\begin{aligned}
\forall\varepsilon > 0, \mathbb{P}\left(\int_S^T |f_t - \varphi_t^n|^2\, dt > \varepsilon\right) &= \mathbb{P}\left(\int_S^T \left|\sum_j (f_t - f_{t_j})\mathbb{I}_{[t_j,t_{j+1})}(t)\right|^2\, dt > \varepsilon\right) &(128)\\
&\le \mathbb{P}\left(\sum_j \int_{t_j}^{t_{j+1}} (f_t - f_{t_j})^2\, dt > \varepsilon\right) &(129)\\
&\le \mathbb{P}\left(\varepsilon'^2(T-S) > \varepsilon\right) \to 0\ (n\to\infty) &(130)
\end{aligned}$$

note that here ε is fixed first, and when $n \to \infty$, the partition is fine enough and a smaller ε' can always be found such that it controls $|f_t - f_{t_j}|$ and $\varepsilon'^2(T-S) \le \varepsilon$. In other words, ε' can depend on n, making it possible to be much smaller than the fixed ε.

For bounded f_t, still use the convolution with the mollifier, and for general f_t, use a bounded truncation to approximate just as done above, but in the sense of convergence in probability. Note that since we are in the Hilbert space any longer, the convergence in probability has to be proved in an explicit way (much work to do). We can also show that the Ito's isometry still holds if $\mathbb{E}\int_S^T f_t^2\, dt = \infty$.

For all the details, refer to GTM 274.

The price of such extension is to **lose the martingality** and now we can only ensure that **the Ito process is a continuous local martingale** (the continuity of path is maintained). Local martingale X_t is defined in a way that there exists a sequence of stopping time $\tau_n \nearrow \infty$ $(n \to \infty)$ such that the stopped process $X_{t\wedge\tau_n}$ is martingale for each n. Note that there's no local MG in discrete time (countably many), so local MG is a special concept only for continuous time. If a process is a discrete-time local MG, then it must be a true MG.

The classical example of a local MG which is not a MG is the inverse Bessel process $X_t = \frac{1}{||B_t||}$, where B_t is a 3-dim BM does not start from origin. A less complicated example can be $M_t = ZB_t, Z \in \mathscr{F}_0, \mathbb{E}|Z| = \infty$.

Stratonovich Integral

In the definition of Ito integral, the left endpoint is always taken in the approximation to ensure the measurability property. The advantage is that Ito integrals are continuous local MG, but the disadvantage is that chain rule fails for Ito integral (quadratic variation as Ito correction term appears).

It's natural to ask if it's possible to take the right endpoint $f_{t_{j+1}}$ or the midpoint $\frac{f_{t_j}+f_{t_{j+1}}}{2}$ on the interval $[t_j, t_{j+1})$ to do the approximation. The **Fisk-Stratonovich integral** is defined as taking $(1-\varepsilon)f_{t_j} + \varepsilon f_{t_{j+1}}$ on each interval $[t_j, t_{j+1})$. When $\varepsilon = 0$, it's just Ito integral. When $\varepsilon = \frac{1}{2}$, it's the **Stratonovich integral**, denoted

$$\int f_s \circ dB_s \tag{131}$$

It satisfies the usual **chain rule** but we lose the martingale property (in the proof of martingality, $f_{t_j} \in \mathscr{F}_{t_j}$ is critical). To see the chain rule, compute the integral for BM:

$$\int_0^t B_s \circ dB_s = \lim_{n\to\infty} \sum_j \frac{B_{t_j} + B_{t_{j+1}}}{2}(B_{t_{j+1}} - B_{t_j}) \tag{132}$$

$$= \frac{1}{2} \lim_{n\to\infty} \sum_j (B^2_{t_{j+1}} - B^2_{t_j}) = \frac{B_t^2}{2} \tag{133}$$

This works as if we replace B_s with s and conclude that $\int_0^t s\, ds = \frac{t^2}{2}$.

Ito Formula

Ito formula provides a method to know about the behavior of $g(B_t)$, a function of BM or other processes by expanding it into stochastic integrals. The 1-dim Ito formula is formulated as

Theorem 3. *If $g \in C^2 : \mathbb{R} \to \mathbb{R}$, then $dg(B_t) = g'(B_t)\, dB_t + \frac{1}{2} g''(B_t)\, d\langle B, B\rangle_t$, where $\langle B, B\rangle_t$ is the quadratic variation of BM in time interval $[0, t]$ (so $\langle B, B\rangle_t = t$). $\langle M, M\rangle_t$ is generally defined for continuous local MG M_t as the unique increasing process such that $M_t^2 - \langle M, M\rangle_t$ is also a continuous local MG (analogue to Doob's MG decomposition).*

Proof. Let's prove the integral form since the terms are actually defined in the integral form:

$$g(B_t) - g(B_0) = \int_0^t g'(B_s)\, dB_s + \frac{1}{2}\int_0^t g''(B_s)\, ds \tag{134}$$

Let's naturally apply the Taylor expansion with Lagrange remainder with a telescoping form, t_j are truncated dyadic numbers in $[0, t]$

$$g(B_t) - g(0) = \sum_j g(B_{t_{j+1}}) - g(B_{t_j}) \tag{135}$$

$$= \sum_j \left[g'(B_{t_j})(B_{t_{j+1}} - B_{t_j}) + \frac{1}{2} g''(\xi_{t_j})(B_{t_{j+1}} - B_{t_j})^2 \right] \tag{136}$$

for some ξ_{t_j} between $B_{t_j}, B_{t_{j+1}}$.

The first term converges to $\int_0^t g'(B_s)\, dB_s$ in L^2 sense by the definition of Ito integral above directly. Now we prove that the second term converges in L^2 to $\frac{1}{2}\int_0^t g''(B_s)\,ds$:

$$\mathbb{E}\left(\frac{1}{2}\sum_j g''(\xi_{t_j})(B_{t_{j+1}}-B_{t_j})^2-\frac{1}{2}\sum_j g''(B_{t_j})(B_{t_{j+1}}-B_{t_j})^2\right)^2 \tag{137}$$

$$=\frac{1}{4}\mathbb{E}\left(\sum_j[g''(\xi_{t_j})-g''(B_{t_j})](B_{t_{j+1}}-B_{t_j})^2\right)^2 \tag{138}$$

$$=\frac{1}{4}\sum_{i,j}\mathbb{E}[g''(\xi_{t_j})-g''(B_{t_j})](B_{t_{j+1}}-B_{t_j})^2[g''(\xi_{t_i})-g''(B_{t_i})](B_{t_{i+1}}-B_{t_i})^2 \tag{139}$$

$$=\frac{1}{4}\sum_j\mathbb{E}[g''(\xi_{t_j})-g''(B_{t_j})]^2(B_{t_{j+1}}-B_{t_j})^4+\frac{1}{2}\sum_{i<j}\mathbb{E}[g''(\xi_{t_j})-g''(B_{t_j})](B_{t_{j+1}}-B_{t_j})^2\cdot\mathbb{E}[g''(\xi_{t_i})-g''(B_{t_i})](B_{t_{i+1}}-B_{t_i})^2 \tag{140}$$

the uniform continuity of g'' and the uniform continuity of BM path on interval $[0,t]$ ensures that $\forall j, |g''(\xi_{t_j})-g''(B_{t_j})|<\varepsilon$. Notice that $\sum_j\mathbb{E}(B_{t_{j+1}}-B_{t_j})^4=\sum_j 3(t_{j+1}-t_j)^2\to 0\ (n\to\infty)$ and that $\sum_{i<j}(t_{j+1}-t_j)(t_{i+1}-t_i)\le\frac{(t2^n)^2}{2^{2n}}=t^2<\infty$, so

$$\mathbb{E}\left(\frac{1}{2}\sum_j g''(\xi_{t_j})(B_{t_{j+1}}-B_{t_j})^2-\frac{1}{2}\sum_j g''(B_{t_j})(B_{t_{j+1}}-B_{t_j})^2\right)^2 \tag{141}$$

$$\le\frac{\varepsilon^2}{4}\sum_j 3(t_{j+1}-t_j)^2+\frac{\varepsilon^2}{2}\sum_{i<j}(t_{j+1}-t_j)(t_{i+1}-t_i)\to 0\ (n\to\infty) \tag{142}$$

Next we prove that $\frac{1}{2}\sum_j g''(B_{t_j})(B_{t_{j+1}}-B_{t_j})^2$ converges to $\frac{1}{2}\int_0^t g''(B_s)\,ds$ in the L^2 sense. Since g'' is continuous and BM has continuous path, $B_s, 0\le s\le t$ is bounded, so $|g''(B_{t_j})|\le M$ for uniform bound M.

$$\mathbb{E}\left(\frac{1}{2}\sum_j g''(B_{t_j})(t_{j+1}-t_j)-\frac{1}{2}\sum_j g''(B_{t_j})(B_{t_{j+1}}-B_{t_j})^2\right)^2 \tag{143}$$

$$=\frac{1}{4}\mathbb{E}\left(\sum_j g''(B_{t_j})[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]\right)^2 \tag{144}$$

$$=\frac{1}{4}\mathbb{E}\sum_{i,j}g''(B_{t_j})[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]g''(B_{t_i})[(t_{i+1}-t_i)-(B_{t_{i+1}}-B_{t_i})^2] \tag{145}$$

$$=\frac{1}{4}\sum_j\mathbb{E}[g''(B_{t_j})]^2[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]^2 \tag{146}$$

$$+\frac{1}{2}\sum_{i<j}\mathbb{E}g''(B_{t_j})[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]\cdot\mathbb{E}g''(B_{t_i})[(t_{i+1}-t_i)-(B_{t_{i+1}}-B_{t_i})^2] \tag{147}$$

$$\leq \frac{M^2}{4}\sum_j \mathbb{E}[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]^2 \tag{148}$$

$$+\frac{M^2}{2}\sum_{i<j} \mathbb{E}[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]\cdot\mathbb{E}[(t_{i+1}-t_i)-(B_{t_{i+1}}-B_{t_i})^2] \tag{149}$$

with calculations

$$\sum_j \mathbb{E}[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]^2 \tag{150}$$

$$=\sum_j (t_{j+1}-t_j)^2+\mathbb{E}(B_{t_{j+1}}-B_{t_j})^4-2(t_{j+1}-t_j)\mathbb{E}(B_{t_{j+1}}-B_{t_j})^2 \tag{151}$$

$$=\sum_j (t_{j+1}-t_j)^2+3(t_{j+1}-t_j)^2-2(t_{j+1}-t_j)^2\to 0\ (n\to\infty) \tag{152}$$

$$\tag{153}$$

$$\sum_{i<j} \mathbb{E}[(t_{j+1}-t_j)-(B_{t_{j+1}}-B_{t_j})^2]\cdot\mathbb{E}[(t_{i+1}-t_i)-(B_{t_{i+1}}-B_{t_i})^2] \tag{154}$$

$$=\sum_{i<j}[(t_{j+1}-t_j)-(t_{j+1}-t_j)]\cdot[(t_{i+1}-t_i)-(t_{i+1}-t_i)]\to 0\ (n\to\infty) \tag{155}$$

Combining all these estimations, we have proved that

$$\frac{1}{2}\sum_j g''(\xi_{t_j})(B_{t_{j+1}}-B_{t_j})^2 \xrightarrow{L^2} \frac{1}{2}\int_0^t g''(B_s)\,ds\ (n\to\infty) \tag{156}$$

thus the Ito formula holds.

□

Actually, the Ito formula can be extended in a parametric case.

Theorem 4. *If* $g(t,x)\in C^{1,2}$, $dg(t,B_t)=g_t(t,B_t)\,dt+g_x(t,B_t)\,dB_t+\frac{1}{2}g_{xx}(t,B_t)\,d\langle B,B\rangle_t$.

Proof. The structure of the proof is basically the same. Telescope, use 2-dim Taylor expansion and estimate the sums using integrals.

$$g(t,B_t)-g(0,B_0)=\sum_j g(t_{j+1},B_{t_{j+1}})-g(t_j,B_{t_j}) \tag{157}$$

$$=\sum_j\left[g_t(t_j,B_{t_j})(t_{j+1}-t_j)+g_x(t_j,B_{t_j})(B_{t_{j+1}}-B_{t_j})+\frac{1}{2}g_{xx}(t_j,B_{t_j})(B_{t_{j+1}}-B_{t_j})^2+...\right] \tag{158}$$

The reason we have not written all second order terms is that the other terms vanish

$$\mathbb{E}\left(\sum_j (t_{j+1}-t_j)(B_{t_{j+1}}-B_{t_j})\right)^2 = \sum_{i,j} \mathbb{E}(t_{j+1}-t_j)(B_{t_{j+1}}-B_{t_j})(t_{i+1}-t_i)(B_{t_{i+1}}-B_{t_i}) \tag{159}$$

$$= \sum_j (t_{j+1}-t_j)^2 \cdot \mathbb{E}(B_{t_{j+1}}-B_{t_j})^2 \tag{160}$$

$$+ 2\sum_{i<j} (t_{i+1}-t_i)(t_{j+1}-t_j)\cdot \mathbb{E}(B_{t_{j+1}}-B_{t_j})\cdot \mathbb{E}(B_{t_{i+1}}-B_{t_i}) \tag{161}$$

$$= \sum_j (t_{j+1}-t_j)^3 \tag{162}$$

$$\to 0 \ (n\to\infty) \tag{163}$$

$$\mathbb{E}\left(\sum_j (B_{t_{j+1}}-B_{t_j})^2\right)^2 = \sum_{i,j} \mathbb{E}(B_{t_{j+1}}-B_{t_j})^2(B_{t_{i+1}}-B_{t_i})^2 \tag{164}$$

$$= \sum_j \mathbb{E}(B_{t_{j+1}}-B_{t_j})^4 + 2\sum_{i<j} \mathbb{E}(B_{t_{j+1}}-B_{t_j})^2\cdot \mathbb{E}(B_{t_{i+1}}-B_{t_i})^2 \tag{165}$$

$$= \sum_j 3(t_{j+1}-t_j)^2 + 2\sum_{i<j}(t_{j+1}-t_j)^2(t_{i+1}-t_i)^2 \to 0 \ (n\to\infty) \tag{166}$$

similarly, the L^2 convergence still hold:

$$\sum_j g_t(t_j, B_{t_j})(t_{j+1}-t_j) \xrightarrow{L^2} \int_0^t g_t(s, B_s)\, ds \ (n\to\infty) \tag{167}$$

$$\sum_j g_x(t_j, B_{t_j})(B_{t_{j+1}}-B_{t_j}) \xrightarrow{L^2} \int_0^t g_x(s, B_s)\, dB_s \ (n\to\infty) \tag{168}$$

$$\sum_j \frac{1}{2} g_{xx}(t_j, B_{t_j})(B_{t_{j+1}}-B_{t_j})^2 \xrightarrow{L^2} \frac{1}{2}\int_0^t g_{xx}(s, B_s)\, d\langle B, B\rangle_s \ (n\to\infty) \tag{169}$$

□

Remark. *It's easy to see that in notation, we can write $dtdt = 0, dtdB_t = 0, dB_t dB_t = dt$. This is due to the fact that $f_t = t$ is of finite variation, so the second variation must be 0 and that the cross variation of $f_t = t$ and B_t is 0. This also explains why there are no higher order terms in the Ito formula.*

An immediate generalization of the Ito formula with a time parameter is that we can replace the time t with ***any finite variation process*** *f_t to let the Ito formula work for things like $g(f_t, B_t)$.*

Example

The first example is to compute

$$d(e^{bt+\sigma B_t}) \tag{170}$$

where b, σ are constants for drift and diffusion, set $g(t,x) = e^{bt+\sigma x}$ and apply the Ito formula

$$d(e^{bt+\sigma B_t}) = be^{bt+\sigma B_t}\,dt + \sigma e^{bt+\sigma B_t}\,dB_t + \frac{\sigma^2}{2}e^{bt+\sigma B_t}\,dt \tag{171}$$

As a result, if define $X_t = e^{bt+\sigma B_t}X_0$, then this X_t is just the solution to the SDE

$$dX_t = \left(b + \frac{\sigma^2}{2}\right)X_t\,dt + \sigma X_t\,dB_t \tag{172}$$

which defines a geometric BM and is closely related to the Black-Scholes model (μ as mean return and σ as volatility)

$$dX_t = \mu X_t\,dt + \sigma X_t\,dB_t \tag{173}$$

Ito formula for Ito Process

The **Ito process** is defined as

$$X_t = x_0 + \int_0^t \psi_s\,ds + \int_0^t \varphi_s\,dB_s \tag{174}$$

$$\psi_s, \varphi_s \in \mathscr{F}_s, \mathbb{E}\int_0^t \varphi_s^2\,ds < \infty, \mathbb{E}\int_0^t |\psi_s|\,ds < \infty \tag{175}$$

a constant x_0 plus the Stieljes integral of a process and plus the stochastic integral of another process. These two processes are both adapted, with φ_s in $L^2([0,t]\times\Omega)$ where the stochastic integral is defined and ψ_s such that the Stieljes integral part has finite expectation.

To go into more details, note that

$$\sum_j \left|\int_0^{t_{j+1}} \psi_s\,ds - \int_0^{t_j} \psi_s\,ds\right| \le \sum_j \int_{t_j}^{t_{j+1}} |\psi_s|\,ds = \int_0^t |\psi_s|\,ds < \infty \ a.s. \tag{176}$$

so the $\int_0^t \psi_s\,ds$ part is a **finite variation process**, i.e. it contributes nothing to the quadratic variation of the whole process. As proved above, the $\int_0^t \varphi_s\,dB_s$ part is a **continuous MG** which is typically not finite variation (if a finite variation process is continuous local MG, it must be constant almost surely). In simple words, **the Ito process is made up of constant part, finite variation part, continuous MG part**.

Theorem 5. *If $g(t,x) \in C^{1,2}$, then $dg(t,X_t) = g_t(t,X_t)\,dt + g_x(t,X_t)\,dX_t + \frac{1}{2}g_{xx}(t,X_t)\,d\langle X,X\rangle_t$ with $d\langle X,X\rangle_t = \varphi_t^2\,dt$.*

Proof. The proof is the same as that for BM above. The only thing to verify now is the quadratic variation of X_t. By previous calculations, only the term $\int_0^t \varphi_s \, dB_s$ contributes to the quadratic variation.

$$\mathbb{E}\sum_j (X_{t_{j+1}} - X_{t_j})^2 = \sum_j \mathbb{E}\left(\int_{t_j}^{t_{j+1}} \varphi_s \, dB_s\right)^2 \tag{177}$$

$$= \sum_j \mathbb{E}\int_{t_j}^{t_{j+1}} \varphi_s^2 \, ds \tag{178}$$

$$= \mathbb{E}\int_0^t \varphi_s^2 \, ds \tag{179}$$

$$\forall t, \mathbb{E}\langle X, X\rangle_t = \mathbb{E}\int_0^t \varphi_s^2 \, ds \tag{180}$$

by the Ito's isometry and the monotone convergence theorem. So it's reasonable to guess that

$$\langle X, X\rangle_t = \int_0^t \varphi_s^2 \, ds \tag{181}$$

The proof can be given in the following sense that when φ_s is a bounded process,

$$\mathbb{E}\left(\sum_j \varphi_{t_j}^2 (B_{t_{j+1}} - B_{t_j})^2 - \sum_j \varphi_{t_j}^2 (t_{j+1} - t_j)\right)^2 \tag{182}$$

$$= \sum_{i,j} \mathbb{E}\left(\varphi_{t_j}^2 [(B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)]\varphi_{t_i}^2 [(B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)]\right) \tag{183}$$

$$= \sum_j \mathbb{E}\left(\varphi_{t_j}^4 [(B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)]^2\right) \tag{184}$$

$$+ 2\sum_{i<j} \mathbb{E}\left[\varphi_{t_i}^2 [(B_{t_{i+1}} - B_{t_i})^2 - (t_{i+1} - t_i)] \cdot \mathbb{E}\left(\varphi_{t_j}^2 [(B_{t_{j+1}} - B_{t_j})^2 - (t_{j+1} - t_j)] \Big| \mathscr{F}_{t_i}\right)\right] \tag{185}$$

$$\to 0 \ (n \to \infty) \tag{186}$$

since there are only terms having order no lower than $dtdt, dtdB_t$. Use truncation and dominated convergence theorem to prove the result for general φ_s.

□

Example

Consider the exponential MG of BM:

$$M_t = e^{\int_0^t h_s \, dB_s - \frac{1}{2}\int_0^t h_s^2 \, ds} \tag{187}$$

with h_s to be a bounded process. Specify $M_t = e^{X_t}, X_t = \int_0^t h_s \, dB_s - \frac{1}{2}\int_0^t h_s^2 \, ds$ and set $g(x) = e^x$ to get

$$dM_t = e^{X_t} \, dX_t + \frac{1}{2} e^{X_t} \, d\langle X, X\rangle_t \tag{188}$$

recall that $\langle X, X\rangle_t = \int_0^t h_s^2 \, ds$, and

$$dX_t = h_t \, dB_t - \frac{1}{2} h_t^2 \, dt \tag{189}$$

so

$$dM_t = e^{X_t} h_t \, dB_t \tag{190}$$

This is telling us that this M_t is generally a **continuous local MG**

$$M_t = M_0 + \int_0^t e^{X_s} h_s \, dB_s, \; M_0 = 1 \tag{191}$$

with

$$\langle M, M\rangle_t = \int_0^t e^{2X_s} h_s^2 \, ds \tag{192}$$

As a result, if $\forall t, \mathbb{E}\int_0^t e^{2X_s} h_s^2 \, ds < \infty$ (h_s is bounded suffices), the continuous local MG satisfies $\forall t, \mathbb{E}\langle M, M\rangle_t < \infty$ so such M_t must be a L^2 MG. In such case, the M_t is a **natural extension of the exponential MG** (in the original setting, h_s is constant but now it can be a bounded process).

Remark. *For continuous local MG M_t, $\forall t, \mathbb{E}\langle M, M\rangle_t < \infty$ is equivalent to M_t being L^2 MG and $\mathbb{E}\langle M, M\rangle_\infty < \infty$ is equivalent to M_t being L^2 bounded MG. For more detailed conditions on M_t being a MG, refer to Kazamaki and Novikov conditions.*

Multi-dimensional Ito Formula

Since Ito process is a more general setting than a function of BM or a function of both time and BM (Ito process is semi-MG), we only describe the Ito formula for Ito process. First set up **the d-dimensional Ito process as the integral w.r.t. m-dimensional BM** as following:

$$X_t^i = x_0^i + \int_0^t \psi_s^i \, ds + \sum_{k=1}^m \int_0^t \varphi_s^{i,k} \, dB_s^k \; (k = 1, ..., m, \; i = 1, 2, ..., d) \tag{193}$$

with the explanation that such Ito process lives in the space $\mathbb{R}^d$ and is constructed by the stochastic integral w.r.t. a m-dimensional BM $B_s = (B_s^1, ..., B_s^m)$. To write it in a more compact form, introduce the notation that

$$X_t = x_0 + \int_0^t \psi_s \, ds + \int_0^t \varphi_s \cdot dB_s \tag{194}$$

$$X_t, x_0, \psi_s \in \mathbb{R}^d, \varphi_s \in \mathbb{R}^{d\times m} \tag{195}$$

here $\varphi_s^{i,k}$ stands for the process φ used to construct the i-th coordinate of Ito process as a stochastic integral w.r.t. the k-th coordinate of m-dimensional BM.

Theorem 6. *If vector-valued function $g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^p \in C^{1,2}$, then multi-dimensional Ito formula holds*

$$dg^i(t, X_t) = \partial_t g^i(t, X_t) \, dt + \nabla_x g^i(t, X_t) \cdot dX_t + \frac{1}{2} Tr(\varphi_t^T H \varphi_t) \, dt \tag{196}$$

$$= \partial_t g^i(t, X_t) \, dt + \sum_{k=1}^d \partial_{x_k} g^i(t, X_t) \, dX_t^k + \frac{1}{2} \sum_{j,k=1}^d \partial_{x_j, x_k} g^i(t, X_t) \, d\langle X^j, X^k\rangle_t \tag{197}$$

where $\varphi_t \in \mathbb{R}^{d\times m}$ is a matrix and $H_{d\times d}$ is the Hessian of g restricted on its action on $x \in \mathbb{R}^d$ at (t, X_t). The bracket $d\langle X^j, X^k\rangle_t = \sum_{l=1}^m \varphi_t^{j,l} \varphi_t^{k,l} \, dt$.

Proof. The structure of the proof is still exactly the same as it is in the 1-dimensional case. The only two things to be verified is that $dB_t^p \, dB_t^q = 0$ $(p \neq q)$ and $d\langle X^j, X^k\rangle_t = \sum_{l=1}^m \varphi_t^{j,l} \varphi_t^{k,l} \, dt$.

$$\mathbb{E}\left(\sum_j (B_{t_{j+1}}^p - B_{t_j}^p)(B_{t_{j+1}}^q - B_{t_j}^q)\right)^2 \tag{198}$$

$$= \mathbb{E} \sum_{i,j} (B_{t_{j+1}}^p - B_{t_j}^p)(B_{t_{j+1}}^q - B_{t_j}^q)(B_{t_{i+1}}^p - B_{t_i}^p)(B_{t_{i+1}}^q - B_{t_i}^q) \tag{199}$$

$$= \sum_j \mathbb{E}(B_{t_{j+1}}^p - B_{t_j}^p)^2 (B_{t_{j+1}}^q - B_{t_j}^q)^2 + 2\sum_{i<j} \mathbb{E}(B_{t_{j+1}}^p - B_{t_j}^p)(B_{t_{j+1}}^q - B_{t_j}^q)(B_{t_{i+1}}^p - B_{t_i}^p)(B_{t_{i+1}}^q - B_{t_i}^q) \tag{200}$$

$$= \sum_j \mathbb{E}(B_{t_{j+1}}^p - B_{t_j}^p)^2 \cdot \mathbb{E}(B_{t_{j+1}}^q - B_{t_j}^q)^2 + 2\sum_{i<j} \mathbb{E}(B_{t_{j+1}}^p - B_{t_j}^p) \cdot \mathbb{E}(B_{t_{j+1}}^q - B_{t_j}^q) \cdot \mathbb{E}(B_{t_{i+1}}^p - B_{t_i}^p) \cdot \mathbb{E}(B_{t_{i+1}}^q - B_{t_i}^q) \tag{201}$$

$$= \sum_j \mathbb{E}^2(B_{t_{j+1}}^p - B_{t_j}^p)^2 + 2\sum_{i<j} \mathbb{E}^2(B_{t_{j+1}}^p - B_{t_j}^p) \cdot \mathbb{E}^2(B_{t_{i+1}}^p - B_{t_i}^p) \tag{202}$$

$$= \sum_j (t_{j+1} - t_j)^2 \to 0 \ (n \to \infty) \tag{203}$$

For the bracket of X_t, it's still true that only $\varphi_s^{i,k}$ contributes to the quadratic variation.

$$\left\langle \sum_{l=1}^m \int_0^t \varphi_s^{j,l}\, dB_s^l, \sum_{p=1}^m \int_0^t \varphi_s^{k,p}\, dB_s^p \right\rangle_t \tag{204}$$

$$= \sum_{l,p=1}^m \left\langle \int_0^t \varphi_s^{j,l}\, dB_s^l, \int_0^t \varphi_s^{k,p}\, dB_s^p \right\rangle_t \tag{205}$$

$$= \sum_{l=1}^m \left\langle \int_0^t \varphi_s^{j,l}\, dB_s^l, \int_0^t \varphi_s^{k,l}\, dB_s^l \right\rangle_t = \sum_{l=1}^m \int_0^t \varphi_s^{j,l} \varphi_s^{k,l}\, ds \tag{206}$$

by using the property just derived that independent BM has bracket 0. □

Remark. *The isometry property of stochastic integral and the bracket may make it much easier to calculate. For general semi-MG M, N and $f_s, g_s \in L^2([0,t]\times\Omega)$(process for which stochastic integral is well-defined),*

$$\left\langle \int_0^\cdot f_s\, dM_s, \int_0^\cdot g_s\, dN_s \right\rangle_t = \int_0^t f_s g_s\, d\langle M, N\rangle_s \tag{207}$$

(See GTM 274, P101, Theorem 5.4) .Using this property, all calculations on the brackets of stochastic integrals are trivial.

Applications

The integration by parts is an application of multi-dimensional Ito formula. Consider the 2-dim Ito process constructed using the stochastic integral w.r.t. 2-dim BM.

$$X_t^1 = x_0^1 + \int_0^t \psi_s^1\, ds + \int_0^t \varphi_s^{1,1}\, dB_s^1 + \int_0^t \varphi_s^{1,2}\, dB_s^2 \tag{208}$$

$$X_t^2 = x_0^2 + \int_0^t \psi_s^2\, ds + \int_0^t \varphi_s^{2,1}\, dB_s^1 + \int_0^t \varphi_s^{2,2}\, dB_s^2 \tag{209}$$

calculate $d(X_t^1 X_t^2)$ setting $g(x_1, x_2) = x_1 x_2$ to find

$$d(X_t^1 X_t^2) = X_t^2\, dX_t^1 + X_t^1\, dX_t^2 + \frac{1}{2}\, dX_t^1\, dX_t^2 + \frac{1}{2}\, dX_t^2\, dX_t^1 \tag{210}$$

$$= X_t^2\, dX_t^1 + X_t^1\, dX_t^2 + \varphi_t^{1,1}\varphi_t^{2,1}\, dt + \varphi_t^{1,2}\varphi_t^{2,2}\, dt \tag{211}$$

the **integration by parts** formula.

Another trivial example is to calculate the moments of 1-dim BM B_t (the superscripts here are powers).

To get $\mathbb{E}B_t^2$, we can think about expanding dB_t^2 using Ito formula with $g(x) = x^2$:

$$dB_t^2 = 2B_t\,dB_t + dt \tag{212}$$

$$B_t^2 = 2\int_0^t B_s\,dB_s + t \tag{213}$$

taking expectation on both sides to get:

$$\mathbb{E}B_t^2 = 2\mathbb{E}\int_0^t B_s\,dB_s + t = t \tag{214}$$

note that $\mathbb{E}\int_0^t B_s\,dB_s = 0$ follows from the fact that Ito integral process (denoted M_t above) is MG. As a result, **for any $\int_0^t f_s\,dB_s, f_s \in V(0,t)$, its expectation is always 0.**

Ito formula can also be applied to B_t^4:

$$dB_t^4 = 4B_t^3\,dB_t + 6B_t^2\,dt \tag{215}$$

$$B_t^4 = 4\int_0^t B_s^3\,dB_s + 6\int_0^t B_s^2\,ds \tag{216}$$

taking expectation on both sides to get

$$\mathbb{E}B_t^4 = 6\mathbb{E}\int_0^t B_s^2\,ds = 6\int_0^t s\,ds = 3t^2 \tag{217}$$

with the interchange of expectation and integral ensured by Fubini.

The last example is the **n-dimensional Bessel process for** $n \geq 2$. Consider $B_t = (B_t^1, ..., B_t^n)$ to be n-dimensional BM and $R_t = ||B_t||_2$ is the Euclidean distance to the origin. Ito formula is applied for $g(x) = ||x||_2$.

$$dR_t = \sum_{j=1}^n \frac{B_t^j}{R_t}\,dB_t^j + \frac{1}{2}\sum_{j=1}^n \frac{(R_t)^2 - (B_t^j)^2}{(R_t)^3}\,dt \tag{218}$$

$$= \sum_{j=1}^n \frac{B_t^j}{R_t}\,dB_t^j + \frac{n-1}{2R_t}\,dt \tag{219}$$

Remark. *The C^2 assumption in Ito formula can be weakened (as long as the process almost surely does not touch the singularity it's fine).*

In the example above, since for $n \geq 2$, almost surely BM never hits the origin, the Ito formula still holds. For the same reason, as long as X_t is positive almost surely, we can also apply Ito formula for $\log X_t$ (the situation in solving Black-Scholes model).

Martingale Representation Theorem

The motivation of MG Rep Thm is natural: since we have already proved that the process of Ito integral for good enough $f_s \in L^2([0,t]\times\Omega)$

$$M_t = \int_0^t f_s\,dB_s \tag{220}$$

is a L^2 continuous MG, can we represent any L^2 continuous MG as the Ito integral of some process f_s? A slight detail is that by martingality, $\forall t, \mathbb{E}\int_0^t f_s\,dB_s = 0$. So if the MG M_t we want to represent is not starting from 0 at time 0, at least we shall subtract the starting point, i.e. to find $f_s \in L^2([0,t]\times\Omega)$ such that

$$\forall t, M_t - M_0 = \int_0^t f_s\,dB_s \tag{221}$$

The MG Rep Thm starts with a weakened version, which is the Ito Representation theorem, stating that any L^2 random variable can be represented as the stochastic integral of a process.

Theorem 7. *If $F \in L^2(\mathscr{F}_T)$ for a fixed time T, then unique $\exists f_t \in L^2([0,T]\times\Omega)$ such that $F = \mathbb{E}F + \int_0^T f_s\,dB_s$.* ***(Ito's Representation Theorem)***

Proof. The proof starts by considering a special family of random variables as extended exponential MG of BM

$$F = e^{\int_0^T h(s)\,dB_s - \frac{1}{2}\int_0^T h^2(s)\,ds} \tag{222}$$

for a deterministic function $h(s)$ (such h should be such that F is well-defined).

Let's consider the integral process $Y_t = e^{\int_0^t h(s)\,dB_s - \frac{1}{2}\int_0^t h^2(s)\,ds}$ changing with time t, Ito formula gives

$$X_t = \int_0^t h(s)\,dB_s - \frac{1}{2}\int_0^t h^2(s)\,ds \tag{223}$$

$$dY_t = e^{X_t}\,dX_t + \frac{1}{2}e^{X_t}\,d\langle X,X\rangle_t \tag{224}$$

$$dX_t = h(t)\,dB_t - \frac{1}{2}h^2(t)\,dt \tag{225}$$

$$d\langle X,X\rangle_t = h^2(t)\,dt \tag{226}$$

so change it into the integral form

$$dY_t = Y_t h(t)\,dB_t \tag{227}$$

$$Y_t - Y_0 = \int_0^t Y_s h(s)\,dB_s \tag{228}$$

$$Y_0 = 1 \tag{229}$$

Setting $f_s = Y_s h(s)$ ends the proof since by martingality $\mathbb{E}F = \mathbb{E}Y_0 = 1$, so $F = \mathbb{E}F + \int_0^T Y_s h(s)\,dB_s$.

Now let's notice that

$$span\left\{e^{\int_0^T h(s)\,dB_s-\frac{1}{2}\int_0^T h^2(s)\,ds}, h\in L^2([0,T])\right\} \tag{230}$$

is a dense subset in $L^2(\Omega, \mathscr{F}_T)$ (with h deterministic). By admitting this, a general F can be approximated by linear combinations of the random variables having the form of exponential MG.

$$\exists c_i\in\mathbb{R}, F_n=\sum_{i=1}^n c_iM_i \xrightarrow{L^2} F\ (n\to\infty) \tag{231}$$

$$M_i\in\left\{e^{\int_0^T h(s)\,dB_s-\frac{1}{2}\int_0^T h^2(s)\,ds}, h\in L^2([0,T])\right\} \tag{232}$$

linearity ensures that the Ito representation theorem still holds for F_i (note that F_i does not necessarily have the form $e^{\int_0^T h(s)\,dB_s-\frac{1}{2}\int_0^T h^2(s)\,ds}$ for deterministic h)

$$\exists f_s^n, F_n=\mathbb{E}F_n+\int_0^T f_s^n\,dB_s \tag{233}$$

it's natural to think about taking the limit of f_s^n as the process that represents F. So we have to figure out whether this sequence of process converge in the $L^2([0,T]\times\Omega)$ sense.

$$\mathbb{E}(F_m-F_n)^2=\mathbb{E}\left(\mathbb{E}F_m-\mathbb{E}F_n+\int_0^T f_s^m-f_s^n\,dB_s\right)^2 \tag{234}$$

$$=\mathbb{E}^2(F_m-F_n)+2\mathbb{E}(F_m-F_n)\cdot\mathbb{E}\left(\int_0^T(f_s^m-f_s^n)\,dB_s\right)+\mathbb{E}\left(\int_0^T(f_s^m-f_s^n)\,dB_s\right)^2 \tag{235}$$

$$=\mathbb{E}^2(F_m-F_n)+\mathbb{E}\left(\int_0^T(f_s^m-f_s^n)\,dB_s\right)^2 \tag{236}$$

notice that $F_n\xrightarrow{L^2}F\ (n\to\infty)$, this is telling us that

$$\mathbb{E}(F_m-F_n)^2\to 0\ (m,n\to\infty) \tag{237}$$

to know immediately with Ito's isometry that

$$\mathbb{E}\left(\int_0^T(f_s^m-f_s^n)\,dB_s\right)^2\to 0\ (m,n\to\infty) \tag{238}$$

$$\mathbb{E}\int_0^T(f_s^m-f_s^n)^2\,ds\to 0\ (m,n\to\infty) \tag{239}$$

$$||f_s^m-f_s^n||_{L^2([0,T]\times\Omega)}\to 0\ (m,n\to\infty) \tag{240}$$

so the completeness of Hilbert space ensures that $f_s^n \overset{L^2([0,T]\times\Omega)}{\to} f_s$ $(n \to \infty)$ converges to some limit process $f_s \in L^2([0,T]\times\Omega)$.

Verify that this limit gives the representation for F.

$$\mathbb{E}F + \int_0^T f_s\,dB_s = \mathbb{E}\lim_{n\to\infty} F_n + \int_0^T \lim_{n\to\infty} f_s^n\,dB_s \tag{241}$$

$$= \lim_{n\to\infty}\mathbb{E}F_n + \lim_{n\to\infty}\int_0^T f_s^n\,dB_s \tag{242}$$

$$= \lim_{n\to\infty} F_n = F \tag{243}$$

where the limit holds in L^2 sense and the second equation is due to the L^2 convergence of F_n and the convergence of stochastic integral.

Eventually, let's show that

$$\int_0^T f_s^n\,dB_s \overset{L^2}{\to} \int_0^T f_s\,dB_s \ (n\to\infty) \tag{244}$$

to complete the proof. By Ito's isometry,

$$\mathbb{E}\left(\int_0^T (f_s^n - f_s)\,dB_s\right)^2 = \mathbb{E}\int_0^T (f_s^n - f_s)^2\,ds \to 0 \ (n\to\infty) \tag{245}$$

and that's the end of the proof for existence (except the dense subset proposition).

For uniqueness, still use Ito's isometry

$$\int_0^T f_s^1\,dB_s = \int_0^T f_s^2\,dB_s \tag{246}$$

$$\int_0^T (f_s^1 - f_s^2)\,dB_s = 0 \tag{247}$$

$$||f_s^1 - f_s^2||_{L^2([0,T]\times\Omega)} = 0 \tag{248}$$

$$f_t^1 = f_t^2 \ a.a.\ (t,\omega) \tag{249}$$

$\square$

Theorem 8. *For any L^2 continuous MG M_t, there always exists $f_t \in L^2([0,T]\times\Omega)$ such that $M_t = M_0 + \int_0^t f_s\,dB_s$.* ***(Martingale Representation Theorem)***

Proof. By Ito's representation theorem, for any fixed time t, always exists $f_s^t \in L^2([0,t]\times\Omega)$ such that

$$M_t - M_0 = \int_0^t f_s^t\,dB_s \tag{250}$$

with the process f_s^t depending on the fixed time point t. The next work is to prove that such process actually doesn't need to depend on t. Since we have not yet used martingality of M_t, try to apply it for $\forall 0 \leq t_1 \leq t_2$ to get

$$
\begin{aligned}
M_{t_1} &= \mathbb{E}(M_{t_2}|\mathscr{F}_{t_1}) && (251)\\
&= M_0 + \mathbb{E}\left(\int_0^{t_2} f_s^{t_2}\, dB_s \middle| \mathscr{F}_{t_1}\right) && (252)\\
&= M_0 + \int_0^{t_1} f_s^{t_2}\, dB_s && (253)
\end{aligned}
$$

Compared with the representation of M_{t_1} to conclude

$$\int_0^{t_1} f_s^{t_1}\, dB_s \overset{L^2}{=} \int_0^{t_1} f_s^{t_2}\, dB_s \tag{254}$$

and by Ito's isometry get

$$
\begin{aligned}
\mathbb{E}\int_0^{t_1} (f_s^{t_1} - f_s^{t_2})^2\, ds = 0 && (255)\\
f_s^{t_1} = f_s^{t_2} \ a.a.\ (s,\omega) && (256)
\end{aligned}
$$

so this f_s^t can be modified such that it does not depend on time t, and the proof is done.

□

A Lemma for Dense Subset

In the proof above, a crucial criterion is that the linear span of the set consisting of exponential MG with deterministic L^2 function $h(s)$, i.e.

$$span\left\{e^{\int_0^T h(s)\, dB_s - \frac{1}{2}\int_0^T h^2(s)\, ds}, h \in L^2([0,T])\right\} \tag{257}$$

is a dense subset in $L^2(\Omega, \mathscr{F}_T)$ for fixed time T and deterministic h. Let's prove this lemma here to complete the whole proof for Ito's representation theorem.

The first lemma shows that for fixed time T, $L^2(\Omega, \mathscr{F}_T)$ has a dense subset consisting of all smooth and compactly supported functionals of finitely many time points of BM.

Lemma 2. *For fixed time T,*

$$\{\phi(B_{t_1}, ..., B_{t_n}) : t_i \in [0,T], \phi \in C_0^\infty, \forall n \in \mathbb{N}\} \tag{258}$$

is a dense subset of $L^2(\Omega, \mathscr{F}_T)$.

Proof. $\forall g \in L^2(\Omega, \mathscr{F}_T)$, the main thought is to consider its projection onto the filtration spanned by the time points of BM. Denote $\mathscr{H}_n = \sigma(B_{t_1}, ..., B_{t_n})$, then $\mathscr{H}_n$ is a filtration with the projection of g to be $X_n = \mathbb{E}(g|\mathscr{H}_n)$.

It's obvious that X_n is a closed MG, so it converges $a.s.$ and in L^1 to $X_\infty = \mathbb{E}(g|\mathscr{F}_T) = g$. Since g is L^2, $\sup_n \mathbb{E}^2(g|\mathscr{H}_n) \leq \mathbb{E}g^2 < \infty$, by MG converegence theorem, this convergence is actually in L^2. So we just have to prove that $\mathbb{E}(g|\mathscr{H}_n)$ can be approximated in L^2 by the elements in the set.

Now that $\mathbb{E}(g|\mathscr{H}_n) \in \mathscr{H}_n$, so exists g_n Borel measurable such that $\mathbb{E}(g|\mathscr{H}_n) = g_n(B_{t_1}, ..., B_{t_n})$. Now consider compactly supported function $h_n(x) = g_n(x)\mathbb{I}_{||x||<n}$ as approximation of g_n at $(B_{t_1}, ..., B_{t_n})$. To get smoothness, just find a mollifier and take the convolution to conclude.

□

Lemma 3. *For fixed time T,*

$$span\left\{e^{\int_0^T h(s)\,dB_s - \frac{1}{2}\int_0^T h^2(s)\,ds}, h \in L^2([0,T])\right\} \tag{259}$$

for deterministic h is dense in $L^2(\Omega, \mathscr{F}_T)$.

Proof. By Hilbert space theory, to prove that it's a dense subset, just has to prove that its orthogonal complement is trivial. If $g \in L^2(\Omega, \mathscr{F}_T)$ is orthogonal to all elements in this set, want to prove that $g = 0$ $a.s.$.

Such g should satisfy

$$\forall \lambda, G(\lambda) = \int e^{\lambda_1 B_{t_1} + ... + \lambda_n B_{t_n}} g\,d\mathbb{P} = 0 \tag{260}$$

and extend all these variables as complex variables to see that

$$\forall z \in \mathbb{C}^n, G(z) = \int e^{z_1 B_{t_1} + ... + z_n B_{t_n}} g\,d\mathbb{P} \tag{261}$$

is actually holomorphic and is always 0 when all components are real, so G is always 0 on the whole $\mathbb{C}^n$ (isolation of the zeros of holomorphic function).

Let $\hat{\phi}$ be the Fourier transform of ϕ, so for $\forall \phi \in C_0^\infty, \forall y \in \mathbb{R}^n$,

$$\int \phi(B_{t_1}, ..., B_{t_n}) g\,d\mathbb{P} = (2\pi)^{-\frac{n}{2}} \int \hat{\phi}(y) G(iy)\,dy = 0 \tag{262}$$

by replacing ϕ by its inverse Fourier transform, using Fubini to change the integration order and replacing once again with G. (just the form of c.f. and Levy's inversion formula) Since all $\phi(B_{t_1}, ..., B_{t_n})$ form a dense subset, $g = 0$ and the lemma is proved.

□

Week 4

Stochastic Differential Equation

A general SDE in $\mathbb{R}^n$ with initial condition looks like

$$\begin{cases} dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \\ X_0 = x \end{cases} \tag{263}$$

where $x, X_t \in \mathbb{R}^n, b : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ and B_t is m-dim BM. Here b is the **drift coefficient** and σ is the **volatility coefficient**. Notice the difference between SDE and a general Ito process

$$X_t = x + \int_0^t \psi_s\,dB_s + \int_0^t \varphi_s\,ds \tag{264}$$

lies in the fact that b, σ are functions of the unknown process X_t while ψ_s, φ_s are known.

To prove the existence and uniqueness of the solution to SDE under special conditions, let's first state the Grownwall's inequality as a tool.

Theorem 9. ***(Grownwall's inequality)*** *Assume v is defined on interval $[a, +\infty)$ and is continuous with $A, F \in \mathbb{R}$. If $v(t) \le F + A\int_a^t v(s)ds$, then*

$$v(t) \le Fe^{A(t-a)} \tag{265}$$

Assume v, β are defined on interval $[a, +\infty)$ and v is differentiable in the interior. If $v'(t) \le \beta(t)\nu(t)$, then

$$v(t) \le v(a)e^{\int_a^t \beta(s)ds} \tag{266}$$

Proof. Let's first prove the differential form. Set $u(t) = e^{\int_a^t \beta(s)\,ds}$ to see that $u'(t) = u(t)\beta(t)$ with $u(a) = 1, u(t) > 0$. Conside the derivative of their quotient

$$\frac{d}{dt}\frac{v(t)}{u(t)} = \frac{v'(t)u(t) - v(t)u(t)\beta(t)}{u^2(t)} \le 0 \tag{267}$$

to find $\forall t \ge a$, $\frac{v(t)}{u(t)} \le \frac{v(a)}{u(a)} = v(a)$. That's why

$$v(t) \le v(a)e^{\int_a^t \beta(s)\,ds} \tag{268}$$

For the integral form, start by constructing (WLOG assume $A > 0$)

$$u(t) = e^{-A(t-a)}\int_a^t Av(s)\,ds \tag{269}$$

to find that $u(a) = 0$ and

$$u'(t) = Ae^{-A(t-a)} \left(v(t) - \int_a^t Av(s)\, ds \right) \leq AFe^{-A(t-a)} \tag{270}$$

integrate from a to t on both sides to get

$$u(t) - u(a) \leq AF \int_a^t e^{-A(s-a)}\, ds \tag{271}$$

$$u(t) \leq F(1 - e^{-A(t-a)}) \tag{272}$$

now turn back to v and take the derivative to conclude

$$A \int_a^t v(s)\, ds \leq F(e^{A(t-a)} - 1) \tag{273}$$

$$v(t) \leq Fe^{A(t-a)} \tag{274}$$

□

Remark. *A special case is when v is continuous function and $F = 0$, so if $v(t) \leq A \int_0^t v(s)\, ds$, then $v \leq 0$.*

Basically the spirit of Grownwall is that if the derivative of a function $v'(t)$ is bounded by a multiple of its function value $\beta(t)v(t)$, then the function actually has an upper bound which is just the solution to the ODE $v'(t) = \beta(t)v(t)$.

The next theorem states the existence and uniqueness of the solution to a general SDE with some conditions.

Theorem 10. *Fix time $T > 0$ and assume $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are deterministic functions with bounded time variable T and satisfy*

$$\exists c > 0, \forall t \in [0, T], \forall x \in \mathbb{R}^n, |b(t, x)| + |\sigma(t, x)| \leq c(1 + |x|) \tag{275}$$

$$\exists D > 0, \forall t \in [0, T], \forall x, y \in \mathbb{R}^n, |b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y| \tag{276}$$

*then the SDE has unique solution in $L^2([0, T] \times \Omega)$ that has continuous sample path. The first condition is called **growth condition** and the second condition is called **Lipschitz condition**.*

Proof. First prove the **uniqueness**. If there are two solutions to the SDE: $X_t, \tilde{X}_t$. Consider plugging them in the function b, σ and form the difference process

$$\alpha_s = b(s, X_s) - b(s, \tilde{X}_s) \tag{277}$$

$$\gamma_s = \sigma(s, X_s) - \sigma(s, \tilde{X}_s) \tag{278}$$

and consider

$$dX_t - d\tilde{X}_t = \alpha_t\, dt + \gamma_t\, dB_t \tag{279}$$

to prove that these two solutions are the same in L^2, turn it into the integral form

$$X_t - \tilde{X}_t = \int_0^t \alpha_s \, ds + \int_0^t \gamma_s \, dB_s \tag{280}$$

and compute the L^2 norm to prove that it converges to 0

$$\begin{aligned}
\mathbb{E}(X_t - \tilde{X}_t)^2 &= \mathbb{E}\left(\int_0^t \alpha_s \, ds + \int_0^t \gamma_s \, dB_s\right)^2 &(281)\\
&\leq 2\mathbb{E}\left(\int_0^t \alpha_s \, ds\right)^2 + 2\mathbb{E}\left(\int_0^t \gamma_s \, dB_s\right)^2 &(282)\\
&= 2\mathbb{E}\left(\int_0^t \alpha_s \, ds\right)^2 + 2\mathbb{E}\left(\int_0^t \gamma_s^2 \, ds\right) &(283)\\
& &(284)
\end{aligned}$$

by the Ito's isometry and recall Cauchy's inequality on L^2 space that $\int_0^t \alpha_s \, ds \leq \sqrt{t \cdot \int_0^t \alpha_s^2 \, ds}$, so

$$\leq 2t \cdot \mathbb{E}\left(\int_0^t \alpha_s^2 \, ds\right) + 2\mathbb{E}\left(\int_0^t \gamma_s^2 \, ds\right) \tag{285}$$

The reason to turn $\int_0^t \alpha_s \, ds$ into $\int_0^t \alpha_s^2 \, ds$ is that the Lipschitz condition would then allow us to bound the square integral by the square difference of two solutions. To see that, notice that D is uniform: $|\alpha_s| \leq D|X_s - \tilde{X}_s|, |\gamma_s| \leq D|X_s - \tilde{X}_s|$

$$\begin{aligned}
&\leq (2D^2 t + 2D^2) \cdot \mathbb{E}\left(\int_0^t (X_s - \tilde{X}_s)^2 \, ds\right) &(286)\\
&= (2D^2 t + 2D^2) \cdot \int_0^t \mathbb{E}(X_s - \tilde{X}_s)^2 \, ds &(287)
\end{aligned}$$

Denote $v(s) = \mathbb{E}(X_s - \tilde{X}_s)^2$, so

$$v(t) \leq (2D^2 + 2D^2) \cdot \int_0^t v(s) \, ds \tag{288}$$

by Grownwall's inequality, conclude that

$$\begin{aligned}
&\forall t \in [0, T], v(t) = 0 &(289)\\
&\forall t \in [0, T], \mathbb{E}(X_t - \tilde{X}_t)^2 = 0 &(290)\\
&\forall t \in [0, T], X_t = \tilde{X}_t \ a.s. &(291)
\end{aligned}$$

Then prove the **existence** of such solution by **Picard iteration**. Similar to that in the ODE theory, construct

$$X_t^0 = x \tag{292}$$

$$X_t^{k+1} = x + \int_0^t b(s, X_s^k)\, ds + \int_0^t \sigma(s, X_s^k)\, dB_s \ (k = 0, 1, ...) \tag{293}$$

we want to prove that X_t^n actually converges in L^2 sense to some limit as n goes to infinity and the limit is just a solution to the SDE. First prove convergence by showing that it's Cauchy.

Notice the fact that

$$\begin{aligned}
\mathbb{E}(X_t^{k+1} - X_t^k)^2 &= \mathbb{E}\left(\int_0^t [b(s, X_s^k) - b(s, X_s^{k-1})]\, ds + \int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s^{k-1})]\, dB_s\right)^2 &(294)\\
&\leq 2\mathbb{E}\left(\int_0^t [b(s, X_s^k) - b(s, X_s^{k-1})]\, ds\right)^2 + 2\mathbb{E}\left(\int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s^{k-1})]\, dB_s\right)^2 &(295)\\
&= 2\mathbb{E}\left(\int_0^t [b(s, X_s^k) - b(s, X_s^{k-1})]\, ds\right)^2 + 2\mathbb{E}\left(\int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s^{k-1})]^2\, ds\right) &(296)\\
&\leq 2t \cdot \mathbb{E}\int_0^t [b(s, X_s^k) - b(s, X_s^{k-1})]^2\, ds + 2\mathbb{E}\int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s^{k-1})]^2\, ds &(297)\\
&\leq (2D^2T + 2D^2) \cdot \int_0^t \mathbb{E}(X_s^k - X_s^{k-1})^2\, ds &(298)
\end{aligned}$$

by applying this iteratively, it's easy to see that

$$\begin{aligned}
\mathbb{E}(X_t^{k+1} - X_t^k)^2 &\leq (2D^2T + 2D^2) \cdot \int_0^t \mathbb{E}(X_s^k - X_s^{k-1})^2\, ds &(299)\\
&\leq (2D^2T + 2D^2)^2 \cdot \int_0^t \int_0^{s_1} \mathbb{E}(X_{s_2}^{k-1} - X_{s_2}^{k-2})^2\, ds_2\, ds_1 &(300)\\
&\leq ... &(301)\\
&\leq (2D^2T + 2D^2)^k \cdot \int_{0<s_k<...<s_1<t} \mathbb{E}(X_{s_k}^1 - X_{s_k}^0)^2\, ds_k ... ds_2\, ds_1 &(302)\\
& &(303)
\end{aligned}$$

The problem turns into getting an upper bound of $\mathbb{E}(X_t^1 - X_t^0)^2$

$$\begin{aligned}
\mathbb{E}(X_t^1 - X_t^0)^2 &= \mathbb{E}\left(\int_0^t b(s, X_s^0)\, ds + \int_0^t \sigma(s, X_s^0)\, dB_s\right)^2 &(304)\\
&\leq 2t \cdot \mathbb{E}\int_0^t b^2(s, x)\, ds + 2\mathbb{E}\int_0^t \sigma^2(s, x)\, ds &(305)
\end{aligned}$$

the techniques are exactly the same as above (Cauchy, Ito's isometry...), then apply the growth condition for uniform

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

$$\begin{aligned}
\mathbb{E}(X_t^1 - X_t^0)^2 &\le 2c^2 t \cdot \mathbb{E}\int_0^t (1+|X_s^0|)^2\,ds + 2c^2\cdot\mathbb{E}\int_0^t (1+|X_s^0|)^2\,ds && (306)\\
&\le 2c^2t^2\cdot\mathbb{E}(1+|X_s^0|)^2 + 2c^2t\cdot\mathbb{E}(1+|X_s^0|)^2 && (307)\\
&\le 2c^2T^2\cdot\mathbb{E}(1+|X_s^0|)^2 + 2c^2T\cdot\mathbb{E}(1+|X_s^0|)^2 && (308)\\
&\le A && (309)
\end{aligned}$$

for some fixed constant A. As a result, the estimation of the upper bound is

$$\begin{aligned}
\mathbb{E}(X_t^{k+1} - X_t^k)^2 &\le A(2D^2T+2D^2)^k \cdot \int_{0<s_k<\ldots<s_1<t} ds_k \ldots ds_2\, ds_1 && (310)\\
&= A(2D^2T+2D^2)^k\cdot\frac{t^k}{k!} && (311)
\end{aligned}$$

Now compute the L^2 norm to prove that the sequence is Cauchy for $n<m$ by telescoping:

$$\begin{aligned}
||X_t^m - X_t^n||_{L^2([0,T]\times\Omega)} &\le \sum_{k=n}^{m-1} ||X_t^{k+1} - X_t^k||_{L^2([0,T]\times\Omega)} && (312)\\
&= \sum_{k=n}^{m-1}\sqrt{\int_0^T \mathbb{E}(X_t^{k+1}-X_t^k)^2\,dt} && (313)\\
&\le \sum_{k=n}^{m-1}\sqrt{\int_0^T A(2D^2T+2D^2)^k\cdot\frac{t^k}{k!}\,dt} && (314)\\
&= \sum_{k=n}^{m-1}\sqrt{\frac{AT^{k+1}(2D^2T+2D^2)^k}{(k+1)!}} \to 0\ (m,n\to\infty) && (315)
\end{aligned}$$

since $\sum_{k=0}^{\infty}\sqrt{\frac{AT^{k+1}(2D^2T+2D^2)^k}{(k+1)!}} < \infty$, so it's proved that this Picard sequence is Cauchy and its limit exists in L^2 space (actually it's **uniformly Cauchy in L^2 sense** since the upper bound does not depend is uniform in t)

$$\exists X_t \in L^2([0,T]\times\Omega), X_t^n \overset{L^2([0,T]\times\Omega)}{\rightarrow} X_t\ (n\to\infty) \tag{316}$$

To see that this limit is actually a solution to the SDE with continuous sample path, let's prove the convergence of the integrals. Now for $\forall t\in[0,T]$, set $k\to\infty$

$$\begin{aligned}
\mathbb{E}\left(\int_0^t [b(s,X_s^k) - b(s,X_s)]\,ds\right)^2 &\le t\cdot\mathbb{E}\int_0^t [b(s,X_s^k)-b(s,X_s)]^2\,ds && (317)\\
&\le D^2t\cdot\int_0^t \mathbb{E}(X_s^k - X_s)^2\,ds \to 0\ (k\to\infty) && (318)
\end{aligned}$$

for the stochastic integral, use Ito's isometry and Lipschitz

$$\mathbb{E}\left(\int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s)]\, dB_s\right)^2 = \mathbb{E}\int_0^t [\sigma(s, X_s^k) - \sigma(s, X_s)]^2\, ds \tag{319}$$

$$\leq D^2 \int_0^t \mathbb{E}(X_s^k - X_s)^2\, ds \to 0 \ (k \to \infty) \tag{320}$$

the convergence of $\int_0^t \mathbb{E}(X_s^k - X_s)^2\, ds$ is due to dominated convergence theorem since $\sup_k \mathbb{E}(X_s^k - X_s)^2$ is bounded and does not depend on s since it's uniformly Cauchy, thus integrable on bounded interval $[0, T]$. So we proved that the limit is actually a solution to this SDE.

Note that

$$\mathbb{E}\int_0^t \sigma^2(s, X_s)\, ds \leq c^2 \cdot \mathbb{E}\int_0^t (1 + |X_s|)^2\, ds < \infty \tag{321}$$

since $X_t \in L^2([0, T] \times \Omega)$, so $\mathbb{E}\int_0^t X_s^2\, ds < \infty$. This tells us that $\int_0^t \sigma(s, X_s)\, dB_s$ is always a continuous local MG, ensuring that there exists a modification of the solution with continuous sample paths.

$\square$

Remark. *Note that this is actually the definition of a* ***strong solution*** *to this SDE. There is always a condition*

$$\mathbb{P}\left(\int_0^T |b(s, X_s)| + \sigma^2(s, X_s)\, ds < \infty\right) = 1 \tag{322}$$

added for the general definition of strong solution to ensure the continuous modification, but it's not necessary in the theorem above since we are operating in L^2 space.

Note that the uniqueness of the solution in the theorem above is ***in the sense of modification*** *but not in the sense of indistinguishability. The existence of the solution is in the* ***global*** *sense.*

Example

Consider the SDE

$$\begin{cases} dX_t = X_t^2\, dt \\ X_0 = 1 \end{cases} \tag{323}$$

which is actually an ODE. The unique solution is

$$X_t = \frac{1}{1-t}\ t \in [0, 1) \tag{324}$$

notice that $b(t, x) = x^2$ which is not Lipschitz in x and violates the growth condition, so this SDE does not have any global solutions (violates the existence).

Consider the SDE

$$\begin{cases} dX_t = 3X_t^{\frac{2}{3}}\,dt \\ X_0 = 0 \end{cases} \tag{325}$$

which is also an ODE. The solution is

$$X_t = \begin{cases} 0 & t \le a \\ (t-a)^3 & t > a \end{cases} \tag{326}$$

for $\forall a > 0$, so there are multiple solutions in the sense of modification (violates the uniqueness). Since $b(t,x) = 3x^{\frac{3}{2}}$ which is not Lipschitz in x and violates the growth condition.

Example

Consider the Black-Scholes model

$$dX_t = X_t(\mu\,dt + \sigma\,dB_t)\ (\mu, \sigma \in \mathbb{R}) \tag{327}$$

where $b(t,x) = \mu x, \sigma(t,x) = \sigma x$ both Lipschitz in x and satisfy the growth condition. By the theorem, the solution exists globally in L^2 space with a modification with continuous sample paths, also unique in the sense of modification.

To solve it, notice that if there's no stochastic terms, $dX_t = \mu X_t\,dt$ is an ODE with solution $X_t = X_0 \cdot e^{\mu t}$. As a result, consider changing the variables with $Y_t = \log(X_t)$

$$d\log(X_t) = \frac{1}{X_t}\,dX_t - \frac{1}{2}\frac{1}{X_t^2}\,d\langle X, X\rangle_t \tag{328}$$

$$= \mu\,dt + \sigma\,dB_t - \frac{1}{2}\frac{1}{X_t^2}\,d\langle X, X\rangle_t \tag{329}$$

where the bracket can be computed from the intergal from $X_t = X_0 + \mu\int_0^t X_s\,ds + \sigma\int_0^t X_s\,dB_s$ that

$$d\langle X, X\rangle_t = \sigma^2 X_t^2\,dt \tag{330}$$

plug in to get the solution

$$d\log(X_t) = \sigma\,dB_t + \left(\mu - \frac{\sigma^2}{2}\right)\,dt \tag{331}$$

$$\log(X_t) = \log(X_0) + \sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right)t \tag{332}$$

$$X_t = X_0 \cdot e^{\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right)t} \tag{333}$$

The parameters μ, σ in the model can be estimated using quadratic variation of the observed data $X_t,\ t \in [0, T]$

since

$$\log(X_t) = \log(X_0) + \sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right) t \tag{334}$$

$$\langle \log(X), \log(X) \rangle_t = \sigma^2 t \tag{335}$$

$$\mathbb{E}(\log(X_t) - \log(X_0)) = \left(\mu - \frac{\sigma^2}{2}\right) t \tag{336}$$

As a result, if there's enough data with equal time gap, the empirical quadratic variation over interval $[0, T]$ provides estimation of diffusion σ

$$\hat{\sigma^2} = \frac{\sum_j (\log X_{t_{j+1}} - \log X_{t_j})^2}{T} \overset{L^2, a.s.}{\rightarrow} \sigma^2 (\Delta t \rightarrow 0) \tag{337}$$

and the drift coefficient μ is always harder to estimate.

Example

Consider the Ornstein-Uhlenbeck process defined by SDE with a deterministic initial value condition

$$\begin{cases} dX_t = \alpha(m - X_t)\, dt + \sigma\, dB_t \\ X_0 = x \end{cases} \tag{338}$$

where m is the mean reversion level and α is the speed of mean reversion. The dynamics described by this SDE is that no matter what value X_t takes, it goes toward m with a stochastic noise of size σ. This model can be used to describe the fluctuation of interest rate around the mean interest rate m with the speed of regression described by α.

To solve the SDE, change the variable to set the regression level to 0: $Z_t = X_t - m$, so the SDE becomes

$$dZ_t = -\alpha Z_t\, dt + \sigma\, dB_t \tag{339}$$

Consider the case where $\sigma = 0$, i.e. there is no stochastic noise, then SDE turn into an ODE with solution

$$Z_t = Z_0 \cdot e^{-\alpha t} \tag{340}$$

turn the constant into a process and assume $Z_t = C_t e^{-\alpha t}$, apply Ito formula and compare to original SDE to get

$$dZ_t = -\alpha Z_t\, dt + e^{-\alpha t}\, dC_t \tag{341}$$

$$dC_t = \sigma e^{\alpha t}\, dB_t \tag{342}$$

$$C_0 = Z_0 = X_0 - m = x - m \tag{343}$$

solve out the **OU process**

$$C_t = x - m + \sigma \int_0^t e^{\alpha s}\, dB_s \tag{344}$$

$$Z_t = (x-m)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s}\, dB_s \tag{345}$$

$$X_t = (x-m)e^{-\alpha t} + m + \sigma e^{-\alpha t} \int_0^t e^{\alpha s}\, dB_s \tag{346}$$

It's easy to see that since $\int_0^t e^{\alpha s}\, dB_s$ is a Wiener integral, it's a centered Gaussian random variable. In the case where $X_0 = x$ is deterministic, X_t is also Gaussian with expectation $(x-m)e^{-\alpha t} + m$ and variance

$$\mathbb{E}\left(\sigma e^{-\alpha t} \int_0^t e^{\alpha s}\, dB_s\right)^2 = \sigma^2 e^{-2\alpha t}\mathbb{E}\left(\int_0^t e^{2\alpha s}\, ds\right) = \sigma^2 e^{-2\alpha t} \int_0^t e^{2\alpha s}\, ds = \frac{\sigma^2(1-e^{-2\alpha t})}{2\alpha} \tag{347}$$

So the OU process has Gaussian distribution at each time point

$$X_t \sim N\left((x-m)e^{-\alpha t} + m, \frac{\sigma^2(1-e^{-2\alpha t})}{2\alpha}\right) \tag{348}$$

Since it's clear that the L^2 limit of Gaussian random variables is still Gaussian with the mean and variance just the respective limits of the mean and variance sequences (from characteristic function). As a result,

$$X_t \xrightarrow{L^2} N\left(m, \frac{\sigma^2}{2\alpha}\right)\ (t \to \infty) \tag{349}$$

providing the inspiration of taking the Gaussian $N\left(m, \frac{\sigma^2}{2\alpha}\right)$ as the **invariant distribution** of this SDE. The invariant distribution is defined in a way that **if the initial value X_0 is a random variable with the invariant distribution, then according to the dymanics defined by the SDE, at each time the underlying solution still follows such invariant distribution**.

Let's now prove that **the invariant distribution of OU process is Gaussian** $N\left(m, \frac{\sigma^2}{2\alpha}\right)$.

Remark. *The setting for initial value X_0 to be a given random variable is always that X_0 is independent of the whole BM and the filtration is set as ($\vee$ denotes the sigma field generated by the union of two sigma fields)*

$$\mathscr{F}_t = \sigma(B_s, 0 \le s \le t) \vee \sigma(X_0) \tag{350}$$

In general cases, denote X_0 as the initial value of the solution, a random variable. The solution to the SDE is now

$$X_t = (X_0 - m)e^{-\alpha t} + m + \sigma e^{-\alpha t} \int_0^t e^{\alpha s}\, dB_s \tag{351}$$

since X_0 is Gaussian and the Wiener integral is also Gaussian and they are independent, X_t must be Gaussian, only need to calculate the expectation and variance. By previous calculations,

$$\mathbb{E}X_t = (\mathbb{E}X_0 - m)e^{-\alpha t} + m = m \tag{352}$$

$$\begin{aligned}
Var(X_t) &= Var((X_0 - m)e^{-\alpha t}) + Var\left(\sigma e^{-\alpha t}\int_0^t e^{\alpha s}\, dB_s\right) &(353)\\
&= e^{-2\alpha t}\frac{\sigma^2}{2\alpha} + \sigma^2 e^{-2\alpha t}\cdot \mathbb{E}\left(\int_0^t e^{\alpha s}\, dB_s\right)^2 &(354)\\
&= e^{-2\alpha t}\frac{\sigma^2}{2\alpha} + \sigma^2 e^{-2\alpha t}\cdot \int_0^t e^{2\alpha s}\, ds &(355)\\
&= \frac{\sigma^2}{2\alpha} &(356)
\end{aligned}$$

so the invariant distribution is proved.

The OU process is a **continuous Markov Gaussian process** and it's **generally not a MG**. The auto-correlation function (with invariant initial value condition) is

$$cov(X_t, X_s) = \frac{\sigma^2}{2\alpha}e^{-\alpha|s-t|} \tag{357}$$

The calculation for general initial condition goes like

$$\begin{aligned}
\forall s < t, cov(X_t, X_s) &= cov\left((X_0 - m)e^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha p}\, dB_p, (X_0 - m)e^{-\alpha s} + \sigma e^{-\alpha s}\int_0^s e^{\alpha q}\, dB_q\right) &(358)\\
&= e^{-\alpha(s+t)}Var(X_0) + \sigma^2 e^{-\alpha(s+t)}cov\left(\int_0^t, \int_0^s\right) &(359)\\
&= e^{-\alpha(s+t)}Var(X_0) + \sigma^2 e^{-\alpha(s+t)}cov\left(\int_0^s + \int_s^t, \int_0^s\right) &(360)\\
&= e^{-\alpha(s+t)}Var(X_0) + \sigma^2 e^{-\alpha(s+t)}\mathbb{E}\left(\int_0^s e^{\alpha p}\, dB_p\right)^2 &(361)\\
&= e^{-\alpha(s+t)}Var(X_0) + \sigma^2 e^{-\alpha(s+t)}\int_0^s e^{2\alpha p}\, dp &(362)\\
&= e^{-\alpha(s+t)}Var(X_0) + \frac{\sigma^2}{2\alpha}\left(e^{-\alpha(t-s)} - e^{-\alpha(t+s)}\right) &(363)\\
\forall t, s \geq 0, cov(X_t, X_s) &= e^{-\alpha(s+t)}Var(X_0) + \frac{\sigma^2}{2\alpha}\left(e^{-\alpha|t-s|} - e^{-\alpha(t+s)}\right) &(364)
\end{aligned}$$

There are similar problems for estimating parameters m, σ, α. One can notice that

$$\langle X, X\rangle_t = \sigma^2 e^{-2\alpha t}\int_0^t e^{2\alpha s}\, ds = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right) \tag{365}$$

Strong solution and Weak Solution

The **strong solution** is defined as the solution $X_t \in \mathscr{F}_t$ adapted to the filtration generated by BM and the initial condition (if it's random)

$$\mathscr{F}_t = \sigma(B_s, 0 \le s \le t) \vee \sigma(X_0) \tag{366}$$

The uniqueness and global existence theorem stated for SDE is just proving that for strong solutions.

The **weak solution**, on the other hand, refers to the solution pair $(\tilde{X}_t, \tilde{B}_t, \mathscr{H}_t)$ such that $\tilde{X}_t \in \mathscr{H}_t$ is adapted to a specific filtration and $\tilde{B}_t \in \mathscr{H}_t$ is also BM under such filtration. The important point is that different weak solutions can live in the same or different filtered probability spaces. We will see that the uniqueness for weak solutions cannot be discussed in the pathwise sense unless two different weak solutions are living in the same probability space. Instead, the uniqueness is in the sense of finite-dimensional distribution.

Example

Consider SDE

$$\begin{cases} dX_t = dB_t \\ X_0 = 0 \end{cases} \tag{367}$$

and let B_t^1, B_t^2 be two BM living in two different probability spaces $(\Omega^1, \mathscr{F}^1, \mathscr{F}_t^1, \mathbb{P}^1)$ and $(\Omega^2, \mathscr{F}^2, \mathscr{F}_t^2, \mathbb{P}^2)$. For example, $(\Omega^1, \mathscr{F}^1, \mathscr{F}_t^1, \mathbb{P}^1) = ([0,1], \mathscr{B}_{[0,1]}, \mathscr{F}_t^1, \lambda), (\Omega^2, \mathscr{F}^2, \mathscr{F}_t^2, \mathbb{P}^2) = ([3,4], \mathscr{B}_{[3,4]}, \mathscr{F}_t^2, \lambda)$, both probability space equipped with Lebesgue measure but the sample space has no intersections.

By the definition of weak solution, $(\tilde{X}_t = B_t^1, \tilde{B}_t = B_t^1, \mathscr{F}_t = \sigma(B_s^1, 0 \le s \le t)), (\tilde{X}_t = B_t^2, \tilde{B}_t = B_t^2, \mathscr{F}_t = \sigma(B_s^2, 0 \le s \le t))$ are both weak solutions to this SDE. In this situation, $\mathbb{P}\left(B_t^1 = B_t^2\right)$ is not well-defined so no pathwise uniqueness argument can be made.

However, if we consider now the BM B_t^3 defined on a filtered probability space $(\Omega^3, \mathscr{F}^3, \mathscr{F}_t^3, \mathbb{P}^3)$, pairs $(\tilde{X}_t = B_t^3, \tilde{B}_t = B_t^3, \mathscr{F}_t = \sigma(B_s^3, 0 \le s \le t))$ and $(\tilde{X}_t = -B_t^3, \tilde{B}_t = -B_t^3, \mathscr{F}_t = \sigma(B_s^3, 0 \le s \le t))$ are both weak solutions, but now they are in the same filtered probability space and pathwise uniqueness arguments are well-defined. However,

$$\mathbb{P}\left(B_t^3 = -B_t^3\right) = \mathbb{P}\left(B_t^3 = 0\right) = 0 \tag{368}$$

so the sample paths of these two weak solutions has 0 probability of looking the same. Although the pathwise argument generally cannot work for weak solutions, the uniqueness in the sense of finite-dimensional distribution works since $(B_{t_1}^3, ..., B_{t_d}^3) \overset{d}{=} (-B_{t_1}^3, ..., -B_{t_d}^3)$.

Week 5

Example

Consider the following **Tanaka's equation**

$$\begin{cases} dX_t = sign(X_t)\, dB_t \\ X_0 = x \end{cases} \tag{369}$$

By **Tanaka's formula**, one knows that

$$d|B_t| = sgn(B_t) dB_t + dL_t \tag{370}$$

$$L_t = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \{s \in [0,t] : |B_s| < \varepsilon\} \tag{371}$$

where L_t is the local time of BM at 0.

Let's prove that this equation **has no strong solution, has a unique weak solution in the distribution sense and has no pathwise uniqueness**.

Firstly, assume that X_t is a strong solution. This means that X_t satisfies the SDE and is also adapted to the filtration $\mathscr{F}_t$ generated by BM B_t. Then

$$d\langle X, X\rangle_t = sign^2(X_t)\, dt = dt \tag{372}$$

and the Levy's characterization of BM tells us that X_t is a BM. This is actually telling us the uniqueness of weak solution, since each solution shall have the same finite-dimensional distribution as that of BM. To see the non-existence of strong solution, we need a **measurability argument**. Tanaka's formula for X_t implies

$$|X_t| = |x| + \int_0^t sign(X_s)\, dX_s + L_t^X \tag{373}$$

$$= |x| + B_t + L_t^X \tag{374}$$

$$B_t = |X_t| - |x| - L_t^X \tag{375}$$

where $|X_t| - |x| - L_t^X \in \mathscr{F}_t^{|X|} = \sigma(|X_s|, 0 \le s \le t)$ so $\mathscr{F}_t^B \subset \mathscr{F}_t^{|X|}$. However, by the definition of strong solution, $\mathscr{F}_t^X \subset \mathscr{F}_t^B$. As a result,

$$\forall t \ge 0, \mathscr{F}_t^X \subset \mathscr{F}_t^{|X|} \tag{376}$$

This can't be true since for a BM, the filtration generated by its absolute value process is a strict subset of the filtration generated by itself, a contradiction!

Remark. *An analogue is that x always contains more information than $|x|$. By knowing x one can compute $|x|$ but one can never infer x by knowing the value of $|x|$. Note that the measurability argument only holds for $\overset{a.s.}{=}$ but not* *for* $\stackrel{d}{=}$.

To be rigorous, $|X|$ *is a function of* X*, so* $\mathscr{F}_t^{|X|} \subset \mathscr{F}_t^X$*. However, since* X *is BM which is not trivial, there exists event like* $\{T_1 \leq t\} \in \mathscr{F}_t^X$ *where* T_1 *is the first hitting time to 1 of* X_t*, but* $\{T_1 \leq t\} \notin \mathscr{F}_t^{|X|}$*. Another example is* $\{X_t > 0\}$*, if it is in* $\mathscr{F}_t^{|X|}$*, then since* $\{|X_1| = 1\} \in \mathscr{F}_t^{|X|}$*, their intersection is* $\{X_1 = 1\} \in \mathscr{F}_t^{|X|}$ *which is a contradiction since* $\{1\}$ *is not symmetric w.r.t. 0.*

The following theorem characterizes BM by a calculation of the quadratic variation, continuity of sample paths is even not required as long as X_t has the structure of local martingale.

Theorem 11. *If* X_t *is local martingale with* $X_0 = 0$ *adapted to filtration* $\mathscr{F}_t$*, then* X *is BM under filtration* $\mathscr{F}_t$ *iff* $X_t^2 - t$ *is* $\mathscr{F}_t$ *adapted continuous local martingale iff* $\forall t \geq 0, \langle X, X \rangle_t = t$ ***(Levy's characterization of BM)***

Let's then construct a weak solution to the Tanaka's equation. Since we already know that X_t has to be BM, set $X_t = x + \hat{B}_t$ with $\hat{B}$ to be any BM. Define

$$\tilde{B}_t = \int_0^t sign(X_s)\, dX_s \tag{377}$$

then the pair $(X_t, \tilde{B}_t)$ is the solution to the SDE because $dX_t = d\hat{B}_t$ and $d\tilde{B}_t = sign(X_t)\, dX_t$, so $dX_t = sign(X_t)\, d\tilde{B}_t$ satisfies the SDE and $\langle \tilde{B}, \tilde{B} \rangle = \langle X, X \rangle_t = t$, proving that $\tilde{B}$ is a BM.

Remark. *Although there is no way to find a solution* X_t *adapted to the filtration generated by* B_t*, which is the BM in the SDE, there is a way to specify another BM* $\tilde{B}_t$ *and* X_t *such that they have some connections and work as the solution to this SDE. The "weak" refers to replacing a general BM with a specific chosen BM in the SDE (allows connections with the constructed solution).*

We can also see that the weak solution to this SDE has no pathwise uniqueness (the uniqueness for strong solution). The definition of **pathwise uniqueness** is that the solution has pathwise uniqueness if any two solutions in the same filtered probability space almost surely have the same path.

Assume that $(X_t, \tilde{B}_t)$ is a weak solution pair of Tanaka's equation, define $\tau = \inf\{t \geq 1, X_t = 0\}$ be the first hitting time after time 1 that hits 0. Define the reflected process

$$\tilde{X}_t = \begin{cases} X_t & t \leq \tau \\ -X_t & t > \tau \end{cases} \tag{378}$$

to find that the pair $(\tilde{X}_t, \tilde{B}_t)$ is still a weak solution.

$$x + \int_0^t sign(\tilde{X}_s)\, d\tilde{B}_s = x + \int_0^{t\wedge\tau} sign(X_s)\, d\tilde{B}_s - \int_{t\wedge\tau}^t sign(X_s)\, d\tilde{B}_s \tag{379}$$

$$= x + \int_0^{t\wedge\tau} dX_s - \int_{t\wedge\tau}^t dX_s \tag{380}$$

$$= 2X_{t\wedge\tau} - X_t = \tilde{X}_t \tag{381}$$

since $X_\tau = 0$. However, $\tilde{X}_t$ and X_t have totally different sample paths, so pathwise uniqueness fails.

The following theorem asserts the reason why pathwise uniqueness is crucial to consider in solving SDE.

Theorem 12. *If a SDE has weak solution and the solution has pathwise uniqueness, then its strong solution exists.* ***(Yamada-Watanabe)***

For Tanaka's equation, since weak solution exists, the non-existence of strong solution directly implies the failure of pathwise uniqueness.

Ito Diffusion

Let's consider the **Ito diffusion**, which is defined by a time-homogeneous SDE

$$\begin{cases} dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t \ (t \geq s) \\ (X_t \in \mathbb{R}^n, B_t \in \mathbb{R}^m, b : \mathbb{R}^n \to \mathbb{R}^n, \sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}) \\ X_s = x \end{cases} \tag{382}$$

a n-dim process generated by the SDE with given initial condition at time s. In addition, it is assumed that b, σ are both Lipschitz on $\mathbb{R}^n$. In the following context, whenever the condition for the unique existence of strong solution holds, **it's always assumed that the solution X_t is the modification with continuous sample paths**.

Remark. *Ito diffusion refers to* ***the drift and diffusion coefficient*** b, σ ***not depending on time*** t ***and the Lipschitz condition*** *is here to ensure the existence and uniqueness of the strong solution of such SDE (proved above). Note that the growth condition is naturally satisfied since Lipschitz condition implies*

$$\exists D, \forall x, y \in \mathbb{R}^n, ||b(x) - b(y)|| + ||\sigma(x) - \sigma(y)|| \leq D||x - y|| \tag{383}$$

so the growth condition

$$\exists C, \forall x \in \mathbb{R}^n, ||b(x)|| + ||\sigma(x)|| \leq C(1 + ||x||) \tag{384}$$

is always true. In the context below, always denote X_t as the Ito diffusion which is the solution to this SDE with ***continuous sample path****.*

Next we talk about **properties** of such Ito diffusion process. The first property is the time homogeneity naturally implied by the fact that b, σ does not directly contain time t.

Theorem 13. ***(Time-Homogeneity)*** *Denote $X^{s,x}_{s+h}$ as the solution to the SDE above with initial condition $X_s = x$ as an Ito diffusion,*

$$\forall s \geq 0, x \in \mathbb{R}^n, \left\{X^{s,x}_{s+t}\right\}_{t\geq 0} \overset{d}{=} \left\{X^{0,x}_t\right\}_{t\geq 0} \tag{385}$$

Proof. By the definition of the solution,

$$X_{s+h}^{s,x} = x + \int_s^{s+h} b(X_u^{s,x})\, du + \int_s^{s+h} \sigma(X_u^{s,x})\, dB_u \tag{386}$$

change the variables by $v = u - s$ to drag the initial condition to time 0 to get

$$X_{s+h}^{s,x} = x + \int_0^h b(X_{v+s}^{s,x})\, dv + \int_0^h \sigma(X_{v+s}^{s,x})\, dB_{v+s} \tag{387}$$

$$= x + \int_0^h b(X_{v+s}^{s,x})\, dv + \int_0^h \sigma(X_{v+s}^{s,x})\, d(B_{v+s} - B_s) \tag{388}$$

note that here we make use of the fact that

$$\int_0^h \sigma(X_{v+s}^{s,x})\, dB_s = 0 \tag{389}$$

since the integral is w.r.t. v but not s. The motivation is to add another part to B_{s+v} such that it becomes another BM by Markov property.

$$X_{s+h}^{s,x} = x + \int_0^h b(X_{v+s}^{s,x})\, dv + \int_0^h \sigma(X_{v+s}^{s,x})\, d\tilde{B}_v \tag{390}$$

Notice that $X_t^{0,x}$, which is the solution to the SDE with initial condition $X_0 = x$ satisfies

$$X_t^{0,x} = x + \int_0^t b(X_s^{0,x})\, ds + \int_0^t \sigma(X_s^{0,x})\, dB_s \tag{391}$$

by comparing the two equations above, we conclude that

$$(X_{s+t}^{s,x}, \tilde{B}_t), (X_t^{0,x}, B_t) \tag{392}$$

are both weak solutions to the same SDE for Ito diffusion. Since the Lipschitz condition of original SDE holds, the weak solution has uniqueness in distribution (see the theorem below)

$$\forall s \geq 0, \forall x \in \mathbb{R}^n, \left\{X_{s+t}^{s,x}\right\}_{t\geq 0} \overset{d}{=} \left\{X_t^{0,x}\right\}_{t\geq 0} \tag{393}$$

$\square$

Remark. *The equality in distribution does not generally hold in the almost sure sense. Intuitively, those two solutions has something to do with different BMs, one is B_t, the BM in the SDE and the other is $\tilde{B}_t$, the BM derived by shifting the time of B_t.*

Here we have made use of the theorem that if the existence and uniqueness condition of a SDE holds, weak solution is unique in the sense of distribution.

Theorem 14. *Consider the SDE with initial condition*

$$\begin{cases} dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \\ X_0 = x \end{cases} \tag{394}$$

If the existence and uniqueness condition holds (growth condition and Lipschitz condition), any two weak solutions have the same distribution.

Proof. Assume $(\hat{X}_t, \hat{B}_t, \hat{\mathscr{H}_t}), (\tilde{X}_t, \tilde{B}_t, \tilde{\mathscr{H}_t})$ are two weak solution pairs, then by the existence of strong solution, construct strong solutions for filtration $\hat{\mathscr{H}_t}$ and BM $\hat{B}_t$ denoted $\hat{Z}_t$ and strong solution for filtration $\tilde{\mathscr{H}_t}$ and BM $\tilde{B}_t$ denoted $\tilde{Z}_t$. By the uniqueness of strong solution in the sense of modification, $\forall t, \hat{X}_t = \hat{Z}_t, \tilde{X}_t = \tilde{Z}_t$ *a.s.* To prove that $\hat{X}_t \overset{d}{=} \tilde{X}_t$, only need to prove that $\hat{Z}_t \overset{d}{=} \tilde{Z}_t$.

Recall that the strong solutions $\hat{Z}_t, \tilde{Z}_t$ are constructed as the L^2 limit of the Picard iteration sequence

$$\hat{Z}_t^n \overset{L^2}{\to} \hat{Z}_t \ (n \to \infty) \tag{395}$$

$$\tilde{Z}_t^n \overset{L^2}{\to} \tilde{Z}_t \ (n \to \infty) \tag{396}$$

since the underlying SDE of the construction of $\hat{Z}_t^n, \tilde{Z}_t^n$ are the same and the initial conditions are also the same

$$\forall n, \hat{Z}_t^n \overset{d}{=} \tilde{Z}_t^n \tag{397}$$

their L^2 limits should also be the same. As a result, the weak solutions in the same probability space have the same distribution.

□

Remark. *These two weak solutions don't have to be in the same probability space. Note that $\hat{Z}_t, \hat{X}_t$ are in the same probability space w.r.t. the same BM, that's why we can use the uniqueness argument to conclude that at any time they are equal almost surely.*

The second property is the flow property. It's stating the natural fact that the Ito diffusion at time $t+h$ starting with initial condition $X_0 = x$ is always almost surely equal to the Ito diffusion at time $t+h$ starting with the same initial condition $X_0 = x$, stopped at time t and restarted with the initial condition X_t at time t.

Theorem 15. ***(Flow Property)*** *For Ito diffusion X_t,*

$$\forall t, h \geq 0, X_h^{0, X_t^{0,x}} = X_{t+h}^{0,x} \ a.s. \tag{398}$$

Proof. By construction, these two processes are both strong solutions to the same SDE:

$$X_{t+h}^{0,x} = X_t^{0,x} + \int_t^{t+h} b(X_s^{0,x})\,ds + \int_t^{t+h} \sigma(X_s^{0,x})\,dB_s \tag{399}$$

By the uniqueness in the sense of modification, we conclude the flow property that

$$\forall t, h \geq 0, X_h^{0,X_t^{0,x}} = X_{t+h}^{0,x} \ a.s. \tag{400}$$

□

The Markov property and strong Markov property of Ito diffusion can be implied from the flow property. For the proof, refer to Oksendal P120.

Theorem 16. ***(Markov Property)*** *For Ito diffusion X_t, any bounded Borel measurable $f : \mathbb{R}^n \to \mathbb{R}$ and $\forall t, h \geq 0$,*

$$\mathbb{E}_x(f(X_{t+h})|\mathscr{F}_t) = \mathbb{E}_{X_t} f(X_h) \tag{401}$$

where $\mathscr{F}_t$ is the filtration generated by the BM B_t in the SDE and $\mathbb{E}_x$ denotes the expectation under the condition that the SDE has initial condition starting from x.

Theorem 17. ***(Strong Markov Property)*** *For Ito diffusion X_t, any bounded Borel measurable $f : \mathbb{R}^n \to \mathbb{R}$ and τ as a stopping time w.r.t. filtrations $\mathscr{F}_t$ such that $\tau < \infty$ a.s., for $\forall h \geq 0$*

$$\mathbb{E}_x(f(X_{\tau+h})|\mathscr{F}_\tau) = \mathbb{E}_{X_\tau} f(X_h) \tag{402}$$

where $\mathscr{F}_t$ is the filtration generated by the BM B_t in the SDE and $\mathbb{E}_x$ denotes the expectation under the condition that the SDE has initial condition starting from x.

Remark. *Although the Markov and strong Markov property are stated for $\mathscr{F}_t$, the filtration generated by BM in the SDE, it also holds for the filtration generated by Ito diffusion, namely $\mathscr{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$. To see this, note that X_t is a strong solution to the SDE so $\mathscr{F}_t^X \subset \mathscr{F}_t$*

$$\mathbb{E}_x(f(X_{t+h})|\mathscr{F}_t^X) = \mathbb{E}[\mathbb{E}_x(f(X_{t+h})|\mathscr{F}_t)|\mathscr{F}_t^X] \tag{403}$$

$$= \mathbb{E}[\mathbb{E}_{X_t} f(X_h)|\mathscr{F}_t^X] = \mathbb{E}_{X_t} f(X_h) \tag{404}$$

Generator of Ito Diffusion

Consider for $f \in C_b^2$, the Ito diffusion as the solution of the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t \tag{405}$$

define **the semi-group operator** P_t as

$$P_t f(x) = \mathbb{E}\left(f(X_t)|X_0 = x\right) = \mathbb{E}_x f(X_t) \tag{406}$$

Then the Markov property of the Ito diffusion implies that

$$\begin{aligned}
P_{t+s}f(x) &= \mathbb{E}_x f(X_{t+s}) &(407)\\
&= \mathbb{E}_x[\mathbb{E}_x\left(f(X_{t+s})|\mathscr{F}_t^X\right)] &(408)\\
&= \mathbb{E}_x\mathbb{E}_{X_t}f(X_s) = \mathbb{E}_x P_s f(X_t) &(409)\\
&= P_t P_s f(x) &(410)
\end{aligned}$$

for the same reasoning, $P_{t+s} = P_s P_t = P_t P_s$, so we conclude that this generator is **commutative**.

For example, if there's a BM B_t, its semi-group generator is

$$\begin{aligned}
P_t f(x) &= \mathbb{E}(f(B_t)|B_0 = x) &(411)\\
&= \mathbb{E}f(x + N(0,t)) &(412)\\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} f(x+y)e^{-\frac{y^2}{2t}}\,dy &(413)\\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} f(y)e^{-\frac{(y-x)^2}{2t}}\,dy &(414)
\end{aligned}$$

The next natural question to ask is that how shall we use P_t to characterize the underlying SDE for Ito diffusion X_t. By Ito formula,

$$\begin{aligned}
P_h f(x) &= \mathbb{E}_x f(X_h) &(415)\\
&= \mathbb{E}_x\left(f(X_0) + \int_0^h f'(X_s)\,dX_s + \frac{1}{2}\int_0^h f''(X_s)\,d\langle X, X\rangle_s\right) &(416)\\
&= f(x) + \mathbb{E}_x\left(\int_0^h f'(X_s)b(X_s)\,ds + \int_0^h f'(X_s)\sigma(X_s)\,dB_s + \frac{1}{2}\int_0^h f''(X_s)\sigma^2(X_s)\,ds\right) &(417)
\end{aligned}$$

since $f \in C_b^2$, σ has linear growth rate and X_t is a strong solution in the L^2 space

$$\mathbb{E}\int_0^h f'^2(X_s)\sigma^2(X_s)\,ds \le C\mathbb{E}\int_0^h X_s^2\,ds < \infty \tag{418}$$

as a result, the stochastic integral

$$\int_0^h f'(X_s)\sigma(X_s)\,dB_s \tag{419}$$

is a MG (since $f'(X_s)\sigma(X_s) \in L^2([0,h]\times\Omega)$), this term disappears from the expansion of $P_h f(x)$ after taking expectation. Organize the terms left behind and use the intermediate value theorem for integral

$$\frac{P_h f(x) - f(x)}{h} = \frac{1}{h}\mathbb{E}_x\left(\int_0^h f'(X_s)b(X_s) + \frac{1}{2}f''(X_s)\sigma^2(X_s)\,ds\right) \tag{420}$$

$$= \mathbb{E}_x\left(f'(X_\xi)b(X_\xi) + \frac{1}{2}f''(X_\xi)\sigma^2(X_\xi)\right) \tag{421}$$

for some $\xi \in [0, h]$. Setting $h \to 0$ and use the continuity of sample path of X_t, one can conclude that

$$\frac{P_h - id}{h}f(x) = \frac{P_h f(x) - f(x)}{h} = \mathbb{E}_x\left(f'(X_\xi)b(X_\xi) + \frac{1}{2}f''(X_\xi)\sigma^2(X_\xi)\right) \to f'(x)b(x) + \frac{1}{2}\sigma^2(x)f''(x) \tag{422}$$

This gives the natural definition of the **infinitesimal generator**

$$L = \lim_{h\to 0}\frac{P_h - id}{h} = b\partial_x + \frac{1}{2}\sigma^2\partial_{xx} \tag{423}$$

note that the infinitesimal generator has close connections with the derivative of the operator P_t in that

$$\frac{dP_t}{dt} = \lim_{h\to 0}\frac{P_{t+h} - P_t}{h} = P_t\lim_{h\to 0}\frac{P_h - id}{h} = P_t L \tag{424}$$

$$\frac{dP_t}{dt} = \lim_{h\to 0}\frac{P_{t+h} - P_t}{h} = \lim_{h\to 0}\frac{P_h - id}{h}P_t = LP_t \tag{425}$$

telling us that **the semi-group generator P_t and the infinitesimal generator L commutes, moreover, their product is the derivative of P_t**.

Consider the case for higher dimensions, where we have the settings $X_t \in \mathbb{R}^n, b \in \mathbb{R}^n, \sigma \in \mathbb{R}^{n\times m}, B_t \in \mathbb{R}^m$. Let's calculate the infinitesimal generator in the similar style.

Theorem 18. ***(Infinitesimal Generator in General Case)*** *For the setting $X_t \in \mathbb{R}^n, b \in \mathbb{R}^n, \sigma \in \mathbb{R}^{n\times m}, B_t \in \mathbb{R}^m$, the infinitesimal generator of the Ito diffusion starting from $X_0 = x$ is given by*

$$Lf = b\cdot\nabla f + \frac{1}{2}Tr(\sigma\sigma^T H) \tag{426}$$

$$Lf(x) = \sum_{i=1}^n b_i(x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^n(\sigma\sigma^T)_{i,j}(x)\frac{\partial^2 f}{\partial x_i\partial x_j} \tag{427}$$

where H is the Hessian of $f : \mathbb{R}^n \to \mathbb{R}$ at x.

Proof. The proof is pure calculations with multi-dimensional Ito formula

$$P_h f(x) = \mathbb{E}_x(f(X_h)) \tag{428}$$

$$= \mathbb{E}_x(f(X_0) + \int_0^h \nabla f(X_s) \cdot dX_s + \frac{1}{2}\sum_{i,j=1}^n \int_0^h \frac{\partial^2 f(X_s)}{\partial x_i \partial x_j} d\langle X^i, X^j\rangle_s) \tag{429}$$

$$= f(x) + \mathbb{E}_x(\int_0^h \nabla f(X_s) \cdot b(X_s)\, ds + \int_0^h \nabla f(X_s) \cdot \sigma(X_s) \cdot dB_s + \frac{1}{2}\sum_{i,j=1}^n \int_0^h \frac{\partial^2 f(X_s)}{\partial x_i \partial x_j} \sum_{k=1}^m \sigma_{ik}\sigma_{jk}(X_s)\, ds) \tag{430}$$

as what we have proved above, the stochastic integral terms disappears since it's a MG. Simplify the remaining terms to get

$$\frac{P_h f(x) - f(x)}{h} = \frac{1}{h}\mathbb{E}_x(\int_0^h \nabla f(X_s) \cdot b(X_s)\, ds + \frac{1}{2}\sum_{i,j=1}^n \int_0^h \frac{\partial^2 f(X_s)}{\partial x_i \partial x_j} \sum_{k=1}^m \sigma_{ik}\sigma_{jk}(X_s)\, ds) \tag{431}$$

$$= \mathbb{E}_x(\nabla f(X_\xi) \cdot b(X_\xi) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f(X_\xi)}{\partial x_i \partial x_j} \sum_{k=1}^m \sigma_{ik}\sigma_{jk}(X_\xi)) \tag{432}$$

for some $\xi \in [0, h]$. Set $h \to 0$ and use the continuity of the sample paths of X_t to get

$$\frac{P_h f(x) - f(x)}{h} \to \nabla f(x) \cdot b(x) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \sum_{k=1}^m \sigma_{ik}\sigma_{jk}(x) \ (h \to 0) \tag{433}$$

□

Examples

The easiest example is for $n = m = 1$ and to take the Ito diffusion X_t as 1-dim BM, satisfying the following SDE

$$dX_t = dB_t \tag{434}$$

with $b = 0, \sigma = 1$. We immediately know that $L = \frac{1}{2}\partial_{xx}$.

If now $n = m$ are integers larger than 1, take the Ito diffusion X_t as multi-dimensional BM, satisfying the following SDE

$$dX_t = dB_t \tag{435}$$

with $b = 0, \sigma = I_n$. We immediately know that $L = \frac{1}{2}Tr(H) = \frac{1}{2}\Delta$ shows **the connection between BM and Laplacian**.

Consider the graph of BM, i.e. $X_t = (t, B_t) \in \mathbb{R}^2$ for 1-dim BM B_t. It's easy to see that such X_t should be the strong solution to SDE

$$\begin{cases} dX_t^1 = dt \\ dX_t^2 = dB_t \\ X_0^1 = 0, X_0^2 = 0 \end{cases} \tag{436}$$

with $b = (1,0)^T, \sigma = (0,1)^T$. The infinitesimal generator of this Ito diffusion is

$$L = \partial_t + \frac{1}{2}\partial_{xx} \tag{437}$$

for function $f(t,x) : \mathbb{R}^2 \to \mathbb{R}$ which is **the heat operator**.

In the upper examples, we are calculating the infinitesimal generator given the Ito diffusion. Actually, for a given infinitesimal generator, we can also construct correspondent SDEs (obviously it may not be unique). For example, if the infinitesimal generator is known as

$$L = \frac{1+x^2}{2}\partial_{xx} \tag{438}$$

then one may construct the following SDEs

$$dX_t = \sqrt{1+X_t^2}\, dB_t \tag{439}$$

$$dX_t = -\sqrt{1+X_t^2}\, dB_t \tag{440}$$

$$dX_t = dB_t^1 + X_t\, dB_t^2 \tag{441}$$

with this specific L. Note that we are not restricted to 1-dimension, so $\frac{1+x^2}{2}$ can be torn apart into $\frac{1}{2} + \frac{x^2}{2}$ for two independent BM to produce stochastic noise.

Remark. *Although one has much freedom constructing an SDE with the given infinitesimal generator, one has to notice that the drift and diffusion coefficients have to satisfy the Lipschitz condition. This is because all discussions above are made on the basis of the existence and uniqueness of strong solutions.*

For the examples above, $\sigma(x) = \pm\sqrt{1+x^2}$ and $\sigma(x) = (1,x)$ are Lipschitz so these SDEs are fine. However one cannot construct an SDE like

$$dX_t = (X_t)^2\, dB_t^1 + \sqrt{1+(X_t)^2 - (X_t)^4}\, dB_t^2 \tag{442}$$

although $\frac{1}{2}(x^4 + 1 + x^2 - x^4) = \frac{1+x^2}{2}$. This is because x^2 is not Lipschitz.

Remark. *When $f = \mathbb{I}_B$ for some Borel set $B \subset \mathbb{R}^n$,*

$$P_t f = \mathbb{E}_x f(X_t) = \mathbb{P}\left(X_t \in B | X_0 = x\right) \tag{443}$$

gives the Markov transition kernel. That's why P_t is called the generator of a Markov process.

Dynkin Formula

Theorem 19. *(Dynkin Formula) For Ito diffusion X_t with initial condition $X_0 = x$, $f \in C_c^2(\mathbb{R}^n)$ and a stopping time τ w.r.t. the filtration $\mathscr{F}_t$ generated by BM in the SDE satisfying $\mathbb{E}_x \tau < \infty$,*

$$\mathbb{E}_x f(X_\tau) = f(x) + \mathbb{E}_x \int_0^\tau Lf(X_s)\, ds \tag{444}$$

Proof. Let's only prove for 1-dimension X_t. By Ito formula and the calculations above,

$$\forall t \geq 0, f(X_t) = f(x) + \int_0^t Lf(X_s)\, ds + \int_0^t f'(X_s)\sigma(X_s)\, dB_s \tag{445}$$

to show the Dynkin formula, we just have to show that

$$\mathbb{E}_x \int_0^\tau f'(X_s)\sigma(X_s)\, dB_s = 0 \tag{446}$$

where $Y_t = \int_0^t f'(X_s)\sigma(X_s)\, dB_s$ is actually a MG with continuous sample paths, so it's equivalent to saying that we are trying to prove the optional stopping theorem holds for stopping time τ, i.e. $\mathbb{E}Y_\tau = 0$. For any fixed integer k, OST applies to show

$$\tau \wedge k \leq k \ a.s.,\ \mathbb{E}Y_{\tau\wedge k} = 0 \tag{447}$$

note that $f \in C_c^2$ and σ is Lipschitz so $|f'(X_s)\sigma(X_s)| \leq M$ has to be bounded, by Ito's isometry

$$\begin{aligned}
\mathbb{E}_x \left(Y_\tau - Y_{\tau\wedge k}\right)^2 &= \mathbb{E}_x \left(\int_{\tau\wedge k}^\tau f'(X_s)\sigma(X_s)\, dB_s\right)^2 && (448)\\
&= \mathbb{E}_x \left(\int_{\tau\wedge k}^\tau (f'(X_s)\sigma(X_s))^2\, ds\right) && (449)\\
&\leq M^2 \cdot \mathbb{E}_x(\tau - \tau\wedge k) \to 0\ (k \to \infty) && (450)
\end{aligned}$$

by monotone convergence theorem and $\mathbb{E}\tau < \infty$. As a result, we have proved that $\mathbb{E}_x Y_\tau = 0$ so Dynkin formula holds.

□

Remark. *Actually Dynkin's formula still holds for the time inhomogeneous diffusion, which is the solution to*

$$\begin{cases} dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \\ X_0 = x \end{cases} \tag{451}$$

under the growth condition and the Lipschitz condition to ensure the uniqueness and existence of the strong solution. The proof is the same and the conclusion is still

$$\mathbb{E}_x f(X_\tau) = f(x) + \mathbb{E}_x \int_0^\tau Lf(X_s)\, ds \tag{452}$$

with the infinitesimal generator as

$$Lf = b(t,\cdot)\cdot\nabla f + \frac{1}{2}Tr(\sigma(t,\cdot)\sigma^T(t,\cdot)H) \tag{453}$$

However, the derivative of the generator $\frac{dP_t}{dt}$ *is not well-defined since now* $P_tP_h \neq P_hP_t$ *and* $P_tL \neq LP_t$.

Example

Consider n-dim BM B_t starting at $x, ||x|| < R$ and define τ_R as the first exit time of B_t from the sphere with radius R. Assume that we are not knowing that $\mathbb{E}\tau_R < \infty$, then we would take the truncation $\tau_R \wedge k$ and apply the Dynkin formula for $f(x) = ||x||^2$ ($||x|| < R$) and the Ito diffusion $X_t = B_t$ to get that

$$Lf(X_s) = \frac{1}{2}\Delta f(X_s) = n \tag{454}$$

the infinitesimal generator and that

$$\begin{aligned}\mathbb{E}_x f(B_{\tau_R\wedge k}) &= f(x) + \mathbb{E}_x \int_0^{\tau_R\wedge k} Lf(X_s)\, ds &(455)\\ &= ||x||^2 + n\cdot \mathbb{E}_x\tau_R\wedge k &(456)\end{aligned}$$

setting $k\to\infty$ to see

$$\begin{aligned}\mathbb{E}_x f(B_{\tau_R}) &= ||x||^2 + n\cdot\mathbb{E}_x\tau_R &(457)\\ &= R^2 &(458)\\ \mathbb{E}_x\tau_R &= \frac{R^2 - ||x||^2}{n} &(459)\end{aligned}$$

gives **the expected exit time of BM from a sphere with fixed radius**.

Remark. *If* τ *is the first exit time of* X_t *from a bounded set with* $\mathbb{E}_x\tau < \infty$*, then Dynkin formula holds for any* $f \in C^2$*. (Compact support condition is not necessary) The proof is straightforward. Since* X_t *has continuous sample path,* X_t *can't exit the bounded set before stopping time* τ*. Assume* τ *is the first exit time from a bounded set* B*, by setting* $f_B(x) = f(x)\mathbb{I}_B(x)$ *with compact support* $\overline{B}$*, one can still apply Dynkin formula.*

Remark. *The Dynkin formula requires the function* f *to be* C^2 ***on the whole space****. Some examples of violations may show the essence of the Dynkin formula. Consider taking* f *as* Γ*, the fundamental solution to the Laplacian* *equation with dimension $n \geq 2$, taking X_t as BM B_t. If Dynkin formula holds,*

$$\mathbb{E}_x f(B_{\tau_R}) = \Gamma(R) = \Gamma(x) \tag{460}$$

since the infinitesimal generator for BM is just $L = \frac{1}{2}\Delta$, the fundamental solution is harmonic in $\mathbb{R}^n - \{0\}$ and its value $\Gamma(x)$ only depends on the radial value $||x||$. This is a contradiction since $||x|| < R$. The crucial point here is that ***even the singularity at one point can result in the failure of Dynkin formula*** *(since BM can actually hit 0 before stopping time τ_R).*

On the other hand, if we take X_t as BM B_t but have $f \in C^2(\mathbb{R}^n), \Delta f = 0$ as a harmonic function in the whole space, by Dynkin formula,

$$\forall R > 0, \forall ||x|| < R, \mathbb{E}_x f(B_{\tau_R}) = f(x) \tag{461}$$

showing us a generalized version of ***the mean-value property of harmonic functions****. Recall that the mean-value property of harmonic functions is saying that $f(x)$ is the integral average of f on a sphere of any radius centered at x. Here we are saying that $f(x)$ can be specified as the average of f on a sphere of radius $R > ||x||$ centered at origin. The difference is that now we have to start the BM at x to see where it exits the sphere. By collecting the function values at all those exit points and taking an average under probability measure $\mathbb{E}_x$, one can recover $f(x)$.*

By setting $x = 0$, we get

$$\forall R > 0, \mathbb{E}_0 f(B_{\tau_R}) = f(0) \tag{462}$$

notice that B_{τ_R} is uniform on the sphere $\partial B_R(0)$, so

$$\forall R > 0, \frac{1}{|\partial B_R(0)|} \int_{\partial B_R(0)} f(x)\, dx = f(0) \tag{463}$$

a ***probabilistic proof for the classical mean-value property****.*

In addition, if we take X_t as the graph of n-dimensional BM and pick an appropriate stopping time for X_t, we can also get a generalized version of the mean-value property for the solution to the heat equation.

In the remark above, we have shown some interesting points for Dynkin formula and we see that Dynkin formula fails if the Ito diffusion is taken as the BM and the f is specified as the fundamental solution to the Laplacian equation. However, one may realize that the failure in this example results from the fact the BM is possible to hit the singularity at 0. To create a "nice" region where the fundamental solution to the Laplacian equation draws some conclusions, we need to ensure that (i): this area does not contain the origin (ii): this area has some symmetricity in the radial direction (fundamental solution of Laplacian equation Γ only depends on the radial value). It's then natural to take the **annulus region** $A_k^R = \left\{x \in \mathbb{R}^n : R < ||x|| < 2^k R\right\}$ and set the stopping time τ_k^R as the first exit time of X_t from A_k^R

$$\tau_k^R = \tau_R \wedge \tau_{2^k R} \tag{464}$$

The Ito diffusion is taken as BM $X_t = B_t$ starting from $x, R < ||x|| < 2^k R$.

By Dynkin formula and the harmonic property of Γ on $\mathbb{R}^n - \{0\}$,

$$\forall R, k > 0, \forall R < ||x|| < 2^k R, \mathbb{E}_x \Gamma\left(B_{\tau_k^R}\right) = \Gamma(x) \tag{465}$$

we know that $||B_{\tau_k^R}||$ either takes value R or $2^k R$. The decomposition w.r.t. exiting place gives

$$\forall R, k > 0, \forall R < ||x|| < 2^k R, \mathbb{P}_x\left(||B_{\tau_k^R}|| = R\right) \cdot \Gamma(R) + \mathbb{P}_x\left(||B_{\tau_k^R}|| = 2^k R\right) \cdot \Gamma\left(2^k R\right) = \Gamma(x) \tag{466}$$

recall the expressions of fundamental solutions that

$$\Gamma(x) = \begin{cases} -\frac{1}{2\pi} \log ||x|| & n = 2 \\ \frac{1}{n(n-2)V_n(1)} ||x||^{2-n} & n \geq 3 \end{cases} \tag{467}$$

where $V_n(1)$ is the volume of unit ball in $\mathbb{R}^n$.

For $n = 2$,

$$\forall R, k > 0, \forall R < ||x|| < 2^k R, \mathbb{P}_x\left(\tau_R < \tau_{2^k R}\right) = \frac{k \log 2 + \log R - \log ||x||}{k \log 2} \tag{468}$$

set $k \to \infty$

$$\forall R > 0, \forall ||x|| > R, \mathbb{P}_x\left(\tau_R < \infty\right) = 1 \tag{469}$$

For $n = 3$,

$$\forall R, k > 0, \forall R < ||x|| < 2^k R, \mathbb{P}_x\left(\tau_R < \tau_{2^k R}\right) = \frac{||x||^{2-n} - \left(2^k R\right)^{2-n}}{R^{2-n} - \left(2^k R\right)^{2-n}} \tag{470}$$

set $k \to \infty$

$$\forall R > 0, \forall ||x|| > R, \mathbb{P}_x\left(\tau_R < \infty\right) = \left(\frac{||x||}{R}\right)^{2-n} < 1 \tag{471}$$

This is saying that if a BM starts from any point outside of ball $B_R(0)$, in 2-dimension it almost surely hits sphere $\partial B_R(0)$ in finite time but in higher dimension it hits sphere $\partial B_R(0)$ in finite time with probability less than 1. As a result, we have shown that **BM is recurrent in dimension 2 but transient in higher dimension**.

Remark. *The conclusion is the same to that of simple random walk. For simple random walk S_n, it's recurrent if and only if*

$$\mathbb{P}\left(S_n = 0 \ i.o.\right) = 1 \tag{472}$$

if and only if

$$\sum_{m=0}^{\infty} \mathbb{P}\left(S_m = 0\right) = \infty \tag{473}$$

since $\sum_{m=0}^{\infty} \mathbb{P}\left(S_m = 0\right) = \sum_{m=0}^{\infty} \mathbb{P}\left(T_0^m < 0\right) = \sum_{m=0}^{\infty} \left(\mathbb{P}\left(T_0^1 < 0\right)\right)^m = \frac{1}{1-\mathbb{P}\left(T_0^1 < \infty\right)}$ *for* T_0^i *to be the i-th hitting time of* S_n *to 0. Then simple combinatorics and approximation proves the conclusion.*

Week 6

In the last section, the discussion of Ito diffusion is under the setting of the time-homogeneous case, where the drift and diffusion coefficients do not depend on time. Now we consider the general diffusion process X_t as the solution to the SDE with continuous sample path

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{474}$$

where the existence and uniqueness of the solution is still ensured (growth condition and Lipschitz condition). We would be able to get the forward and backward Kolmogorov equations for such diffusion process. These two equations show the connection between SDE and PDE and can be understood as the necessary consistency conditions of the solution X_t.

Backward Kolmogorov Equation (BKE)

Now we fix the end time T and consider $t \leq T$, notice that X_t is Markov (similar to that in the time homogeneous case), so it's natural to think about

$$u(t, x) = \mathbb{E}(f(X_T)|X_t = x) \tag{475}$$

where $f \in C_c^2(\mathbb{R})$ as the expectation at a fixed ending time with the initial condition to be with value x given at time t. Here we are varying the condition $X_t = x$, both the initial time and the initial value in order to get a PDE for such u.

Apply tower property and perturb the time by h such that $t + h \leq T$,

$$\begin{aligned}
u(t, x) &= \mathbb{E}(\mathbb{E}[f(X_T)|\mathscr{F}_{t+h}]|X_t = x) &(476)\\
&= \mathbb{E}(u(t + h, X_{t+h})|X_t = x) &(477)\\
&= \mathbb{E}\left(u(t, X_t) + \int_t^{t+h} u_s(s, X_s)\, ds + \int_t^{t+h} u_x(s, X_s)\, dX_s + \frac{1}{2}\int_t^{t+h} u_{xx}(s, X_s)\, d\langle X, X\rangle_s \Big| X_t = x\right) &(478)\\
&= u(t, x) + \mathbb{E}\left(\int_t^{t+h} (\partial_s + L)u(s, X_s)\, ds + \int_t^{t+h} u(s, X_s)\sigma(s, X_s)\, dB_s \Big| X_t = x\right) &(479)
\end{aligned}$$

as a result, we know that $\forall t + h \leq T$,

$$\mathbb{E}\left(\int_t^{t+h} (\partial_s + L)u(s, X_s)\, ds + \int_t^{t+h} u(s, X_s)\sigma(s, X_s)\, dB_s \Big| X_t = x\right) = 0 \tag{480}$$

for the stochastic integral term, by Markov property,

$$\mathbb{E}\left(\int_t^{t+h} u(s,X_s)\sigma(s,X_s)\,dB_s\Big|X_t=x\right)=\mathbb{E}\left(\int_0^h u(s,X_s)\sigma(s,X_s)\,dB_s\Big|X_0=x\right) \tag{481}$$

when $Y_h=\int_0^h u(s,X_s)\sigma(s,X_s)\,dB_s$ is a MG, such conditional expectation has value 0, so it will disappear in the equation. Obviously, one sufficient condition to satisfy such that Y_h is a MG is that $u(s,X_s)\sigma(s,X_s)\in L^2([0,T]\times\Omega)$. In other words, $\mathbb{E}\left(\int_0^T u^2(s,X_s)\sigma^2(s,X_s)\,ds\right)<\infty$. Now divide this quantity by h and apply the intermediate value theorem for integral to find that

$$\frac{1}{h}\mathbb{E}\left(\int_t^{t+h}(\partial_s+L)u(s,X_s)\,ds\Big|X_t=x\right)=0 \tag{482}$$

$$\exists\xi\in[t,t+h],\mathbb{E}\left((\partial_s+L)u(\xi,X_\xi)\Big|X_t=x\right)=0 \tag{483}$$

setting $h\to 0$ and notice the continuity of sample path of X_t

$$(\partial_t+L)u(t,x)=0 \tag{484}$$

as a result, the **backward Kolmogorov equation** is given by

$$\begin{cases}(\partial_t+L)u=0\\ u(T,x)=f(x)\end{cases} \tag{485}$$

Conversely, let's prove that if the backward Kolmogorov equation holds for u such that $(\partial_t+L)u=0,u(T,x)=f(x)$, then such u must be the conditional expectation of $f(X_T)$ given $X_t=x$. Apply Ito formula to get

$$\begin{aligned}\mathbb{E}(f(X_T)|X_t=x)&=\mathbb{E}(u(T,X_T)|X_t=x) &(486)\\ &=u(t,x)+\mathbb{E}\left(\int_t^T\partial_s u(s,X_s)\,ds+\int_t^T\partial_x u(s,X_s)\,dX_s+\frac{1}{2}\int_t^T\partial_{xx}u(s,X_s)\,d\langle X,X\rangle_s\Big|X_t=x\right) &(487)\\ &=u(t,x)+\mathbb{E}\left(\int_t^{t+h}(\partial_s+L)u(s,X_s)\,ds+\int_t^{t+h}u(s,X_s)\sigma(s,X_s)\,dB_s\Big|X_t=x\right) &(488)\\ &=u(t,x) &(489)\end{aligned}$$

under the same condition that $Y_h=\int_0^h u(s,X_s)\sigma(s,X_s)\,dB_s$ is a MG.

Remark. *The backward Kolmogorov equation is* ***a PDE with backward time flow*** *derived from* ***an SDE with forward time flow****. Moreover, such equation characterizes the function u, providing connections with the diffusion process X_t.*

One perspective to understand the backward equation is that we first fix the initial condition as (t_1,x_1) and *generate the diffusion X_t based on the dynamics of the SDE, then at fixed time T by applying the function f for the value of the diffusion and taking expectation we would be able to get $u(t_1, x_1)$. The similar approach can be taken to set the initial condition as (t_2, x_2) $(t_1 < t_2)$ and generate the diffusion X_t again based on the dynamics of the SDE to recover $u(t_2, x_2)$. Then u must follow the backward Kolmogorov equation such that those two u values are consistent with the dynamics of the SDE. As a result, the backward Kolmorogov equation can be interpreted as a* ***consistency condition****.*

Forward Kolmogorov Equation (FKE)

To get the forward time flow of a PDE derived from such SDE, set $X_0 \sim \rho_0$ with ρ_0 as an initial probability density. Denote $\rho(t,x)$ as the density of X_t given $X_0 \sim \rho_0$ where t is the time variable and x is the space variable, i.e. $\mathbb{P}(X_t \in A | X_0 \sim \rho_0) = \int_A \rho(t,x)\, dx$. We are expecting that there exists a PDE w.r.t. $\rho(t,x)$ with forward time flow describing the evolution of the diffusion X_t.

Let's start with its connection with BKE that for fixed time T and $t \leq T$,

$$\begin{aligned}
\mathbb{E}(f(X_T)|X_0 \sim \rho_0) &= \mathbb{E}(\mathbb{E}(f(X_T)|X_t)|X_0 \sim \rho_0) && (490)\\
&= \mathbb{E}(u(t, X_t)|X_0 \sim \rho_0) && (491)\\
&= \int u(t,x)\rho(t,x)\, dx && (492)
\end{aligned}$$

it's clear that the quantity on the left hand side shall be independent of variable t, so by taking derivatives w.r.t. t, we get

$$\begin{aligned}
\partial_t \int u(t,x)\rho(t,x)\, dx = 0 && (493)\\
\int \partial_t u \cdot \rho\, dx + \int u \cdot \partial_t \rho\, dx = 0 && (494)
\end{aligned}$$

now plug in the BKE we have just got to find

$$-\int Lu \cdot \rho\, dx + \int u \cdot \partial_t \rho\, dx = 0 \tag{495}$$

denote L^* to be the adjoint operator of the infinitesimal generator under the inner product $\langle f, g\rangle = \int f \cdot g\, dx$ to get

$$\begin{aligned}
-\int u \cdot L^* \rho\, dx + \int u \cdot \partial_t \rho\, dx = 0 && (496)\\
\int u \cdot (-L^* + \partial_t)\rho\, dx = 0 && (497)\\
(-L^* + \partial_t)\rho = 0 && (498)
\end{aligned}$$

The **forward Kolmogorov equation** is then given by

$$\begin{cases}(\partial_t - L^*)\rho = 0\\ \rho(0,x) = \rho_0(x)\end{cases} \tag{499}$$

Remark. *The FKE is a PDE with forward time flow and a **Fokker-Planck equation** that describes the time evolution of probability density.*

*One might be curious about the use of the function f since we are introducing such function into u but such function has no appearances in both BKE and FKE. Actually, this f is an **auxiliary function** that can be selected to have good enough properties.*

For the last step of the derivation from $\int u\cdot(-L^*+\partial_t)\rho\,dx = 0$ to $(-L^*+\partial_t)\rho = 0$, we are making use of such f since we can vary f such that u traverses through all elements in a dense subset of the L^2 function space. By doing this, the derivation becomes natural. (Details are not presented here)

At last, let's derive the explicit form of the adjoint of the infinitesimal generator. By the definition of adjoint operator, for any density f, g

$$\langle Lf, g\rangle_{L^2} = \langle f, L^* g\rangle_{L^2} \tag{500}$$

$$\begin{aligned}\langle Lf, g\rangle_{L^2} &= \int \left(b\cdot\partial_x f + \frac{\sigma^2}{2}\cdot\partial_{xx} f\right)\cdot g\,dx &(501)\\ &= -\int \partial_x(b\cdot g)\cdot f\,dx - \frac{1}{2}\int \partial_x(\sigma^2\cdot g)\cdot\partial_x f\,dx &(502)\\ &= -\int \partial_x(b\cdot g)\cdot f\,dx + \frac{1}{2}\int \partial_{xx}(\sigma^2\cdot g)\cdot f\,dx &(503)\\ &= \int \left[-\partial_x(b\cdot g) + \frac{1}{2}\partial_{xx}(\sigma^2\cdot g)\right]\cdot f\,dx &(504)\end{aligned}$$

where all boundary terms disappear because f, g are densities and shrink to 0 at ∞. We conclude that **the adjoint of the infinitesimal generator** has the action

$$\forall f, L^* f = -\partial_x(b\cdot f) + \frac{1}{2}\partial_{xx}(\sigma^2\cdot f) \tag{505}$$

this provides all the details for the FKE.

Week 7

Remark. *We know that for general diffusion process X_t (could be time inhomogeneous, b,σ could depend on t), fix $T>0$ and set $u(t,x)=\mathbb{E}(f(X_T)|X_t=x)$ then we would get*

$$\begin{cases}\partial_t u + Lu = 0\\ u(T,x)=f(x)\end{cases} \tag{506}$$

when X_t is ***time homogeneous****, however, by setting $\tau = T-t$ and $\tilde{u}(\tau,x)=u(T-\tau,x)=u(t,x)$, the BKE becomes*

$$\begin{cases}\partial_\tau \tilde{u} = L\tilde{u}\\ \tilde{u}(0,x)=f(x)\end{cases} \tag{507}$$

since $\partial_t u = \partial_\tau \tilde{u}, Lu = L\tilde{u}$. By time homogeneity, $\tilde{u}(\tau,x)=\mathbb{E}(f(X_T)|X_{T-\tau}=x)=\mathbb{E}(f(X_\tau)|X_0=x)$ still provides the same probabilistic interpretation of BKE. Note that it's a PDE with ***initial condition and forward time flow!***

More about BKE and FKE

In the time homogeneous case, consider s as the solution to BKE. An interesting $s(x)$ would be the one such that it's independent of time t, so by BKE, $Ls=0$. Such s is called a **scale function**. By its definition,

$$b(x)\cdot s' + \frac{1}{2}\sigma^2(x)\cdot s'' = 0 \tag{508}$$

we can solve out the form of such s that

$$s'(x) = s'(x_0)\cdot e^{-\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)}\,dy} \tag{509}$$

$$s(x) = s(x_0) + \int_{x_0}^x s'(t_0)\cdot e^{-\int_{t_0}^t \frac{2b(y)}{\sigma^2(y)}\,dy}\,dt \tag{510}$$

such $s(X_t)$ is always a **continuous local MG** since $Ls=0$

$$ds(X_t) = s'(X_t)\,dX_t + \frac{1}{2}s''(X_t)\,d\langle X,X\rangle_t \tag{511}$$

$$= s'(X_t)\sigma(X_t)\,dB_t \tag{512}$$

Similarly, we can look into the function $\Phi(x)$ as the solution to FKE independent of time t, by FKE, $L^*\Phi=0$ and such Φ is defined as the **invariant distribution**. To see why this definition of invariant distribution is consistent with the one we defined in the earlier context, notice that if $X_0\sim\rho_0(x)=\Phi(x)$, then $X_t\sim\rho_(t,x)=\Phi(x)$. The reason is that Φ is a solution to BKE and is independent of time t, so it characterizes the evolution of the density of the diffusion process. As a result, if the diffusion starts with initial density $\Phi(x)$ and evolves following the dynamics of the SDE, at any time the diffusion process still follows density $\Phi(x)$.

Example

For OU process, it's defined as the solution to the SDE

$$dX_t = \alpha(m - X_t)\, dt + \sigma\, dB_t \tag{513}$$

in previous context, we have proved by solving the SDE and doing calculations that $N\left(m, \frac{\sigma^2}{2\alpha}\right)$ is the invariant distribution of such OU process. Now we use the adjoint infinitesimal generator to compute its invariant distribution.

$$L^*\Phi(x) = 0 \tag{514}$$

$$-(\alpha(m-x)\cdot\Phi(x))' + \frac{1}{2}\sigma^2\Phi''(x) = 0 \tag{515}$$

$$\alpha\Phi(x) - \alpha(m-x)\Phi'(x) + \frac{1}{2}\sigma^2\Phi''(x) = 0 \tag{516}$$

one can verify that $\Phi(x) = C\cdot e^{-\frac{\alpha(x-m)^2}{\sigma^2}}$ is the solution to this ODE. Since FKE is the equation for density function, adding the condition that $\int \Phi(x)\, dx = 1$, one would be able to conclude that $\Phi(x)$ is just the density of $N\left(m, \frac{\sigma^2}{2\alpha}\right)$.

FKE may help us prove things relevant with the invariant distribution. For BM, it's natural to realize that there's no invariant distribution exists. To prove this, let's consider 1-dimensional BM B_t, and it's clear that $L = L^* = \frac{1}{2}\partial_{xx}$. If there exists $\Phi(x)$ as the density of the invariant distribution,

$$\Phi''(x) = 0 \tag{517}$$

$$\Phi(x) = C_1x + C_2 \tag{518}$$

must be a linear density function on $\mathbb{R}$. However, this is impossible since if $C_1 \neq 0$ then it's not integrable and if $C_1 = 0$, the integral is always 0.

To generalize it and prove that invariant distribution does not exists for n-dimensional BM, such $\Phi(x)$ shall satisfy

$$\Delta\Phi = 0 \tag{519}$$

as a harmonic function on $\mathbb{R}^n$. However, since $\int \Phi(x)\, dx = 1$, such function must be bounded and thus constant by Louville theorem. However, all constants won't satisfy $\int \Phi(x)\, dx = 1$, so invariant distribution does not exist.

The Resolvent Operator

It's easy to observe that the infinitesimal generator L has no inverse since $Lg = 0$ for any constant function g. This gives rise to the definition of the **resolvent operator** R_α as $(\alpha - L)^{-1}$ for $\forall \alpha > 0$. The definition of the resolvent operator for Ito diffusion (time homogeneous) X_t is

$$R_\alpha g(x) = \mathbb{E}\left(\int_0^\infty e^{-\alpha t} g(X_t)\, dt \Big| X_0 = x\right),\ (\alpha > 0, g \in C_b) \tag{520}$$

multiplying an extra exponential decaying term and integrate. Firstly let's prove some properties of such resolvent operator.

Theorem 20. *$R_\alpha g$ is bounded and continuous in x.*

Proof. To prove the continuity, we first prove that for l.s.c. lower bounded g and fixed time $t > 0$ with

$$u(x) = \mathbb{E}_x g(X_t) = \mathbb{E}(g(X_t)|X_0 = x) \tag{521}$$

if g is lower semi-continuous (l.s.c.), then u is also l.s.c.

Now for fixed time $t > 0$, we want to see how the change in the initial condition affects the solution. Replicate the calculations for the uniqueness of strong solution to SDE to get

$$\begin{aligned}
\mathbb{E}(X_t^x - X_t^y)^2 &= \mathbb{E}\left(x - y + \int_0^t [b(s, X_s^x) - b(s, X_s^y)]\, ds + \int_0^t [\sigma(s, X_s^x) - \sigma(s, X_s^y)]\, dB_s\right)^2 &(522)\\
&\le 3(x-y)^2 + 3\mathbb{E}\left(\int_0^t [b(s, X_s^x) - b(s, X_s^y)]\, ds\right)^2 + 3\mathbb{E}\left(\int_0^t [\sigma(s, X_s^x) - \sigma(s, X_s^y)]\, dB_s\right)^2 &(523)\\
&\le 3(x-y)^2 + 3t \cdot \mathbb{E}\int_0^t [b(s, X_s^x) - b(s, X_s^y)]^2\, ds + 3\mathbb{E}\int_0^t [\sigma(s, X_s^x) - \sigma(s, X_s^y)]^2\, ds &(524)\\
&\le 3(x-y)^2 + (3D^2 t + 3D^2) \cdot \int_0^t \mathbb{E}(X_s^x - X_s^y)^2\, ds &(525)
\end{aligned}$$

where X_t^x denotes the solution to the SDE $dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$ starting from x at time 0. Apply Grownwall's inequality for $v(t) = \mathbb{E}(X_t^x - X_t^y)^2$ to get

$$\begin{aligned}
v(t) &\le 3(x-y)^2 + (3D^2 t + 3D^2)\int_0^t v(s)\, ds &(526)\\
v(t) &\le 3(x-y)^2 \cdot C(t) &(527)
\end{aligned}$$

if $t \in [0, T]$ for $\forall T > 0$, where $C(t) > 0$ is a constant that only depends on t but contains no x, y.

As a result, for any y_n such that $y_n \to x$ $(n \to \infty)$, we have $\mathbb{E}(X_t^x - X_t^{y_n})^2 \to 0$ $(n \to \infty)$ and

$$\forall T > 0, \forall t \in [0, T], X_t^{y_n} \xrightarrow{L^2} X_t^x\ (n \to \infty) \tag{528}$$

there exists a subsequence y_{n_k} such that

$$y_{n_k} \to x \ (k \to \infty) \tag{529}$$

$$\forall T > 0, \forall t \in [0,T], X_t^{y_{n_k}} \xrightarrow{a.s.} X_t^x \ (k \to \infty) \tag{530}$$

By Fatou's lemma and knowing g is l.s.c.,

$$u(x) = \mathbb{E}g(X_t^x) \leq \mathbb{E}\underline{\lim}_{k\to\infty} g(X_t^{y_{n_k}}) \leq \underline{\lim}_{k\to\infty}\mathbb{E}g(X_t^{y_{n_k}}) = \underline{\lim}_{k\to\infty} u(y_{n_k}) \tag{531}$$

it's proved that u is also l.s.c.

To see why the resolvent operator is continuous and bounded, consider $h(x) = \int_0^\infty e^{-\alpha t} g(x)\,dt$. When g is continuous and bounded, h is obviously continuous bounded. Apply the proposition above for h and $-h$ to conclude that

$$R_\alpha g(x) = \mathbb{E}(h(X_t)|X_0 = x) \tag{532}$$

is bounded and both l.s.c. and u.s.c. in x, so $R_\alpha g$ is continuous bounded.

□

Remark. *For the resolvent operator, there's no need to worry about the asymptotic growth of X_t in time t since we only care about $g(X_t)$ and g is always bounded. Instead, one should be careful with the growth in x when the initial condition of the SDE is varying.*

The motivation of the resolvent operator is that it's actually the representation of $(\alpha - L)^{-1}$ for $\alpha > 0$.

Theorem 21. ***(Inverse Representation of Resolvent Operator)*** *Let D_L be the set of functions whose action under the infinitesimal generator is well-defined, then $\forall f \in C_c^2, R_\alpha(\alpha - L)f = f$ and $\forall g \in C_b, R - \alpha g \in D_L, (\alpha - L)R_\alpha g = g$.*

Proof. First let's verify that

$$\begin{aligned}\forall f \in C_c^2, R_\alpha(\alpha - L)f(x) &= \alpha R_\alpha f(x) - R_\alpha Lf(x) &(533)\\ &= \alpha\mathbb{E}_x\left(\int_0^\infty e^{-\alpha t} f(X_t)\,dt\right) - \mathbb{E}_x\left(\int_0^\infty e^{-\alpha t}\cdot Lf(X_t)\,dt\right) &(534)\end{aligned}$$

apply integration by parts to get

$$\begin{aligned}&= \mathbb{E}f(X_0) + \mathbb{E}_x\left(\int_0^\infty e^{-\alpha t}\,df(X_t)\right) - \mathbb{E}_x\left(\int_0^\infty e^{-\alpha t}\cdot Lf(X_t)\,dt\right) &(535)\\ &= \mathbb{E}f(X_0) + \int_0^\infty e^{-\alpha t}\frac{d}{dt}\mathbb{E}_x f(X_t)\,dt - \int_0^\infty e^{-\alpha t}\cdot\mathbb{E}_x Lf(X_t)\,dt &(536)\end{aligned}$$

note that for Ito diffusion (time homogeneous) X_t, we have proved that

$$\frac{dP_t}{dt} = P_t L \tag{537}$$

$$\frac{d}{dt}\mathbb{E}_x f(X_t) = \frac{dP_t}{dt} f(x) = P_t L f(x) = \mathbb{E}_x(Lf(X_t)) \tag{538}$$

so we have proves that $R_\alpha(\alpha - L)f(x) = \mathbb{E}f(X_0) = f(x)$.

To verify the other equation, let's calculate

$$LR_\alpha g(x) = \lim_{t\to 0} \frac{P_t - id}{t} R_\alpha g(x) \tag{539}$$

$$= \lim_{t\to 0} \frac{\mathbb{E}_x(R_\alpha g(X_t)) - R_\alpha g(x)}{t} \tag{540}$$

first we have to calculate the expectation

$$\mathbb{E}_x(R_\alpha g(X_t)) = \mathbb{E}_x \mathbb{E}_{X_t}\left(\int_0^\infty e^{-\alpha s} g(X_s)\, ds\right) \tag{541}$$

$$= \mathbb{E}_x \mathbb{E}_x\left(\theta_t \circ \int_0^\infty e^{-\alpha s} g(X_s)\, ds \middle| \mathscr{F}_t\right) \tag{542}$$

where θ_t is the shift operator for time t mapping $\{\omega_s\}_{s\geq 0}$ to $\{\omega_s\}_{s\geq t}$ for sample point $\{\omega_s\}$ in the probability space of the whole process. So here we are actually applying the Markov property in a reversed way.

$$= \mathbb{E}_x \mathbb{E}_x\left(\int_0^\infty e^{-\alpha s} g(X_{s+t})\, ds \middle| \mathscr{F}_t\right) \tag{543}$$

$$= \mathbb{E}_x \int_0^\infty e^{-\alpha s} g(X_{s+t})\, ds \tag{544}$$

$$= \int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_{s+t})\, ds \tag{545}$$

Now we can continue to calculate $LR_\alpha g(x)$

$$LR_\alpha g(x) = \lim_{t\to 0} \frac{\int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_{s+t})\, ds - \int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_s)\, ds}{t} \tag{546}$$

$$= \lim_{t\to 0} \frac{\int_0^\infty \left(\int_u^\infty \alpha e^{-\alpha s}\, ds\right) \mathbb{E}_x g(X_{u+t})\, du - \int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_s)\, ds}{t} \tag{547}$$

$$= \lim_{t\to 0} \frac{\int_0^\infty \alpha e^{-\alpha s} \int_0^s \mathbb{E}_x g(X_{u+t})\, du\, ds - \int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_s)\, ds}{t} \tag{548}$$

$$= \lim_{t\to 0} \frac{\int_0^\infty \alpha e^{-\alpha s} \int_t^{s+t} \mathbb{E}_x g(X_u)\, du\, ds - \int_0^\infty e^{-\alpha s} \cdot \mathbb{E}_x g(X_s)\, ds}{t} \tag{549}$$

by writing $e^{-\alpha s}$ as an integral and apply Fubini to interchange the order (a frequently used trick). Now add and subtract a same term $\int_0^\infty \alpha e^{-\alpha s}\int_t^s \mathbb{E}_x g(X_u)\,du\,ds$ to simplify the expression

$$= \lim_{t\to 0}\frac{\int_0^\infty \alpha e^{-\alpha s}\int_s^{s+t}\mathbb{E}_x g(X_u)\,du\,ds + \int_0^\infty e^{-\alpha s}[\alpha\int_t^s \mathbb{E}_x g(X_u)\,du - \mathbb{E}_x g(X_s)]\,ds}{t} \tag{550}$$

$$= \alpha\mathbb{E}_x\int_0^\infty e^{-\alpha s} g(X_s)\,ds + \lim_{t\to 0}\frac{\int_0^\infty e^{-\alpha s}[\alpha\int_t^s \mathbb{E}_x g(X_u)\,du - \mathbb{E}_x g(X_s)]\,ds}{t} \tag{551}$$

by applying the intermediate value theorem for the first term on the numerator and using the continuity of g and X_t. For the term left on the numerator, apply the integration by parts to find

$$\lim_{t\to 0}\frac{\int_0^\infty e^{-\alpha s}[\alpha\int_t^s \mathbb{E}_x g(X_u)\,du - \mathbb{E}_x g(X_s)]\,ds}{t} \tag{552}$$

$$= \lim_{t\to 0}\frac{\alpha\int_0^\infty e^{-\alpha s}\int_t^s \mathbb{E}_x g(X_u)\,du\,ds - \int_0^\infty e^{-\alpha s}\mathbb{E}_x g(X_s)\,ds}{t} \tag{553}$$

$$= \lim_{t\to 0}\frac{-\int_0^t \mathbb{E}_x g(X_u)\,du}{t} \tag{554}$$

$$= -\lim_{t\to 0}\mathbb{E}_x g(X_t) \tag{555}$$

$$= -\mathbb{E}_x g(X_0) = -g(x) \tag{556}$$

is the consequence of the continuity of g and X_t once again.

As a result, one can conclude that

$$LR_\alpha g(x) = \alpha R_\alpha g(x) - g(x) \tag{557}$$

and the conclusion is proved.

□

Remark. *Note that the resolvent operator is defined only for* ***time homogeneous*** *Ito diffusion X_t since in the prove of the theorem above we have applied the property of semi-group generator and the infinitesimal generator that $\frac{dP_t}{dt} = P_t L$, which is not true in the time inhomogeneous case.*

Feynman-Kac Formula

Now let's consider a generalization of BKE which is for fixed $T > 0$ and $u = u(t,x)$

$$\begin{cases} \partial_t u + Lu + f = 0 \\ u(T,x) = \psi(x) \end{cases} \tag{558}$$

note that an extra potential term f is added to the BKE $\partial_t u + Lu = 0$.

Theorem 22. ***(Feynman-Kac Formula)*** *Let $u = u(t,x)$ be the solution to the PDE above with $f \in C_c^2$, then* *there exists a probabilistic representation*

$$u(t,x) = \mathbb{E}\left(\int_t^T f(u,X_u)\,du + \psi(X_T)\Big|X_t = x\right) \tag{559}$$

that characterizes the solution.

Proof. Set $Y_s = u(s,X_s) + \int_t^s f(u,X_u)\,du$ and apply Ito formula to get

$$dY_s = \partial_s u\,ds + \partial_x u\,dX_s + \frac{1}{2}\partial_{xx}u\,d\langle X,X\rangle_s + f(s,X_s)\,ds \tag{560}$$

$$= \partial_s u\,ds + Lu\,ds + \partial_x u\cdot\sigma(X_s)\,dB_s + f(s,X_s)\,ds \tag{561}$$

If u is the solution to the PDE,

$$dY_s = \partial_x u\cdot\sigma(X_s)\,dB_s \tag{562}$$

assume that such Y_s is a MG (which does not necessarily hold), then since $Y_t = u(t,X_t), \forall t < T, \mathbb{E}(Y_T|X_t) = Y_t$,

$$u(t,x) = \mathbb{E}\left(u(T,X_T) + \int_t^T f(u,X_u)\,du\Big|X_t = x\right) \tag{563}$$

$$= \mathbb{E}\left(\psi(X_T) + \int_t^T f(u,X_u)\,du\Big|X_t = x\right) \tag{564}$$

so it has the probabilistic representation.

Conversely, if u has such probabilistic representation, then by tower property, $\forall h > 0$,

$$u(t,x) = \mathbb{E}\left[\mathbb{E}\left(\int_t^T f(u,X_u)\,du + \psi(X_T)\Big|\mathscr{F}_{t+h}\right)\Big|X_t = x\right] \tag{565}$$

$$= \mathbb{E}\left[u(t+h,X_{t+h}) + \int_t^{t+h} f(u,X_u)\,du\Big|X_t = x\right] \tag{566}$$

$$= \mathbb{E}\left[u(t,X_t) + \int_t^{t+h}\partial_t u\,dt + \int_t^{t+h}\partial_x u\,dX_t + \int_t^{t+h} f(u,X_u)\,du + \frac{1}{2}\int_t^{t+h}\partial_{xx}u\,d\langle X,X\rangle_t\Big|X_t = x\right] \tag{567}$$

$$= u(t,x) + \mathbb{E}\left[\int_t^{t+h}[(\partial_t + L)u + f]\,dt + \int_t^{t+h}\partial_x u\cdot\sigma(t,X_t)\,dB_t\Big|X_t = x\right] \tag{568}$$

assume that $\int_t^{t+h}\partial_x u\cdot\sigma(s,X_s)\,dB_s$ is a MG in h, then

$$(\partial_t + L)u + f = 0 \tag{569}$$

$$u(T,x) = \psi(x) \tag{570}$$

which gives the PDE with initial condition. This has the similar derivation as BKE. $\square$

Remark. *Actually in the derivation of BKE, FKE and Feymann-Kac formula, we are all assuming that the Ito integral part is a martingale. This assumption is necessary for those equations to hold since Ito integral is generally a local martingale.*

Remark. *By setting f as 0, we are getting the BKE. In such case, the Feymann-Kac formula becomes $u(t,x) = \mathbb{E}(\psi(X_T)|X_t = x)$, which is exactly the function we have defined to derive BKE.*

In the most general case, Feynmnn-Kac formula can work not only with a potential f but also with a linear term $V(t,x)u$ added to the equation. The Feynman-Kac formula is stated in the following form.

Theorem 23. ***(Feynman-Kac Formula in the General Case)*** *Let $u = u(t,x)$ be the solution to the PDE*

$$\begin{cases} (\partial_t + L)u - Vu + f = 0 \\ u(T,x) = \psi(x) \end{cases} \tag{571}$$

where $f \in C_c^2$ and $V(t,x)$ is a continuous lower bounded function, then there exists a probabilistic representation

$$u(t,x) = \mathbb{E}\left(\int_t^T e^{-\int_t^r V(s,X_s)\,ds} f(r,X_r)\,dr + e^{-\int_t^T V(s,X_s)\,ds}\psi(X_T)\Big| X_t = x\right) \tag{572}$$

that characterizes the solution, the coefficient of u in the PDE will appear as an exponential factor.

Proof. Let's do the similar thing to what we have done above. Set

$$Y_s = \int_t^s e^{-\int_t^r V(p,X_p)\,dp} f(r,X_r)\,dr + e^{-\int_t^s V(r,X_r)\,dr} u(s,X_s) \tag{573}$$

and apply Ito formula to see

$$dY_s = h_s f(s,X_s)\,ds + u(s,X_s)\,dh_s + h_s\,du(s,X_s) \tag{574}$$

where $h_u = e^{-\int_t^u V(r,X_r)\,dr}$ is the exponential process. Apply Ito formula once more to find

$$dh_s = -h_s V(s,X_s)\,ds \tag{575}$$

$$du(s,X_s) = (\partial_t + L)u\,ds + \partial_x u \cdot \sigma(s,X_s)\,dB_s \tag{576}$$

plug in to get

$$dY_s = h_s \cdot [(\partial_t + L)u - Vu + f]\,ds + h_s \sigma \partial_x u\,dB_s \tag{577}$$

If u is the solution to the PDE, we get

$$dY_s = h_s \sigma \partial_x u \, dB_s \tag{578}$$

assume such Y_s is a MG, since $Y_t = u(t, X_t)$

$$u(t,x) = \mathbb{E}(Y_T | X_t = x) \tag{579}$$

the representation is proved.

The converse direction is omitted, but the spirit is still to apply tower property and perturb time t by $\forall h > 0$. Use Ito formula to expand terms and to conclude.

$\square$

Remark. *The BKE, FKE, Feynman-Kac formula shows the connection between diffusion process and PDE. The connection is important because it shows us the possibility to solve PDE by simulations and the way to compute conditional expectations by solving PDE.*

If we have an equation

$$\begin{cases} (\partial_t + L)u - Vu + f = 0 \\ u(T,x) = \psi(x) \end{cases} \tag{580}$$

by its probabilistic representation, we can simulate the diffusion X_s to let it start from time t with initial value x. Then Monte Carlo methods tell us an estimate of $\mathbb{E}\left(\int_t^T e^{-\int_t^r V(s,X_s)\,ds} f(r, X_r)\,dr + e^{-\int_t^T V(s,X_s)\,ds}\psi(X_T)\middle| X_t = x\right)$ for each (t,x) pair. By varying the pair (t,x), one would get the numerical solution to such PDE.

Week 8

Ito Process as Diffusion

The Ito formula is telling us that for any $f \in C^2(\mathbb{R}^n)$, if X_t is an Ito process, then $f(X_t)$ is also an Ito process. However, if X_t is an Ito diffusion, then $f(X_t)$ is not necessarily an Ito diffusion. The motivation of recognizing Ito process as a diffusion process comes from the the Bessel process.

Set B_t as n-dimensional BM, then $R_t = ||B_t||$ is the Bessel process. Apply Ito formula to find

$$dR_t = \sum_{j=1}^n \frac{B_t^j}{R_t}\, dB_t^j + \frac{1}{2}\sum_{j=1}^n \frac{||B_t||^2 - (B_t^i)^2}{||B_t||^3}\, dt \tag{581}$$

$$= \sum_{j=1}^n \frac{B_t^j}{R_t}\, dB_t^j + \frac{1}{2}\frac{nR_t^2 - R_t^2}{R_t^3}\, dt \tag{582}$$

$$= \sum_{j=1}^n \frac{B_t^j}{R_t}\, dB_t^j + \frac{n-1}{2R_t}\, dt \tag{583}$$

where B_t^j is the j-th component of B_t. It seems that dR_t cannot be written in the form $b(t, R_t)\, dt + \sigma(t, R_t) \cdot dB_t$ since now $\sigma(t, R_t)$ also has something to do with B_t. Setting up some theorems to judge whether an Ito process is a diffusion process is then necessary since sometimes we cannot judge directly by the form of the SDE.

The following theorem shows the equivalent condition an Ito process is an Ito diffusion.

Theorem 24. ***(Condition for Ito Process to be an Ito Diffusion)***

X_t is an Ito diffusion defined as the solution to

$$\begin{cases} dX_t = b(X_t)\, dt + \sigma(X_t)\, dB_t \\ X_0 = x \end{cases} \tag{584}$$

and Y_t is an Ito process defined by

$$\begin{cases} dY_t = \varphi_t\, dt + \psi_t\, dB_t \\ Y_0 = x \end{cases} \tag{585}$$

where $X_t, Y_t \in \mathbb{R}^n$ and the BM in the SDE B_t is m-dimensional with $x, b, \varphi_t \in \mathbb{R}^n, \sigma, \psi_t \in \mathbb{R}^{n\times m}$.

Then $\{X_t\}_{t\geq 0} \overset{d}{=} \{Y_t\}_{t\geq 0}$ if and only if

$$\mathbb{E}(\varphi_t | \mathscr{F}_t^Y) = b(Y_t), \psi_t\psi_t^T = \sigma\sigma^T(Y_t)\ a.e.(\omega, t) \tag{586}$$

Remark. *By setting the Ito diffusion X_t as BM, one would know that $b = 0, \sigma\sigma^T = I_n$, plugging the theorem above* *to know that Ito process Y_t is a BM if and only if*

$$\mathbb{E}(\varphi_t|\mathscr{F}_t^Y) = 0, \psi_t\psi_t^T = I_n \tag{587}$$

as already implied by Levy's characterization of BM. The process has to be continuous local MG so there's no drift term, the process shall have the same quadratic variation as the BM so $\psi_t\psi_t^T$ is just the identity matrix.

At this point, one can already argue that the Bessel process is actually an Ito diffusion. The Bessel process can be defined as

$$dR_t = \sum_{j=1}^n \frac{B_t^j}{R_t}\, dB_t^j + \frac{n-1}{2R_t}\, dt \tag{588}$$

and it's obviously an Ito process with

$$\varphi_t = \frac{n-1}{2R_t} \tag{589}$$

$$\psi_t = \left[\frac{B_t^1}{R_t}, ..., \frac{B_t^m}{R_t}\right] \tag{590}$$

one would be able to see that the diffusion part satisfies

$$\psi_t\psi_t^T = \sum_{i=1}^m \frac{(B_t^i)^2}{(R_t)^2} = 1 \tag{591}$$

as a result, this is telling us that

$$\sum_{j=1}^n \frac{B_t^j}{R_t}\, dB_t^j = d\tilde{B}_t \tag{592}$$

is a new 1-dimensional BM $\tilde{B}_t$. Rewrite the SDE as

$$dR_t = d\tilde{B}_t + \frac{n-1}{2R_t}\, dt \tag{593}$$

in the weak solution sense. So **the Bessel process is actually an Ito diffusion in the weak sense**.

Random Time Change

An example for time change is for the Y_t generated by the following SDE

$$dY_t = \sqrt{c(t)}\, dB_t \tag{594}$$

where $c(t) > 0$ is deterministic function and we want to find a process α_t such that $\{Y_{\alpha_t}\}_{t\geq 0} \overset{d}{=} \{B_t\}_{t\geq 0}$.

Observe that since Y_t is a Wiener integral, it's Gaussian

$$Y_t \sim N\left(0, \int_0^t c(s)\, ds\right) \tag{595}$$

as a result, we would expect to see $\{Y_t\}_{t\geq 0} \stackrel{d}{=} \left\{B_{\int_0^t c(s)\, ds}\right\}_{t\geq 0}$. In order to find the time change rate α_t such that Y is a BM under α_t, just plug in to see

$$\{Y_{\alpha_t}\}_{t\geq 0} \stackrel{d}{=} \left\{B_{\int_0^{\alpha_t} c(s)\, ds}\right\}_{t\geq 0} \tag{596}$$

and now to set it as a BM,

$$\int_0^{\alpha_t} c(s)\, ds = t \tag{597}$$

as a result, it's natural to set

$$\alpha_t = \inf\left\{s : \int_0^s c(u)\, du \geq t\right\} \tag{598}$$

then $\{Y_{\alpha_t}\}_{t\geq 0} \stackrel{d}{=} \{B_t\}_{t\geq 0}$.

Remark. *It's easy to see that here the $\int_0^t c(s)\, ds$ is just the quadratic variation of Y in the time interval $[0,t]$. Actually, any continuous local MG M_t is always a time-changed BM with $M_t = B_{\langle M,M\rangle_t}$ if $\langle M,M\rangle_\infty = \infty$. That's why in order to represent the BM as a time changed process of Y, just take the inverse of $\beta_t = \langle Y,Y\rangle_t$ as α_t and Y_{α_t} is just what we want. (just be careful with the filtration)*

In general, Y_t is still generated by the following SDE

$$dY_t = \sqrt{c_t}\, dB_t \tag{599}$$

but here $c_t \geq 0$ is $\mathscr{F}_t$ adapted process and set

$$\beta_t = \int_0^t c_s\, ds \tag{600}$$

as the quadratic variation process of Y. Define

$$\alpha_t = \inf\{s : \beta_s > t\} \tag{601}$$

as the right inverse of β. By construction, α_t is right-continuous in t and $\alpha_{\beta_t} = t$ while $\beta_{\alpha_t} \geq t$. Note that when $c_t > 0$ is strictly positive, β_t is strictly increasing so α_t is the true inverse.

Theorem 25. ***(Time Changed BM)*** *Define the process*

$$\tilde{B}_t = \int_0^{\alpha_t} \sqrt{c_s}\, dB_s \tag{602}$$

for a_t, c_t continuous such that $\forall t, \mathbb{E}\alpha_t < \infty$ then it's a $\mathscr{F}_{\alpha_t}$ BM and for any bounded continuous process v_t adapted to $\mathscr{F}_t$,

$$\int_0^{\alpha_t} v_s\, dB_s = \int_0^t v_{\alpha_s} \sqrt{\alpha'_s}\, d\tilde{B}_s \tag{603}$$

where $\alpha'_s = \frac{1}{c_{\alpha_s}}$ is the derivative w.r.t. s.

Proof. Set $Y_t = \int_0^t \sqrt{c_s}\, dB_s$ and notice that $\{\alpha_t < s\} = \{t < \beta_s\} \in \mathscr{F}_s$ so α_t is a stopping time. The main thought of the proof is to use the fact mentioned in week 1 that **a process Z_t has the same finite-dimensional distribution as BM if and only if**

$$\forall u \in \mathbb{R}, \forall s < t, \mathbb{E}\left(e^{iu(Z_t - Z_s)}\middle|\mathscr{F}_s\right) = e^{-\frac{1}{2}(t-s)u^2} \tag{604}$$

That's why here we consider the exponential local MG of a complex-value process in order to prove that Y_{α_t} satisfies the property above (intuition from characteristic function). Define

$$M_u = e^{i\lambda Y_{\alpha_t \wedge u} + \frac{\lambda^2}{2}\langle Y, Y\rangle_{\alpha_t \wedge u}} = e^{i\lambda Y_{\alpha_t\wedge u} + \frac{\lambda^2}{2}\beta_{\alpha_t \wedge u}} \tag{605}$$

since Y_t is local MG, such M_u is the **exponential local MG** of $h_u = i\lambda Y_{\alpha_t \wedge u}$

$$Y_{\alpha_t \wedge u} = \int_0^{\alpha_t \wedge u} \sqrt{c_s}\, dB_s = \int_0^u \sqrt{c_s} \cdot \mathbb{I}_{s < \alpha_t}\, dB_s \tag{606}$$

$$\langle Y, Y\rangle_{\alpha_t \wedge u} = \int_0^{\alpha_t \wedge u} c_s\, ds \tag{607}$$

By the definition of α_t, $\langle Y, Y\rangle_{\alpha_t \wedge u} \leq t$, so $|M_u| = e^{\frac{\lambda^2}{2}\beta_{\alpha_t \wedge u}} \leq e^{\frac{\lambda^2}{2}t}$ is an upper bound uniform in u.

Now for $\forall t \geq 0$ fixed, $\sup_u |M_u| < \infty$. Since M_u is bounded, it has to be a true MG (bounded convergence theorem holds) and optional stopping theorem applies (U.I. MG) for $\forall s < t, \alpha_s \leq \alpha_t$ *a.s.* gives

$$\mathbb{E}(M_{\alpha_t}|\mathscr{F}_{\alpha_s}) = M_{\alpha_s} \tag{608}$$

plug in the definition of α_t to know

$$\mathbb{E}(e^{i\lambda Y_{\alpha_t} + \frac{\lambda^2 t}{2}}|\mathscr{F}_{\alpha_s}) = e^{i\lambda Y_{\alpha_s} + \frac{\lambda^2 s}{2}} \tag{609}$$

and this is telling us that

$$\mathbb{E}(e^{i\lambda(Y_{\alpha_t}-Y_{\alpha_s})}|\mathscr{F}_{\alpha_s}) = e^{-\frac{\lambda^2}{2}(t-s)} \tag{610}$$

so Y_{α_t} has independent and stationary Gaussian increments. Note that Y_t, α_t are both continuous, so Y_{α_t} has continuous sample path and it's BM with $Y_{\alpha_t} \in \mathscr{F}_{\alpha_t}$.

To show the second property, only need to show that $\int_0^{\alpha_t} v_s\sqrt{c_s}\,dB_s = \int_0^t v_{\alpha_s}\,d\tilde{B}_s$. Let's show only for $v_s = \mathbb{I}_{(0,\alpha]}(s)$ now

$$\begin{aligned}
\int_0^t v_{\alpha_s}\,d\tilde{B}_s &= \int_0^t \mathbb{I}_{\alpha_s<\alpha}\,d\tilde{B}_s && (611)\\
&= \int_0^t \mathbb{I}_{\beta_\alpha>s}\,d\tilde{B}_s && (612)\\
&= \int_0^{\beta_\alpha\wedge t} d\tilde{B}_s && (613)\\
&= \tilde{B}_{t\wedge\beta_\alpha} && (614)\\
&= \int_0^{\alpha_{t\wedge\beta_\alpha}} \sqrt{c_s}\,dB_s && (615)\\
&= \int_0^{\alpha_t\wedge\alpha} \sqrt{c_s}\,dB_s && (616)\\
&= \int_0^{\alpha_t} \mathbb{I}_{(0,\alpha]}(s)\sqrt{c_s}\,dB_s && (617)\\
&= \int_0^{\alpha_t} v_s\sqrt{c_s}\,dB_s && (618)
\end{aligned}$$

the conclusion follows by setting $\tilde{v}_s = v_s\sqrt{c_s}$. It's easy to see that such conclusion holds for any elementary process v_s. By noting that elementary process is dense, this proves the conclusion for all v_s.

□

Remark. *Actually Y_{α_t} is only defined for $t < \beta_\infty$. As a result, $\beta_\infty = \infty$* ***is needed if time change is expected to work at all time****, otherwise the time changed BM works only for $t < \beta_\infty$.*

To make life simpler, just understand α_t as the inverse of β_t, which is the quadratic variation. ***The spirit is that quadratic variation is always the rate of random time change.***

The derivative α'_t can be understood in the classical way that

$$\begin{aligned}
\alpha(t) &= \beta^{-1}(t) && (619)\\
\alpha'(t) &= [\beta'(\beta^{-1}(t))]^{-1} = \frac{1}{c(\beta^{-1}(t))} = \frac{1}{c(\alpha_t)} && (620)
\end{aligned}$$

consistent with classical definition of derivative. Random time change is often used in finding weak solutions solving *the SDE with the form*

$$dX_t = \sigma(X_t)\, dB_t \tag{621}$$

and it provides a way to absorb the diffusion coefficient σ into the BM. However, random time change only works for continuous local MG so it does not allow the drift coefficient to appear.

Girsanov Theorem

In contrast to random time change, Girsanov theorem **absorbs the drift coefficient**, telling us that if Ito process Y_t satisfies

$$dY_t = \alpha_t\, dt + dB_t \tag{622}$$

with drift coefficient α_t, then it is actually a new BM under a new probability measure with the given Radon-Nikodym derivative. Before building up the Girsanov theorem, let's first state a technical lemma for the exponential local MG to be a MG.

Lemma 4. ***(Exponential Martingale Condition)*** *Consider exponential local MG $M_t = e^{\int_0^t \alpha_s\, dB_s - \frac{1}{2}\int_0^t \alpha_s^2\, ds}$, if $\mathbb{E}M_\infty = 1$, then it's actually a U.I. MG.*

Proof. It's easy to notice that the stochastic integral part $\int_0^t \alpha_s\, dB_s$ is not controlled, so it's natural to define the stopping time that reduces M_t

$$\tau_n = \inf\left\{s : \left|\int_0^s \alpha_u\, dB_u\right| \geq n\right\} \nearrow \infty\ (n \to \infty) \tag{623}$$

such that the stochastic integral in the stopped process $M_{t\wedge\tau_n}$ is controlled

$$|M_{t\wedge\tau_n}| \leq e^n \tag{624}$$

so for each fixed n, $M_{t\wedge\tau_n}$ is a bounded local MG, thus a bounded MG.

Now that

$$\forall s < t, \mathbb{E}(M_{t\wedge\tau_n}|\mathscr{F}_s) = M_{s\wedge\tau_n} \tag{625}$$

by Fatou's lemma for conditional expectation,

$$M_s = \underline{\lim}_{n\to\infty} M_{s\wedge\tau_n} = \underline{\lim}_{n\to\infty}\mathbb{E}(M_{t\wedge\tau_n}|\mathscr{F}_s) \geq \mathbb{E}(\lim_{n\to\infty} M_{t\wedge\tau_n}|\mathscr{F}_s) = \mathbb{E}(M_t|\mathscr{F}_s) \tag{626}$$

so M_t is a non-negative super-MG with $\forall t > 0, \mathbb{E}M_t \leq \mathbb{E}M_0 = 1$ and $M_t \overset{a.s.}{\to} M_\infty$ $(t \to \infty)$, so by Fatou's lemma,

$$\mathbb{E}M_\infty \leq 1 \tag{627}$$

By Fatou's lemma again, $\forall 0 < s < t$,

$$M_s = \underline{\lim}_{t\to\infty} M_s \geq \underline{\lim}_{t\to\infty} \mathbb{E}(M_t|\mathscr{F}_s) \geq \mathbb{E}(\lim_{t\to\infty} M_t|\mathscr{F}_s) = \mathbb{E}(M_\infty|\mathscr{F}_s) \tag{628}$$

Combine those results with $\mathbb{E}M_\infty = 1$, we are knowing that

$$\forall t > 0, \mathbb{E}M_t = 1 \tag{629}$$

so

$$M_t = \mathbb{E}(M_\infty|\mathscr{F}_t) \tag{630}$$

and M_t is a closed MG, thus it's a U.I. MG. □

Remark. *In general, it's not easy to ensure that $\mathbb{E}M_\infty = 1$, so we always truncate the time w.r.t. a fixed time $T > 0$ and consider M_t with time truncation $t \leq T$. In this case, we can apply the lemma for local MG $N_t = M_{t\wedge T}$ to see that **if** $\mathbb{E}M_T = 1$ **then the exponential local MG** M_t **is a U.I. MG**.*

Theorem 26. ***(Girsanov's Theorem)*** *Consider $M_t = e^{\int_0^t \alpha_s\,dB_s - \frac{1}{2}\int_0^t \alpha_s^2\,ds}$ as the exponential local MG for the local MG $\int_0^t \alpha_s\,dB_s$ and a fixed time T, let $\frac{d\tilde{P}}{dP} = M_T$, if $\mathbb{E}M_T = 1$, then $\tilde{P}$ is a probability measure with*

$$\tilde{B}_s = B_s - \int_0^s \alpha_u\,du \tag{631}$$

to be a BM under $\tilde{P}$ for $s \leq T$.

Proof. Since $\mathbb{E}M_T = 1$, by the last lemma, M_t is a U.I. MG.

Now for given bounded process Z_t such that $Z_t \leq B$ *a.s.*, $i\int_0^s Z_u\,d\tilde{B}_u$ is a local MG, consider its exponential local MG

$$N_s = e^{i\int_0^s Z_u\,d\tilde{B}_u + \frac{1}{2}\int_0^s Z_u^2\,du} \tag{632}$$

then by Ito formula,

$$d(M_tN_t) = N_t\,dM_t + M_t\,dN_t + d\langle M, N\rangle_t \tag{633}$$

by previous calculations on the exponential local MG, one know that

$$\begin{aligned}
dM_t &= M_t\alpha_t\,dB_t && (634)\\
dN_t &= iN_tZ_t\,d\tilde{B}_t = iN_tZ_t(dB_t - \alpha_t\,dt) && (635)\\
M_t &= 1 + \int_0^t M_s\alpha_s\,dB_s,\ N_t = 1 + \int_0^t iN_sZ_s\,d\tilde{B}_s && (636)\\
d\langle M, N\rangle_t &= iM_t\alpha_tN_tZ_t\,d\langle B, \tilde{B}\rangle_t = iM_t\alpha_tN_tZ_t\,dt && (637)
\end{aligned}$$

plug in to get

$$d(M_tN_t) = M_tN_t(\alpha_t + iZ_t)\,dB_t \tag{638}$$

and conclude that M_tN_t is a local MG since the drift term cancels out.

Make use of the following stopping time again

$$\tau_n = \inf\left\{s : \left|\int_0^s \alpha_u\,dB_u\right| \geq n\right\} \nearrow \infty\ (n \to \infty) \tag{639}$$

and the stopped local MG $M_{t\wedge\tau_n}N_{t\wedge\tau_n}$ is bounded for each fixed n in that

$$|M_{t\wedge\tau_n}N_{t\wedge\tau_n}| \leq e^{n+\frac{B^2T}{2}} \tag{640}$$

Now we are knowing that M_t is U.I. MG, N_t is bounded MG, $(MN)_t$ is local MG and $(MN)_{t\wedge\tau_n}$ is bounded MG. Let's try to prove that $(MN)_t$ is a MG. Now

$$\forall s < t, \mathbb{E}(M_{t\wedge\tau_n}N_{t\wedge\tau_n}|\mathscr{F}_s) = M_{s\wedge\tau_n}N_{s\wedge\tau_n} \tag{641}$$

and we hope to set $n \to \infty$ to see that the limit can interchange with the conditional expectation. For this to hold, we need L^1 convergence in n and the U.I. in n would suffice.

Notice that we already have $M_{t\wedge\tau_n} \overset{a.s.}{\to} M_t$ $(n \to \infty)$ and $\mathbb{E}M_T = 1$ is implying that $\forall t \leq T, \mathbb{E}M_t = 1$, so $\mathbb{E}M_{t\wedge\tau_n} \to \mathbb{E}M_t$ $(n \to \infty)$ shows the convergence of L^1 norm. The convergence of L^1 norm and the convergence in probability ensures that $M_{t\wedge\tau_n} \overset{L^1}{\to} M_t$ $(n \to \infty)$ and $M_{t\wedge\tau_n}$ is U.I. in n. Since $N_{t\wedge\tau_n}$ is bounded in n, $MN_{t\wedge\tau_n}$ is U.I. in n and we have thus proved that $(MN)_t$ is a MG.

$$\forall s < t, \mathbb{E}(M_tN_t|\mathscr{F}_s) = M_sN_s \tag{642}$$

Finally, take $Z_t = \sum_{j=1}^n \lambda_j \mathbb{I}_{(t_{j-1},t_j]}(t)$ for any $0 = t_0 < t_1 < ... < t_n = T$ and any $\lambda_j \in \mathbb{R}$, then

$$1 = \mathbb{E}M_0 N_0 = \mathbb{E}M_T N_T = \tilde{\mathbb{E}}N_T \tag{643}$$

$$= \tilde{\mathbb{E}}e^{i\sum_j \lambda_j(\tilde{B}_{t_j} - \tilde{B}_{t_{j-1}}) + \frac{1}{2}\sum_j \lambda_j^2(t_j - t_{j-1})} \tag{644}$$

where $\tilde{\mathbb{E}}$ is the expectation under probability measure $\tilde{\mathbb{P}}$. From this (form of characteristic function) we conclude that $\tilde{B}_t$ has same finite-dimensional distribution as BM. Since it has continuous sample path, it's a BM under $\tilde{\mathbb{P}}$.

□

Remark. *The Girsanov theorem is actually telling us that by* ***subtracting the cross quadratic variation***

$$\left\langle B, \int_0^{\cdot} \alpha_s \, dB_s \right\rangle_t = \int_0^t \alpha_s \, ds \tag{645}$$

from the original BM B_t, one can always find a new process which is a BM under a new probability measure.

A simple way to understand Girsanov theorem is that if we want to change the SDE (with α_t nice enough)

$$dX_t = \alpha_t \, dt + \sigma \, dB_t \tag{646}$$

into the SDE

$$dX_t = \sigma \, d\tilde{B}_t \tag{647}$$

with B_t as a BM under $\mathbb{P}$ and $\tilde{B}_t$ as a BM under $\mathbb{Q}$. Then we can think of "putting $\alpha_t \, dt$ into $\sigma \, dB_t$". To complete this task, one need to see

$$\frac{\alpha_t}{\sigma} \, dt + \, dB_t = d\tilde{B}_t \tag{648}$$

so it's natural to set

$$\tilde{B}_t = B_t + \int_0^t \frac{\alpha_s}{\sigma} \, ds \tag{649}$$

note that the Radon-Nikodym derivative should be given by $\frac{d\tilde{P}}{d\mathbb{P}} = M_T = e^{-\int_0^T \frac{\alpha_s}{\sigma} \, dB_s - \frac{1}{2}\int_0^T \frac{\alpha_s^2}{\sigma^2} \, ds}$.

The ***Radon-Nikodym derivative is given by the exponential local MG induced by*** $\int_0^t \alpha_s \, dB_s$ ***for drift coefficient*** α_s ***evaluated at the end of the period, i.e. time*** T*. Note that Girsanov only applies for finite time interval* $[0, T]$*.*

Remark. *The fact that $M_T > 0$ a.s. tells us the relationship between two probability measure $\mathbb{P}, \tilde{\mathbb{P}}$ that $\mathbb{P} << \tilde{\mathbb{P}}$ and $\tilde{\mathbb{P}} << \mathbb{P}$ so they are actually close to each other.*

One simple application of Girsanov theorem lies in the Black-Scholes model, where S_t stands for the stock price at time t following the geometric BM

$$\frac{dS_t}{S_t} = \alpha_t\,dt + \sigma\,dB_t \tag{650}$$

and the risk-free interest rate as r. One wants to change the drift term into $r\,dt$ to price the option, so one wants to see

$$\begin{aligned}\frac{dS_t}{S_t} &= r\,dt + \sigma\,d\tilde{B}_t &(651)\\ &= rdt + (\alpha_t - r)\,dt + \sigma\,dB_t &(652)\end{aligned}$$

for BM under physical measure $\mathbb{P}$ which is B_t and BM under risk-neutral measure $\mathbb{Q}$ which is $\tilde{B}_t$. In order to absorb the drift term $(\alpha - r)\,dt$, one shall set

$$\tilde{B}_t = B_t + \int_0^t \frac{\alpha_s - r}{\sigma}\,ds \tag{653}$$

with the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\int_0^T \frac{\alpha_s - r}{\sigma}\,dB_s - \frac{1}{2}\int_0^T \frac{(\alpha_s - r)^2}{\sigma^2}\,ds} \tag{654}$$

so one changes the SDE into

$$dX_t = r\,dt + \sigma\,d\tilde{B}_t \tag{655}$$

and $\frac{\alpha_s - r}{\sigma}$ is just the Sharpe ratio.

Week 9

In Girsanov's theorem, there are some frequently used conditions to ensure that $\mathbb{E}M_T = 1$ holds. One of them is the following condition.

Theorem 27. ***(Condition for Girsanov to hold)*** *If for $\forall\varepsilon > 0$ small enough, $\mathbb{E}e^{(\frac{1}{2}+\varepsilon)\int_0^T \alpha_t^2 dt} < \infty$, then $\mathbb{E}M_T = 1$.*

Proof. Define local MG $X_t = \int_0^t \alpha_s \, dB_s$ and consider its exponential local MG

$$M_t = e^{X_t - \frac{1}{2}\langle X,X\rangle_t} \tag{656}$$

consider using the stopping time $\tau_n = \inf\{t : |X_t| \geq n\}$ to bound the X_t. Pick $\lambda, p > 1$ as fixed constant but unspecified, consider

$$\mathbb{E}M_{t\wedge\tau_n}^\lambda = \mathbb{E}\left[e^{\lambda X_{\tau_n\wedge t} - \frac{p\lambda^2}{2}\langle X,X\rangle_{\tau_n\wedge t}} \cdot e^{\frac{\lambda(p\lambda-1)}{2}\langle X,X\rangle_{\tau_n\wedge t}}\right] \tag{657}$$

apply Holder's inequality for conjugate $p, q = \frac{p}{p-1}$ to get

$$\mathbb{E}M_{t\wedge\tau_n}^\lambda \leq \mathbb{E}\left[e^{p\lambda X_{\tau_n\wedge t} - \frac{p^2\lambda^2}{2}\langle X,X\rangle_{\tau_n\wedge t}}\right]^{\frac{1}{p}} \cdot \mathbb{E}\left[e^{\frac{q\lambda(p\lambda-1)}{2}\langle X,X\rangle_{\tau_n\wedge t}}\right]^{\frac{1}{q}} \tag{658}$$

notice that the first term is another exponential local MG and it's bounded, so it's a MG and the first term is actually equal to 1. For the second term, notice that $q\lambda(p\lambda-1) = \frac{p\lambda(p\lambda-1)}{p-1}$ approaches 1 when $\lambda, p \to 1$, that's why $\frac{1}{2}+\varepsilon$ with small enough $\varepsilon > 0$ appears.

$$\mathbb{E}\left[e^{\frac{q\lambda(p\lambda-1)}{2}\langle X,X\rangle_{\tau_n\wedge t}}\right]^{\frac{1}{q}} \leq \mathbb{E}\left[e^{(\frac{1}{2}+\varepsilon)\langle X,X\rangle_t}\right]^{\frac{1}{q}} < \infty \tag{659}$$

This is telling us that $M_{t\wedge\tau_n} \in L^\lambda$, so $M_{t\wedge\tau_n}$ is U.I. in n since the upper bound is uniform in n. By Vitali convergence theorem, since $M_{t\wedge\tau_n} \overset{a.s.}{\to} M_t$ $(n\to\infty)$, we have that $M_{t\wedge\tau_n} \overset{L^1}{\to} M_t$ $(n\to\infty)$. Notice that $M_{t\wedge\tau_n}$ is bounded for fixed n, so it's a bounded MG, and $\mathbb{E}M_{t\wedge\tau_n} = 1$, this is telling us that $\forall t > 0, \mathbb{E}M_t = 1$ and it's proved. □

Remark. *Actually, the following **Novikov's condition** is enough to ensure that the exponential local MG is a true MG getting rid of the ε on the exponential*

$$\mathbb{E}e^{\frac{1}{2}\int_0^T \alpha_t^2 \, dt} < \infty \tag{660}$$

refer to Ikeda, Watanabe (1988), section III Theorem 5.3 for details.

Applications of Girsanov Theorem

The first application is to **construct solutions of SDE**. Let B_t be an $\mathscr{F}_t$-BM. In order to absorb the drift coefficient, consider local MG $\int_0^t b(s, B_s)\, dB_s$ and its exponential local MG

$$M_t = e^{\int_0^t b(s,B_s)\, dB_s - \frac{1}{2}\int_0^t b^2(s,B_s)\, ds} \tag{661}$$

assume that we are in the good situation where the Girsanov theorem holds, M_t would be a MG. Consider probability measure $\mathbb{Q}$ with $\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T$, we would know that

$$\tilde{B}_t = B_t - \left\langle B, \int_0^{\cdot} b(s, B_s)\, dB_s \right\rangle_t = B_t - \int_0^t b(s, B_s)\, ds \tag{662}$$

is $\mathscr{F}_t$-BM under $\mathbb{Q}$.

As a result, under probability measure $\mathbb{Q}$, there exists $\mathscr{F}_t$-BM $\tilde{B}_t$ such that $X_t = B_t$ is the weak solution to the SDE

$$dX_t = b(t, X_t)\, dt + d\tilde{B}_t \tag{663}$$

Remark. *In the context above, we are taking it as granted that Girsanov theorem holds. However, this actually requires us to verify the Novikov's condition*

$$\mathbb{E}e^{\frac{1}{2}\int_0^T b^2(s,B_s)\, ds} < \infty \tag{664}$$

one simple condition is that $\forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}, |b(t,x)| \le g(t)$ *where* $g \in L^2([0,T])$*. When* $g \in L^2(\mathbb{R}_+)$*, it's easy to see that Girsanov's theorem applies on time interval* $[0, \infty)$*. An example of this would be that* b *is bounded on* $[0,T] \times \mathbb{R}_+$ *and vanishes on* $(T, \infty) \times \mathbb{R}_+$*.*

By taking the b function as special forms, one can derive the Cameron-Martin formula.

Theorem 28. ***(Cameron-Martin Formula)*** *Consider* $g \in L^2(\mathbb{R}_+)$ *and* $h(t) = \int_0^t g(s)\, ds$ *(all functions* h *that can be written in this form builds up the Cameron-Martin space* $\mathscr{H}$*), then for every non-negative function* $\Phi : C(\mathbb{R}_+, \mathbb{R}) \to \mathbb{R}$,

$$\mathbb{E}_{\mathbb{P}}[M_\infty \cdot \Phi(\{B_t\}_{t\ge 0})] = \mathbb{E}_{\mathbb{P}}[\Phi(\{B_t + h(t)\}_{t\ge 0})] \tag{665}$$

where $C(\mathbb{R}_+, \mathbb{R})$ *is the space consisting of all continuous functions from* $\mathbb{R}_+$ *to* $\mathbb{R}$ *where the sample paths of BM live and*

$$M_t = e^{\int_0^t g(s)\, dB_s - \frac{1}{2}\int_0^t g^2(s)\, ds} \tag{666}$$

$$= e^{\int_0^t h'(s)\, dB_s - \frac{1}{2}\int_0^t [h'(s)]^2\, ds} \tag{667}$$

Proof. Take $b(t,x) = g(t)$ as stated above, so Girsanov theorem applies on time interval $[0,\infty)$ and

$$\tilde{B}_t = B_t - \int_0^t g(s)\,ds \tag{668}$$

is BM under $\mathbb{Q}$, where $\frac{d\mathbb{Q}}{d\mathbb{P}} = M_\infty$. Change the measure to get

$$\mathbb{E}_{\mathbb{P}}[M_\infty \cdot \Phi(\{B_t\}_{t\geq 0})] = \mathbb{E}_{\mathbb{Q}}[\Phi(\{B_t\}_{t\geq 0})] \tag{669}$$

$$= \mathbb{E}_{\mathbb{Q}}[\Phi(\left\{\tilde{B}_t + h(t)\right\}_{t\geq 0})] \tag{670}$$

$$= \mathbb{E}_{\mathbb{P}}[\Phi(\{B_t + h(t)\}_{t\geq 0})] \tag{671}$$

□

Remark. *The cameron-Martin formula shows the quasi-invariance property of the Wiener measure under the translations by functions in the Cameron-Martin space. Notice that each BM trajectory is an element of $C(\mathbb{R}_+,\mathbb{R})$, so the BM can be constructed under the probability measure on such space denoted W such that*

$$\forall A \subset C(\mathbb{R}_+,\mathbb{R}), W(A) = \mathbb{P}\left(B_t(\omega) \in A\right) \tag{672}$$

and the Wiener measure on all cylinder sets can be determined by the finite-dimensional distribution of BM, characterizing the Wiener measure on the whole space.

Rewrite the Cameron-Martin formula in the language of Wiener integral, one would see that

$$\int e^{\int_0^\infty h'(s)\,dw(s) - \frac{1}{2}\int_0^\infty [h'(s)]^2\,ds} \cdot \Phi(w)\,W(dw) = \int \Phi(w+h)\,W(dw) \tag{673}$$

the RHS is translating $\Phi(w)$ by $h \in \mathscr{H}$.

Using these tools, one can now look into **BM with drift**. Let B_t be 1-dimensional BM and T_a be the first hitting time to a of B_t, for given constant c, consider

$$U_a = \inf\{t \geq 0 : B_t + ct = a\} \tag{674}$$

with the drift term ct added to the stopping time.

Let's fix $T > 0$ and apply Cameron-Martin formula for $g(t) = c \cdot \mathbb{I}_{t\leq T}$ and $h(t) = \int_0^t g(s)\,ds = c\cdot(t\wedge T)$. Set the non-negative function $\Phi : C(\mathbb{R}_+,\mathbb{R}) \to \mathbb{R}$ as $\Phi(w) = \mathbb{I}_{\max_{t\in[0,T]} w(t)\geq a}$ to get

$$\mathbb{P}\left(U_a \leq T\right) = \mathbb{E}\Phi(B+h) \tag{675}$$

$$= \mathbb{E}\Phi(B) \cdot e^{\int_0^T g(t)\,dB_t - \frac{1}{2}\int_0^T g^2(t)\,dt} \tag{676}$$

$$= \mathbb{E}\Phi(B) \cdot e^{cB_T - \frac{c^2}{2}T} \tag{677}$$

by forming the probability as the expectation as an indicator and write it as a function of BM sample path under the translation of a function in the Cameron-Martin space. Now plug in the definition of Φ to see

$$\mathbb{P}(U_a \leq T) = \mathbb{E}\left[\mathbb{I}_{\max_{t\in[0,T]} B(t)\geq a} \cdot e^{cB_T - \frac{c^2}{2}T}\right] \tag{678}$$

$$= \mathbb{E}\left[\mathbb{I}_{T_a\leq T} \cdot e^{cB_T - \frac{c^2}{2}T}\right] \tag{679}$$

apply optional stopping theorem for the exponential MG of BM and $T_a \wedge T$ to get $\mathbb{E}(e^{cB_t - \frac{c^2}{2}t}|\mathscr{F}_{t\wedge T_a}) = e^{cB_{t\wedge T_a} - \frac{c^2}{2}(t\wedge T_a)}$

$$\mathbb{P}(U_a \leq T) = \mathbb{E}\left[\mathbb{I}_{T_a\leq T} \cdot e^{cB_{T\wedge T_a} - \frac{c^2}{2}(T\wedge T_a)}\right] \tag{680}$$

$$= \mathbb{E}\left[\mathbb{I}_{T_a\leq T} \cdot e^{ca - \frac{c^2}{2}T_a}\right] \tag{681}$$

now notice that by the reflection principle of BM, we can prove that $T_a \stackrel{d}{=} \frac{a^2}{B_1^2}$ which is following a $\frac{1}{2}$-stable law with density

$$f_{T_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}} \ (t > 0) \tag{682}$$

Combining this conclusion with the calculations above to see that

$$\mathbb{P}(U_a \leq T) = \int_0^T e^{ca - \frac{c^2}{2}t} \cdot \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}} \, dt \tag{683}$$

and the density of U_a is

$$f_{U_a}(t) = \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{(a-ct)^2}{2t}} \ (t > 0) \tag{684}$$

Remark. *One may see that by combining Girsanov theorem / Cameron-Martin formula with classical conclusions, one may prove analogues for a drifted version of the conclusion.*

Optimal Stopping Problem

One of the most well-known situations for the optimal stopping problem is the pricing of American options where one has to decide when to exercise the American option in advance to maximize the profit. The general frame of this problem is formed as: process X_t is the diffusion process as the solution to SDE

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{685}$$

and g is the reward function with time horizon until fixed time $T > 0$ and our objective is to obtain

$$\max_{0\leq\tau\leq T} \mathbb{E}[g(\tau, X_\tau)|X_0 = x] \tag{686}$$

where τ is a stopping time between time 0 and T.

Remark. *In the American option pricing problem, T is the time to maturity, X_t is the stock price at time t, x is the initial stock price and g is the overall profit at time T. We want to find the optimal time τ to early exercise in order to get the maximum profit.*

While the **terminal reward** mentioned above is formed as $g(\tau, X_\tau)$ only relevant with the time and the process at the terminal time τ, one can also consider **running reward**

$$\mathbb{E}\left(\int_0^\tau f(t, X_t)\, dt + g(\tau, X_\tau)\right) \tag{687}$$

that depends on not only the terminal time, but also all the history before the terminal time. However, one can always **transfer running reward into terminal reward by lifting the dimension of the process**

Let's define

$$Y_t = \int_0^t f(u, X_u)\, du \tag{688}$$

$$G(t, x_1, x_2) = x_2 + g(t, x_1) \tag{689}$$

and consider the terminal reward

$$\max_{0\leq\tau\leq T} \mathbb{E}[G(\tau, X_\tau, Y_\tau)] \tag{690}$$

which is equal to the original running reward. The price to pay is that now we have a 2-dimensional process (X_t, Y_t) controlled by the SDE system

$$\begin{cases} dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \\ dY_t = f(t, X_t)\, dt \end{cases} \tag{691}$$

Value Function

For simplicity, let's consider first the discrete-time case $t = 0, 1, ..., T$ and X_n as a **time-homogeneous Markov chain**. Recall that the discrete time optimal stopping problem is the same as pricing American option with binomial tree model. As a result, one considers **backwardly in time**, when it's the time of maturity, one receives some known payoffs, at each time point one compares the value of holding the option (computed using the payoff of the option in the possible future states) with the value of exercising the option immediately (computed using the stock price for the time being) and take the larger one.

Following this idea, when it's already time T, one receives $g(X_T)$ without the necessity of making any choices. At time $T-1$, one compares the value of stopping immediately

$$g(X_{T-1}) \tag{692}$$

and the value of waiting till time T

$$\mathbb{E}[g(X_T)|\mathscr{F}_{T-1}] = \mathbb{E}[g(X_T)|X_{T-1}] \tag{693}$$

since X_n is Markov and take the larger one as the optimal value at time $T-1$

$$\max\{\mathbb{E}[g(X_T)|X_{T-1}], g(X_{T-1})\} \tag{694}$$

by adopting this strategy, one will be able to write down the recurrence relationship of the optimal value process.

Let's define the **optimal value process** V_n as the best possible value one can have at time n, then

$$V_n = \sup_{n\le\tau\le T} \mathbb{E}[g(X_T)|\mathscr{F}_n] \tag{695}$$

is the stopping after time n that maximizes the expected value given all information before time n. For each V_n, there always exists a stopping time $\hat{\tau}_n = \inf\{k \ge n : V_k = g(X_k)\} \wedge T$ that realizes the supreme in the definition.

Let's then consider 3 strategies that gives the recurrence relationship of V_n. The first strategy is to apply $\hat{\tau}_n$, the best stopping time after time n. It's obvious that one would get value V_n at time n based on the definition. The second strategy is to stop immediately at time n. By doing this, one always gets value $g(X_n)$ at time n. The third strategy is not to stop at time n but to wait until time $n+1$ and then behave optimally according to $\hat{\tau}_{n+1}$. By taking the third strategy, one would get value

$$\mathbb{E}[g(X_{\hat{\tau}_{n+1}})|\mathscr{F}_n] = \mathbb{E}(\mathbb{E}[g(X_{\hat{\tau}_{n+1}})|\mathscr{F}_{n+1}]|\mathscr{F}_n) = \mathbb{E}(V_{n+1}|\mathscr{F}_n) \tag{696}$$

at time n because we only have the information until time n but are sticking to the stopping time $\hat{\tau}_{n+1}$.

By the optimality of $\hat{\tau}_n$, one can see that

$$V_n \ge \max\{g(X_n), \mathbb{E}(V_{n+1}|\mathscr{F}_n)\} \ \ a.s. \tag{697}$$

Remark. *Actually, this inequality is an equality because there are only 2 possible options at time n: stop or wait, they make up the best stopping time $\hat{\tau}_n$. If one chooses to stop at time n, he gets the rewards immediately and has nothing to do with the process any longer. If one chooses to wait at time n, he has to wait till time $n+1$, then all possible strategies cannot be strictly better than $\hat{\tau}_{n+1}$. This is the reason why the* ***recurrence relationship of the optimal value process*** *is exactly*

$$\begin{cases} V_n = \max\{g(X_n), \mathbb{E}(V_{n+1}|\mathscr{F}_n)\} \ \ a.s. \\ V_T = g(X_T) \end{cases} \tag{698}$$

and it's easy to see that by taking g to be nice enough, V_n is actually a ***super-MG****.*

In fact, the structure of V_n is more subtle in that it is the tightest super-MG dominating $g(X_n)$ in the stochastic sense. We say that **process** X_n **dominates process** Y_n if

$$\forall n, X_n \geq Y_n \ a.s. \tag{699}$$

and by assuming $\forall n \leq T, \mathbb{E}Y_n < \infty$, the **Snell envelope** S_n of the process Y_n is defined as the smallest super-MG that dominates Y_n. Here the "smallest" refers to the fact that for any super-MG D_n dominating Y_n, it's always the case that D_n dominates S_n.

Theorem 29. ***(Snell Envelope Structure of Optimal Value Process)*** *V_n is the Snell envelope of $g(X_n)$ for nice enough g.*

Proof. We have already proved that V_n is a super-MG and

$$V_n = \max\{g(X_n), \mathbb{E}(V_{n+1}|\mathscr{F}_n)\} \geq g(X_n) \ a.s. \tag{700}$$

so it dominates $g(X_n)$.

If there is another super-MG D_n that dominates $g(X_n)$, then

$$D_T \geq g(X_T) = V_T \tag{701}$$

induct backwardly to find that

$$D_{T-1} \geq \mathbb{E}(D_T|\mathscr{F}_{T-1}) \geq \mathbb{E}(V_T|\mathscr{F}_{T-1}) \tag{702}$$

and that

$$D_{T-1} \geq g(X_{T-1}) \tag{703}$$

so

$$D_{T-1} \geq \max\{g(X_{T-1}), \mathbb{E}(V_T|\mathscr{F}_{T-1})\} = V_{T-1} \tag{704}$$

so this is true by induction for any $0 \leq t \leq T$.

□

Week 10

Although the optimal value process V_n is only ensured to be a sup-MG, the optimally stopped optimal value process $V_t^{\hat{\tau}_n} = V_{t\wedge\hat{\tau}_n}$ is actually a MG.

Theorem 30. *The stopped process $V^{\hat{\tau}_n}$ is a MG on $[n,T]$.*

Proof. WLOG, show the conclusion for $n=0$ and denote $\hat{\tau}_0$ as $\hat{\tau}$. By definition,

$$\begin{aligned}\mathbb{E}(V_{m+1}^{\hat{\tau}}|\mathscr{F}_m) &= \mathbb{E}(V_{m+1}\mathbb{I}_{\hat{\tau}\geq m+1}|\mathscr{F}_m) + \mathbb{E}(V_{\hat{\tau}}\mathbb{I}_{\hat{\tau}\leq m}|\mathscr{F}_m) &(705)\\ &= \mathbb{I}_{\hat{\tau}\geq m+1}\mathbb{E}(V_{m+1}|\mathscr{F}_m) + \mathbb{I}_{\hat{\tau}\leq m}\mathbb{E}(V_{\hat{\tau}}|\mathscr{F}_m) &(706)\\ &= \mathbb{I}_{\hat{\tau}\geq m+1}V_m + \mathbb{I}_{\hat{\tau}\leq m}V_{\hat{\tau}} &(707)\\ &= V_m^{\hat{\tau}} &(708)\end{aligned}$$

the second last equation is due to the fact that when the optimal stopping time happens no earlier than $m+1$, one would never stop at time m, so $V_m = \mathbb{E}(V_{m+1}|\mathscr{F}_m)$. □

Continuous-time Optimal Stopping

Let's then consider the optimal stopping problem is continuous-time setting. X_t is a diffusion process with terminal reward $g(\tau, X_\tau)$ for stopping time $0\leq\tau\leq T$. The **optimal value process** is defined as

$$V_t = \sup_{t\leq\tau\leq T}\mathbb{E}[g(\tau,X_\tau)|\mathscr{F}_t] \tag{709}$$

the largest possible reward achieved and **the optimal stopping time** is defined as

$$\hat{\tau}_t = \inf\{s\geq t: V_s = g(s,X_s)\} \tag{710}$$

the first time such that value function is equal to the terminal reward when the person chooses to stop immediately. Under some mild conditions like $\sup_{0\leq\tau\leq T}|g(\tau,X_\tau)|<\infty$, the V_t is still the Snell envelope of $g(X_t)$ and $V^{\hat{\tau}_t}$ is still a MG on time $[t,T]$ as done in discrete time.

Note that since X_t is a diffusion process as the solution to the SDE

$$dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dB_t \tag{711}$$

it's Markov and one can make sure that

$$v(t,x) = \sup_{t\leq\tau\leq T}\mathbb{E}[g(\tau,X_\tau)|X_t=x] \tag{712}$$

is well-defined since conditioning on $\mathscr{F}_t$ is the same as conditioning on X_t. To figure out the consistency conditions satisfied by such $v(t,x)$, we compare three strategies at time t when $X_t = x$.

The first one is to use $\hat{\tau}_t$ to get the optimal value $v(t,x)$. The second strategy is to stop at time t immediately and receive $g(t,x)$. The third strategy is not to stop at time t but to wait until time $t+h$ and follows $\hat{\tau}_{t+h}$ from then on. This gives reward $\mathbb{E}[g(t+h, X_{\hat{\tau}_{t+h}})|X_t = x]$. By using tower property,

$$\begin{aligned}\mathbb{E}[g(t+h, X_{\hat{\tau}_{t+h}})|X_t = x] &= \mathbb{E}[\mathbb{E}(g(t+h, X_{\hat{\tau}_{t+h}})|\mathscr{F}_{t+h})|X_t = x] &(713)\\ &= \mathbb{E}[v(t+h, X_{t+h})|X_t = x] &(714)\end{aligned}$$

One immediately get that

$$\begin{cases} v(t,x) \geq g(t,x) \\ v(t,x) \geq \mathbb{E}[v(t+h, X_{t+h})|X_t = x] \end{cases} \tag{715}$$

in order to connect $v(t+h, X_{t+h})$ with $v(t, X_t)$, apply Ito formula to see

$$\mathbb{E}[v(t+h, X_{t+h})|X_t = x] = v(t,x) + \mathbb{E}\left[\int_t^{t+h} [(\partial_t + L)v](s, X_s)\, ds + \int_t^{t+h} \sigma(s, X_s) \cdot \partial_x v(s, X_s)\, dB_s \Big| X_t = x\right] \tag{716}$$

assume $\hat{\tau}_t$ exists for each pair (t,x), $v \in C^{1,2}$ smooth enough and all process are integrable enough such that the stochastic integral is MG. (One always needs to turn back and verify that those conditions do hold after solving out the final result!)

By those assumptions, dividing both sides by h and taking $h \to 0^+$, one can simplify the inequalities to get

$$\begin{cases} v(t,x) \geq g(t,x) \\ (\partial_t + L)v(t,x) \leq 0 \end{cases} \tag{717}$$

To be more careful with the details here, note that if it's optimal to stop at time t when $X_t = x$, then $V(t,x) = g(t,x)$ and $(\partial_t + L)v(t,x) < 0$. However, if it's optimal not to stop, then $v(t,x) \geq g(t,x)$, $(\partial_t + L)v(t,x) = 0$. It's like some complementary conditions where exactly one equality is attained. Therefore, let's define **the continuation region** (the region in which one chooses not to stop at time t when $X_t = x$)

$$C = \{(t,x) : v(t,x) > g(t,x)\} \tag{718}$$

and it's now clear that **the optimal stopping problem can be characterized by a set of variational in-**

equalities.

$$\begin{cases} v(T,x) = g(T,x) \\ \forall (t,x), v(t,x) \geq g(t,x) \\ \forall (t,x) \in C^c, v(t,x) = g(t,x) \\ \forall (t,x), (\partial_t + L)v(t,x) \leq 0 \\ \forall (t,x) \in C, (\partial_t + L)v(t,x) = 0 \end{cases} \tag{719}$$

the first one identifies the terminal condition, other two inequalities are the same as we have inferred and the remaining two are complimentary conditions. To write it in **the compact form**,

$$\max\{g - v, (\partial_t + L)v\} = 0 \tag{720}$$

Remark. *The problem here is actually a* ***free boundary problem****. This refers to the fact that when solving this problem the continuation region C is unknown and will be obtained simultaneously as figuring out v.*

Example: American Option Pricing

As we have stated above, the most natural problem setting of the optimal stopping is the American option pricing, where one has to decide whether to exercise early or not. In this model, the stock price at time t which is S_t follows the geometric BM (after applying Girsanov)

$$\frac{1}{S_t}\, dS_t = r\, dt + \sigma\, dB_t^{\mathbb{Q}} \tag{721}$$

where $\mathbb{Q}$ is the risk-neutral measure and $B_t^{\mathbb{Q}}$ is the BM under $\mathbb{Q}$. For the American call option, one's goal is to consider

$$\sup_{0\leq\tau\leq T} \mathbb{E}_{\mathbb{Q}}[e^{r(T-\tau)}(S_\tau - K)_+] \tag{722}$$

and for the American put option, one's goal is to consider

$$\sup_{0\leq\tau\leq T} \mathbb{E}_{\mathbb{Q}}[e^{r(T-\tau)}(K - S_\tau)_+] \tag{723}$$

in order to achieve the maximum profit (future value).

Consider the case for the put option, $g(S_t) = (K - S_t)_+$, the optimal value function is defined as

$$v(t,s) = \sup_{t\leq\tau\leq T} \mathbb{E}_{\mathbb{Q}}[e^{-r\tau}(K - S_\tau)_+ | S_t = s] \cdot e^{rt} \tag{724}$$

discounted as the value at time t.

By previous calculations, the first three variational inequalities that characterize the optimal stopping problem are

$$\begin{cases} v(T,s) = (K-s)_+ \\ \forall (t,s), v(t,s) \geq (K-s)_+ \\ \forall (t,s) \in C^c, v(t,s) = (K-s)_+ \end{cases} \tag{725}$$

note that here the infinitesimal generator is

$$Lv(t,s) = rs \cdot \partial_s v + \frac{\sigma^2 s^2}{2} \partial_{ss} v \tag{726}$$

and since a **discount factor** $e^{r(t-\tau)}$ exists in v, the PDE is a little bit different from what we have seen in the previous non-discounted case. Start from the comparison of two strategies, one to follow the optimal strategy at time t and the other to wait till time $t+h$ and follow the optimal strategy at time $t+h$ to get

$$v(t,s) \geq e^{-rh} \cdot \mathbb{E}[v(t+h, S_{t+h})|S_t = s] \tag{727}$$

apply Ito formula for the RHS to know

$$e^{-rh} \cdot \mathbb{E}[v(t+h, S_{t+h})|S_t = s] \tag{728}$$

$$= e^{-rh} v(t,s) + e^{-rh} \cdot \mathbb{E}\left[\int_t^{t+h} [(\partial_t + L)v](p, S_p)\, dp + \int_t^{t+h} \sigma(p, S_p) \cdot \partial_s v(p, S_p)\, dB_p \Big| S_t = s\right] \tag{729}$$

and assume $\int_t^{t+h} \sigma S_p \cdot \partial_s v(p, X_p)\, dB_p$ is a MG in h to get

$$v(t,s) \geq e^{-rh} \cdot \mathbb{E}\left[\frac{1}{h}\int_t^{t+h} [(\partial_t + L)v](p, S_p)\, dp \Big| S_t = s\right] \tag{730}$$

set $h \to 0^+$ to find

$$r \cdot v(t,s) \geq [(\partial_t + L)v](t,s) \tag{731}$$

plug in the infinitesimal generator to find

$$\partial_t v + rs \cdot \partial_s v + \frac{\sigma^2 s^2}{2} \partial_{ss} v - r \cdot v \leq 0 \tag{732}$$

this is **the PDE for the optimal stopping problem when there is a discount factor**. (Note the new term $r \cdot v$ due to the discount factor!)

As a result, we may get the variational inequalities that characterize the American put option pricing problem

$$\begin{cases} v(T,s) = (K-s)_+ \\ \forall (t,s), v(t,s) \geq (K-s)_+ \\ \forall (t,s) \in C^c, v(t,s) = (K-s)_+ \\ \forall (t,s), \partial_t v + rs \cdot \partial_s v + \frac{\sigma^2 s^2}{2} \partial_{ss} v - r \cdot v \leq 0 \\ \forall (t,s) \in C, \partial_t v + rs \cdot \partial_s v + \frac{\sigma^2 s^2}{2} \partial_{ss} v - r \cdot v = 0 \end{cases} \tag{733}$$

actually in practice we often add another boundary condition to the PDE such that it can be solved out. The condition is formed as

$$\lim_{s \searrow b_t} \partial_s v(t,s) = -1 \tag{734}$$

and is called the **smooth fit condition**, where b_t is the boundary of the continuation region C. By adding such boundary condition, one is able to solve out the solution to the PDE numerically.

Remark. *To understand this boundary condition, notice that when the stock price is high enough, the put is out-of-the-money and when the stock price is low enough, the put is in-the-money. The condition is saying that when the stock price goes close enough to the boundary of continuation region from the out-of-the-money side (the side out of C), the partial derivative of value w.r.t. the stock price is always -1.*

One might notice that $\partial_s v$ is actually the Delta of this option. So this condition is assuming that when early exercising is as profitable as holding the option, the Delta of the put is -1, which is saying that the put is now in-the-money (since it will be exercised immediately).

Example: Perpetual American Put

The only difference between traditional American put and perpetual American put is that the latter one lasts forever without a maturity date T. One still has S_t as the stock price at time t following geometric BM

$$\frac{1}{S_t}\, dS_t = r\, dt + \sigma\, dB_t \tag{735}$$

and the perpetual American put has price

$$\sup_{\tau \geq 0} \mathbb{E}_{\mathbb{Q}}[e^{-r\tau}(K - S_\tau)_+] \tag{736}$$

Note that the reward given for immediate exercising at time t is $g(S_t) = (K - S_t)_+$ and under this Markovian setting the optimal value function under the condition $S_t = s$ is defined as

$$v(t,s) = \sup_{\tau \geq t} \mathbb{E}[e^{-r\tau}(K - S_\tau)_+ | S_t = s] \cdot e^{rt} \tag{737}$$

However, one has to notice the important difference between American put and perpetual American put that perpetual American put has no maturity date. This is telling us that **the value function $v(t,s)$ is actually just a function of s, independent of time t.** To argue this point, notice that S_t is Markov and assuming that the initial stock price is $S_0 = s_0$, then

$$\{S_r\}_{r\geq 0}\,|_{S_0=s_0} \overset{d}{=} \{S_{r+t}\}_{r\geq 0}\,|_{S_t=s_0} \tag{738}$$

that's why we conclude that

$$v(0,s) = \sup_{\tau\geq 0}\mathbb{E}[e^{-r\tau}(K-S_\tau)_+|S_0=s] = \sup_{\tau\geq t}\mathbb{E}[e^{-r(\tau-t)}(K-S_\tau)_+|S_t=s] = v(t,x) \tag{739}$$

so we would denote **the value function** as

$$v(s) = \sup_{\tau\geq 0}\mathbb{E}[e^{-r\tau}(K-S_\tau)_+|S_0=s] \tag{740}$$

only depending on the initial stock price. It's easy to see that the optimal stopping strategy $\hat{\tau}$ also only depends on the stock price one observes (thus easy to obtain).

Consider two strategies, where the first one is to follow the optimal strategy when observing stock price $S_t = s$ to get value $v(s)$. The second one is to exercise immediately when observing stock price $S_t = s$ to get value $(K-s)_+$. The third one is to wait till time $t+h$ for the stock price to change and then follow the optimal strategy to get value $\mathbb{E}[v(S_{t+h})|\mathscr{F}_t]$.

Of course we would see that

$$\begin{cases} v(s) \geq (K-s)_+ \\ v(s) \geq e^{-rh}\cdot\mathbb{E}[v(S_{t+h})|S_t=s] \end{cases} \tag{741}$$

apply Ito formula for $\mathbb{E}[v(S_{t+h}|\mathscr{F}_t)]$ to see

$$e^{-rh}\cdot\mathbb{E}[v(S_{t+h})|S_t=s] = e^{-rh}v(s) + e^{-rh}\cdot\mathbb{E}\left[\int_t^{t+h} v'(p)\,dS_p + \frac{1}{2}\int_t^{t+h} v''(p)\,d\langle S,S\rangle_p\Big|S_t=s\right] \tag{742}$$

$$= e^{-rh}v(s) + e^{-rh}\cdot\mathbb{E}\left[\int_t^{t+h} Lv(p)\,dp + \int_t^{t+h}\sigma S_p\cdot v'(p)\,dB_p\Big|S_t=s\right] \tag{743}$$

assume that $\int_t^{t+h}\sigma S_p\cdot v'(p)\,dB_p$ is a MG in h (verified after solving $v'(p)$ in the following context), multiply $\frac{1}{h}$ on both sides and take $h\to 0^+$ to get

$$Lv(s) - r\cdot v(s) \leq 0 \tag{744}$$

As a result, we may get the variational inequalities that characterize the perpetual American put option

$$\begin{cases} v(\infty) = 0 \\ \forall s, v(s) \geq (K-s)_+ \\ \forall s \in C^c, v(s) = (K-s)_+ \\ \forall s, Lv(s) - r \cdot v(s) \leq 0 \\ \forall s \in C, Lv(s) - r \cdot v(s) = 0 \\ \lim_{s \searrow b} v'(s) = -1 \end{cases} \tag{745}$$

with the last condition as the smooth fit condition to ensure that on the boundary of the continuation region the Delta of the put is -1. Now we know that the continuation region C is defined as the region where $Lv(s) - r \cdot v(s) = 0$. However, this ODE is easy to solve. Write in the explicit form to get

$$rs \cdot v'(s) + \frac{\sigma^2 s^2}{2} v''(s) - r \cdot v(s) = 0 \tag{746}$$

solve it to know

$$v(s) = C_1 \cdot s + C_2 \cdot s^{-\frac{2r}{\sigma^2}} \tag{747}$$

for some constant C_1, C_2. By $v(\infty) = 0$, one immediately finds out that $C_1 = 0$. To apply the other boundary condition, compute

$$v'(s) = -\frac{2rC_2}{\sigma^2} \cdot s^{-\frac{2r}{\sigma^2} - 1} \tag{748}$$

and find out that since $C_2 > 0$, $v'(s) < 0$, **the value is monotone decreasing w.r.t. the stock price**. Since if $b \in \partial C$ then $(K - b)_+ = -\frac{2rC_2}{\sigma^2} \cdot b^{-\frac{2r}{\sigma^2} - 1}$, this is telling us that **the boundary of continuation region consists of only one point denoted** s^*. When $s > s^*$, one always chooses to hold the option and when $s < s^*$ one always chooses to exercise immediately.

Now we finally know that

$$\begin{cases} v'(s^*) = -\frac{2rC_2}{\sigma^2} \cdot (s^*)^{-\frac{2r}{\sigma^2} - 1} = -1 \\ v(s^*) = C_2 \cdot (s^*)^{-\frac{2r}{\sigma^2}} = (K - s^*)_+ = K - s^* \end{cases} \tag{749}$$

one for the smooth fit condition and the other for the fact that two actions are indifferent on the boundary of the continuation region (note that when the stock price is exactly s^* it must be true that $K > s^*$ since $\forall s > 0, v(s) > 0$).

Solve out C_2 and s^* to find

$$\begin{cases} s^* = \frac{2rK}{2r+\sigma^2} \\ C_2 = \frac{\sigma^2}{2r}\left(\frac{2rK}{2r+\sigma^2}\right)^{\frac{2r}{\sigma^2}+1} \end{cases} \tag{750}$$

Remark. *This ODE is the **Cauchy-Euler equation**. To solve it, apply transformation $t = \log s, y = f(t) = f(\log s)$ and solve for f.*

Notice that

$$\frac{dy}{ds} = \frac{df}{dt}\frac{1}{s} \tag{751}$$

$$\frac{d^2y}{ds^2} = \frac{1}{s^2}\left(\frac{d^2f}{dt^2} - \frac{df}{dt}\right) \tag{752}$$

so combine those two equations to get the following

$$\frac{\sigma^2}{2}f'' + \left(r - \frac{\sigma^2}{2}\right)f' - rf = 0 \tag{753}$$

a linear ODE with constant coefficient. Consider its characteristic equation

$$\frac{\sigma^2}{2}\lambda^2 + \left(r - \frac{\sigma^2}{2}\right)\lambda - r = 0 \tag{754}$$

and solve out the roots $\lambda_1 = 1, \lambda_2 = -\frac{2r}{\sigma^2}$ so

$$f(t) = C_1 \cdot e^t + C_2 \cdot e^{-\frac{2r}{\sigma^2}t} \tag{755}$$

and

$$v(s) = C_1 \cdot e^{\log s} + C_2 \cdot e^{-\frac{2r}{\sigma^2}\log s} \tag{756}$$

$$= C_1 \cdot s + C_2 \cdot s^{-\frac{2r}{\sigma^2}} \tag{757}$$

Now we have completely solved the perpetual American put problem. To conclude, **the optimal strategy is to exercise the option immediately if and only if one observes that the stock price is larger than**

$$s^* = \frac{2rK}{2r+\sigma^2} \tag{758}$$

one's **value function** is always

$$v(s) = \frac{\sigma^2}{2r}\left(\frac{2rK}{2r+\sigma^2}\right)^{\frac{2r}{\sigma^2}+1} s^{-\frac{2r}{\sigma^2}} \tag{759}$$

Stochastic Control

Now let's consider the controlled dynamics given by the SDE

$$dX_t^\alpha = b(t, X_t^\alpha, \alpha_t)\,dt + \sigma(t, X_t^\alpha, \alpha_t)\,dB_t \tag{760}$$

where α_t is the **strategy/policy** one takes at time t and the solution X_t^α depends on the strategy one takes at each time. The reward is given by

$$\int_0^T f(t, X_t^\alpha, \alpha_t)\,dt + g(X_T^\alpha) \tag{761}$$

where $\int_0^T f(t, X_t^\alpha, \alpha_t)\,dt$ is the running reward in time interval $[0, T]$ and the terminal reward is $g(X_T^\alpha)$ (this is a finite time horizon problem that ends at time T). Let $\mathscr{A}$ be the **admissible set**, i.e. the set of all possible strategy to choose from (this set is usually equipped with measurability and integrability conditions). The goal in stochastic control problem is to **maximize the expectation of the reward** and to **find out the best strategy** α.

There are two main approaches to solve the problem. The **PDE approach** is applied for the setting where $\mathscr{A}$ is the set of Markovian strategies $\alpha(t, x)$ or even stronger condition like $\alpha_t \in \mathscr{F}_t^X$. The **probabilistic approach** uses FBSDE (2 SDE coupled) and $\mathscr{A}$ is the set of adapted strategy $\alpha_t \in \mathscr{F}_t^B$ where $\mathscr{F}_t^B$ is the filtration of BM.

Let's put emphasis on the PDE approach here. Define the **problem value following strategy α at time t observing $X_t = x$**

$$J(t, x; \alpha) = \mathbb{E}\left[\int_t^T f(s, X_s, \alpha_s)\,ds + g(X_T)\Big| X_t = x\right] \tag{762}$$

and the **value function** as

$$V(t, x) = \sup_\alpha J(t, x; \alpha) \tag{763}$$

the PDE approach derives a PDE for $V(t, x)$ and solves it. After we find out the best strategy $\alpha(t, x)$, remember to verify that it's actually in the admissible set $\mathscr{A}$.

Before setting up a PDE for V, let's explain the action of the strategy. Let $\tau_{t,T}$ denote the set of all stopping time between t and T. If we introduce a stopping time $\theta \in \tau_{t,T}$ such that the game ends and rewards are given at time θ, one would see the following consistency equation satisfies by the problem value.

Theorem 31. ***(Consistency of Problem Value in Stochastic Control)***

$$\forall \theta \in \tau_{t,T}, J(t, x; \alpha) = E\left[\int_t^\theta f(s, X_s, \alpha_s)\,ds + J(\theta, X_\theta; \alpha)\Big| X_t = x\right] \tag{764}$$

Proof. Apply tower property to find

$$J(t,x;\alpha) = \mathbb{E}\left[\int_t^T f(s,X_s,\alpha_s)\,ds + g(X_T)\Big|X_t = x\right] \tag{765}$$

$$= \mathbb{E}\left[\mathbb{E}\left(\int_t^T f(s,X_s,\alpha_s)\,ds + g(X_T)\Big|\mathscr{F}_\theta\right)\Big|X_t = x\right] \tag{766}$$

$$= \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + \mathbb{E}\left(\int_\theta^T f(s,X_s,\alpha_s)\,ds + g(X_T)\Big|\mathscr{F}_\theta\right)\Big|X_t = x\right] \tag{767}$$

$$= \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + J(\theta,X_\theta;\alpha)\Big|X_t = x\right] \tag{768}$$

by the definition of problem value. □

Now after choosing the stopping time θ (according to some criteria), we choose a strategy α that depends on θ to maximize the problem value. We would thus expect to see that **the selection of the strategy has already taken the selection of stopping time into full consideration**, i.e. no matter what criteria based on which the stopping time is chosen, the best strategy α always maximizes the problem value and attains the same value function. This is described by the following dynamic programming principle (DPP).

Theorem 32. ***(Dynamic Programming Principle)***

$$V(t,x) = \sup_\alpha \sup_{\theta\in\tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big|X_t = x\right] \tag{769}$$

$$= \sup_\alpha \inf_{\theta\in\tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big|X_t = x\right] \tag{770}$$

Proof. Since $\forall\alpha\in\mathscr{A}, V(t,x)\geq J(t,x;\alpha)$, one see that

$$\forall\theta\in\tau_{t,T}, J(t,x;\alpha) = E\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + J(\theta,X_\theta;\alpha)\Big|X_t = x\right] \tag{771}$$

$$\leq \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big|X_t = x\right] \tag{772}$$

take inf on both sides w.r.t. the stopping time to see

$$J(t,x;\alpha) \leq \inf_{\theta\in\tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big|X_t = x\right] \tag{773}$$

take sup on both sides w.r.t. the strategy to see

$$V(t,x) \le \sup_{\alpha} \inf_{\theta \in \tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s, \alpha_s)\, ds + V(\theta, X_\theta) \Big| X_t = x\right] \tag{774}$$

On the other hand, by the definition of V,

$$\forall \varepsilon > 0, \theta \in \tau_{t,T}, , \exists \alpha^\varepsilon, V(\theta, X_\theta) - \varepsilon \le J(\theta, X_\theta, \alpha^\varepsilon) \tag{775}$$

where α^ε is called the ε-optimal strategy. Consider the strategy

$$\hat{\alpha}_s = \begin{cases} \alpha_s & s \in [t, \theta] \\ \alpha_s^\varepsilon & s \in [\theta, T] \end{cases} \tag{776}$$

that behaves as α before stopping time θ but behaves optimally after θ. One would see that

$$V(t,x) \ge J(t,x;\hat{\alpha}) \tag{777}$$

$$= \mathbb{E}\left[\int_t^\theta f(s, X_s, \hat{\alpha}_s) + J(\theta, X_\theta, \hat{\alpha}) \Big| X_t = x\right] \tag{778}$$

$$= \mathbb{E}\left[\int_t^\theta f(s, X_s, \alpha_s) + J(\theta, X_\theta, \alpha^\varepsilon) \Big| X_t = x\right] \tag{779}$$

$$\ge \mathbb{E}\left[\int_t^\theta f(s, X_s, \alpha_s) + V(\theta, X_\theta) \Big| X_t = x\right] - \varepsilon \tag{780}$$

now take sup w.r.t. strategy and stopping time on both sides to see that

$$V(t,x) \ge \sup_{\alpha} \sup_{\theta \in \tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s, \alpha_s) + V(\theta, X_\theta) \Big| X_t = x\right] \tag{781}$$

so the theorem is proved. $\square$

Remark. *The DPP is telling us that when one is in a stochastic control problem with a random terminal time $\theta \in \tau_{t,T}$, no matter what criteria the stopping time is constructed upon, one can always find a good enough strategy α such that the optimal value function reaches the same optimal level.*

Let's then consider the **Hamilton-Jacobi-Bellman (HJB) equation** of this control problem. Let's take the stopping time as $\theta = t + h$ with h positive but small enough to know that according to DPP

$$V(t,x) \ge \mathbb{E}\left[\int_t^{t+h} f(s, X_s, \alpha_s) + V(t+h, X_{t+h}) \Big| X_t = x\right] \tag{782}$$

apply Ito formula (assume $V \in C^{1,2}$ is smooth enough, verified after solving out) to get

$$V(t+h, X_{t+h}) = V(t, X_t) + \partial_t V(t, X_t)\, dt + \partial_x V(t, X_t)\, dX_t + \frac{1}{2}\partial_{xx} V(t, X_t)\, d\langle X, X\rangle_t \tag{783}$$

assume $\int_t^{t+h} \partial_x V(s, X_s) \cdot \sigma(s, X_s, \alpha_s)\, dB_s$ is a MG in h (assume and verify after solving), divide both sides by h and set $h \to 0^+$ to get

$$(\partial_t + L^\alpha)V + f \leq 0 \tag{784}$$

note that the infinitesimal generator L^α depends on the strategy α since the drift and diffusion coefficients depend on α. Now note that one can now take the best strategy α^* (assume this exist and verify after solving) such that all inequalities in the context above become equalities and $\partial_t V + L^{\alpha^*} V + f = 0$, combining those two conclusions we get

$$\partial_t V + \sup_\alpha \{L^\alpha V + f\} = 0 \tag{785}$$

which is **the HJB equation** of the control problem. By this step, we have proved that the value function V has to satisfy the HJB equation. Actually, one can also prove that if HJB equation holds and has solution $W \in C^{1,2}$, then $W = V$ must be the value function. So **the HJB equation characterizes the value function of the stochastic control problem**.

Example: Linear Stochastic Regulator Problem

Consider the evolution of the state X_t is given by

$$dX_t = (H_t X_t + M_t \alpha_t)\, dt + \sigma_t\, dB_t \tag{786}$$

where $\alpha(t, X_t) \in \mathbb{R}^k$ is the strategy we are applying, $X_t \in \mathbb{R}^n$ and B_t is m-dimensional BM. The problem loss (one wants to minimize) is given by

$$J(t, x; \alpha) = \mathbb{E}\left[\int_t^T (X_s^T C_s X_s + \alpha_s^T D_s \alpha_s)\, ds + X_T^T R X_T \middle| X_t = x\right] \ (t \leq T) \tag{787}$$

To clarify, here $H_t \in \mathbb{R}^{n\times n}, M_t \in \mathbb{R}^{n\times k}, \sigma_t \in \mathbb{R}^{n\times m}, C_t \in \mathbb{R}^{n\times n}, D_t \in \mathbb{R}^{k\times k}, R \in \mathbb{R}^{n\times n}$ are continuous in t and deterministic with $f(t, X_t, \alpha_t) = X_t^T C_t X_t + \alpha_t^T D_t \alpha_t, g(X_T) = X_T^T R X_T$. Here T is a fixed time and C_t, R, D_t are symmetric, positive-definite at every time t.

The goal of this control problem is to find the best strategy $\alpha = \alpha(t, X_t) \in \mathbb{R}^k$ to minimize the loss $J(t, x; \alpha)$. Since the goal in now minimization, we naturally define the value function to be

$$V(t, x) = \inf_\alpha J(t, x; \alpha) \tag{788}$$

and let's derive the HJB equation again from scratch. The DPP now looks like

$$V(t,x) = \inf_{\alpha} \sup_{\theta\in\tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big| X_t = x\right] \tag{789}$$

$$= \inf_{\alpha} \inf_{\theta\in\tau_{t,T}} \mathbb{E}\left[\int_t^\theta f(s,X_s,\alpha_s)\,ds + V(\theta,X_\theta)\Big| X_t = x\right] \tag{790}$$

and by setting $\theta = t+h$ for positive and small enough h, the HJB equation shall look like

$$(\partial_t + L^\alpha)V + f \geq 0 \tag{791}$$

by taking the optimal strategy α^* one would know that

$$\partial_t V + \inf_\alpha \{L^\alpha V + f\} = 0 \tag{792}$$

Plug in the setting of this problem, the infinitesimal generator is

$$L^\alpha V = (H_t x + M_t\alpha)\cdot\nabla_x V + \frac{1}{2}Tr(\sigma_t\sigma_t^T H) \tag{793}$$

where H is the Hessian of V, so **the HJB equation** is

$$\partial_t V + \inf_\alpha \left\{ x^T C_t x + \alpha^T D_t \alpha + \sum_{i=1}^n (H_t x + M_t\alpha)_i \partial_{x_i} V + \frac{1}{2}\sum_{i,j=1}^n (\sigma_t\sigma_t^T)_{ij}\partial_{x_i,x_j} V \right\} = 0 \ (0\leq t\leq T) \tag{794}$$

with boundary value condition $V(T,x) = x^T R x$. One could then apply **the verification theorem** to find the solution.

Example: Merton's Problem

Now assume the stock price follows geometric BM

$$\frac{1}{S_t}\,dS_t = \mu\,dt + \sigma\,dB_t \tag{795}$$

and α_t denotes the dollar amount invested in S_t at time t (risky asset) and all remaining balance invested into riskless asset earning interest at rate r (continuous-time). Assume the strategy is self-financing (one needs no exterior endowment to change the position in risky and riskless assets), consider the total wealth X_t at time t.

It's immediate that from time t to $t+h$, assume that one is still putting α_t in stock and $X_t - \alpha_t$ in riskless asset

$$X_{t+h} = \alpha_t\frac{S_{t+h}}{S_t} + (X_t - \alpha_t)\cdot e^{rh} \tag{796}$$

divide by h on both sides and set $h \to 0^+$ to know

$$dX_t = \alpha_t \left(\frac{dS_t}{S_t} - r\,dt \right) + rX_t\,dt \tag{797}$$

$$dX_t = \alpha_t(\mu\,dt + \sigma\,dB_t) + r(X_t - \alpha_t)\,dt \tag{798}$$

and we get **the SDE of the wealth process**.

Now define the problem value and value function as

$$J(t,x;\alpha) = \mathbb{E}\left[U(X_T)\middle|X_t = x\right] \quad (t \le T) \tag{799}$$

$$V(t,x) = \sup_\alpha J(t,x;\alpha) \tag{800}$$

note that the game end at time T and our goal is to make the utility as high as possible at the end of this game with the utility function U has variable as the amount of money we have. Plug into the general HJB equation we derive above to see the HJB equation for this problem is

$$\partial_t V + \sup_\alpha \{L^\alpha V\} = 0 \tag{801}$$

with infinitesimal generator

$$L^\alpha V(t,x) = [(\mu - r)\alpha + rx]\partial_x V + \frac{\sigma^2\alpha^2}{2}\partial_{xx} V \tag{802}$$

so let's write

$$\partial_t V + \sup_\alpha \left\{ [(\mu - r)\alpha + rx]\partial_x V + \frac{\sigma^2\alpha^2}{2}\partial_{xx} V \right\} = 0 \tag{803}$$

which is **the HJB equation with boundary value** $V(T,x) = U(x)$.

Merton's Problem with No Interest Rate

For the simplicity of calculations, let's now consider the case where $r = 0$ with power utility $U(x) = \frac{x^\gamma}{\gamma}$ $(0 < \gamma < 1)$. **The HJB equation** now becomes

$$\begin{cases} \partial_t V + \sup_\alpha \left\{ \mu\alpha\partial_x V + \frac{\sigma^2\alpha^2}{2}\partial_{xx} V \right\} = 0 \\ V(T,x) = \frac{x^\gamma}{\gamma} \end{cases} \tag{804}$$

To solve the PDE, let's first consider removing the supreme in the PDE. This can be done by taking the derivative of $\mu\alpha\partial_x V + \frac{\sigma^2\alpha^2}{2}\partial_{xx}V$ w.r.t. α to solve for α^*

$$\mu\partial_x V + \sigma^2\alpha^*\partial_{xx}V = 0 \tag{805}$$

$$\alpha^* = -\frac{\mu\partial_x V}{\sigma^2\partial_{xx}V} \tag{806}$$

now plug it back into the PDE to get rid of the supreme

$$\partial_t V - \frac{\mu^2(\partial_x V)^2}{2\sigma^2\partial_{xx}V} = 0 \tag{807}$$

to solve this PDE, one can find an ansatz

$$V(t,x) = f(t)\frac{x^\gamma}{\gamma} \tag{808}$$

such that $f(T) = 1$ to satisfy the boundary condition that $V(T,x) = U(x)$, then plug in the PDE above to see

$$f'(t) = \frac{\gamma\mu^2}{2\sigma^2(\gamma-1)}f(t) \tag{809}$$

so combining with the terminal value condition $f(T) = 1$ we know

$$f(t) = e^{\frac{\gamma\mu^2}{2\sigma^2(1-\gamma)}(T-t)} \tag{810}$$

and this solves the whole Merton's problem.

To conclude, we know that **the solution to Merton's problem with $r = 0$ and power utility** is that **the value function** is

$$V(t,x) = e^{\frac{\gamma\mu^2}{2\sigma^2(1-\gamma)}(T-t)} \cdot \frac{x^\gamma}{\gamma}\ (t \leq T) \tag{811}$$

and **the optimal strategy** is

$$\alpha^* = -\frac{\mu\partial_x V}{\sigma^2\partial_{xx}V} = \frac{\mu}{\sigma^2(1-\gamma)}x \tag{812}$$

in other words, always put $\frac{\mu}{\sigma^2(1-\gamma)}x$ in risky asset when having total wealth x (a fixed proportion).

Remark. *When solving stochastic control problem, always solve out α^* and V at the same time since they are both unknown at the beginning. For HJB equation, when the function in the supreme is a good enough function of α, one can always mimic what we have done above and turn the problem into solving an easier PDE. (find ansatz to guess the solution)*