CONSTRUCTION OF DIFFUSION PROCESSES II

GUOHUAN ZHAO

1. MOTIVATION

Kolmogorov's approach to constructing diffusion processes is purely analytical. A natural question that arises is whether it is possible to provide a "microscopic" construction at the level of trajectories. This was one of the original motivations for Itô's introduction of stochastic integrals and stochastic differential equations.

Assume $d = 1$. Intuitively, the diffusion process can be constructed as follows: Let Δt be a fixed, small time interval, and consider the following approximation process:

$$X_t = X_0 + b(X_0)t + \sqrt{2a(X_0)}(W_t - W_0), \qquad\qquad t \in [0, \Delta t);$$

$$X_t = X_{\Delta t} + b(X_{\Delta t})(t - \Delta t) + \sqrt{2a(X_{\Delta t})}(W_t - W_{\Delta t}), \qquad\qquad t \in [\Delta t, 2\Delta t);$$

$$\vdots$$

$$X_t = X_{k\Delta t} + b(X_{k\Delta t})(t - k\Delta t) + \sqrt{2a(X_{k\Delta t})}(W_t - W_{k\Delta t}), \qquad t \in [k\Delta t, (k+1)\Delta t).$$

If, as $\Delta t \to 0$, the process X_t (which depends on Δt) converges (in some sense) to a stochastic process, then formally, the limiting process (still denoted as X_t) satisfies:

$$X_t = X_0 + \int_0^t b(X_t)\mathrm{d}t + \underbrace{\int_0^t \sqrt{2a(X_t)}\mathrm{d}W_t}_{?}. \tag{1.1}$$

One issue is how to understand the last term in the above equation.

Exercise 1.1. *Prove that*

$$\mathbf{P}\left(\limsup_{t \to 0} \frac{|W_t|}{t^{1/2}} = \infty\right) = 1.$$

When h_t is β-Hölder continuous and x_t is α-Hölder continuous, if $\alpha + \beta > 1$, then we can prove that the following Riemann sum converges:

$$\sum_{k=0}^{n-1} h_k(x_{k+1} - x_k).$$

However, when $\alpha + \beta < 1$, in general, we cannot mathematically prove the convergence of the above Riemann sum.

Note that the paths of Brownian motion are only α-Hölder continuous for $\alpha < 1/2$. It is expected that the Hölder exponent of the paths of X will not exceed $1/2$ either. Therefore, unless we uncover more information about the paths of Brownian motion, we cannot define the second term on the right-hand side of (1.1) pathwise.

1.5 weeks.

Itô considered a more general problem: Suppose H is an adapted process, how can we define the following integral

$$\int_0^1 H_s \mathrm{d}W_s.$$

Assume $F_t = \sigma(W_s : s \in [0,t])$, and H is a bounded simple process with respect to F_t, i.e., $H_t = H_0 I_{\{0\}}(t) + \sum_{i=0}^{n-1} H_{t_i} I_{(t_i,t_{i+1}]}(t)$, where $0 = t_0 < t_1 < \cdots < t_n = 1$, and $H_{t_i} \in F_{t_i}$. Naturally, we can define $\int_0^t H_s \mathrm{d}W_s = \sum_{i=0}^{n-1} H_{t_i}(W_{t \wedge t_{i+1}} - W_{t \wedge t_i})$. Itô observed that $t \mapsto \int_0^t H_s \mathrm{d}W_s$ is a martingale and satisfies the isometry property

$$\mathbf{E}\left(\int_0^t H_s \mathrm{d}W_s\right)^2 = \mathbf{E}\int_0^t H_s^2 \mathrm{d}s.$$

If H is a general adapted process, and there exists a sequence of simple processes $(H^n)_{n \in \mathbb{N}}$ such that

$$\mathbf{E}\left(\int_0^1 (H_s^n - H_s)^2 \mathrm{d}s\right) \to 0, \quad n \to \infty,$$

then by Doob's inequality,

$$\mathbf{E}\left(\sup_{t \in [0,1]} \left|\int_0^t H_s^n \mathrm{d}W_s - \int_0^t H_s^m \mathrm{d}W_s\right|^2\right) \leqslant C\mathbf{E}\int_0^1 (H_s^n - H_s^m)^2 \mathrm{d}s \to 0.$$

Thus, $\int_0^t H_s^n \mathrm{d}W_s$ converges to a continuous martingale, which we define as $\int_0^t H_s \mathrm{d}W_s$. In fact, later we can argue that we can define the integral of a very general adapted process H with respect to Brownian motion W. Once the integral with respect to Brownian motion is defined, under very general conditions, we can solve equation (1.1), providing a probabilistic construction of diffusion processes.

Essentially, Itô utilized the adaptability of the integrand and the martingale property of Brownian motion. Thus, stochastic analysis injected new vitality into the development of martingale theory. Kunita and Watanabe extended the theory of stochastic integrals from the case of Brownian motion to general square-integrable martingales using the Doob-Meyer decomposition. The Strasbourg school in France further generalized it to the most general case of semimartingales and established a general theory of stochastic processes.

Basic results in stochastic analysis will be used in this note are presented in the Appendix.

2. Stochastic Differential Equations

One of the main object in this note is the following SDE:

$$\mathrm{d}X_t^i = \sigma_k^i(X_t)\mathrm{d}W_t^k + b^i(X_t)\mathrm{d}t, \quad X_0 = \xi \in \mathcal{F}_0. \tag{2.1}$$

Given $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t \geqslant 0}, W_t)$, we say (2.1) has a **pathwise solution** if there exists a continuous $\mathcal{F}_t$-adapted process X_t satisfying (2.1). We say that we have **pathwise uniqueness** for (2.1) if whenever X_t and Y_t are two solutions, then there exists a set $\mathcal{N}$ such that $\mathbf{P}(\mathcal{N}) = 0$ and for all $\omega \notin \mathcal{N}$, we have $(X_t(\omega))_{t \geqslant 0} = (Y_t(\omega))_{t \geqslant 0}$.

2.1. **Lipschitz conditions.**

Theorem 2.1 (Itô). *Suppose σ and b are Lipschitz. Then there exists a unique pathwise solution to the SDE* (2.1) *for any $X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbf{P})$.*

Proof. Let $\mathcal{B}$ denote the set of all continuous processes ξ that are adapted to the filtration $\mathcal{F}_t$ and satisfy

$$\|\xi\|_{\mathcal{B}} := \left(\mathbf{E} \sup_{t \in [0,T]} |\xi_t|^2 \right)^{1/2} < \infty.$$

Here T is a positive number which will be determined later. It is not hard to verify that $\mathcal{B}$ is a Banach space. Define a map $\mathscr{A}$ on $\mathcal{B}$ by

$$(\mathscr{A}(\xi))_t := X_0 + \int_0^t \sigma(\xi_s) \cdot \mathrm{d}W_s + \int_0^t b(\xi_s)\mathrm{d}s, \quad t \in [0,T].$$

By (B.2) (or Doob's inequality) and Lipschitz condition on the coefficients,

$$\|\mathscr{A}(\xi) - \mathscr{A}(\eta)\|_{\mathcal{B}}^2 = \mathbf{E} \sup_{t \in [0,T]} |\mathscr{A}(\xi)_t - \mathscr{A}(\eta)_t|^2$$

$$\leqslant 2\mathbf{E} \sup_{t \in [0,T]} \left| \int_0^t (\sigma(\xi_s) - \sigma(\eta_s))\mathrm{d}W_s \right|^2 + 2\mathbf{E} \sup_{t \in [0,T]} \left| \int_0^t (b(\xi_s) - b(\eta_s))\mathrm{d}s \right|^2$$

$$\overset{\text{(B.2)}}{\leqslant} C\mathbf{E} \int_0^T |\sigma(\xi_s) - \sigma(\eta_s)|^2 \mathrm{d}s + C\mathbf{E} \left(\int_0^T |b(\xi_s) - b(\eta_s)|\mathrm{d}s \right)^2$$

$$\leqslant C(T + T^2)\mathbf{E} \sup_{t \in [0,T]} |\xi_t - \eta_t|^2 = C_1(T + T^2)\|\xi - \eta\|_B^2.$$

Choosing $T > 0$ sufficiently small such that $C_1(T + T^2) \leqslant 1/2$, then $\mathscr{A}$ is a Contraction mapping on $\mathcal{B}$. Banach fixed-point theorem yields that $\mathscr{A}$ has a unique fixed point, which is the unique pathwise solution to (2.1). We can extend the same result to arbitrarily time intervals. $\qquad\square$

2.2. **Definitions of solutions.**

(1) **strong solution exists** to (2.1): if given the Brownian motion W_t there exists a process X_t satisfying (2.1) such that X_t is adapted to the filtration generated by W_t.

(2) **weak solution exists** to (2.1): if there exists $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t\geqslant 0}; X_t, W_t)$ such that W_t is a $(\mathcal{F}_t)_{t\geqslant 0}$-Brownian motion and the equation (2.1) holds.

(3) **weak uniqueness**: if whenever $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t\geqslant 0}; X_t, W_t)$ and $(X, \mathcal{G}, \mathbf{Q}, (\mathcal{G})_{t\geqslant 0}; Y_t, B_t)$ are two weak solutions, then the laws of the processes X and Y are equal; **Joint uniqueness in law** means the joint law of (X, W) and (Y, B) are equal.

A fundamental result is

Theorem 2.2 (Yamada-Watanabe-Engelbert [Eng91]). *The following two conditions are equivalent.*

(i) *For every initial distribution, there exists a weak strong solution to* (2.1) *and the solution to* (2.1) *is pathwise unique.*

(ii) *For every initial distribution, there exists a strong strong solution to* (2.1) *and the solution to* (2.1) *is jointly unique in law.*

If one (and therefore both) of these conditions is satisfied then every solution to (2.1) *is a strong solution.*

2.3. SDEs with Hölder drifts.

For strong well-posedness, if the diffusion coefficient σ is non-degenerate, then the condition on b can be weakened.

Let $\delta \in (0,1)$. Define

$$\mathbb{S}_\delta^d = \left\{ A \in S(d) : \delta I_d \leqslant A \leqslant \delta^{-1} I_d \right\}$$

Theorem 2.3 (Krylov [Kry21]). *Suppose that $a \in \mathbb{S}_\delta^d$ and $\nabla\sigma, b \in L^d(\mathbb{R}^d)$, then equation* (2.1) *admits a unique strong solution.*

Of course, we will not to prove such a strong result here, but a simper one below.

Theorem 2.4 (Flandoli-Gubinelli-Priola [FGP10]). *Equation* (2.1) *admits a unique strong solution, provided that $a \in \mathbb{S}_\delta^d$, σ is Lipschitz, and $b \in C^\alpha(\mathbb{R}^d)$ ($\forall \alpha > 0$).*

Let

$$Lu = a_{ij}\partial_{ij}u + b_i\partial_i u.$$

We will consider

$$\lambda u - Lu = f, \quad \lambda > 0. \tag{2.2}$$

We need the following apriori estimate.

Lemma 2.5. *Suppose $a \in \mathbb{S}_\delta^d$ and $a, b \in C^\alpha$. Then for any $\lambda > 0$ and $u \in C^{2,\alpha}$, it holds that*

$$\lambda\|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C\|\lambda u - Lu\|_\alpha, \tag{2.3}$$

where C only depends on d, δ, α, and $\|a\|_\alpha$ and $\|b\|_\alpha$.

Proof. The proof for the above lemma for $L = \Delta$ can be founded in Appendix C.

For general elliptic operators, let $\zeta \in C_c^\infty(B_2)$ such that $\zeta \geqslant 0$, $\zeta \equiv 1$ in B_1. Set $\zeta_\varepsilon^z = \zeta((x-z)/\varepsilon)$, and $f = \lambda u - Lu$. Then

$$\lambda(u\zeta_\varepsilon^z) - a_{ij}(z)\partial_{ij}(u\zeta_\varepsilon^z)$$
$$= (a_{ij} - a_{ij}(z))\partial_{ij}(u\zeta_\varepsilon^z) - 2a_{ij}\partial_i u\partial_j\zeta_\varepsilon^z - a_{ij}\partial_{ij}\zeta_\varepsilon^z u$$
$$+ b_i\partial_i u\,\zeta_\varepsilon^z + f\zeta_\varepsilon^z,$$

In virtue of Lemma C.2, we have

$$\lambda[u\zeta_\varepsilon^z]_\alpha + [\nabla^2(u\zeta_\varepsilon^z)]_\alpha \leqslant C\varepsilon[\nabla^2(u\zeta_\varepsilon^z)]_\alpha + C[f]_\alpha + C\varepsilon^{-\alpha}\|f\|_0$$
$$+ C\|\nabla^2 u\|_0 + C\varepsilon^{-1}[\nabla u]_\alpha + C\varepsilon^{-1-\alpha}\|\nabla u\|_0 + C\varepsilon^{-2}[u]_\alpha + C\varepsilon^{-2-\varepsilon}\|u\|_0.$$

Here we use the fact that

$$[fg]_\alpha \leqslant \|f\|_0[g]_\alpha + [f]_\alpha\|g\|_0.$$

Choosing $\varepsilon_0 > 0$ sufficiently small so that $C\varepsilon_0 \leqslant 1/2$, we get

$$\lambda[u]_\alpha + [\nabla^2 u]_\alpha \leqslant \sup_{z \in \mathbb{R}^d} \left(\lambda[u\zeta_{\varepsilon_0}^z]_\alpha + [\nabla^2(u\zeta_{\varepsilon_0}^z)]_\alpha\right) \leqslant C_{\varepsilon_0}(\|f\|_\alpha + \|u\|_2).$$

Noting that $\|u\|_2 \leqslant \delta[\nabla^2 u]_\alpha + C_\delta\|u\|_0$, $\delta > 0$, we obtain that

$$\lambda[u]_\alpha + [\nabla^2 u]_\alpha \leqslant C(\|f\|_\alpha + \|u\|_0).$$

Since $\lambda\|u\|_0 \leqslant \|f\|_0$, by interpolation, one sees that

$$\lambda\|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C(1 + \lambda^{-1})\|f\|_\alpha.$$

So we obtain our desired assertion. $\qquad\square$

Theorem 2.6. *Suppose $a \in \mathbb{S}_\delta^d$ and $a, b \in C^\alpha$. Then for any $\lambda > 0$ and $f \in C^\alpha$, equation (2.2) admits a unique solution in $C^{2,\alpha}$. Moreover,*

$$\lambda \|u\|_\alpha + \|u\|_{2+\alpha} \leqslant C(1 + \lambda^{-1})\|f\|_\alpha, \tag{2.4}$$

where C only depends on d, δ, α, and $\|a\|_\alpha$ and $\|b\|_\alpha$.

Sketch of the proof for Lemma Theorem 2.6:

(i) If $L = \Delta$ and $f \in \mathscr{S}(\mathbb{R}^d)$, then for each $\lambda > 0$, one can use Fourier transformation to solve (2.2), i.e. $u = \mathcal{F}^{-1}\left[\mathcal{F}(f) \cdot (\lambda + 4\pi^2|\cdot|^2)\right] \in \cap_{s>0} H^s \subseteq C_b^\infty$. Moreover, (2.4) can also be proved by Fourier analysis method (see Appendix C);

(ii) For any L satisfying the conditions in Theorem 2.6, and any $u \in C^{2,\alpha}$, by Lemma 2.5, (2.4) holds true for any $\lambda > 0$;

(iii) Let χ be a cutoff function and ζ be a mollifier. For any $f \in C^\alpha$, we set $f_\varepsilon = \chi_\varepsilon(f * \zeta_\varepsilon)$. Here $\chi_\varepsilon(x) = \chi(x/\varepsilon)$ and $\zeta_\varepsilon(x) = \varepsilon^{-d}\zeta(x/\varepsilon)$. Using (i), for each $\varepsilon > 0$, there is a smooth soluiton, say u_ε, to (2.2) with L and f replaced by Δ and f_ε. The limit of (u_ε), u, satisfies $\lambda u - \Delta u = f$, and u also satisfies (2.4);

(iv) In the light of (2.4) and the method of continuity (see Lemma 2.7 below), one can obtain the solvability of (2.2) in $C^{2,\alpha}$.

Lemma 2.7 (Method of continuity). *Let B be a Banach space, V a normed vector space, and T_t a norm continuous family of bounded linear operators from B into V. Assume that there exists a positive constant C such that for every $t \in [0,1]$ and every $x \in B$,*

$$\|x\|_B \leqslant C\|T_t x\|_V.$$

Then T_0 is a surjective if and only if T_1 is surjective as well.

Proof of Theorem 2.4. Since σ and b are bounded continuous, weak solution exists to (2.1) (see [Hua01]). Thanks to Theorem 2.2, we only need to prove the pathwise uniqueness.

Let $\lambda \gg 1$ Consider the following equation

$$\lambda \mathbf{u}_\lambda - L\mathbf{u}_\lambda = b.$$

By Lemma 2.5 and interpolation theorem

$$\|\nabla \mathbf{u}\|_\alpha \leqslant \|\mathbf{u}\|_\alpha^{\frac{1}{2}}\|\nabla^2 \mathbf{u}\|_\alpha^{\frac{1}{2}} \leqslant C\lambda^{-\frac{1}{2}}\|b\|_\alpha.$$

Choosing λ sufficiently large so that $C\lambda^{-\frac{1}{2}} < 1/2$. Set $\phi(x) = x + \mathbf{u}(x)$, then $\phi : \mathbb{R}^d \to \mathbb{R}^d$ is a $C^{1,\alpha}$-homeomorphism.

Assume that X and X' are two solutions to (2.1). Set $Y_t = \phi(X_t)$ and $Y_t' = \phi(X_t')$. Then by Itô's formula,

$$\mathrm{d}Y_t^i = (\delta_j^i + \partial_j \mathbf{u}^i)(X_t)\sigma_{jk}(X_t)\mathrm{d}W_t^k + \left[a_{jk}(X_t)\partial_{jk}\mathbf{u}^i(X_t) + (\delta_j^i + \partial_j\mathbf{u}^i)(X_t)b^j(X_t)\right]\mathrm{d}t$$

i.e.

$$\begin{aligned}
\mathrm{d}Y_t =&[(I + \nabla\mathbf{u})\sigma] \circ \phi^{-1}(Y_t)\mathrm{d}W_t + \left[a : \nabla^2\mathbf{u} + (I + \nabla\mathbf{u})b\right] \circ \phi^{-1}(Y_t)\mathrm{d}t \\
=&\underbrace{[(I + \nabla\mathbf{u})\sigma] \circ \phi^{-1}}_{=:\widetilde{\sigma}}(Y_t)\mathrm{d}W_t + \underbrace{\lambda u \circ \phi^{-1}}_{=:\widetilde{b}}(Y_t)\mathrm{d}t.
\end{aligned}$$

Similarly, $\mathrm{d}Y_t' = \widetilde{\sigma}(Y_t')\mathrm{d}W_t + \widetilde{b}(Y_t')\mathrm{d}t$. Since $\widetilde{\sigma}$ and $\widetilde{b}$ are both $C^{1,\alpha}$ functions, as in the proof for Theorem 2.1, we have

$$\mathbf{E}|Y_t - Y_t'|^2 \leqslant C \int_0^t \mathbf{E}|Y_s - Y_s'|^2 \mathrm{d}s.$$

This yields $Y_t = Y_t'$, due to Gronwall's inequality. Since ϕ is one-to-one, $X_t = X_t'$. $\qquad\square$

2.4. **Stochastic Flow.** Consider (2.1) .

Theorem 2.8. *If σ and b are Lipschitz, then there exist versions of $X_t(x)$ that are jointly continuous in t and x a.s.*

Proof.

$$X_t(x) - X_t(y) = x - y + \int_0^t [\sigma(X_s(x)) - \sigma(X_s(y))]\,\mathrm{d}W_s + \int_0^t [b(X_s(x)) - b(X_s(y))]\,\mathrm{d}s$$

By the Burkholder-Davis-Gundy inequalities, for any $t \in [0,1]$,

$$\mathbf{E} \sup_{s \in [0,t]} \left| \int_0^s [\sigma(X_r(x)) - \sigma(X_r(y))]\,\mathrm{d}W_r \right|^p$$

$$\leqslant C\mathbf{E} \left(\int_0^t |X_s(x) - X_s(y)|^2 \mathrm{d}s \right)^{p/2}$$

$$\leqslant C\mathbf{E} \int_0^t |X_s(x) - X_s(y)|^p \mathrm{d}s.$$

Set $g(t) = \mathbf{E}\sup_{s\in[0,t]} |X_s(x) - X_s(y)|^p$. Then we have

$$g(t) \leqslant C|x - y|^p + \int_0^t g(s)\mathrm{d}s, \quad t \in [0,1].$$

Gronwall's inequality yields

$$\mathbf{E} \sup_{t\in[0,1]} |X_t(x) - X_t(y)|^p \leqslant C|x - y|^p, \quad \forall p \geqslant 2.$$

Further, one can verify that

$$\mathbf{E}\,|X_t(x) - X_s(y)|^p \leqslant C \left(|x - y| + |t - s|^{\frac{1}{2}} \right)^p, \quad x, y \in \mathbb{R}^d,\ t, s \in [0,1],\ p \geqslant 2.$$

This together with Kolmogorov's continuity theorem implies that there is a continuous version of $(t, x) \mapsto X_t(x)$ such that

$$\|X(\omega)\|_{C^\alpha([0,1];\dot{C}^\beta(B_R))} \leqslant K(\omega)$$

with $\alpha \in (0, 1/2)$ and $\beta \in (0, 1)$, and $K \in L^p$ for all $p \geqslant 1$. $\qquad\square$

Remark 2.1. *The above result also holds if σ and b are ω-dependent and $\|\sigma\|_{C^1} + \|b\|_{C^1} \leqslant L$ a.s., for some constant L.*

The collection of processes $X_t(x)$ is called a flow. If σ and b are smoother functions, then $X_t(x)$ will be smoother in x. Taking derivative, and using the chain rule, formally we have

$$\partial_j X_t^i(x) = \delta_j^i + \int_0^t \partial_l \sigma_k^i(X_s(x))\partial_j X_s^l(x)\mathrm{d}W_s^k$$
$$+ \int_0^t \partial_l b^i(X_s(x))\partial_j X_s^l(x)\mathrm{d}s.$$

Suppose that σ and b are in C^2 and are bounded with bounded first and second derivatives, we consider the SDE

$$\mathrm{d}J_t(x) = \partial_l\sigma(X_t(x))J_t^l(x)\mathrm{d}W_t + \nabla b(X_t(x))J_t(x)\mathrm{d}t, \quad Y_0 = I \tag{2.5}$$

Follow the proof of Theorem 2.8, we have

Proposition 2.9. *Assume $\sigma, b \in C_b^2$. A pathwise solution to (2.5) exists and is unique. The solution has moments of all orders. If $J_t(x)$ denotes the solution, versions of $J_t(x)$ exist that are jointly Hölder continuous in t and x, and*

$$\mathbf{E} \sup_{t \in [0,1]} |J_t(x) - J_t(y)|^p \leqslant C|x-y|^p, \quad x, y \in \mathbb{R}^d, \ p \geqslant 1.$$

Exercise 2.1. *Prove Proposition 2.9.*

We now prove the differentiability of $X_t(x)$.

Theorem 2.10. *Suppose $\sigma, b \in C_b^k$. Then $x \mapsto X_t(x)$ is $C^{k-1,\alpha}$ a.s., and $\nabla X_t(x) = J_t(x)$.*

Proof. For simplicity we take $b = 0$ and $k = 2$. Set

$$D_j^h X_t^i(x) := \frac{1}{|h|}\left[X_t^i(x+e_jh) - X_t^i(x)\right], \quad h \in (-1, 1)$$

and

$$Z_t(x, h) := D^h X_t(x) - J_t(x).$$

Noting that

$$D^h X_t(x) = I + \int_0^t \underbrace{\left[\int_0^1 \partial_l\sigma(\tau X_s(x+h) + (1-\tau)X_s(x))\mathrm{d}\tau\right]}_{\leqslant \|\nabla\sigma\|_0} D^h X_s^l(x)\mathrm{d}W_s,$$

as we done in the proof of Theorem 2.8, it is not hard to prove

$$\sup_{x \in \mathbb{R}^d; h \in (-1,1)} \mathbf{E} \sup_{t \in [0,1]} |D^h X_t(x)|^p < \infty, \quad p \geqslant 1. \tag{2.6}$$

By Taylor expansion,

$$D^h X_t(x) - D^h X_t(y)$$
$$= \int_0^t \underbrace{\left[\int_0^1 \partial_l\sigma(\tau X_s(x+h) + (1-\tau)X_s(x))\mathrm{d}\tau\right]}_{\leqslant \|\nabla\sigma\|_0} \left[D^h X_t(x) - D^h X_t(y)\right]\mathrm{d}W_s$$
$$+ \int_0^t \underbrace{\left[\int_0^1 \partial_l\sigma(\tau X_s(x+h) + (1-\tau)X_s(x)) - \partial_l\sigma(\tau X_s(y+h) + (1-\tau)X_s(y))\mathrm{d}\tau\right]}_{\leqslant \|\nabla^2\sigma\|_0(|X_s(x+h)-X_s(y+h)|+|X_s(x)-X_s(y)|)}$$
$$D^h X_s^l(y)\mathrm{d}W_s.$$

Let

$$g(t) = \mathbf{E} \sup_{s \in [0,t]} |D^h X_s(x) - D^h X_s(y)|^p.$$

Then BDG inequality and Gronwall's inequality yield

$$g(t) \leqslant C \int_0^t \mathbf{E} \left[\left(|X_s(x+h) - X_s(y+h)| + |X_s(x) - X_s(y)| \right) D^h X_s^l(y) \right]^p \mathrm{d}s$$

$$\leqslant C \left[\mathbf{E} \int_0^t \left(|X_s(x+h) - X_s(y+h)| + |X_s(x) - X_s(y)| \right)^{2p} \mathrm{d}s \right]^{1/2}$$

$$\left[\mathbf{E} \int_0^t |D^h X_s(y)|^{2p} \mathrm{d}s \right]^{1/2} \stackrel{(2.6)}{\leqslant} C|x-y|^p, \quad p \geqslant 2.$$

This implies

$$\mathbf{E} \sup_{t \in [0,1]} |Z_t(x,h) - Z_t(y,h)|^p \leqslant C|x-y|^p, \quad p \geqslant 1. \tag{2.7}$$

On the other hand,

$$Z_t(x,h) = \int_0^t \left\{ D^h \left[\sigma(X_s(x)) \right] - \partial_l \sigma(X_s(x)) J_s^l(x) \right\} \mathrm{d}W_s$$

$$= \int_0^t \left\{ \partial_l \sigma(X_s(x)) \left[D^h X_s^l(x) - J_s^l(x) \right] + R_s^h(x) \right\} \mathrm{d}W_s$$

$$= \int_0^t \left[\partial_l \sigma(X_s(x)) Z_s(x,h) + R_s^h(x) \right] \mathrm{d}W_s,$$

where

$$R_s^h(x) = D^h \left[\sigma(X_s(x)) \right] - \partial_l \sigma(X_s(x)) D^h X_s^l(x),$$

and

$$|R_s^h(x)| \leqslant |h| \, \|\nabla^2 \sigma\|_0 |D^h X_s(x)|.$$

Again utilizing BDG inequality and Gronwall's inequality, we see that for each $p \geqslant 2$,

$$\mathbf{E} \sup_{\tau \in [0,t]} |Z_\tau(x,h)|^p \leqslant C \mathbf{E} \left\{ \int_0^t \left[|Z_s(x,h)|^2 + \left(R_s^h(x) \right)^2 \right] \mathrm{d}s \right\}^{p/2}$$

$$\leqslant C \int_0^t \mathbf{E} \sup_{\tau \in [0,s]} |Z_\tau(x,h)|^p \mathrm{d}s + C|h|^p \int_0^t \mathbf{E} |D^h X_s(x)|^p \mathrm{d}s,$$

which together with (2.6) implies

$$\mathbf{E} \sup_{t \in [0,1]} |Z_t(x,h)|^p \leqslant C|h|^p, \quad |h| < 1. \tag{2.8}$$

Combining (2.7) and (2.8), we obtain

$$\mathbf{E} \sup_{t \in [0,1]} |Z_t(x,h) - Z_t(y,h)|^p \leqslant C|x-y|^{\theta p} |h|^{(1-\theta)p}, \quad \theta \in [0,1], \ p \geqslant 1 \ |h| < 1.$$

Thanks to Kolmogorov's continuity theorem, for any $\alpha \in (0,1)$,

$$\lim_{|h| \to 0} \mathbf{E} \sup_{t \in [0,1]} \|Z_t(\cdot, h)\|_{C^\alpha}^p = 0, \quad p \geqslant 1.$$

Therefore, $X_t(\cdot) \in C^{1,\alpha}$ a.s., and $\nabla X_t(x) = J_t(x)$. $\square$

One can also show (see Ikeda and Watanabe [IW14]) that the map $x \mapsto X_t(x)$ is one-to-one and onto $\mathbb{R}^d$.

Appendix A. Basic Stochastic Analysis

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a standard probability space. Let $\mathcal{F}_n$ $(n \in \mathbb{N})$ be an increasing sequence of σ-fields. A sequence of random variables X_n is **adapted** to $\mathcal{F}_n$ if for each n, X_n is $\mathcal{F}_n$ measurable. Similarly a collection of random variables X_t $(t \in \mathbb{R}_+)$ is adapted to $\mathcal{F}_t$ if each X_t is $\mathcal{F}_t$ measurable. We say the filtration $\mathcal{F}_t$ satisfies the usual conditions if $\mathcal{F}_t$ is **right continuous** (i.e., $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, where $\mathcal{F}_{t+} = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$) and each $\mathcal{F}_t$ is **complete** (i.e., $\mathcal{F}_t$ contains all **P**-null sets).

We say $\tau : \Omega \to \mathbb{N}$ $(\mathbb{R}_+) \cup \{\infty\}$ is a **stopping time** if τ satisfying $\{\tau \leqslant n\} \in \mathcal{F}_n$ $(\{\tau \leqslant t\} \in \mathcal{F}_t)$, for each $n \in \mathbb{N}$ $(t \in \mathbb{R}_+)$.

$\mathcal{F}_\tau$ is a σ-field containing all measurable sets $A \in cF$ such that $A \cap \{\tau \leqslant n\} \in \mathcal{F}_n$ $(A \cap \{\tau \leqslant t\} \in \mathcal{F}_t)$ for all $n \in \mathbb{N}$ $(t \in \mathbb{R}_+)$.

Definition A.1. *Let X_t be a real-valued $\mathcal{F}_t$-adapted processes. If for each t and $s < t$, X_t is integrable and $\mathbf{E}(X_t|\mathcal{F}_s) \geqslant (\leqslant)X_s$ a.s., then we call X_t is a* **submartingale (supermartingale)**. *We say X_t is a* **martingale** *if it is both a submartingale and a supermartingale.*

Example 1. *Let $\xi_1, \xi_2, \cdots$ be a sequence of i.i.d random variable. Set $X_n := \sum_{i=0}^n \xi_i$ and $\mathcal{F}_n := \sigma(\xi_0, \cdots \xi_n)$.*

Below we recall the results about discrete time martingales and submartingales that will be used. The proof of the subsequent statements can be found in Durrett's book [Dur19], and in many other books dealing with discrete time martingales.

Theorem A.1 (Doob). *If $X_n \in \mathcal{F}_n$ is a submartingale then it can be uniquely decomposed as $X_n = M_n + A_n$, where $M_n \in \mathcal{F}_n$ is martingale, $A_n = 0$, $A_{n+1} \geqslant A_n$ almost surely and A_n is $\mathcal{F}_{n-1}$-measurable.*

The following theorem lies at the basis of all other results for martingales.

Theorem A.2 (Doob's Optional stopping theorem). *Assume that σ and τ are two bounded stopping time, and X_t is a submartingale, then $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) \geqslant X_{\sigma \wedge \tau}$.*

Lemma A.3. *Let X_n be a submartingale, and τ be a bounded stopping time and $\tau \leqslant K$ (constant). Then*

(i) $\mathbf{E}(X_K|\mathcal{F}_\tau) \geqslant X_\tau$;
(ii) $X_{\tau \wedge n}$ is a $\mathcal{F}_n$-submartingale.

Proof. (i). for each $A \in \mathcal{F}_\tau$, we will show that $\mathbf{E}(X_K; A) \geqslant \mathbf{E}(X_\tau; A)$. In fact,

$$\mathbf{E}(X_\tau; A) = \sum_{k=0}^K \mathbf{E}(X_k; \underbrace{A \cap \{\tau = k\}}_{\in \mathcal{F}_k}) \leqslant \sum_{k=0}^K \mathbf{E}(X_K; A \cap \{\tau = k\}) = \mathbf{E}(X_K; A).$$

(ii). For each $A \in \mathcal{F}_{n-1}$,

$$\begin{aligned}
\mathbf{E}(X_{\tau \wedge n}; A) =& \mathbf{E}(X_{\tau \wedge n}; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_{\tau \wedge n}; A \cap \{\tau > n-1\}) \\
=& \mathbf{E}(X_\tau; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_n; \underbrace{A \cap \{\tau > n-1\}}_{\in \mathcal{F}_{n-1}}) \\
\geqslant& \mathbf{E}(X_\tau; A \cap \{\tau \leqslant n-1\}) + \mathbf{E}(X_{n-1}; A \cap \{\tau > n-1\}) \\
=& \mathbf{E}(X_{\tau \wedge (n-1)}; A).
\end{aligned}$$

$\square$

Proof of Theorem A.2. By the above lemma, we have $\mathbf{E}(X_\tau|\mathcal{F}_\sigma) = \mathbf{E}(X_{K\wedge\tau}|\mathcal{F}_\sigma) \geqslant X_{\sigma\wedge\tau}.$ $\square$

Theorem A.4 (Doob's inequality)**.** *Let M_n be a martingale. If $M_n^* := \sup_{k\leqslant n}|M_k|$, then*

$$\mathbf{P}(M_n^* > \lambda) \leqslant \lambda^{-1}\mathbf{E}(|M_n|; M_n^* > \lambda).$$

Proof. Let $\tau = \inf\{k : |M_k| > \lambda\}$. Noting that $\{M_n^* > \lambda\}\} = \{\tau \leqslant n\}$, we have

$$\lambda\mathbf{P}(M_n^* > \lambda) = \lambda\mathbf{P}(\tau \leqslant n) \leqslant \mathbf{E}(|M_\tau|; \tau \leqslant n)$$
$$\leqslant \mathbf{E}(|M_{\tau\wedge n}|; \tau \leqslant n) \leqslant \mathbf{E}(|M_n|; M_n^* > \lambda).$$

$\square$

Corollary A.5. *Let M_n be a martingale and T be a stopping time. For each $p > 1$, $\mathbf{E}|M_T^*|^p \leqslant C_p\mathbf{E}|M_T|^p$.*

Let $a \leqslant b$. Set $\sigma_1 = \inf\{n \geqslant 0 : X_n \leqslant a\}$, $\tau_1 = \inf\{n > \sigma_1 : X_n \geqslant b\}$, $\sigma_2 = \inf\{n > \tau_1 : M_n \leqslant a\}$, $\tau_2 = \inf\{n > \sigma_2 : X_n \geqslant b\}$, ..., and $U_N := \max\{k : \tau_k \leqslant N\}$.

Lemma A.6 (Upcrossing inequality)**.** *Suppose that X_N is a submartingale, then*

$$(b - a)\mathbf{E}U_N(a, b) \leqslant \mathbf{E}(X_N - a)^+.$$

Proof. We only prove the case that $a = 0$ and $X_k \geqslant 0$.

$$X_N = \underbrace{X_{S_1\wedge N}}_{\geqslant 0} + \underbrace{\sum_{i=1}^\infty X_{T_i\wedge N} - X_{S_i\wedge N}}_{\geqslant bU_N(0,b)} + \sum_{i=1}^\infty \underbrace{X_{S_{i+1}\wedge N} - X_{T_i\wedge N}}_{\text{positive expectation}}.$$

$\square$

Upcrossing inequality leads to

Theorem A.7. *If X_n is a submartingale such that $\sup_n \mathbf{E}X_n^+ < \infty$, then X_n converges a.s. as $n \to \infty$.*

Corollary A.8. *Suppose that $X \in L^1(\mathbf{P}, \Omega)$, $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, then*

$$\lim_{n\to\infty} \mathbf{E}(X|\mathcal{F}_n) = \mathbf{E}(X|\mathcal{F}_\infty), \quad a.s. \text{ and in } L^1.$$

For an example of a discrete martingale, let $\Omega = [0, 1], \mathbf{P}$ Lebesgue measure, and f an integrable function on $[0, 1]$. Let $\mathcal{F}_n$ be the σ-field generated by the sets

$$\{[k/2^n, (k + 1)/2^n), k = 0, 1, \ldots, 2^n - 1\}.$$

Let $f_n = \mathbf{E}[f \mid \mathcal{F}_n]$. If I is an interval in $\mathcal{F}_n$, shows that

$$f_n(x) = \frac{1}{|I|}\int_I f(y)dy \quad \text{if } x \in I.$$

f_n is a particular example of what is known as a dyadic martingale. Of course, $[0, 1]$ could be replaced by any interval as long as we normalize so that the total mass of the interval is 1. We could also divide cubes in $\mathbb{R}^d$ into 2^d subcubes at each step and define f_n analogously. Such martingales are called dyadic martingales. In fact, we could replace Lebesgue measure by any finite measure μ, and instead of decomposing into equal subcubes, we could use any nested partition of sets we like, provided none of these sets had μ measure 0.

All of the above results also hold for all right continuous martingale (submartingales) (see [Hua01]***).***

Theorem A.9. *Assume X is a continuous submartingale, then there exists a unique martingale M and a unique continuous increasing adapted process A such that*

$$A_0 = 0, \quad X_t = M_t + A_t.$$

If M is a continuous square integrable martingale, then M^2 is a submartingale. Thus, there exists a continuous increasing process, denoted by $\langle M \rangle$, the **quadratic variation** of M, such that $M^2 - \langle M \rangle$ is a martingale. Particularly, $\mathbf{E}M_t^2 - \mathbf{E}M_0^2 = \mathbf{E}\langle M \rangle_t$.

A.1. **Stochastic Integral.** From now on, unless stated otherwise, our processes have continuous paths.

Lemma A.10. *Let M_t be a square integrable martingale (that is, $M_t \in L^2$ for every $t \geqslant 0$). Let $0 \leqslant s < t$ and let $s = t_0 < t_1 < \cdots < t_n = t$ be a division of the interval $[s,t]$. Then,*

$$\mathbf{E}\left[\sum_{i=1}^n \left(M_{t_i} - M_{t_{i-1}}\right)^2 \mid \mathcal{F}_s \right] = \mathbf{E}\left[M_t^2 - M_s^2 \mid \mathcal{F}_s \right] = \mathbf{E}\left[(M_t - M_s)^2 \mid \mathcal{F}_s \right].$$

Proof. For every $i = 1, \ldots, n$,

$$\begin{aligned}
\mathbf{E}\left[\left(M_{t_i} - M_{t_{i-1}}\right)^2 \mid \mathcal{F}_s \right] &= \mathbf{E}\left[\mathbf{E}\left[\left(M_{t_i} - M_{t_{i-1}}\right)^2 \mid \mathcal{F}_{t_{i-1}} \right] \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[\mathbf{E}\left[M_{t_i}^2 \mid \mathcal{F}_{t_{i-1}} \right] - 2M_{t_{i-1}} \mathbf{E}\left[M_{t_i} \mid \mathcal{F}_{t_{i-1}} \right] + M_{t_{i-1}}^2 \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[\mathbf{E}\left[M_{t_i}^2 \mid \mathcal{F}_{t_{i-1}} \right] - M_{t_{i-1}}^2 \mid \mathcal{F}_s \right] \\
&= \mathbf{E}\left[M_{t_i}^2 - M_{t_{i-1}}^2 \mid \mathcal{F}_s \right]
\end{aligned}$$

and the desired result follows by summing over i. $\qquad\square$

We say that M_t if a **local martingale** if there exist stopping times $\tau_n \uparrow \infty$ such that $X_{\tau_n \wedge t}$ is a martingale for each $n \in \mathbb{N}$.

Theorem A.11. *Let M_t be a continuous local martingale. There exists an increasing process denoted by $\langle M \rangle_t$, which is unique up to indistinguishability, such that $M_t^2 - \langle M \rangle_t$ is a continuous local martingale. Furthermore, for every fixed $t > 0$, if $\pi^n = \{(t_0^n, \cdots, t_{k_n}^n) : 0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t\}$ is an increasing sequence of subdivisions of $[0,t]$ with mesh going to 0 , then we have*

$$\langle M \rangle_t = \lim_{n \to \infty} \sum_{i=1}^{k_n} \left(M_{t_i^n} - M_{t_{i-1}^n} \right)^2$$

in probability. The process $\langle M \rangle_t$ is called the quadratic variation of M_t.

Lemma A.12. *Let M_t be a continuous bounded martingale. Let $\pi^n = \{(t_0^n, \cdots, t_{k_n}^n) : 0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$ be an increasing sequence of subdivisions of $[0,T]$ with mesh going to 0 , then for each n,*

$$N_t^n := \sum_{i=1}^{k_n} M_{t_{i-1}}(M_{t_i \wedge t} - M_{t_{i-1} \wedge t})$$

is a martingale, and N_t^n convergent uniformly on compacts, with probability one to some square integrable martingale N_t.

Proof. It is easy to verify that N_t^n is a martingale. Let us fix $n \leqslant m$ and evaluate the product $\mathbf{E}\left(N_T^n N_T^m\right)$. This product is equal to

$$\sum_{i=1}^{k_n} \sum_{j=1}^{k_m} \mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n} - M_{t_{i-1}^n}\right) M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)\right].$$

In this double sum, the only terms that may be nonzero are those corresponding to indices i and j such that the interval $\left(t_{j-1}^m, t_j^m\right]$ is contained in $\left(t_{i-1}^n, t_i^n\right]$. Indeed, suppose that $t_i^n \leqslant t_{j-1}^m$ (the symmetric case $t_j^m \leqslant t_{i-1}^n$ is treated in an analogous way).

Then, conditioning on the σ-field $\mathscr{F}_{t_{j-1}^m}$, we have

$$\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n} - M_{t_{i-1}^n}\right) M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)\right]$$
$$= \mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n} - M_{t_{i-1}^n}\right) M_{t_{j-1}^m} \mathbf{E}\left[M_{t_j^m} - M_{t_{j-1}^m} \mid \mathscr{F}_{t_{j-1}^m}\right]\right] = 0.$$

For every $j = 1, \ldots, k_m$, write $i_{n,m}(j)$ for the unique index i such that $\left(t_{j-1}^m, t_j^m\right] \subset \left(t_{i-1}^n, t_i^n\right]$. It follows from the previous considerations that

$$\mathbf{E}\left[N_T^n N_T^m\right] = \sum_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n} - M_{t_{i-1}^n}\right) M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)\right].$$

In each term $\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_i^n} - M_{t_{i-1}^n}\right) M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)\right]$, we can now decompose

$$M_{t_i^n} - M_{t_{i-1}^n} = \sum_{k: i_{n,m}(k) = i}\left(M_{t_k^m} - M_{t_{k-1}^m}\right)$$

and we observe that, if k is such that $i_{n,m}(k) = i$ but $k \neq j$,

$$\mathbf{E}\left[M_{t_{i-1}^n}\left(M_{t_k^m} - M_{t_{k-1}^m}\right) M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)\right] = 0$$

(condition on $\mathscr{F}_{t_{k-1}^m}$ if $k > j$ and on $\mathscr{F}_{t_{j-1}^m}$ if $k < j$). The only case that remains is $k = j$, and we have thus obtained

$$\mathbf{E}\left[N_T^n N_T^m\right] = \sum_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \mathbf{E}\left[M_{t_{i-1}^n} M_{t_{j-1}^m}\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right].$$

As a special case of this relation, we have

$$\mathbf{E}\left[\left(N_T^m\right)^2\right] = \sum_{1 \leqslant j \leqslant k_m} \mathbf{E}\left[M_{t_{j-1}^m}^2\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right].$$

Furthermore,

$$
\begin{aligned}
\mathbf{E}\left[(N_T^n)^2\right] &= \sum_{1 \leqslant i \leqslant n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \left(M_{t_i^n} - M_{t_{i-1}^n}\right)^2\right] \\
&= \sum_{1 \leqslant i \leqslant n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \, \mathbf{E}\left[\left(M_{t_i^n} - M_{t_{i-1}^n}\right)^2 \mid \mathscr{F}_{i_{i-1}^n}\right]\right] \\
&= \sum_{1 \leqslant i \leqslant k_n} \mathbf{E}\left[M_{t_{i-1}^n}^2 \sum_{j : i_{n,m}(j)=i} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mid \mathscr{F}_{t_{i-1}^n}\right]\right] \\
&= \sum_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \mathbf{E}\left[M_{t_{i-1}^n}^2 \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right],
\end{aligned}
$$

If we combine the last three displays, we get

$$
\mathbf{E}\left[(N_T^n - N_T^m)^2\right] = \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^2 \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right].
$$

Using the Cauchy-Schwarz inequality, we then have

$$
\begin{aligned}
\mathbf{E}\left[(N_T^n - N_T^m)^2\right] \leqslant &\, \mathbf{E}\left[\sup_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^4\right]^{1/2} \\
&\times \mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right]^{1/2}.
\end{aligned}
$$

By the continuity of sample paths (together with the fact that the mesh of our subdivisions tends to 0) and dominated convergence, we have

$$
\lim_{n,m \to \infty, n \leqslant m} \mathbf{E}\left[\sup_{1 \leqslant j \leqslant k_m, i = i_{n,m}(j)} \left(M_{t_{i-1}^n} - M_{t_{j-1}^m}\right)^4\right] = 0.
$$

To complete the proof of the lemma, it is then enough to prove the existence of a finite constant C such that, for every m,

$$
\mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right] \leqslant C.
$$

Let A be a constant such that $|M_t| \leqslant A$ for every $t \geqslant 0$. Expanding the square and using Proposition 3.14 twice, we have

$$
\mathbf{E}\left[\left(\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right)^2\right]
$$

$$
= \mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^4\right] + 2\mathbf{E}\left[\sum_{1 \leqslant j < k \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \left(M_{t_k^m} - M_{t_{k-1}^m}\right)^2\right]
$$

$$
\leqslant 4A^2\mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right]
$$

$$
+ 2\sum_{j=1}^{k_m-1} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mathbf{E}\left[\sum_{k=j+1}^{k_m} \left(M_{t_k^m} - M_{t_{k-1}^m}\right)^2 \mid \mathscr{F}_{t_j^m}\right]\right]
$$

$$
= 4A^2\mathbf{E}\left[\sum_{1 \leqslant j \leqslant k_m} \left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2\right]
$$

$$
+ 2\sum_{j=1}^{k_m-1} \mathbf{E}\left[\left(M_{t_j^m} - M_{t_{j-1}^m}\right)^2 \mathbf{E}\left[\left(M_T - M_{t_j^m}\right)^2 \mid \mathscr{F}_{t_j^m}\right]\right]
$$

$\square$

Proof of Theorem A.11. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Let M_t be a square integrable martingale, $0 = t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n = T$ and $H_s(\omega) = \sum_{i=0}^{n-1} H_{t_i}(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(s)$, where F_i is bounded and $\mathcal{F}_{t_i}$-measurable. Define

$$
\int_0^t H_s \mathrm{d}M_s := \sum_{i=0}^{n-1} H_{t_i}(M_{t \wedge t_{i+1}} - M_{t \wedge t_i}).
$$

Then

Lemma A.13. $t \mapsto \int_0^t H_s \mathrm{d}M_s$ *is a L^2-martingale. Moreover, we have the following Itô isometry:*

$$
\mathbf{E}\left(\int_0^t H_s \mathrm{d}M_s\right)^2 = \mathbf{E}\int_0^t H_s^2 \mathrm{d}\langle M\rangle_s. \tag{A.1}
$$

Proof.

$$
\mathbf{E}\left(\int_0^1 H_s \mathrm{d}M_s\right)^2 = \mathbf{E}\sum_i H_{t_i}^2 (M_{t_{i+1}} - M_{t_i})^2 + 2\mathbf{E}\sum_{i<j} H_{t_i}H_{t_j}(M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})
$$

$$
=: I_1 + I_2.
$$

$$
I_1 = \sum_i \mathbf{E}\mathbf{E}\left(H_{t_i}^2(M_{t_{i+1}} - M_{t_i})^2 \big| \mathcal{F}_{t_i}\right) = \sum_i \mathbf{E}\left[H_{t_i}^2 \mathbf{E}\left((M_{t_{i+1}} - M_{t_i})^2 \big| \mathcal{F}_{t_i}\right)\right]
$$

$$
= \sum_i \mathbf{E}H_{t_i}^2(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}) = \mathbf{E}\int_0^1 H_s^2 \mathrm{d}\langle M\rangle_s,
$$

$$I_2 = 2 \sum_{i<j} \mathbf{E} \left[H_{t_i} H_{t_j} (M_{t_{i+1}} - M_{t_i}) \mathbf{E} \left((M_{t_{j+1}} - M_{t_j}) | \mathcal{F}_{t_j} \right) \right] = 0.$$

Therefore,

$$\mathbf{E} \left(\int_0^1 H_s \mathrm{d} M_s \right)^2 = \mathbf{E} \int_0^1 H_s^2 \mathrm{d} \langle M \rangle_s.$$

$t \mapsto \int_0^t H_s \mathrm{d} W_s = \sum_{i=0}^{n-1} H_{t_i} (W_{t \wedge t_{i+1}} - W_{t \wedge t_i})$ is a continuous martingale. $\qquad \square$

We then can use this to extend the above construction to more general H_s satisfying $\int_0^t H_s^2 \mathrm{d} \langle M \rangle_s < \infty$ by taking limits in L^2. For general continuous local martingale, we can employ standard localization argument to define the above integral. For $X_t = M_t + A_t$, a semimartingale, $\int_0^t H_s \mathrm{d} X_s$ is given by

$$\int_0^t H_s \mathrm{d} X_s = \int_0^t H_s \mathrm{d} M_s + \int_0^t H_s \mathrm{d} A_s,$$

where the first integral on the right is a stochastic integral and the second integral on the right is a Riemann-Stieltjes integral.

Proposition A.14.

$$\left\langle \int_0^\cdot H_s \mathrm{d} M_s \right\rangle_t = \int_0^t H_s^2 \mathrm{d} \langle M \rangle_s.$$

Let $N_t = \int_0^t H_s \mathrm{d} M_s$

$$\int_0^t K_s \mathrm{d} N_s = \int_0^t K_s H_s \mathrm{d} M_s$$

Appendix B. Itô's formula and its applications

B.1. **Applications in martingale theory.** We list some important results in stochastic calculus.

Theorem B.1 (Itô's formula). *If each X_t^i (for each $i \in 1, \cdots d\}$) is a continuous semimartingale and $f \in C^2(\mathbb{R}^d)$, then*

$$
\begin{aligned}
& f\left(X_t\right) - f\left(X_0\right) \\
& = \int_0^t \sum_{i=1}^d \partial_i f\left(X_s\right) \mathrm{d} X_s^i + \frac{1}{2} \int_0^t \sum_{i,j=1}^d \partial_{ij} f\left(X_s\right) \mathrm{d} \left\langle X^i, X^j \right\rangle_s
\end{aligned}
\tag{B.1}
$$

(see [Hua01, Theorem 13.5]).

It is often useful to use the language of Stratonovitch's integration to study stochastic differential equations because the Itô's formula takes a much nicer form. If M_t is an $\mathcal{F}_t$-adapted real valued local martingale and if H_t is an $\mathcal{F}_t$-adapted continuous semimartingale satisfying $\mathbf{P} \left(\int_0^T H_s \mathrm{d} \langle M \rangle_s < \infty \right) = 1$, then by definition the Stratonovitch integral of H_t with respect to M_t is defined as

$$\int_0^T H_t \circ \mathrm{d} M_t = \int_0^T H_t \mathrm{d} M_t + \frac{1}{2} \langle H, M \rangle_T.$$

By using Stratonovitch integral instead of Itô's, the Itô formula reduces to the classical change of variable formula.

Theorem B.2. *Let M_t be a d–dimensional continuous semimartingale. Let now f be a C^2 function. We have*

$$f(M_t) = f(M_0) + \int_0^t \partial_i f(X_s) \circ \mathrm{d}M_s^i, \quad t \geqslant 0.$$

Theorem B.3 (Burkholder-Davis-Gundy inequalities)**.** *If M_t is a continuous martingale with $M_0 = 0$, and τ is a stopping time, then*

$$\mathbf{E} \sup_{t \in [0,\tau]} |M_t|^p \asymp_p \mathbf{E}\langle M\rangle_\tau^{p/2}, \quad p \in (0, \infty) \tag{B.2}$$

Proof. Step 1: for any $p \geqslant 2$, by Itô's formula

$$|M_\tau|^p = p \int_0^\tau \mathrm{sgn}(M_t)|M_t|^{p-2} M_t \mathrm{d}M_t + \frac{p(p-1)}{2} \int_0^T |M_t|^{p-2} \mathrm{d}\langle M\rangle_t;$$

By Doob's inequality and Hölder's inequality,

$$\mathbf{E}(M_\tau^*)^p \lesssim_p \mathbf{E}|M_\tau|^p \lesssim_p \mathbf{E}((M_\tau^*)^{p-2}\langle M\rangle_\tau)$$
$$\leqslant (\mathbf{E}(M_\tau^*)^p)^{1-\frac{2}{p}} (\mathbf{E}\langle M\rangle_\tau^{\frac{p}{2}})^{\frac{2}{p}};$$

Step 2: using Lenglart's domination inequality, we can get the proof for the case $p \in (0,2)$.

We proceed now to the proof of the left hand side inequality. We have,

$$M_t^2 = \langle M\rangle_t + 2 \int_0^t M_s dM_s.$$

Therefore, we get

$$\mathbf{E}\left(\langle M\rangle_T^{\frac{p}{2}}\right) \lesssim \mathbf{E}(M_T^*)^p + \mathbf{E}\left(\sup_{0 \leq t \leq T} \left|\int_0^t M_s dM_s\right|^{p/2}\right).$$

By using the previous argument, we now have

$$2^{\frac{p}{2}}\mathbf{E}\left(\sup_{0 \leq t \leq T} \left|\int_0^t M_s dM_s\right|^{p/2}\right) \leq C\mathbf{E}\left(\left(\int_0^T M_s^2 d\langle M\rangle_s\right)^{p/4}\right)$$
$$\leqslant C\mathbf{E}\left((M_T^*)^{p/2}\langle M\rangle_T^{p/4}\right) \leqslant C\left(\mathbf{E}(M_T^*)^p\right)^{1/2}\left(\mathbf{E}\langle M\rangle_T^{p/2}\right)^{1/2}$$
$$\leqslant \varepsilon'\mathbf{E}(M_T^*)^p + C_{\varepsilon'}\mathbf{E}\langle M\rangle_T^{p/2} \leqslant \varepsilon.$$

As a conclusion, we obtained that d $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Proposition B.4 (Lenglart)**.** *Let X_t be a positive adapted right-continuous process and A_t be an increasing process. Assume that for every bounded stopping time τ, $\mathbf{E}(X_\tau \mid \mathcal{F}_0) \leqslant \mathbf{E}(A_\tau \mid \mathcal{F}_0)$. Then, for every $\kappa \in (0,1)$,*

$$\mathbf{E}\left(X_T^*\right)^\kappa \leqslant \frac{2-\kappa}{1-\kappa}\mathbf{E}\left(A_T^\kappa\right).$$

We shall use this lemma to prove the following

Another approach to proving (B.2) is utilizing "good-λ" inequality (cf. [RY13]).

Theorem B.5 (Lévy's theorem)**.** *If X_t is a d-dimensional $(\mathcal{F}_t)_{t \geqslant 0}$-adapted process, each of whose coordinates is a continuous local martingale, and $\langle X^i, X^j\rangle_t = \delta_{ij}t$, then X_t is a d-dimensional $(\mathcal{F}_t)_{t \geqslant 0}$-Brownian motion.*

Proof. Let $\xi \in \mathbb{R}^d$. Then $\xi \cdot X_t$ is a continuous local martingale with quadratic variation $\langle \xi \cdot X \rangle_t = |\xi|^2 t$. By Itô's formula, $\exp(i\xi \cdot X_t + \frac{1}{2}|\xi|^2 t)$ is a continuous local martingale. This complex continuous local martingale is bounded on every finite interval and is therefore a (true) martingale, in the sense that its real and imaginary parts are both martingales. Hence, for every $s < t$,

$$\mathbf{E}\left[\exp\left(i\xi \cdot X_t + \frac{1}{2}|\xi|^2 t\right)\middle| \mathcal{F}_s\right] = \exp\left(i\xi \cdot X_s + \frac{1}{2}|\xi|^2 s\right)$$

Thus,

$$\mathbf{E}\left[\exp\left(i\xi \cdot (X_t - X_s)\right)\middle| \mathcal{F}_s\right] = \exp\left(\frac{1}{2}|\xi|^2(t - s)\right).$$

This implies $X_t - X_s$ is independent with $\mathcal{F}_s$ and $X_t - X_s \sim \mathcal{N}(0, t - s)$.

Finally, X is adapted and has independent increments with respect to the filtration $(\mathcal{F}_t)_{t\geqslant 0}$ so that X is a s-dimensional $(\mathcal{F}_t)_{t\geqslant 0}$-Brownian motion. $\square$

Let M_t be a continuous local martingale with $M_0 = 0$. Set $\mathscr{E}(M)_t := \exp(M_t - \langle M \rangle_t/2)$.

Proposition B.6. *$\mathscr{E}(M)_t$ is a continuous local martingale, and is the unique solution to*

$$\mathrm{d}X_t = X_t \mathrm{d}M_t, \quad X_0 = 1.$$

Theorem B.7 (Girsanov theorem). *Let X_t and M_t be two continuous local martingales under $\mathbb{P}$ with $M_0 = 0$ $\mathbb{P}$-a.s.. Assume that $\mathscr{E}(M)_t$ is a martingale, we define a new probability measure $\mathbb{Q}$ by setting the restriction of $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ to $\mathcal{F}_t$ to be $\mathscr{E}(M)_t$, then $X_t - \langle X, M \rangle_t$ is a martingale under $\mathbb{Q}$ and the quadratic variation of X_t is the same under $\mathbb{P}$ and $\mathbb{Q}$.*

Proof. By localization, we can assume X is a martingale. Set $Y_t = X_t - \langle X, M \rangle_t$. We only need to verify that $Y_t \mathscr{E}(M)_t$ is a martingale under $\mathbb{P}$. By Itô's formula,

$$\begin{aligned}
\mathrm{d}Y_t \mathscr{E}(M)_t &= \mathscr{E}(M)_t \mathrm{d}X_t - \mathscr{E}(M)_t \mathrm{d}\langle X, M \rangle_t + Y_t \mathscr{E}(M)_t \mathrm{d}M_t + \mathrm{d}\langle X, \mathscr{E}(M) \rangle_t \\
&= \mathscr{E}(M)_t \mathrm{d}X_t + Y_t \mathscr{E}(M)_t \mathrm{d}M_t.
\end{aligned}$$

Therefore, $Y_t \mathscr{E}(M)_t$ is a martingale, which implies

$$\mathbf{E}_{\mathbb{Q}}(Y_t; A) = \mathbf{E}_{\mathbb{Q}}(Y_s; A), \quad \forall A \in \mathcal{F}_s,$$

i.e.

$$\mathbf{E}_{\mathbb{Q}}(Y_t | \mathcal{F}_s) = Y_s.$$

$\square$

Theorem B.8 (Dambis–Dubins–Schwarz). *Let M be a continuous local martingale such that $\langle M \rangle_\infty = \infty$. There exists a Brownian motion B such that*

$$M_t = B_{\langle M \rangle_t}.$$

Appendix C. Schauder estimate

Let $\mathscr{S}$ be the Schwartz space of all rapidly decreasing functions, and $\mathscr{S}'$ the dual space of $\mathscr{S}$ called Schwartz generalized function (or tempered distribution) space. Given $f \in \mathscr{S}$, let $\mathscr{F}f = \hat{f}$ be the Fourier transform defined by

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} e^{-i2\pi\xi \cdot x} f(x) \mathrm{d}x.$$

Let $\chi : \mathbb{R}^d \to [0, 1]$ be a smooth radial function with

$$\chi(\xi) = 1, \ |\xi| \leqslant 1, \ \chi(\xi) = 0, \ |\xi| \geqslant 3/2.$$

Define

$$\varphi(\xi) := \chi(\xi) - \chi(2\xi).$$

It is easy to see that $\varphi \geqslant 0$ and $\operatorname{supp} \varphi \subset B_{3/2} \setminus B_{1/2}$ and formally

$$\sum_{j=-k}^{k} \varphi(2^{-j}\xi) = \chi(2^{-k}\xi) - \chi(2^{k+1}\xi) \overset{k\to\infty}{\to} 1. \tag{C.1}$$

In particular, if $|j - j'| \geqslant 2$, then

$$\operatorname{supp}\varphi(2^{-j}\cdot) \cap \operatorname{supp}\varphi(2^{-j'}\cdot) = \varnothing.$$

From now on we shall fix such χ and φ and define

$$\Delta_j f := \mathscr{F}^{-1}(\varphi(2^{-j}\cdot)\mathscr{F}f), \quad j \in \mathbb{Z}.$$

Set $h := \mathscr{F}^{-1}(\varphi)$, then $h_j := \mathscr{F}^{-1}(\varphi(2^{-j}\cdot)) = 2^{jd}h(2^j\cdot)$. Noting that we have

$$\fint_{\mathbb{R}^d} h_j = \varphi(0) = 0.$$

We first recall the following useful lemmas.

Lemma C.1. *Let $\alpha \in (0, 1)$. For any $u \in C^\alpha$, it holds that*

$$\frac{1}{C} \sup_{j\in\mathbb{Z}} 2^{j\alpha} \|\Delta_j u\|_0 \leqslant [u]_\alpha \leqslant C \sup_{j\in\mathbb{Z}} 2^{j\alpha} \|\Delta_j u\|_0, \tag{C.2}$$

where

$$[u]_\alpha := \sup_{x\neq y} \frac{|u(x) - u(y)|}{|x - y|^\alpha},$$

and C only depends on d and α.

Proof. 1) For any $x \neq y$, we have

$$|u(x) - u(y)| \leqslant \sum_j |\Delta_j u(x) - \Delta_j u(y)| \leqslant \sum_j (|x - y| \|\nabla\Delta_j u\|_0) \wedge (2\|\Delta_j u\|_0)$$

$$\leqslant C \sum_j (2^j |x - y| \wedge 1)\|\Delta_j u\|_0$$

$$\leqslant C 2^{j\alpha} \|\Delta_j u\|_0 \left(|x - y| \sum_{j\leqslant \log_2 |x-y|} 2^{j(1-\alpha)} + \sum_{j\geqslant \log_2 |x-y|} 2^{-j\alpha} \right)$$

$$\leqslant C|x - y|^\alpha \sup_j 2^{j\alpha} \|\Delta_j u\|_0.$$

2) For any $j \in \mathbb{Z}$ and $x \in \mathbb{R}^d$, we have

$$|\Delta_j u(x)| = \left| \fint_{\mathbb{R}^d} u(x - y)h_j(y)\mathrm{d}y \right| = \left| \fint_{\mathbb{R}^d} [u(x - y) - u(x)]h_j(y)\mathrm{d}y \right|$$

$$\leqslant [u]_\alpha 2^{jd} \int_{\mathbb{R}^d} |y|^\alpha |h(2^j y)|\mathrm{d}y \leqslant C 2^{-j\alpha}[u]_\alpha.$$

$\square$

Lemma C.2. *There is a constant $C = C(d, \alpha)$, such that for any $u \in C^{2,\alpha}$ and $\lambda \geqslant 0$,*

$$\lambda \|u\|_0 \leqslant C \|\lambda u - \Delta u\|_0 \tag{C.3}$$

and

$$\lambda [u]_\alpha + [\nabla^2 u]_\alpha \leqslant C [\lambda u - \Delta u]_\alpha \tag{C.4}$$

Proof. 1) If there exits $x_0 \in \mathbb{R}^d$ such that $u(x_0)($ or $- u(x_0)) = \|u\|_0$, then $\Delta u(x_0) \leqslant 0$ $(\Delta u(x_0) \geqslant 0)$. This implies $|\lambda u(x_0)| \leqslant |\lambda u(x_0) - \Delta u(x_0)|$; If such x_0 does not exist, then we can consider function $u_R = u \chi(\cdot / R)$ $(R \gg 1)$.

2) We only prove the case $\lambda = 0$. Let $f = \Delta u$. Define

$$\varphi^{kl}(\xi) := \frac{\xi_k \xi_l}{|\xi|^2} \varphi(\xi), \quad h^{kl}(x) := \mathscr{F}^{-1}(\varphi^{kl})(x); \quad \varphi_j^{kl}(\xi) := \varphi^{kl}(2^{-j}\xi), \quad h_j^{kl}(x) := 2^{jd} h^{kl}(2^j x).$$

It is easy to see

$$\partial_{kl} u = \sum_{j \in \mathbb{Z}} u_j^{kl} := \sum_{j \in \mathbb{Z}} \varphi_j^{kl}(D) f = \sum_{j \in \mathbb{Z}} h_j^{kl} * f.$$

For any $x \in \mathbb{R}^d$, noticing $h^{kl} \in \mathscr{S}(\mathbb{R}^d)$ and $\int h^{kl} = \varphi(0) = 0$, we get

$$|u_j^{kl}(x)| = \left| \int_{\mathbb{R}^d} h_j^{kl}(y) f(x - y) \mathrm{d}y \right| = \left| \int_{\mathbb{R}^d} h^{kl}(z) (f(x - 2^{-j}z) - f(x)) \mathrm{d}z \right|$$

$$\leqslant \int_{\mathbb{R}^d} |h^{kl}(z)| \cdot [f]_\alpha |2^{-j}z|^\alpha \mathrm{d}z \leqslant C[f]_\alpha 2^{-j\alpha}$$

This together with Lemma C.1 yields that

$$[\nabla^2 u]_\alpha \leqslant C \sup_{j,k,l} 2^{j\alpha} \|\Delta_j \partial_{kl} u\|_0 \leqslant C[f]_\alpha.$$

$\square$

References

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Eng91] Hans-Jürgen Engelbert. On the theorem of T. Yamada and S. Watanabe. *Stochastics: An International Journal of Probability and Stochastic Processes*, 36(3-4):205–216, 1991.

[FGP10] Franco Flandoli, Massimiliano Gubinelli, and Enrico Priola. Well-posedness of the transport equation by stochastic perturbation. *Inventiones mathematicae*, 180(1):1–53, 2010.

[Hua01] Zhiyuan Huang. *Fundamentals of stochastic analysis, version 2 (in Chinese)*. Science Press, 2001.

[IW14]   Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24. Elsevier, 2014.

[Kry21] N. V. Krylov. On strong solutions of Itô's equations with $\sigma \in W_d^1$ and $b \in L_d$. *Ann. Probab.*, 49(6):3142–3167, 2021.

[RY13]   Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.