

[SDE II] Introduction to Itô Integrals and Numerical Methods

Anqiao Ouyang, Xuecheng Liu

August 29, 2025

1 From ODEs to SDEs

1.1 Deterministic Systems

In classical differential equation theory, we often consider deterministic dynamical systems of the form

$$\frac{dX_t}{dt} = f(X_t, t), \quad X(0) = x_0$$

Here X_t describes the state evolving over time, and f is the drift term of the system, which determines the deterministic trajectory. Under this setting, given the initial value x_0 , the solution X_t is uniquely determined by f .

However, in many practical problems, the system is affected by noise, environmental disturbances, or uncontrollable factors, and a pure ODE is insufficient to capture these uncertainties.

1.2 From Determinism to Stochastic Dynamics

Within the framework of classical ordinary differential equations (ODEs), the evolution of the system is fully governed by a deterministic drift term. For example,

$$dX_t = f(X_t, t) dt, \quad X(0) = x_0,$$

its solution X_t is uniquely determined as a function once the initial value is specified. Such a model is suitable for describing idealized, noise-free systems.

In practice, however, systems are often subject to uncontrollable external factors or microscopic uncertainties, and a pure ODE cannot capture such “unpredictable fluctuations.” Therefore, it is necessary to introduce stochastic perturbations on top of deterministic dynamics.

A natural approach is to add a noise-driven term to the drift term of the ODE. For example,

$$dX_t = f(X_t, t) dt + \sigma dB_t,$$

where B_t is a standard Brownian motion (a continuous martingale process with mean zero and variance t), and $\sigma > 0$ denotes the noise intensity. The additional term σdB_t describes the persistent random disturbances to the trajectory on top of the average trend f . Intuitively, the drift term determines the “overall trend,” while the Brownian noise determines the “local fluctuations.”

More generally, the strength of the noise itself may depend on the current state and time. In that case, one writes

$$dX_t = f(X_t, t) dt + g(X_t, t) dB_t,$$

where $g(X_t, t)$ is called the **diffusion coefficient**, which modulates the amplitude of the stochastic disturbance. This leads to the most common form of the **stochastic differential equation** (SDE).

From a mathematical perspective, the key difference between SDEs and ODEs lies in the nature of their solutions: the solution X_t to an ODE is a deterministic function, while the solution to an SDE is a stochastic process defined on a probability space. Here $f(X_t, t)$ is still the “drift term,” determining the average direction of the system, while $g(X_t, t) dB_t$ is the “diffusion term,” which introduces non-smooth, uncertain path structures through Brownian motion. Since Brownian paths are almost surely nowhere differentiable, the term $g(X_t, t) dB_t$ must be interpreted within the framework of Itô or Stratonovich integration.

SDEs not only generalize ODEs in form, but also fundamentally make the solution a family of trajectories in a probabilistic sense. This transition requires stochastic integration theory (Itô integrals, Itô’s lemma, etc.) to be rigorously defined and analyzed.

2 Itô Integral and Itô’s Lemma

2.1 Non-differentiability of Brownian Motion and the Integration Problem

One can prove the non-differentiability of Brownian motion paths by showing the following probability result. Consider the right-hand derivative:

$$P(\{\lim_{t \rightarrow 0^+} \frac{B_{t+h} - B_t}{h} \text{ exists}\}) = 0$$

However, directly describing the event $\{\lim_{t \rightarrow 0^+} \frac{B_{t+h} - B_t}{h} \text{ exists}\}$ is very difficult. A common trick in probability theory is that if we want to prove $P(A) = 0$ but the event A is hard to handle, we relax it to a slightly larger event $B \supset A$, and then prove $P(B) = 0$. From analysis, when limits are difficult, one may first look at boundedness. Having boundedness is easier than having a limit. First consider the absence of a right-hand derivative at $t = 0$.

If a function $f(t)$ is differentiable at 0, then

$$\exists n, k \in \mathbb{N}_+, \forall t \in [0, 1/k), \text{ s.t. } |B_t| \leq nt$$

Fixing n, k , denote the event $A(n, k) = \{\forall t \in [0, 1/k), |B_t| \leq nt\}$. Then we must have

$$\{\lim_{h \rightarrow 0^+} \frac{B_{t+h} - B_t}{h} \text{ exists}\} \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A(n, k)$$

However, $A(n, k)$ may not be measurable. Thus, we need to find an even larger null set containing it. To relax the condition, one can proceed cleverly: $A(n, k)$ requires the property for all $t < 1/k$. We can weaken this to a countable or even finite set.

First, by using the countability of rational numbers, we can restrict to $t \in \mathbb{Q} \cap [0, 1/k)$. Then,

$$A(n, k) \subset C(n, k) = \{\forall t \in [0, 1/k) \cap \mathbb{Q}, |B_t| \leq nt\} = \bigcap_{p \in \mathbb{Q} \cap [0, 1/k)} \{|B_p| \leq np\}$$

It is easy to calculate that $P(|B_p| \leq np) \leq \sqrt{(1 - e^{-n^2 p})/2\pi}$, hence $P(C(n, k)) = 0$.

By properties of Brownian motion, we know it has no right-hand derivative at any point in $[0, 1)$. But this does not imply that it is almost surely nowhere differentiable on $[0, 1)$. The essential reason is that

$$\forall t \in [0, 1], P(A_t) = 0 \not\Rightarrow P\left(\bigcup_{t \in [0, 1]} A_t\right) = 0.$$

This only holds when the index set is countable. A simple example: if X is a continuous random variable, then $P(X = x) = 0$ for all $x \in \mathbb{R}$; but $P(X \in \mathbb{R}) = 1$.

Thus, we must focus on the entire interval $[0, 1)$. Redefine

$$A(n, k) = \{\exists t \in [0, 1), \forall h \in [0, 1/k), |B_{t+h} - B_t| \leq nh\}.$$

Again, we need to relax the restriction and find $B(n, k) \supset A(n, k)$. The difficulty is that measure theory has poor properties with continuum indices (like $t \in [0, 1)$). We continue by discretizing the parameter:

- Partition $[0, 1)$ into intervals. Then for any $t \in [0, 1)$, t lies in some $[t_j, t_{j+1})$. If the partition is fine enough, then for $\omega \in A(n, k)$ we must have

$$|B_{t_{j+2}}(\omega) - B_{t_{j+1}}(\omega)| \leq |B_{t_{j+1}}(\omega) - B_{t_j}(\omega)| + |B_{t_{j+2}}(\omega) - B_{t_j}(\omega)| \leq n(t_{j+1} + t_{j+2} - 2t_j).$$

For equal partitions with spacing Δ , we have

$$|B_{t_{j+2}}(\omega) - B_{t_{j+1}}(\omega)| \leq 3n\Delta.$$

- Suppose $[0, 1)$ is divided into m equal parts. From above, it suffices to have $1/m < 1/2k$. Define

$$C(m, n, k) = \bigcup_{j=0}^{m-1} \{|B_{t_{j+1}} - B_{t_j}| \leq 3n/m\}.$$

- But $C(m, n, k)$ is not yet a null set. Still, for all $m \geq 3k$, we have $A(n, k) \subset C(m, n, k)$. Taking the intersection over all such m , we are likely to obtain a null set. Let

$$B(n, k) = \bigcap_{m \geq 3k} C(m, n, k).$$

It is not hard to prove this is a null set.

- In summary,

$$A(n, k) \subset C(m, n, k) \Rightarrow A(n, k) \subset B(n, k) \Rightarrow \bigcup_n \bigcup_k A(n, k) \subset \bigcup_n \bigcup_k B(n, k),$$

and since $P(B(n, k)) = 0$ for all n, k , the path of B_t is almost surely non-differentiable.

2.2 Definition of the Itô Integral

2.2.1 Itô Integral of Predictable Simple Processes

Definition of a predictable simple process:

Definition 2.1. A stochastic process $H = \{H_t\}$ is called a predictable simple process if there exists a partition of the interval $[0, T]$

$$0 = t_0 < t_1 < \dots < t_N = T$$

such that H is constant on each interval $t_n < t \leq t_{n+1}$. That is, there exists a sequence of random variables $\{H_n\}_{n=1}^N$ such that

$$H_t = \sum_{n=0}^{N-1} H_n \mathbf{1}_{(t_n, t_{n+1}]}(t) + H_0 \mathbf{1}_{\{0\}}(t),$$

where $E(|H_n|^2) < \infty$, H_t is \mathcal{F}_{t_n} -measurable for $t \in (t_n, t_{n+1}]$, and H_0 is \mathcal{F}_0 -measurable.

Intuitively, the trajectory of such a process is a step function. Predictability means it is right-continuous with left limits. We can then define the Itô integral for simple processes:

Definition 2.2 (Itô Integral for Simple Processes). Let H be an adapted simple process. Define

$$I_T(H) = \sum_{n=1}^N H_n \Delta B_{t_n}.$$

Then $I_T(H)$ is called the Itô integral with respect to H , denoted by

$$\int_0^T H_t dB_t.$$

The construction of the Itô integral is similar to that of the Lebesgue integral: first define the integral for a class of simple objects, then extend it to a broader class of functions. The next step is to introduce several properties of $I_t(H)$: linearity, additivity over intervals, adaptedness, and the martingale property.

More importantly, the following two properties hold:

1. Zero mean:

$$E\left(\int_0^T H_t dB_t\right) = 0.$$

2. Isometry:

$$\mathbb{E}\left[\left(\int_0^T H_t dB_t\right)^2\right] = \mathbb{E}\int_0^T |H_t|^2 dt.$$

The proofs are straightforward once one knows how to compute the expectation and variance of the normal distribution and handle independence, and are omitted here.

2.2.2 Itô Integral on L_T^2

To extend the Itô integral to a larger space, we define the space $L_T^2(\Omega)$:

$$L_T^2(\Omega) = \left\{ X = \{X_t\}_{0 \leq t \leq T} : \begin{aligned} &X(t, \omega) \text{ measurable with respect to } \mathcal{B}(0, T) \otimes \mathcal{F}, \\ &X_t \text{ measurable with respect to } \mathcal{F}_t, \\ &E\left(\int_0^T |X_t|^2 dt\right) < \infty \end{aligned} \right\}.$$

Some sources also require X to be **progressively measurable**. Define the norm

$$\|X\|_{L_T^2} = \sqrt{E\left(\int_0^T |X_t|^2 dt\right)}.$$

This is similar to $L^2([0, T] \times \Omega)$. As expected, simple processes are included in $L_T^2(\Omega)$. One can show that $L_T^2(\Omega)$ is a **Banach** space. With the inner product

$$\langle X, Y \rangle = \mathbb{E} \int_0^T X_t Y_t dt,$$

L_T^2 becomes a **Hilbert** space. By standard arguments in analysis, simple processes are dense in $L_T^2(\Omega)$:

Theorem 2.3. *For any stochastic process $X \in L_T^2(\Omega)$ and any $\epsilon > 0$, there exists a simple process H such that*

$$\|X - H\|_{L_T^2} < \epsilon.$$

Sketch of proof in three steps:

1. Assume $X(t, \omega)$ is continuous in t and bounded. Use the dominated convergence theorem and continuity to prove density.
2. Assume $X(t, \omega)$ is bounded but not necessarily continuous. To use the continuous case, apply **convolution** to smooth it. Choose a family of kernel functions $K_\lambda(t)$, and define

$$(X * K_\lambda)(t) = \int_0^t K_\lambda(t-s) X(s) ds.$$

3. For general $X \in L_T^2$, truncate X to obtain $X^{(n)}$, apply the result from 2, and then approximate X .

Thus, by property 2, $I_t(H)$ is a linear isometry from $\{H \text{ simple} : H \in L_T^2\}$ to $L^2(\Omega)$, and hence a bounded linear operator. If $H^{(n)}$ is a Cauchy sequence in L_T^2 , then $I_t(H^{(n)})$ is a Cauchy sequence in $L^2(\Omega)$. Therefore, we can define the Itô integral on $L_T^2(\Omega)$:

Definition 2.4. For $X \in L_T^2(\Omega)$, there exists a sequence of simple processes $H^{(n)}$ such that $H^{(n)} \rightarrow X$ in the L_T^2 norm. Define the Itô integral of X by

$$\int_0^T X_s dB_s = \lim_n \int_0^T H_t^{(n)} dB_t,$$

where the limit is taken in the L^2 sense.

The extended Itô integral still satisfies linearity, additivity over intervals, adaptedness, and the martingale property, as well as the following formulas:

1. Zero mean:

$$E \left(\int_0^T X_t dB_t \right) = 0.$$

2. Isometry:

$$E \left(\int_0^T X_t dB_t \right)^2 = \int_0^T E |X_t|^2 dt.$$

3. Covariance formula:

$$E \left(\int_0^T X_t dB_t \int_0^T Y_t dB_t \right) = \int_0^T E(X_t Y_t) dt.$$

These follow directly from the continuity of approximation by simple processes.

2.2.3 Quadratic Variation

Since Brownian motion paths are continuous everywhere but nowhere differentiable, they are not of bounded variation on $[0, T]$. However, they do have finite quadratic variation.

Lemma 2.5 (Quadratic Variation of Brownian Motion).

$$[B, B]_t = t.$$

This proof is straightforward. Motivated by this, we can define the quadratic covariation of any two processes on $[0, T]$. Let $t_0 = 0, t_N = T, \Delta = \max_n \{\Delta t_n\}$.

Definition 2.6. For processes X, Y , if the following limit converges in probability:

$$\lim_{\Delta \rightarrow 0} \sum_{n=0}^{N-1} (Y_{t_{n+1}} - Y_{t_n})(X_{t_{n+1}} - X_{t_n}),$$

then the limiting process is called the quadratic covariation of X, Y on $[0, T]$, denoted by $[X, Y]_T$.

If we take $X = B$ (Brownian motion) and $Y = t$ (a deterministic function), then

$$\lim_{\Delta \rightarrow 0} \sum_{n=1}^{N-1} (B_{t_{n+1}} - B_{t_n})(t_{n+1} - t_n) = 0.$$

The limit converges in the L^2 norm, hence also in probability. In fact, the variance of the limit is of the same order as $\lim \sum_n (t_{n+1} - t_n)^{3/2}$, i.e., order $1/2$.

We are also interested in the quadratic variation of Itô integrals. On $[0, t], t \leq T$, let $I_t = \{\int_0^t X_s dB_s\}_{0 \leq t \leq T}$ be viewed as a stochastic process. For simple processes, its quadratic variation is

$$[I, I]_T = \lim \sum_{n=0}^{N-1} \left(\int_{t_n}^{t_{n+1}} X_s dB_s \right)^2 = \lim \sum_{n=0}^{N-1} X_n^2 (B_{t_{n+1}} - B_{t_n})^2 = \int_0^T |X_s|^2 ds.$$

By approximating general processes with simple ones, the same expression is obtained. Again, convergence is in probability.

2.3 Itô's Lemma and Examples

2.3.1 Itô's Formula for Brownian Motion

Assume f is twice continuously differentiable; then $f(B_T) - f(B_0)$ can be expanded by **Taylor**. To handle this more precisely, partition $[0, T]$:

$$f(B_T) - f(B_0) = \sum_{n=1}^N (f(B_{t_n}) - f(B_{t_{n-1}})).$$

Isolate one summand and apply the **Taylor** expansion; write $\Delta B_{n-1} = B_{t_n} - B_{t_{n-1}}$:

$$f(B_{t_n}) - f(B_{t_{n-1}}) = f'(B_{t_{n-1}}) \Delta B_{n-1} + \frac{1}{2} f''(\xi_n) (\Delta B_{n-1})^2,$$

where $\xi_n = B_{t_{n-1}} + \theta \Delta B_{n-1}$, $\theta \in (0, 1)$. Hence,

$$f(B_T) - f(B_0) = \sum_n f'(B_{t_{n-1}}) \Delta B_{n-1} + \frac{1}{2} \sum_n f''(\xi_n) (\Delta B_{n-1})^2.$$

The first term on the right converges in L^2 to the Itô integral $\int_0^T f'(B_t) dB_t$; the second term converges **in probability** to $\frac{1}{2} \int_0^T f''(B_t) dt$.

Theorem 2.7. Assume $f \in C^2$ and $E \int_0^T |f'(B_t)|^2 dt < \infty$, $E \int_0^T |f''(B_t)|^2 dt < \infty$. Then for any $T \geq 0$,

$$f(B_T) - f(B_0) = \int_0^T f'(B_t) dB_t + \frac{1}{2} \int_0^T f''(B_t) dt.$$

Proof. We only prove the second-order term $\int_0^T f''(B_t) dt$. That is, we need to show

$$\sum_n f''(\xi_n) (\Delta B_{n-1})^2 \xrightarrow{P} \int_0^T f''(B_t) dt.$$

1. First show

$$\sum_n f''(\xi_n) (\Delta B_{n-1})^2 \stackrel{\text{a.s.}}{\sim} \sum_n f''(B_{n-1}) (\Delta B_{n-1})^2,$$

where \sim means the difference of the two sides is infinitesimal in a certain sense (L^2 , a.s., or in probability).

2. Then show

$$\sum_n f''(B_{n-1}) (\Delta B_{n-1})^2 \stackrel{L^2}{\sim} \sum_n f''(B_{n-1}) \Delta t_{n-1}.$$

3. Next, prove

$$\sum_n f''(B_{n-1}) \Delta t_{n-1} \xrightarrow{\text{a.s.}} \int_0^T f''(B_t) dt.$$

4. Finally, note that convergence in L^2 or almost surely implies convergence in probability. ■

2.3.2 Itô Processes and Their Itô Formula

Definition 2.8 (Itô Process).

$$X_t = X_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dB_s \quad (1)$$

where $\mu, \sigma \in L_T^2(\Omega)$.

Similarly, consider a C^2 function $u(t, x)$ in two variables, and examine the difference over $[t_n, t_{n+1}]$:

$$\begin{aligned} u(t_{n+1}, X_{n+1}) - u(t_n, X_n) &= \partial_t u(t_n, X_n) \Delta t_n + \partial_x u(t_n, X_n) \Delta X_n \\ &\quad + \frac{1}{2} \partial_{xx}^2 u(\lambda_n, \xi_n) (\Delta X_n)^2 \\ &\quad + \frac{1}{2} \partial_{tt}^2 u(\lambda_n, \xi_n) (\Delta t_n)^2 \\ &\quad + \partial_{tx}^2 u(\lambda_n, \xi_n) \Delta t_n \Delta X_n, \end{aligned}$$

where $(\lambda_n, \xi_n) = (t_n, X_n) + \theta(\Delta t_n, \Delta X_n)$ with $\theta \in (0, 1)$. We find that the ∂_{tx}^2 and ∂_{tt}^2 terms should converge to 0 in probability. Moreover, ΔX_n and $(\Delta X_n)^2$ are of interest. From the difference viewpoint, as $\Delta t_n \rightarrow 0$ we heuristically have

$$\Delta X_n = \int_{t_n}^{t_{n+1}} \mu(s) ds + \int_{t_n}^{t_{n+1}} \sigma(s) dB_s \approx \mu_n \Delta t_n + \sigma_n \Delta B_n,$$

where μ_n, σ_n abbreviate $\mu(t_n), \sigma(t_n)$. For the quadratic term,

$$(\Delta X_n)^2 \approx \mu_n^2 (\Delta t_n)^2 + \sigma_n^2 (\Delta B_n)^2 + \mu_n \sigma_n \Delta t_n \Delta B_n.$$

To clarify the picture, ignore the higher-order terms $(\Delta t_n)^2$ and $\Delta t_n \Delta B_n$, which vanish in the limit. Thus,

$$\begin{aligned} \Delta u(t_n, X_n) &\approx (\partial_t + \mu_n \partial_x) u(t_n, X_n) \Delta t_n \\ &\quad + \frac{1}{2} \sigma_n^2 \partial_{xx}^2 u(\lambda_n, \xi_n) (\Delta B_n)^2 \\ &\quad + \partial_x u(t_n, X_n) \Delta B_n. \end{aligned}$$

Theorem 2.9 (Itô's Formula for Itô Processes).

$$\begin{aligned} u(T, X_T) &= u(0, X_0) + \int_0^T \partial_t u(s, X_s) ds + \int_0^T \partial_x u(s, X_s) dX_s + \frac{1}{2} \int_0^T \partial_{xx}^2 u(s, X_s) (dX_s)^2 \\ &= u(0, X_0) + \int_0^T \left(\partial_t u(s, X_s) + \mu(s) \partial_x u(s, X_s) + \frac{1}{2} \sigma^2(s) \partial_{xx}^2 u(s, X_s) \right) ds \\ &\quad + \int_0^T \sigma(s) \partial_x u(s, X_s) dB_s \end{aligned}$$

The second equality has already been clarified by the discussion above; the key point is the first equality, which defines stochastic integrals with respect to X_t and its quadratic variation. If we ignore the first term in (1), i.e., the one with μ_s , then X_t is a stochastic integral with respect to Brownian motion. The term μ is commonly called the drift. By properties of the Itô integral, this yields a martingale. Since $\sigma_s \in L_T^2$, it is a **square-integrable martingale**.

Definition 2.10. A process X on $[0, T]$ is called a square-integrable martingale if it is a martingale and

$$\sup_{t \leq T} E|X_t|^2 < +\infty.$$

Consider first the case $\mu_s = 0$, i.e., $X_t = \int_0^t \sigma(s) dB_s$. Its quadratic variation is $[X, X]_t = \int_0^t |\sigma_s|^2 ds$, abbreviated $[X]_t$. It is easy to see that this process has sample paths that are monotonically increasing; moreover, it is continuous with bounded variation. We call such a process a **nondecreasing process**.

Definition 2.11 (Nondecreasing Process). A process F on $[0, T]$ is called nondecreasing if its sample paths are almost surely nondecreasing and right-continuous.

For a nondecreasing process, one can define the Riemann–Stieltjes integral on each path. The following limit converges a.s.:

$$\lim_{\Delta \rightarrow 0} \sum_n g(\xi_n, \omega) \Delta F_n(\omega),$$

provided the process g satisfies suitable conditions. We denote the limit by $\int_0^T g_t dF_t$. In our discussion limited to Itô's formula (2.9), take $g = \sigma^2$ and $F_t = [X]_t = \int_0^t \sigma_t^2 dt$. Since such an integral exists, we also consider its differential, denoted $d[X]_t$, or heuristically $(dX_t)^2$. Formally, $d[X]_t = \sigma_t^2 dt$. That is, we need to verify

$$\int_0^T Y_t d[X]_t = \int_0^T Y_t \sigma_t^2 dt$$

for any integrable Y_t . The proof is straightforward and omitted.

On the other hand, note that X has continuous sample paths. Analogously to the Itô integral, we can define stochastic integrals with respect to continuous square-integrable martingales. Let

M denote such a process. Define the process space

$$\mathcal{M}_T^2(\Omega) = \left\{ Y = \{Y_t\}_{0 \leq t \leq T} : Y(t, \omega) \text{ measurable with respect to } \mathcal{B}(0, T) \otimes \mathcal{F}, \right. \\ \left. Y_t \text{ measurable with respect to } \mathcal{F}_t, \right. \\ \left. E\left(\int_0^T |Y_t|^2 d[M]_t\right) < \infty \right\},$$

with Y progressively measurable. The construction details are analogous to those for the Itô integral. When $M_t = X_t = \int_0^t \sigma_s dB_s$, it is easy to show

$$\int_0^T Y_t dX_t = \int_0^T Y_t \sigma_t dB_t.$$

When the drift is nonzero, X need not be a square-integrable martingale, or even a martingale at all. But $\int_0^t \mu_s ds$ is certainly a **bounded-variation process**, i.e., a process with right-continuous paths of bounded variation. Hence X can be written as the sum of a bounded-variation process and a martingale, commonly called a semimartingale. One can then define stochastic integrals with respect to semimartingales, but we will not elaborate here. It suffices to note that

$$\int_0^T Y_t dX_t = \int_0^T Y_t \mu_t dt + \int_0^T Y_t \sigma_t dB_t.$$

3 Basic Solution Methods for SDEs

3.1 Itô Integral Equation Form

$$X_t = X_s + \int_s^t a(u, X_u) du + \int_s^t b(u, X_u) dB_u$$

3.2 Euler–Maruyama Discretization Method

The stochastic differential equation

$$dX_t = a(X_t) dt + b(X_t) dB_t \quad (2)$$

Itô formula:

$$df(X_t) = f'(X_t) dX_t + \frac{1}{2} f''(X_t) (dX_t)^2$$

with $(dX_t)^2 = b^2(X_t) dt$. Furthermore,

$$df(X_t) = \left(a(X_t) f'(X_t) + \frac{1}{2} b^2(X_t) f''(X_t) \right) dt + f'(X_t) b(X_t) dB_t \quad (3)$$

The full integral form of 2 is

$$X_t - X_s = \int_s^t a(X_\tau) d\tau + \int_s^t b(X_\tau) dB_\tau.$$

Let $s = t_n$, $t = t_{n+1}$, then

$$X_{t_{n+1}} - X_{t_n} = a(X_{t_n}) \Delta t_n + b(X_{t_n}) \Delta B_{t_n} \\ + \int_{t_n}^{t_{n+1}} [a(X_\tau) - a(X_{t_n})] d\tau + \int_{t_n}^{t_{n+1}} [b(X_\tau) - b(X_{t_n})] dB_\tau \quad (4)$$

Apply 3 to $a(X_\tau) - a(X_{t_n})$ and $b(X_\tau) - b(X_{t_n})$. Of course, a, b need to satisfy smoothness conditions. Write the operators $L_1 f(x) = a(x)f'(x) + \frac{1}{2}b(x)f''(x)$ and $L_2 f(x) = b(x)f'(x)$. The two formulas above can be written as

$$a(X_\tau) - a(X_{t_n}) = \int_{t_n}^{\tau} L_1 a(X_s) ds + \int_{t_n}^{\tau} L_2 a(X_s) dB_s \quad (5)$$

$$b(X_\tau) - b(X_{t_n}) = \int_{t_n}^{\tau} L_1 b(X_s) ds + \int_{t_n}^{\tau} L_2 b(X_s) dB_s \quad (6)$$

It is not hard to see that L_1, L_2 are both linear. Let the sum of the last two terms in 4 be R , then

$$\begin{aligned} R(N, X) &= \int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} L_1 a(X_s) ds + \int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} L_2 a(X_s) dB_s \\ &+ \int_{t_n}^{t_{n+1}} dB_\tau \int_{t_n}^{\tau} L_1 b(X_s) ds + \int_{t_n}^{t_{n+1}} dB_\tau \int_{t_n}^{\tau} L_2 b(X_s) dB_s \end{aligned} \quad (7)$$

We may regard the term R as an error term. Then 4 can be abbreviated as

$$\Delta X_{t_n} = a(X_{t_n})\Delta t_n + b(X_{t_n})\Delta B_{t_n} + R(N, X) \quad (8)$$

By analogy with converting an **ODE** to a difference equation, to numerically solve 2 we may consider its difference form

$$\Delta X_{t_n} = a(X_{t_n})\Delta t_n + b(X_{t_n})\Delta B_{t_n}.$$

Notice that if $R(N, X) = 0$ in 8, then it has exactly the **difference** form. Thus $R(N, X)$ can be viewed as an error term. As long as $R(N, X) \rightarrow 0$ when $\Delta \rightarrow 0$, then under certain conditions (e.g., **Lipschitz** conditions on a, b), the solution of the difference equation may gradually approach the true solution. In fact, this is the **Euler–Maruyama scheme**:

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a(\bar{X}_{t_n})\Delta_n + b(\bar{X}_{t_n})\Delta B_{t_n} \quad (9)$$

In practice, the first two conditions are used to control inequalities during the proof. Earlier, when discussing the **EM scheme**, we considered an equal partition of $[0, T]$. Here we generalize and allow a general partition, and denote $\delta = \max_n |\Delta_n|$.

Since $R(N, X)$ is an error term, we should estimate its order. Assume an equal partition of $[0, T]$, i.e., $\Delta t_n = \Delta = T/N$. Then

$$\int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} L_1 a(X_s) ds \approx \int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} L_1 a(X_{t_n}) ds \approx \frac{1}{2} L_1 a(X_{t_n}) \Delta^2,$$

and similarly

$$\int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} L_2 a(X_s) dB_s \approx \frac{1}{2} L_1 a(X_{t_n}) \Delta^{3/2} \quad (10)$$

$$\int_{t_n}^{t_{n+1}} dB_\tau \int_{t_n}^{\tau} L_1 b(X_s) ds \approx \frac{1}{2} L_1 b(X_{t_n}) \Delta^{3/2},$$

$$\int_{t_n}^{t_{n+1}} dB_\tau \int_{t_n}^{\tau} L_2 b(X_s) dB_s \approx \frac{1}{2} L_2 b(X_{t_n}) (B_\Delta^2 - \Delta) \approx \frac{1}{2} L_2 b(X_{t_n}) \Delta.$$

See details in A. Note that the leading term among the orders above is $O(\Delta)$, corresponding to the term $\int dB_\tau \int dB_s$. Although the **E–M scheme** is simple in form, its convergence is somewhat weaker. To improve the convergence of the **E–M scheme**, keeping the $O(\Delta)$ term yields the **Milstein scheme**.

3.3 Milstein Scheme

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a(\bar{X}_{t_n})\Delta_n + b(\bar{X}_{t_n})\Delta B_{t_n} + \frac{1}{2}L_2b(X_{t_n})(B_\Delta^2 - \Delta)$$

3.3.1 Higher Order

Recall how we obtained the **Milstein scheme**: apply 5 and 6 to 4. This resembles a **Taylor expansion**. Continuing to expand 7 using 5 and 6 and then neglecting higher-order terms, we obtain

$$\begin{aligned} R(\Delta, X) &= \sum_{i_1, i_2} L_{i_1} a_{i_1}(X_{t_n}) \Delta_{i_1, i_2} \\ &+ \sum_{i_1, i_2, i_3} \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} \int_{t_n}^{s_2} L_{i_1} L_{i_2} a_{i_3}(X_{s_3}) dB_{s_1}^{i_1} dB_{s_2}^{i_2} dB_{s_3}^{i_3} \end{aligned}$$

where $i_1, i_2, i_3 = 0, 1$, $a_1, a_2 = a, b$, and $\Delta_{i_1, i_2} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dB_{s_1}^{i_1} dB_{s_2}^{i_2}$. Moreover,

$$dB_{s_j}^{i_j} = \begin{cases} ds_j, & i_j = 0, \\ dB_{s_j}, & i_j = 1, \end{cases} \quad j = 1, 2, 3.$$

Thus we obtain a higher-order **E–M scheme**:

$$\begin{aligned} \bar{X}_{t_{n+1}} &= \bar{X}_{t_n} + a(\bar{X}_{t_n})\Delta_n + b(\bar{X}_{t_n})\Delta B_{t_n} \\ &+ \sum_{i_1, i_2} L_{i_1} a_{i_1}(\bar{X}_{t_n}) \Delta_{i_1, i_2}. \end{aligned}$$

3.3.2 First-Order Derivative-Free Schemes

Note that in the **Milstein scheme** the term $L_2b(X_{t_n})$ involves derivatives. The goal here is to avoid derivatives. Since $L_2b(X_{t_n}) = bb'(X_{t_n})$, consider

$$\tilde{X}_{t_{n+1}} = X_{t_n} + a(X_{t_n})\Delta + b(X_{t_n})\sqrt{\Delta}.$$

The reason for this form is a crude but practical estimate: $\Delta B_{t_n} \approx \sqrt{\Delta t_n}$. Using an ordinary (non-stochastic) Taylor expansion, we see that

$$\frac{1}{\sqrt{\Delta}}(b(\tilde{X}_{t_{n+1}}) - b(X_{t_{n+1}}))$$

approximates $bb'(X_{t_n})$. Therefore, we obtain the **first-order derivative-free schemes**:

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a(\bar{X}_{t_n})\Delta_n + b(\bar{X}_{t_n})\Delta B_{t_n} + \frac{1}{2} \frac{1}{\sqrt{\Delta}}(b(\tilde{X}_{t_{n+1}}) - b(X_{t_{n+1}}))(B_\Delta^2 - \Delta).$$

3.4 Convergence

3.4.1 Strong vs. Weak Convergence

Definitions

$$\epsilon_t(\delta) = \sup_{0 \leq s \leq t} E|\bar{X}_s^\delta - X_s|$$

$$r_t(\delta) = \sup_{0 \leq s \leq t} |E(\phi(\bar{X}_s^\delta)) - E(\phi_s(X_s))|$$

where $\phi \in C_b^\infty$, i.e., a bounded smooth function whose derivatives of all orders are bounded. $\epsilon_t^\delta, r_t^\delta$ give rise to the notions of strong and weak convergence, respectively.

Definition 3.1 (Strong convergence). If there exist $\alpha, \delta_0, C > 0$ such that for any $0 < \delta < \delta_0$,

$$\epsilon_t(\delta) \leq C\delta^\alpha,$$

then \bar{X}^δ is said to converge strongly to X with order α .

Definition 3.2 (Weak convergence). If there exist $\alpha, \delta_0, C > 0$ such that for any $0 < \delta < \delta_0$,

$$r_t(\delta) \leq C\delta^\alpha, \quad \forall \phi \in C_b^\infty,$$

then \bar{X}^δ is said to converge weakly to X .

Below we only prove the half-order strong convergence and first-order weak convergence of the **Euler–Maruyama scheme**.

3.4.2 Strong Convergence of the EM Scheme

To discuss convergence, we need to strengthen some assumptions. Since symbols were not specified in detail earlier, we now give explicit statements.

- Condition 1 (Lipschitz): $a(t, x), b(t, x)$ are **Lipschitz** in the second variable, i.e.,

$$|a(t, x) - a(t, y)| \leq L|x - y|.$$

- Condition 2 (linear growth): a, b satisfy a linear growth condition in the second variable, i.e.,

$$|a(t, x)|^2 + |b(t, x)|^2 \leq C(1 + |x|^2).$$

- Condition 3 (L^2): a, b are in L^2 .
- Condition 4 (initial value): X_0 is \mathcal{F}_0 -measurable and is in L^2 ($E(\| \cdot \|^2) < \infty$).

See B for detailed statements. The above four conditions are typically necessary, and what follows assumes these conditions hold. We will not repeat them later.

3.4.3 Weak Convergence of the EM Scheme

Before weak convergence, we introduce the stochastic solution of a parabolic operator, which will be used in the proof of weak convergence.

Definition 3.3 (Parabolic operator).

$$\partial_t + \mathcal{L} = \partial_t + a\partial_x + \frac{1}{2}b^2\partial_x^2.$$

Definition 3.4 (Terminal-value parabolic PDE (deterministic)). Assume g is bounded:

$$\begin{cases} \partial_t u + \mathcal{L}u + gu = 0, \\ u(T, x) = f(x), \end{cases} \quad u : [0, T] \times \mathbb{R}$$

Lemma 3.5. *The stochastic solution to B.2 is*

$$u(t, x) = E \left(f(X_T) \exp \left(\int_s^T g(X_u) du \right) \middle| X_s = x \right),$$

where X_t is Itô.

Let C_P^l denote the space of $2l$ -times continuously differentiable functions with polynomial growth. Polynomial growth means that for any $k \leq l$ there exist C, p such that

$$|\partial_x^k u| \leq C(1 + |x|^{2p}).$$

To make the proof work, we need two lemmas.

Lemma 3.6. *Assume $a, b, f \in C_P^{2l}$ and all derivatives of a, b are uniformly bounded. Then $u \in C_P^{2l}$ and u satisfies the differential equation 3.4.*

Lemma 3.7. *Let $\{Y_n\}_{n=0}^{n_T}$ be the sequence generated by the **EM scheme 9**. Assume $E(|X_0|^{2l}) = E(|Y_0|^{2l}) < \infty$. Then the following moment estimate holds:*

$$\sup_n E(|Y_n|^{2l}) < \infty.$$

In fact, this follows by proving the inequality

$$E(|Y_{n+1}|^{2l}) \leq K_1 E(|Y_n|^{2l}) + K_2 ((\Delta t_n)^{2l} + (\Delta t_n)^l).$$

(BTW, general SDEs admit similar moment estimates.)

The core of the theorem's proof is to use the **Gronwall** inequality and the *Itô* formula. We now state the detailed result of the first-order weak convergence of the **EM scheme**.

Theorem 3.8 (First-order weak convergence of the EM scheme). *Suppose $a, b, f \in C_P^{2l}$ and all their derivatives are uniformly bounded. Then the **EM scheme 9** converges weakly to the true solution with order 1.*

See B.2 for details.

3.5 Ornstein–Uhlenbeck Process

The Ornstein–Uhlenbeck (OU) process is the classic mean-reverting model: the state fluctuates under noise but is pulled back toward a long-term mean μ . Its one-dimensional SDE is

$$dX_t = \theta(\mu - X_t) dt + \sigma dB_t, \quad \theta > 0, \sigma > 0, \quad (11)$$

where θ is the rate of mean reversion, μ is the long-term mean, and σ is the diffusion intensity. This model is equivalent to the Langevin equation in physics under a linear friction potential and is also the unique stationary Gaussian Markov process with an exponential autocovariance kernel. Multiplying (11) by the integrating factor $e^{\theta t}$ and integrating yields the explicit solution

$$X_t = \mu + (X_s - \mu)e^{-\theta(t-s)} + \sigma \int_s^t e^{-\theta(t-u)} dB_u,$$

so that, conditional on $X_s = x$, X_t is normally distributed:

$$\mathbb{E}[X_t | X_s = x] = \mu + (x - \mu)e^{-\theta\Delta}, \quad \text{Var}(X_t | X_s = x) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta}), \quad \Delta = t - s.$$

Thus the transition law is

$$X_t | X_s = x \sim \mathcal{N}\left(\mu + (x - \mu)e^{-\theta\Delta}, \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta})\right).$$

If we take the initial value $X_0 \sim \mathcal{N}(\mu, \sigma^2/(2\theta))$, then X_t is stationary with $\mathbb{E}[X_t] = \mu$, variance $\sigma^2/(2\theta)$, and autocovariance

$$\text{Cov}(X_t, X_{t+\tau}) = \frac{\sigma^2}{2\theta} e^{-\theta\tau}, \quad \rho(\tau) = e^{-\theta\tau}.$$

Hence OU is a Gaussian, Markov, stationary process with exponential autocorrelation, whose unique stationary distribution is $\mathcal{N}(\mu, \sigma^2/(2\theta))$. The infinitesimal generator of OU is

$$\mathcal{L}\varphi(x) = \theta(\mu - x)\varphi'(x) + \frac{\sigma^2}{2}\varphi''(x),$$

and the Fokker–Planck equation is

$$\partial_t p = -\partial_x(\theta(\mu - x)p) + \frac{\sigma^2}{2}\partial_{xx}^2 p,$$

whose unique stationary solution is exactly the normal distribution above. Since the OU equation is linear and solvable, it admits an exact discretization for any step size Δ :

$$X_{t+\Delta} = \mu + (X_t - \mu)e^{-\theta\Delta} + \sigma\sqrt{\frac{1 - e^{-2\theta\Delta}}{2\theta}} Z, \quad Z \sim \mathcal{N}(0, 1),$$

which coincides with the AR(1) form and is thus commonly used for parameter estimation and simulation. In contrast, the Euler–Maruyama discretization

$$X_{n+1}^{\text{EM}} = X_n + \theta(\mu - X_n)\Delta + \sigma\sqrt{\Delta} Z_n$$

is only a strong order-1/2 approximation, and exhibits more bias for large step sizes or strong mean reversion. Finally, by Itô's formula the moment equations close: the mean satisfies $m'(t) = \theta(\mu - m(t))$ with solution $m(t) = \mu + (m(0) - \mu)e^{-\theta t}$; the variance satisfies $v'(t) = -2\theta v(t) + \sigma^2$, and if X_0 is deterministic then $v(t) = \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t})$.

A Rough Order Estimates

The estimate in 10 is slightly more involved. Simply put, it amounts to estimating $\int_{t_n}^{t_{n+1}} d\tau \int_{t_n}^{\tau} dB_s$. In fact, this reduces to estimating

$$D_\Delta = \int_0^\Delta B_t dt$$

where I_Δ is an Itô process. It is easy to obtain $E(I_\Delta) = 0$. Applying the Itô formula, we have

$$\begin{aligned} I_\Delta^2 &= 2 \int_0^\Delta \int_0^t B_u B_t du dt \\ &= 2 \int_0^\Delta \int_0^t B_u (B_t - B_u) du dt + 2 \int_0^\Delta \int_0^t B_u^2 du dt \end{aligned}$$

Therefore

$$E(I_\Delta^2) = 2 \int_0^\Delta \int_0^t E(B_u^2) du dt = \Delta^3/3$$

Alternatively, using stochastic integration by parts,

$$\int_0^\Delta B_t dt + \int_0^\Delta t dB_t = \Delta B_\Delta$$

hence

$$I_{\Delta}^2 = \Delta^2 B_T^2 + \left(\int_0^{\Delta} t dB_t \right)^2 - 2\Delta \int_0^{\Delta} dB_t \int_0^{\Delta} t dB_t$$

so $E(I_{\Delta}^2) = \Delta^3 + \Delta^3/3 - \Delta^3$. The last two terms use properties of Itô integrals. Moreover,

$$\int_0^{\Delta} \int_0^t dB_s dB_t = \int_0^{\Delta} B_t dB_t = \frac{1}{2}(B_{\Delta}^2 - \Delta)$$

One can remember several rough but very practical estimates: $dB_t \approx \sqrt{dt}$, $dt dB_t \approx 0$, $(dB_t)^2 \approx dt$. Here “ ≈ 0 ” holds in the sense of neglecting higher-order terms, and the third is an observation about **quadratic variation**.

B Details of the EM-Scheme Convergence Proof

B.1 Strong Convergence

All prerequisites are in 3.4.2.

$$\begin{aligned} E |\bar{X}_{n_s}^{\delta} - X_s|^2 &= E \left| \sum_{n=0}^{n_s-1} \Delta \bar{X}_{t_n}^{\delta} - \int_0^s a(\tau, X_{\tau}) d\tau - \int_0^s b(\tau, X_{\tau}) dW_{\tau} \right|^2 \\ &= E \left| \int_0^s (a(\tau, \bar{X}_{n_{\tau}}^{\delta}) - a(\tau, X_{\tau})) d\tau + \int_0^s (b(\tau, \bar{X}_{n_{\tau}}^{\delta}) - b(\tau, X_{\tau})) dB_{\tau} \right. \\ &\quad \left. + \int_{n_s}^s a(\tau, X_{\tau}) d\tau + \int_{n_s}^s b(\tau, X_{\tau}) dB_{\tau} \right|^2 \\ &\leq C \left\{ E \left| \int_0^s (a(\tau, \bar{X}_{n_{\tau}}^{\delta}) - a(\tau, X_{\tau})) d\tau \right|^2 \right. \\ &\quad \left. + E \left| \int_0^s (b(\tau, \bar{X}_{n_{\tau}}^{\delta}) - b(\tau, X_{\tau})) dB_{\tau} \right|^2 \right. \\ &\quad \left. + E \left| \int_{n_s}^s a(\tau, X_{\tau}) d\tau \right|^2 + E \left| \int_{n_s}^s b(\tau, X_{\tau}) dB_{\tau} \right|^2 \right\} \end{aligned}$$

Also,

$$\begin{aligned} E \left| \int_0^s (a(\tau, \bar{X}_{n_{\tau}}^{\delta}) - a(\tau, X_{\tau})) d\tau \right|^2 &\leq s E \int_0^s |a(\tau, \bar{X}_{n_{\tau}}^{\delta}) - a(\tau, X_{\tau})|^2 d\tau \\ &\leq s \int_0^s E |\bar{X}_{n_{\tau}}^{\delta} - X_{\tau}|^2 d\tau \end{aligned}$$

$$\begin{aligned} E \left| \int_0^s (b(\tau, \bar{X}_{n_{\tau}}^{\delta}) - b(\tau, X_{\tau})) dB_{\tau} \right|^2 &= E \int_0^s |b(\tau, \bar{X}_{n_{\tau}}^{\delta}) - b(\tau, X_{\tau})|^2 d\tau \\ &\leq \int_0^s E |\bar{X}_{n_{\tau}}^{\delta} - X_{\tau}|^2 d\tau \end{aligned}$$

$$\begin{aligned} E \left| \int_{n_s}^s a(\tau, X_{\tau}) d\tau \right|^2 + E \left| \int_{n_s}^s b(\tau, X_{\tau}) dB_{\tau} \right|^2 &\leq (s - n_s) \left\{ E \int_{n_s}^s |a(\tau, X_{\tau})|^2 + |b(\tau, X_{\tau})|^2 d\tau \right\} \\ &\leq 2(s - n_s) \left\{ E \int_{n_s}^s 1 + |X_{\tau}|^2 d\tau \right\} \end{aligned}$$

Now, collecting the infinitesimal terms and neglecting higher-order terms, we have:

$$E|\bar{X}_{n_s}^\delta - X_s|^2 \leq C_1 \int_0^s E|\bar{X}_{n_\tau}^\delta - X_\tau|^2 d\tau + C_2(s - n_s)$$

By the **GronWall** inequality, we must have

$$E|\bar{X}_{n_s}^\delta - X_s|^2 \leq C_2(s - n_s) + C_1 \int_0^s (\tau - n_\tau) e^{C_1(s-\tau)} d\tau, \forall 0 \leq s \leq T$$

In the above, the integral term is in fact higher order. Neglecting higher-order terms yields $\sup_{0 \leq s \leq T} E|\bar{X}_{n_s}^\delta - X_s|^2 \leq C\delta$. Hence

$$\epsilon_T(\delta) \leq \sqrt{\sup_{0 \leq s \leq T} E|\bar{X}_{n_s}^\delta - X_s|^2} \leq C\delta^{1/2},$$

which proves the half-order strong convergence of the **EM-Scheme**.

B.2 Weak Convergence

For convenience, we simplify notation: Y_n denotes the random variable generated by the **EM scheme** at time t_n , and $B_n = B_{t_n}$. Since $f \in C_P^{2l}$, consider the terminal-value PDE

$$\begin{cases} \partial_t u + \mathcal{L}u = 0 \\ u(T, x) = f(x) \end{cases} \quad u \in C_P^2([0, T] \times \mathbb{R})$$

Then

$$u(t, x) = E(f(X_T) | X_t = x).$$

Therefore,

$$\begin{aligned} E(f(Y_T)) - E(f(X_T)) &= E(u(T, Y_T)) - E(u(0, Y_0)) \\ &= E\left(\sum_{n=0}^{n_T-1} (u(t_{n+1}, Y_{n+1}) - u(t_n, Y_n))\right) \\ &= E\left(\sum_{n=0}^{n_T-1} (u(t_{n+1}, Y_{n+1}) - u(t_n, Y_n))\right). \end{aligned}$$

Let $\Delta u(t_n, Y_n) = u(t_{n+1}, Y_{n+1}) - u(t_n, Y_n)$. Using a Taylor expansion,

$$\begin{aligned} \Delta u(t_n, Y_n) &= \sum_{k=1}^3 \frac{1}{k!} \{\Delta t_n (\partial_t + a\partial_x) + \Delta B_n b\partial_x\}^k u(t_n, Y_n) + R_n(Y_{n+1}) \\ &= A(Y_n) + R_n(Y_{n+1}) \end{aligned} \tag{12}$$

$$R_n(Y_{n+1}) = \frac{1}{4!} \{\Delta t_n (\partial_t + a\partial_x) + \Delta B_n b\partial_x\}^4 u(\tau_n, \xi_n), \quad \theta \in (0, 1).$$

By a judicious use of conditional expectation, the odd powers in ΔB_n in 12 vanish:

$$|E(\Delta u(t_n, Y_n))| \leq |E\{E(A(Y_n) | \mathcal{F}_n)\}| + E|R_n(Y_{n+1})|.$$

Now $E(A(Y_n) | \mathcal{F}_n)$ contains no ΔB_n terms (nor their higher powers), but only powers of Δt_n . The first-order term is exactly

$$(\partial_t + \mathcal{L})u(t_n, Y_n) \Delta t_n = 0.$$

Moreover, the lowest-order term in $E|R_n(Y_{n+1})|$ is $(\Delta t_n)^2$. Hence

$$|E(\Delta u(t_n, Y_n))| \leq (\Delta t_n)^2 E(K(a, b, u)) + O((\Delta t_n)^2),$$

where $K(a, b, u)$ denotes combinations of a, b, u and derivatives of u . By the theorem's assumptions, $E(K(a, b, u))$ is bounded. Therefore,

$$\sum_{n=0}^{n_T-1} |E(\Delta u(t_n, Y_n))| \leq C \Delta t_n,$$

which completes the proof: the **EM scheme** has first-order weak convergence.