Itô vs. Stratonovich Stochastic Integration

Gemini Pro 3, with edits by Fei Lu

February 18, 2026

1 Formal Definition of the Stochastic Integral

Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and a standard Brownian motion B_t. For a partition of the interval $[0, T]$ given by $0 = t_0 < t_1 < \cdots < t_n = T$ with mesh size $|\Delta| \to 0$, we define the stochastic integral of an adapted, square-integrable process G_t as the limit in $L^2(\Omega)$ of the Riemann-Stieltjes sum:

$$S_n = \sum_{i=0}^{n-1} G_{\tau_i}(B_{t_{i+1}} - B_{t_i}). \tag{1.1}$$

The distinction between the two frameworks depends on the choice of the evaluation point $\tau_i \in [t_i, t_{i+1}]$:

- **Itô Integral:** Evaluated at the left-hand endpoint, $\tau_i = t_i$.

$$\int_0^T G_t\, dB_t := \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} G_{t_i}(B_{t_{i+1}} - B_{t_i}) \text{ in } L^2(\Omega).$$

- **Stratonovich Integral:** Evaluated at the midpoint time,

$$\int_0^T G_t \circ dB_t := \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} G_{\tau_i}(B_{t_{i+1}} - B_{t_i}) \text{ in } L^2(\Omega), \qquad \tau_i = \frac{t_i + t_{i+1}}{2}.$$

2 Examples and Comparison

2.1 Example: Integral of Brownian Motion

Let $G_t = B_t$. We compute $\int_0^T B_t\, dB_t$ and $\int_0^T B_t \circ dB_t$ using the respective definitions.

Itô Integral: $\tau_i = t_i$.

$$\sum B_{t_i}(B_{t_{i+1}} - B_{t_i}) = \frac{1}{2}\sum(B_{t_{i+1}}^2 - B_{t_i}^2 - (B_{t_{i+1}} - B_{t_i})^2) = \frac{1}{2}B_T^2 - \frac{1}{2}\sum(\Delta B_i)^2.$$

As $\sum(\Delta B_i)^2 \to T$ in L^2, we have

$$\int_0^T B_t\, dB_t = \frac{1}{2}B_T^2 - \frac{1}{2}T.$$

Stratonovich Integral: Use the midpoint-time rule $\tau_i = \frac{t_i + t_{i+1}}{2}$. We decompose the integrand as $B_{\tau_i} = B_{t_i} + (B_{\tau_i} - B_{t_i})$. The sum becomes:

$$\sum_{i=0}^{n-1} B_{\tau_i}\Delta B_i = \sum_{i=0}^{n-1} B_{t_i}\Delta B_i + \sum_{i=0}^{n-1}(B_{\tau_i} - B_{t_i})\Delta B_i.$$

The first term is the Itô sum, converging to $\int_0^T B_t dB_t = \frac{1}{2}B_T^2 - \frac{1}{2}T$. For the second term (the correction), we write $\Delta B_i = (B_{t_{i+1}} - B_{\tau_i}) + (B_{\tau_i} - B_{t_i})$. Then

$$(B_{\tau_i} - B_{t_i})\Delta B_i = (B_{\tau_i} - B_{t_i})(B_{t_{i+1}} - B_{\tau_i}) + (B_{\tau_i} - B_{t_i})^2.$$

Note that $Z_i = (B_{\tau_i} - B_{t_i})(B_{t_{i+1}} - B_{\tau_i})$ has mean 0 and variance $\mathbb{E}|Z_i|^2 = \frac{1}{4}\Delta t_i^2$, thus,the first part $\sum_i Z_i$ vanishes in L^2. The second part is the quadratic variation over the left half-intervals $[t_i, \tau_i]$, summing to $\frac{1}{2}T$. Thus,

$$\int_0^T B_t \circ dB_t = \left(\frac{1}{2}B_T^2 - \frac{1}{2}T\right) + \frac{1}{2}T = \frac{1}{2}B_T^2.$$

2.2 Comparison Table

The choice of τ_i leads to fundamental differences in the resulting calculus.

Property	**Itô Calculus**	**Stratonovich Calculus**
Notation	$\int_0^T G_t\, dB_t$	$\int_0^T G_t \circ dB_t$
Expectation	$\mathbb{E}\left[\int_0^T G_t\, dB_t\right] = 0$	Generally $\neq 0$
Martingale Property	The integral is a martingale	Not a martingale
Chain Rule	Requires Itô's Lemma	Standard Calculus (Leibniz rule)
Application	Finance, Filtering theory	Physics, Manifold-valued SDEs

Table 1: Key differences between Itô and Stratonovich frameworks.

3 The Itô–Stratonovich Correction

We derive the correction term using the same midpoint-time convention

$$\int_0^t H_s \circ dB_s = \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} H_{\tau_i}\Delta B_i, \qquad \tau_i = \frac{t_i + t_{i+1}}{2}, \quad \Delta B_i = B_{t_{i+1}} - B_{t_i}.$$

Theorem 3.1 (Itô–Stratonovich Correction, one-dimensional)**.** *Assume $f, g \in C_b^1(\mathbb{R})$ with sufficient growth control so the following integrals are well-defined, and let X solve the Itô SDE*

$$dX_t = f(X_t)\, dt + g(X_t)\, dB_t. \tag{3.1}$$

Then

$$\int_0^t g(X_s) \circ dB_s \;=\; \int_0^t g(X_s)\, dB_s + \frac{1}{2}\int_0^t g(X_s)g'(X_s)\, ds. \tag{3.2}$$

Consequently, (3.1) *is equivalent to the Stratonovich SDE*

$$dX_t = \left(f(X_t) - \tfrac{1}{2}g(X_t)g'(X_t)\right) dt + g(X_t) \circ dB_t. \tag{3.3}$$

Proof. Let Π be a partition and $\tau_i = (t_i + t_{i+1})/2$. Consider the difference of Riemann sums:

$$S_\Pi^\circ - S_\Pi^{\mathrm{Ito}} := \sum_{i=0}^{n-1} \big(g(X_{\tau_i}) - g(X_{t_i})\big)\Delta B_i.$$

By Taylor expansion,

$$g(X_{\tau_i}) - g(X_{t_i}) = g'(X_{t_i})(X_{\tau_i} - X_{t_i}) + r_i, \qquad r_i = o(|X_{\tau_i} - X_{t_i}|).$$

Hence

$$S_\Pi^\circ - S_\Pi^{\mathrm{Ito}} = \sum_i g'(X_{t_i})(X_{\tau_i} - X_{t_i})\Delta B_i + \sum_i r_i \Delta B_i. \tag{3.4}$$

Under standard moment bounds, $X_{\tau_i} - X_{t_i} = O_{L^2(\mathbb{P})}(\sqrt{\Delta t_i})$.

First, we show that the remainder $R_\Delta := \sum_{i=0}^{n-1} r_i \Delta B_i \xrightarrow{\mathbb{P}} 0$. Recall that $r_i = g(X_{\tau_i}) - g(X_{t_i}) - g'(X_{t_i})(X_{\tau_i} - X_{t_i})$. By the Mean Value Theorem, $r_i = (g'(\xi_i) - g'(X_{t_i}))(X_{\tau_i} - X_{t_i})$ for some ξ_i between X_{t_i} and X_{τ_i}. Let $M_\Delta = \sup_{|s-t|\le|\Delta|} |X_s - X_t|$. Since X is continuous, $M_\Delta \to 0$ almost surely as $|\Delta| \to 0$.

We can bound the magnitude of the remainder sum:

$$|R_\Delta| \le \sum_{i=0}^{n-1} |g'(\xi_i) - g'(X_{t_i})| \cdot |X_{\tau_i} - X_{t_i}| \cdot |\Delta B_i|.$$

Take $K > 0$. Since $g \in C_b^1$ on $[-K, K]$, for any $\varepsilon > 0$, there exists $\delta > 0$ (depending on K) such that $|g'(y) - g'(x)| < \varepsilon$ whenever $|y - x| < \delta$. Let $\Omega_{K,\delta} = \{\sup_{t\in[0,T]} |X_t| \le K\} \cap \{M_\Delta < \delta\}$. On this compact set, g' is uniformly continuous, so we have $|g'(\xi_i) - g'(X_{t_i})| < \varepsilon$. Thus,

$$\mathbb{E}[|R_\Delta| \mathbf{1}_{\Omega_{K,\delta}}] \le \varepsilon \sum_{i=0}^{n-1} \mathbb{E}[|X_{\tau_i} - X_{t_i}| \cdot |\Delta B_i|].$$

By Hölder's inequality and the standard estimate $\mathbb{E}[|X_{\tau_i} - X_{t_i}|^2] \le C\Delta t$,

$$\mathbb{E}[|X_{\tau_i} - X_{t_i}| \cdot |\Delta B_i|] \le \sqrt{\mathbb{E}|X_{\tau_i} - X_{t_i}|^2}\sqrt{\mathbb{E}|\Delta B_i|^2} \le C\Delta t.$$

Therefore, $\mathbb{E}[|R_\Delta| \mathbf{1}_{\Omega_{K,\delta}}] \le \varepsilon C T$. Since ε is arbitrary and $\mathbb{P}(\Omega_{K,\delta}^c)$ can be made arbitrarily small by choosing K large and $|\Delta|$ small, the result follows.

Second, we analyze the main term $\sum_i g'(X_{t_i})(X_{\tau_i} - X_{t_i})\Delta B_i$. Recall the increment on $[t_i, \tau_i]$:

$$X_{\tau_i} - X_{t_i} = \int_{t_i}^{\tau_i} f(X_s)\, ds + \int_{t_i}^{\tau_i} g(X_s)\, dB_s.$$

The drift contribution in (3.4) is negligible: it is of size $O(\Delta t_i) \cdot O(\sqrt{\Delta t_i}) = O((\Delta t_i)^{3/2})$ and sums to $o(1)$. For the martingale term, Taylor expand $g(X_s)$ at t_i and integrate,

$$\int_{t_i}^{\tau_i} g(X_s)\, dB_s = g(X_{t_i})(B_{\tau_i} - B_{t_i}) + o_P(\sqrt{\Delta t_i}).$$

Togehter, we have the leading order approximation

$$(X_{\tau_i} - X_{t_i})\Delta B_i = g(X_{t_i})(B_{\tau_i} - B_{t_i})\Delta B_i + o_P(\Delta t_i).$$

Write $\Delta B_i = (B_{\tau_i} - B_{t_i}) + (B_{t_{i+1}} - B_{\tau_i})$. Then

$$(B_{\tau_i} - B_{t_i})\Delta B_i = (B_{\tau_i} - B_{t_i})^2 + (B_{\tau_i} - B_{t_i})(B_{t_{i+1}} - B_{\tau_i}).$$

The cross term has mean zero and variance $O((\Delta t_i)^2)$, hence its sum is $o_P(1)$. Moreover $\sum_i \Big((B_{\tau_i} - B_{t_i})^2 - \frac{\Delta t_i}{2}\Big) \xrightarrow{P} 0$, so

$$(X_{\tau_i} - X_{t_i})\Delta B_i = \frac{1}{2} g(X_{t_i})\Delta t_i + o_P(\Delta t_i).$$

Plugging into (3.4) yields

$$S_\Pi^\circ - S_\Pi^{\text{Ito}} = \frac{1}{2}\sum_i g'(X_{t_i})g(X_{t_i})\Delta t_i + o_P(1) \xrightarrow{P} \frac{1}{2}\int_0^t g'(X_s)g(X_s)\,ds.$$

Taking limits gives (3.2), and rearranging the drift gives (3.3). □

To describe the same process X_t in both frameworks, we must adjust the drift. The term

$$\frac{1}{2} g(X_t)g'(X_t)$$

is the **Itô–Stratonovich correction**. Heuristically it arises from the quadratic variation of Brownian motion, i.e. the fact that $(\Delta B_i)^2 \approx \Delta t_i$ at leading order, and it represents the interaction between state-dependent diffusion and noise fluctuations.

4 Chain Rule for the Stratonovich Integral

We state and prove the chain rule (the analogue of the classical Leibniz rule) using the midpoint-time definition

$$\int_0^T H_s \circ dB_s \;:=\; \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} H_{\tau_i}\,(B_{t_{i+1}} - B_{t_i}), \qquad \tau_i = \frac{t_i + t_{i+1}}{2}.$$

Theorem 4.1 (Stratonovich Chain Rule, one-dimensional)**.** *Let X be a continuous semimartingale solving the Stratonovich SDE*

$$dX_t = a(X_t)\,dt + b(X_t) \circ dB_t, \tag{4.1}$$

where a, b are locally Lipschitz (and of at most linear growth), and let $\varphi \in C_b^2(\mathbb{R})$. Then

$$d\varphi(X_t) \;=\; \varphi'(X_t) \circ dX_t \;=\; \varphi'(X_t)a(X_t)\,dt + \varphi'(X_t)b(X_t) \circ dB_t. \tag{4.2}$$

Equivalently, for every $t \in [0, T]$,

$$\varphi(X_t) - \varphi(X_0) \;=\; \int_0^t \varphi'(X_s)a(X_s)\,ds + \int_0^t \varphi'(X_s)b(X_s) \circ dB_s. \tag{4.3}$$

Proof. Fix a partition $0 = t_0 < \cdots < t_n = t$ and set $\tau_i = (t_i + t_{i+1})/2$, $\Delta X_i = X_{t_{i+1}} - X_{t_i}$. By Taylor's theorem,

$$\varphi(X_{t_{i+1}}) - \varphi(X_{t_i}) = \varphi'(X_{\tau_i})\,\Delta X_i + R_i, \tag{4.4}$$

where the remainder satisfies $|R_i| \le C|\Delta X_i|^2$, with $C = \|\varphi''\|_\infty$. Summing (4.4) over i gives

$$\varphi(X_t) - \varphi(X_0) = \sum_{i=0}^{n-1} \varphi'(X_{\tau_i})\Delta X_i + \sum_{i=0}^{n-1} R_i. \tag{4.5}$$

For solutions of (4.1) with locally Lipschitz coefficients one has the increment estimate $\Delta X_i = O_{L^2}(\sqrt{\Delta t_i})$. Indeed, writing the equation in Itô form with modified drift $\tilde{a}(x) = a(x) + \frac{1}{2}b(x)b'(x)$, we have

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} \tilde{a}(X_s)\,ds + \int_{t_i}^{t_{i+1}} b(X_s)\,dB_s.$$

Squaring and taking expectations, using the elementary inequality $(A+B)^2 \le 2A^2 + 2B^2$,

$$\mathbb{E}[|X_{t_{i+1}} - X_{t_i}|^2] \le 2\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} \tilde{a}(X_s)\,ds\right)^2\right] + 2\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} b(X_s)\,dB_s\right)^2\right].$$

By Jensen's inequality (for the drift) and Itô isometry (for the diffusion), assuming bounded coefficients for simplicity (or using localization arguments),

$$\mathbb{E}[|X_{t_{i+1}} - X_{t_i}|^2] \le 2(t_{i+1} - t_i)\int_{t_i}^{t_{i+1}} \mathbb{E}[\tilde{a}^2]\,ds + 2\int_{t_i}^{t_{i+1}} \mathbb{E}[b^2]\,ds \le C(t_{i+1} - t_i).$$

Taking the square root gives the L^2 estimate of order $O(\sqrt{\Delta t_i})$.

Hence

$$\mathbb{E}\Big[\sum_i |R_i|\Big] \;\lesssim\; \sum_i \mathbb{E}[|\Delta X_i|^2] \;\lesssim\; \sum_i \Delta t_i \;=\; t,$$

and, more precisely, $\sum_i R_i \to 0$ in probability as $|\Delta| \to 0$ (one can localize to ensure boundedness of φ'' on the range of X). Therefore,

$$\varphi(X_t) - \varphi(X_0) = \lim_{|\Delta|\to 0} \sum_{i=0}^{n-1} \varphi'(X_{\tau_i})\Delta X_i = \int_0^t \varphi'(X_s) \circ dX_s.$$

Finally, substituting $dX_s = a(X_s)\,ds + b(X_s) \circ dB_s$ and using linearity yields (4.3), and thus (4.2). ☐

5 Solution to SDEs: Itô vs. Stratonovich

Example 5.1 (Ornstein–Uhlenbeck Process)**.** *Consider the SDE*

$$dX_t = -\theta X_t dt + \sigma dB_t, \qquad X_0 = x_0.$$

The solution is the same under both interpretations, since the diffusion coefficient is constant and the correction term vanishes. The explicit solution is

$$X_t = x_0 e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dB_s.$$

This process is mean-reverting with mean reverting level 0 *and rate* θ*, and it has stationary distribution* $\mathcal{N}(0, \sigma^2/(2\theta))$*.*

Example 5.2. *Consider the Ito SDE*

$$dX_t = \alpha X_t dt + \beta X_t^2 dB_t, \qquad X_0 = x_0.$$

The Itô solution satisfies

$$X_t = x_0 + \int_0^t \alpha X_s ds + \int_0^t \beta X_s^2 dB_s.$$

The Stratonovich SDE for the process X *is*

$$dX_t = \left(\alpha X_t - \beta^2 X_t^3\right) dt + \beta X_t^2 \circ dB_t.$$

The correction term $-\beta^2 X_t^3$ *significantly alters the drift's nonlinearity. This illustrates how the choice of interpretation can lead to qualitatively different dynamics, especially in nonlinear systems with state-dependent noise.*