STAT 6900: Stochastic Processes in Modern Machine Learning
Lecture 2: Brownian Motion and Itô Integrals

Farzad Sabzikar

Office: 2218 Snedecor Hall sabzikar@iastate.edu

January 24, 2026

Goals of this lecture

- Define Brownian motion and describe its probabilistic behaviors
- Introduce stochastic integrals with respect to Brownian motion
- Construct the Itô integral rigorously
- Establish martingale and continuity properties of stochastic integrals

Gaussian transition density and finite-dimensional distributions

Fix $n \geq 1$ and a starting point $x \in \mathbb{R}^n$. Define, for $t > 0$ and $y \in \mathbb{R}^n$,

$$p(t, x, y) = \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{|y - x|^2}{2t} \right).$$

(Convention: $p(0, x, y)\, dy := \delta_x(dy)$, the point mass at x.)

Finite-dimensional distributions induced by p

For $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$ and Borel sets $F_1, \ldots, F_k \subset \mathbb{R}^n$, define a measure $\nu_{t_1,\ldots,t_k}$ on $(\mathbb{R}^n)^k$ by

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) := \int_{F_1 \times \cdots \times F_k} p(t_1, x, x_1) \prod_{j=2}^{k} p(t_j - t_{j-1}, x_{j-1}, x_j)\, dx_1 \cdots dx_k.$$

Kolmogorov extension theorem (existence statement)

The family $\{\nu_{t_1,\dots,t_k}\}$ satisfies the consistency conditions, hence there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P}^x)$ and a process $\{B_t\}_{t\geq 0}$ such that, for all k and all $0 \leq t_1 \leq \cdots \leq t_k$,

$$\mathbb{P}^x(B_{t_1} \in F_1, \dots, B_{t_k} \in F_k) = \nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k).$$

Definition of Brownian motion in $\mathbb{R}^n$ (starting at 0)

Definition 2.1 Standard n-dimensional Brownian motion

A stochastic process $\{\mathbf{B}(t)\}_{t\geq 0}$ with values in $\mathbb{R}^n$ is called a (standard) Brownian motion (starting at 0) if:

- $\mathbf{B}(0) = \mathbf{0}$ almost surely;
- (*independent stationary increments*) for all $0 \leq t_0 < t_1 < \cdots < t_k$, the increments $\mathbf{B}(t_1) - \mathbf{B}(t_0), \ldots, \mathbf{B}(t_k) - \mathbf{B}(t_{k-1})$ are independent, and the law of $\mathbf{B}(t+h) - \mathbf{B}(t)$ depends only on h;
- (*Gaussian increment law*) for all $0 \leq s < t$,

  $$\mathbf{B}(t) - \mathbf{B}(s) \sim \mathcal{N}\big(\mathbf{0}, (t-s)I_n\big),$$

  where the covariance matrix means

  $$\mathrm{Cov}\big(\mathbf{B}(t) - \mathbf{B}(s)\big) := \mathbb{E}\Big[(\mathbf{B}(t) - \mathbf{B}(s))(\mathbf{B}(t) - \mathbf{B}(s))^\top\Big] = (t-s)I_n;$$

- the process admits a version with continuous sample paths.

Increment structure of Brownian motion

The following properties record key structural consequences of Definition 2.1 that will be repeatedly used throughout the course.

Property 2.1 Independent increments

Let

$$0 \leq t_0 < t_1 < \cdots < t_k.$$

Then the increments

$$\mathbf{B}(t_1) - \mathbf{B}(t_0),\ \mathbf{B}(t_2) - \mathbf{B}(t_1),\ \ldots,\ \mathbf{B}(t_k) - \mathbf{B}(t_{k-1})$$

are independent $\mathbb{R}^n$-valued random vectors.

This means that the evolution of the process over disjoint time intervals is probabilistically independent.

Increment structure of Brownian motion

Property 2.2 Gaussian increments

Let $0 \le s < t$. The increment

$$\mathbf{B}(t) - \mathbf{B}(s)$$

is an $\mathbb{R}^n$-valued Gaussian random vector with mean zero and covariance matrix $(t-s)I_n$.

Equivalently, writing

$$\mathbf{B}(t) - \mathbf{B}(s) = \big(B_1(t) - B_1(s), \dots, B_n(t) - B_n(s)\big),$$

we have

$$\mathbb{E}[B_i(t) - B_i(s)] = 0, \qquad \operatorname{Cov}\big(B_i(t) - B_i(s), B_j(t) - B_j(s)\big) = (t-s)\delta_{ij}.$$

Property 2.2: Characteristic function of increments

Property 2.3

For any $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $0 \le s < t$,

$$\mathbb{E}\left[e^{i\langle \boldsymbol{\lambda}, \mathbf{B}(t)-\mathbf{B}(s)\rangle}\right] = \exp\left(-\frac{1}{2}(t-s)\,|\boldsymbol{\lambda}|^2\right).$$

The idea of the proof: By the Gaussian increment property of Brownian motion,

$$\mathbf{Z} := \mathbf{B}(t) - \mathbf{B}(s) \sim \mathcal{N}\big(\mathbf{0}, (t-s)I_n\big).$$

Hence $\mathbf{Z}$ can be represented as

$$\mathbf{Z} = \sqrt{t-s}\,\mathbf{G}, \qquad \mathbf{G} \sim \mathcal{N}(\mathbf{0}, I_n).$$

In one-dimensional case, Let $\langle \boldsymbol{\lambda}, \mathbf{G}\rangle = \sum_{j=1}^n \lambda_j G_j$. Since $\mathbf{G}$ is centered Gaussian, $\langle \boldsymbol{\lambda}, \mathbf{G}\rangle$ is a one-dimensional normal random variable with

$$\mathbb{E}\langle \boldsymbol{\lambda}, \mathbf{G}\rangle = 0, \qquad \mathrm{Var}\big(\langle \boldsymbol{\lambda}, \mathbf{G}\rangle\big) = |\boldsymbol{\lambda}|^2.$$

Characteristic function of increments

From the previous slide, we have

$$\langle \boldsymbol{\lambda}, \mathbf{B}(t) - \mathbf{B}(s) \rangle = \sqrt{t-s}\, U, \qquad U \sim \mathcal{N}\big(0, |\boldsymbol{\lambda}|^2\big).$$

Characteristic function of a one-dimensional Gaussian. If $U \sim \mathcal{N}(0, \sigma^2)$, then

$$\mathbb{E}\big[e^{iuU}\big] = \exp\left(-\frac{1}{2}\sigma^2 u^2\right), \qquad u \in \mathbb{R}.$$

Applying this with $\sigma^2 = |\boldsymbol{\lambda}|^2$ and $u = \sqrt{t-s}$ yields

$$\mathbb{E}\left[e^{i\langle \boldsymbol{\lambda}, \mathbf{B}(t) - \mathbf{B}(s) \rangle}\right] = \exp\left(-\frac{1}{2}|\boldsymbol{\lambda}|^2 (t-s)\right).$$

This proves the stated formula.

Stationarity of increments

Property 2.4 Stationary increments

For any $0 \leq s < t$, the distribution of the increment

$$\mathbf{B}(t) - \mathbf{B}(s) \in \mathbb{R}^n$$

depends only on the time difference $t - s$.

This is expressed by the notation

$$\mathbf{B}(t) - \mathbf{B}(s) \overset{d}{=} \mathbf{B}(t - s),$$

which means that these two $\mathbb{R}^n$-valued random vectors have the same probability distribution.

Equivalently, for every Borel set $A \subset \mathbb{R}^n$,

$$\mathbb{P}(\mathbf{B}(t) - \mathbf{B}(s) \in A) = \mathbb{P}(\mathbf{B}(t - s) \in A)\,.$$

Second moments

Property 2.5 Second moments of n-dimensional Brownian motion

Let
$$\mathbf{B}(t) = (B_1(t), \ldots, B_n(t))$$
be a standard Brownian motion in $\mathbb{R}^n$. Then for all $s, t \geq 0$ and indices i, j,
$$\mathbb{E}[B_i(t)] = 0, \qquad \mathrm{Cov}(B_i(t), B_j(s)) = (t \wedge s)\, \delta_{ij}.$$

Consequently, using the Euclidean norm
$$|\mathbf{B}(t)|^2 := \sum_{i=1}^{n} B_i(t)^2,$$
we obtain
$$\mathbb{E}|\mathbf{B}(t)|^2 = \sum_{i=1}^{n} \mathbb{E}[B_i(t)^2] = nt.$$

Total variation of a function

Definition (Total variation)

Let $f : [0, T] \to \mathbb{R}$ be a function. The *total variation* of f on $[0, T]$ is defined as

$$V_{[0,T]}(f) := \sup_{\Pi} \sum_j \left| f(t_{j+1}) - f(t_j) \right|,$$

where the supremum is taken over all partitions

$$\Pi : \; 0 = t_0 < t_1 < \cdots < t_n = T.$$

- If $V_{[0,T]}(f) < \infty$, we say that f has *finite variation.*
- Smooth functions and functions with bounded derivatives have finite variation.
- Finite variation is the key regularity condition for classical Riemann–Stieltjes integration.

Why classical integration fails for Brownian motion

Property 2.6 Irregularity of Brownian paths

With probability one, Brownian motion $t \mapsto B(t)$ satisfies:

- infinite total variation on every interval $[0, T]$,
- nowhere differentiable sample paths.

Consequences

- The Riemann–Stieltjes integral

$$\int_0^t f(s)\, dB(s)$$

  cannot be defined pathwise.
- Brownian motion oscillates too violently for classical integration theory.
- A new notion of integration, based on probabilistic structure rather than pathwise regularity, is required.

Roadmap: constructing the Itô integral

We construct the stochastic integral with respect to Brownian motion in steps:

- Step 1: Define the integral for elementary (step) adapted processes
- Step 2: Establish the Itô isometry (second moment identity)
- Step 3: Introduce the space of admissible integrands $V(0, T)$
- Step 4: Extend the integral by L^2 completion

For clarity, we work with one-dimensional Brownian motion.

Step stochastic integrands

Let $0 = t_0 < t_1 < \cdots < t_m = T$ be a partition of $[0, T]$. A *step stochastic process* is a process of the form

$$\varphi(t,\omega) = \sum_{j=0}^{m-1} e_j(\omega)\, \mathbf{1}_{[t_j, t_{j+1})}(t),$$

where each e_j is $\mathcal{F}_{t_j}$-measurable. That is, $\varphi(t)$ is constant on each interval and depends only on information available up to the left endpoint. Let $B(t)$ be a one-dimensional Brownian motion.

Definition 2.2

The stochastic integral of φ with respect to B is defined by

$$\int_0^t \varphi(s)\, dB(s) := \sum_{j=0}^{m-1} e_j \big(B(t_{j+1} \wedge t) - B(t_j \wedge t)\big).$$

Example: stochastic integral of a step process

Example 2.1

Let $(B(t))_{t\geq 0}$ be a one-dimensional Brownian motion and fix

$$0 = t_0 < t_1 < t_2 = T.$$

Define the step stochastic process

$$\varphi(t,\omega) = e_0(\omega)\,\mathbf{1}_{[0,t_1)}(t) + e_1(\omega)\,\mathbf{1}_{[t_1,T)}(t),$$

where

$$e_0 \in \mathcal{F}_0, \qquad e_1 \in \mathcal{F}_{t_1}.$$

By definition of the Itô integral,

$$\int_0^t \varphi(s)\,dB(s) = e_0\big(B(t_1\wedge t) - B(0)\big) + e_1\big(B(T\wedge t) - B(t_1\wedge t)\big).$$

Example: stochastic integral of a step process

Example 2.1

Case 1: $0 \le t \le t_1$**.**

$$\int_0^t \varphi(s)\, dB(s) = e_0\big(B(t) - B(0)\big).$$

Case 2: $t_1 < t \le T$**.**

$$\int_0^t \varphi(s)\, dB(s) = e_0\big(B(t_1) - B(0)\big) + e_1\big(B(t) - B(t_1)\big).$$

Second moment and Itô isometry

Notation. From now on, we write $\varphi(t)$ instead of $\varphi(t, \omega)$, with the dependence on ω understood.

Let φ be a step stochastic process of the form

$$\varphi(t) = \sum_{j=0}^{m-1} e_j \, \mathbf{1}_{[t_j, t_{j+1})}(t), \qquad e_j \in \mathcal{F}_{t_j}.$$

Using independence and variance of Brownian increments,

$$\mathbb{E}\left[\left(\int_0^T \varphi(s)\, dB(s)\right)^2\right] = \mathbb{E}\left[\sum_{j=0}^{m-1} e_j^2 (t_{j+1} - t_j)\right] = \mathbb{E}\left[\int_0^T \varphi(s)^2 \, ds\right].$$

This identity is called the *Itô isometry*.

Space of admissible integrands

Recall that for a stochastic integrand $f(t)$, the quantity

$$\int_0^T f(t)^2\, dt$$

is a random variable.

We therefore measure its size using expectation and define

$$V(0,T) = \left\{ f \text{ adapted } : \mathbb{E} \int_0^T f(t)^2\, dt < \infty \right\}.$$

This space is equipped with the norm

$$\|f\|_V^2 = \mathbb{E} \int_0^T f(t)^2\, dt,$$

which makes $V(0,T)$ a Hilbert space.

Extension of the Itô integral

Let $f \in V(0,T)$. Choose a sequence of step processes $\{\varphi_n\}$ such that

$$\|\varphi_n - f\|_V \to 0.$$

By the Itô isometry,

$$\mathbb{E}\left[\left(\int_0^T (\varphi_n - \varphi_m)\,dB\right)^2\right] = \|\varphi_n - \varphi_m\|_V^2.$$

Since $\{\varphi_n\}$ is Cauchy in $V(0,T)$, the sequence $\left\{\int_0^T \varphi_n\,dB\right\}$ is Cauchy in $L^2(\Omega)$ and therefore converges.

Definition

The stochastic integral of f is defined as

$$\int_0^T f\,dB := \lim_{n\to\infty} \int_0^T \varphi_n\,dB,$$

where the limit is taken in $L^2(\Omega)$.

Itô isometry for general integrands

Property 2.7 Itô isometry

For any $f \in V(0,T)$,

$$\mathbb{E}\left[\left(\int_0^T f(t)\,dB(t)\right)^2\right] = \mathbb{E}\int_0^T f(t)^2\,dt.$$

This identity follows by approximation with step processes and continuity of the stochastic integral in $L^2(\Omega)$.

Remark

If f is deterministic, then

$$\mathbb{E}\left[\left(\int_0^T f(t)\,dB(t)\right)^2\right] = \int_0^T f(t)^2\,dt.$$

Canonical examples

Example 2.2 Square function

Let $f(x) = x^2$. Applying the definition of the Itô integral yields

$$\int_0^t B(s)\,dB_s = \frac{1}{2}B(t)^2 - \frac{1}{2}t.$$

Equivalently,

$$B(t)^2 = 2\int_0^t B(s)\,dB_s + t.$$

Example 2.3 Linear function

Let $f(x) = x$. Then

$$\int_0^t 1\,dB_s = B(t).$$

In particular, there is no correction term in this case.

Example: random integrand and Itô isometry

Example 2.4 Random integrand

Let $B(t)$ be one-dimensional Brownian motion and define $f(t,\omega) := B(t,\omega)$. Then f is adapted and $\mathbb{E}\int_0^T B(t)^2\,dt < \infty$, so $f \in V(0,T)$. The Itô integral $\int_0^T B(t)\,dB(t)$ is therefore well defined. By the Itô isometry,

$$\mathbb{E}\left[\left(\int_0^T B(t)\,dB(t)\right)^2\right] = \mathbb{E}\int_0^T B(t)^2\,dt.$$

Since $\mathbb{E}[B(t)^2] = t$, this yields

$$\mathbb{E}\left[\left(\int_0^T B(t)\,dB(t)\right)^2\right] = \int_0^T t\,dt = \frac{T^2}{2}.$$

Example 2.5 Linear drift with Brownian noise

Consider the process

$$X(t) = at + B(t),$$

where $a \in \mathbb{R}$ is a constant and $B(t)$ is one-dimensional Brownian motion.
Then

$$X(t) - X(0) = at + B(t),$$

so the process consists of a deterministic linear trend plus random fluctuations.

Applying Itô's formula to $f(x) = x^2$, we obtain

$$X(t)^2 = 2\int_0^t X(s)\,dB_s + 2a\int_0^t X(s)\,ds + t.$$

The term t arises solely from the stochastic fluctuations.

Remark: n-dimensional Brownian motion

Let

$$\mathbf{B}(t) = (B_1(t), \ldots, B_n(t))$$

be an $\mathbb{R}^n$-valued Brownian motion, where each B_i is a one-dimensional Brownian motion.
If

$$\boldsymbol{\varphi}(t) = (\varphi_1(t), \ldots, \varphi_n(t))$$

is an $\mathbb{R}^n$-valued integrand, we define

$$\int_0^T \langle \boldsymbol{\varphi}(t), d\mathbf{B}(t) \rangle := \sum_{i=1}^n \int_0^T \varphi_i(t)\, dB_i(t).$$

That is, multidimensional stochastic integrals are defined componentwise.

Martingales vs. martingale differences

Martingale

A stochastic process $\{M_t\}$ is a martingale if its *level* does not drift on average when conditioned on the past:

$$\mathbb{E}[M_s \mid \mathcal{F}_t] = M_t \quad (s \geq t).$$

Martingale difference sequence

A sequence $\{\psi_k\}$ is a martingale difference sequence if

$$\mathbb{E}[\psi_k \mid \mathcal{F}_k] = 0.$$

Martingale differences represent *unbiased noise*; sums of martingale differences form martingales.

Example: Brownian motion is a martingale

Example 2.6

Let $\mathbf{B}(t)$ be an $\mathbb{R}^n$-valued Brownian motion and $\mathcal{F}_t = \sigma(\mathbf{B}(s) : s \leq t)$. We verify that $\{\mathbf{B}(t)\}$ is a martingale.

- **Measurability:** $\mathbf{B}(t)$ is $\mathcal{F}_t$-measurable by definition.
- **Integrability:** $\mathbb{E}|\mathbf{B}(t)| < \infty$ since $\mathbb{E}|\mathbf{B}(t)|^2 = nt$.
- **Conditional expectation:** For $s \geq t$,

$$\mathbb{E}[\mathbf{B}(s) \mid \mathcal{F}_t] = \mathbf{B}(t) + \mathbb{E}[\mathbf{B}(s) - \mathbf{B}(t)] = \mathbf{B}(t),$$

using independent increments and zero mean.

Example: Itô integrals are martingales

Example 2.7

Let $f \in V(0,T)$ and define

$$M_t := \int_0^t f(s)\, d\mathbf{B}_s.$$

Then $\{M_t\}$ is a martingale with respect to $\{\mathcal{F}_t\}$.

- **Measurability:** M_t is $\mathcal{F}_t$-measurable.
- **Integrability:** By Itô isometry,

$$\mathbb{E}|M_t|^2 = \mathbb{E}\int_0^t f(s)^2\, ds < \infty.$$

- **Conditional expectation:** For $s \geq t$,

$$\mathbb{E}[M_s - M_t \mid \mathcal{F}_t] = 0$$

since future Brownian increments have mean zero.

Why martingales matter in this course

The previous examples show that:

- Brownian motion is a martingale,
- stochastic integrals with respect to Brownian motion are martingales.

In this course, martingales appear whenever randomness is introduced without bias:

- stochastic gradient descent is driven by martingale difference noise,
- reinforcement learning algorithms rely on martingale errors,
- continuous-time limits of learning algorithms involve martingale terms.

Martingales provide the mathematical language for describing *unbiased stochastic evolution.*

Sample-path regularity of the Itô integral

Property 2.8 Continuity of the Itô integral

Let $f \in V(0,T)$. There exists a version of the stochastic process

$$\left\{ \int_0^t f(s)\, dB_s \right\}_{0 \le t \le T}$$

whose sample paths are continuous functions of t.

This result ensures that, although the Itô integral is initially defined as an $L^2(\Omega)$-limit for each fixed time t, it can be realized as a continuous-time stochastic process.

From now on, whenever we write $\int_0^t f(s)\, dB_s$, we implicitly work with this continuous version.

Maximal inequality for stochastic integrals

Theorem 2.1 Maximal inequality

Let $f \in V(0,T)$ and $\lambda > 0$. Then

$$\mathbb{P}\left(\sup_{0 \le t \le T} \left| \int_0^t f(s)\, dB_s \right| \ge \lambda \right) \le \frac{1}{\lambda^2} \mathbb{E} \int_0^T f(t)^2\, dt.$$

Remark. This inequality is used later to control stochastic fluctuations uniformly in time, for example when proving stability, convergence, or tightness of learning dynamics in stochastic gradient descent, reinforcement learning, and stochastic neural models.

Quadratic variation: definition

Definition 2.3 Quadratic variation

Let $X = \{X(t)\}_{t\geq 0}$ be a real-valued stochastic process. Given a partition

$$\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$$

of $[0, t]$, define

$$Q_\Pi(X) := \sum_{j=0}^{n-1} \big(X(t_{j+1}) - X(t_j)\big)^2.$$

If, as the mesh $|\Pi| \to 0$, the sum $Q_\Pi(X)$ converges in probability (or in L^2), the limit is called the *quadratic variation* of X on $[0, t]$ and is denoted by $[X]_t$.

Interpretation of quadratic variation

Property 2.9

Quadratic variation measures the accumulation of squared increments and captures the *roughness* of sample paths rather than their length.

In particular:

- Smooth or finite-variation processes have zero quadratic variation.
- Processes with highly irregular paths may have nonzero quadratic variation.

Quadratic variation of Brownian motion

Theorem 2.2

Let $B(t)$ be one-dimensional Brownian motion. Then, for every $t \geq 0$,

$$[B]_t = t \quad \text{almost surely.}$$

Moreover, the convergence holds in $L^2(\Omega)$.

Total variation versus quadratic variation

Proposition 2.1

Brownian motion has:

- infinite total variation on every time interval,
- finite, nonzero quadratic variation on every time interval.

This contrast highlights the fundamental difference between classical paths and stochastic paths.

Partitions and notation

Let $0 = t_0 < t_1 < \cdots < t_n = t$ be a partition of $[0, t]$. Define increments

$$\Delta B_j := B(t_{j+1}) - B(t_j).$$

Quadratic variation is obtained by studying limits of

$$\sum_{j=0}^{n-1} (\Delta B_j)^2$$

as the mesh $\max_j (t_{j+1} - t_j)$ tends to zero.

Why quadratic variation matters

For Brownian motion,

$$[B]_t = t \neq 0.$$

This implies that second-order terms arising from stochastic fluctuations do not vanish, even as time steps become arbitrarily small.

As a consequence:

- classical Taylor expansions fail,
- stochastic calculus requires correction terms,
- this leads directly to Itô's formula.

Contrast with smooth functions

Property 2.10

If f is continuously differentiable on $[0, t]$, then

$$\sum_j \big(f(t_{j+1}) - f(t_j)\big)^2 \longrightarrow 0 \quad \text{as } |\Pi| \to 0.$$

Thus:

- smooth paths have zero quadratic variation,
- Brownian paths have finite, nonzero quadratic variation.

Motivation for Itô's formula

Let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. We aim to understand the increment

$$f(B(t)) - f(B(0)).$$

A classical Taylor expansion yields, for small h,

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + o(h^2).$$

In stochastic settings, this expansion cannot be truncated at first order, since $(\Delta B)^2$ does not vanish due to nonzero quadratic variation.

Summation over partitions

Let $\{0 = t_0 < \cdots < t_n = t\}$ be a partition of $[0, t]$ and
$\Delta B_j = B(t_{j+1}) - B(t_j)$.
Applying the Taylor expansion with $h = \Delta B_j$ and summing,

$$f(B(t)) - f(B(0)) = \sum f'(B(t_j))\,\Delta B_j + \frac{1}{2}\sum f''(B(t_j))\,(\Delta B_j)^2 + \text{error}.$$

The error term vanishes in probability as the mesh of the partition tends to zero.

Identification of the limiting terms

As the partition becomes finer:

- The first-order term converges in L^2:

$$\sum f'(B(t_j))\,\Delta B_j \;\longrightarrow\; \int_0^t f'(B(s))\,dB_s.$$

- Using quadratic variation of Brownian motion,

$$\sum f''(B(t_j))\,(\Delta B_j)^2 \;\longrightarrow\; \int_0^t f''(B(s))\,ds.$$

The second-order term has no analogue in classical calculus.

Itô's formula (one-dimensional)

Theorem 2.3 Itô's formula

Let $f \in C^2(\mathbb{R})$. Then, for all $t \geq 0$,

$$f(B(t)) = f(B(0)) + \int_0^t f'(B(s))\, dB_s + \frac{1}{2} \int_0^t f''(B(s))\, ds.$$

Why Itô's formula has this structure

The form of Itô's formula follows from two fundamental facts:

- Brownian motion has *nonzero quadratic variation*:

$$\sum (\Delta B_j)^2 \longrightarrow t.$$

- All higher-order terms involving $(\Delta B_j)^3$ or higher powers vanish in the limit.

As a result, when Taylor expansions are summed over partitions:

- first-order terms produce a stochastic integral,
- second-order terms produce a deterministic time integral,
- no additional correction terms survive.

This is why Itô's formula contains exactly one stochastic term and one drift correction term.

Example: square function

Example 2.8

Let $f(x) = x^2$. Then $f'(x) = 2x$ and $f''(x) = 2$.
Applying Itô's formula,

$$f(B(t)) = f(B(0)) + \int_0^t 2B(s)\,dB_s + \frac{1}{2}\int_0^t 2\,ds.$$

Since $B(0) = 0$, this simplifies to

$$B(t)^2 = 2\int_0^t B(s)\,dB_s + t.$$

This example shows explicitly how quadratic variation produces the deterministic correction term.

Example: exponential function

Example 2.9

Let $f(x) = e^x$. Then $f'(x) = e^x$ and $f''(x) = e^x$.
Applying Itô's formula,

$$e^{B(t)} = e^{B(0)} + \int_0^t e^{B(s)}\, dB_s + \frac{1}{2}\int_0^t e^{B(s)}\, ds.$$

Since $B(0) = 0$, we have $e^{B(0)} = 1$, and therefore

$$e^{B(t)} = 1 + \int_0^t e^{B(s)}\, dB_s + \frac{1}{2}\int_0^t e^{B(s)}\, ds.$$

This example illustrates how a drift term arises naturally, even when the original function has no explicit time dependence.

Conceptual summary of quadratic variation

- Quadratic variation is a fundamental feature of Brownian motion.
- It explains the appearance of the $\frac{1}{2}f''$ correction term in Itô's formula.
- It marks the essential difference between stochastic calculus and classical calculus.
- It will later govern how random fluctuations accumulate in learning dynamics and continuous-time models.

Summary

- Brownian motion constructed via Gaussian, independent, and stationary increments.
- Itô integrals defined first for elementary processes and extended by L^2 completion.
- Martingales and martingale differences introduced and verified through key examples.
- Continuity and maximal inequalities established for stochastic integrals.
- Quadratic variation defined and shown to equal time for Brownian motion, contrasting with smooth paths.
- Itô's formula derived from second-order expansions using quadratic variation.