MA783 Advanced Stochastic Processes Notes

Instructor: Prof. Solesne Bourguin

Student: Wancheng Lin

Contents

1 Probability Spaces, Random Variables, and Stochastic Processes

Sigma-Algebras and Probability Measures

Definition 1.1 (σ-algebra)**.** *Let Ω be a non-empty set. A family $\mathcal{F}$ of subsets of Ω is a σ-algebra if:*

1. $\Omega \in \mathcal{F}$, $\emptyset \in \mathcal{F}$,
2. $A \in \mathcal{F} \implies A^c := \Omega \setminus A \in \mathcal{F}$,
3. *If* $A_1, A_2, \dots \in \mathcal{F}$, *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

This structure guarantees closure under complements and countable unions, providing the minimal requirement for probability to be well-defined.

Definition 1.2 (Probability Space)**.** *A probability space is a triple $(\Omega, \mathcal{F}, P)$ where*

- $(\Omega, \mathcal{F})$ *is a measurable space,*
- $P : \mathcal{F} \to [0, 1]$ *is a probability measure such that*

$$P(\Omega) = 1, \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

  for all disjoint $A_i \in \mathcal{F}$.

A complete probability space also contains all subsets of null sets, ensuring that negligible events are measurable. This prevents technical issues later when working with almost sure properties.

Random Variables and Distributions

Definition 1.3 (Random Variable)**.** *A mapping $X : \Omega \to \mathbb{R}^n$ is a random variable if it is measurable, i.e.*

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{B}(\mathbb{R}^n),$$

where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-algebra.

This guarantees that events of the form $\{X \in B\}$ are measurable. The σ-algebra generated by X, written $\sigma(X)$, represents the information revealed by knowing X.

Definition 1.4 (Distribution)**.** *The distribution of X is the probability measure μ_X on $\mathcal{B}(\mathbb{R}^n)$ defined by*

$$\mu_X(B) := P\big(X^{-1}(B)\big), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Thus the distribution is the pushforward measure of P under X.

Definition 1.5 (Expectation)**.** *If $\int_\Omega |X(\omega)|\, dP(\omega) < \infty$, the expectation of X is*

$$E[X] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbb{R}^n} x\, d\mu_X(x).$$

L^p Spaces and Equivalence Classes

For $p \geq 1$, the L^p-norm of a random variable X is defined by

$$\|X\|_p := \Big(\int_\Omega |X(\omega)|^p\, dP(\omega)\Big)^{1/p}.$$

Definition 1.6 (L^p Spaces)**.** *The space $L^p(P)$ consists of all random variables $X : \Omega \to \mathbb{R}$ with $\|X\|_p < \infty$.*

Equivalence classes. Strictly speaking, $L^p(P)$ is not the set of all random variables with finite p-moment, but rather the set of their *equivalence classes*. Two random variables X and Y are identified whenever

$$P(X = Y) = 1,$$

i.e. they differ only on a null set. Thus, each element of $L^p(P)$ is an equivalence class

$$[X] := \{Y : P(X = Y) = 1\}.$$

This convention avoids ambiguity: if X and Y agree almost surely, then all probabilistic properties we care about (expectation, variance, stochastic integrals) are the same. Working with equivalence classes ensures that the L^p-norm is a true norm rather than merely a seminorm.

Banach and Hilbert structures. The space $L^p(P)$ with norm $\|\cdot\|_p$ is complete, hence a Banach space. When $p = 2$, the space becomes a Hilbert space with inner product

$$\langle X, Y\rangle := E[XY],$$

well-defined on equivalence classes. In particular,

$$\|X\|_2^2 = \langle X, X\rangle.$$

Why L^2 matters in stochastic calculus. Itô integrals are first defined for elementary processes and then extended to all integrands in L^2. The Hilbert structure is essential: convergence in L^2 allows us to use projection and orthogonality arguments. Moreover, martingale theory and isometries (e.g. Itô's isometry)

$$E\left[\left(\int_0^T f(t)\,dB_t\right)^2\right] = E\left[\int_0^T f(t)^2\,dt\right]$$

rely explicitly on the L^2 framework.

Interpretation. From a probabilistic viewpoint, the use of equivalence classes formalizes the phrase "almost surely." We are not interested in the exact values of random variables on exceptional sets of probability zero, since they do not affect expectations, variances, or distributions. Thus L^p spaces provide a natural analytic setting for stochastic processes.

Stochastic Processes

A *stochastic process* $\{X_t\}_{t\in T}$ is simply a family of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{R}^n$.

Two complementary perspectives. There are always two natural ways to look at such a process:

- If we *fix $t \in T$*, then X_t is just a random variable. We can study its distribution, compute its expectation, variance, etc.
- If we *fix $\omega \in \Omega$*, then $t \mapsto X_t(\omega)$ is a deterministic function of t, called the *sample path of the process for that outcome ω*.

This duality — randomness across ω and dynamics across t — is what makes stochastic processes both rich and subtle.

Intuitive picture. It is often helpful to think of t as representing *time* and ω as representing an individual *"particle" or "experiment."* Then $X_t(\omega)$ is the state of that particle at time t. Sometimes one also writes $X(t, \omega)$ instead of $X_t(\omega)$, emphasizing that the process is really a function

$$(t, \omega) \mapsto X(t, \omega),$$

defined on $T\times\Omega$ and taking values in $\mathbb{R}^n$. This viewpoint will become important later, since joint measurability in (t, ω) is crucial for defining stochastic integrals.

Path space representation. Fixing ω, we can identify it with the entire path $t \mapsto X_t(\omega)$. In this sense, we may think of Ω as a subset of the path space

$$(\mathbb{R}^n)^T = \{\text{all functions } f : T \to \mathbb{R}^n\}.$$

The σ-algebra $\mathcal{F}$ will then contain the cylinder σ-algebra $\mathcal{B}$ generated by sets of the form

$$\{\omega : \omega(t_1) \in F_1, \dots, \omega(t_k) \in F_k\}, \quad F_i \in \mathcal{B}(\mathbb{R}^n).$$

From this point of view, a stochastic process is nothing but a probability measure P on the measurable space $\big((\mathbb{R}^n)^T, \mathcal{B}\big)$.

Finite-dimensional distributions. The family of distributions of the vectors $(X_{t_1}, \dots, X_{t_k})$, for $t_1, \dots, t_k \in T$, are called the *finite-dimensional distributions* of the process. Explicitly,

$$\mu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = P(X_{t_1} \in F_1, \dots, X_{t_k} \in F_k)\,,$$

for Borel sets $F_i \subseteq \mathbb{R}^n$. These distributions capture many important properties of the process — for instance, whether increments are Gaussian or independent — though not everything (they do not by themselves encode path continuity, for example).

Kolmogorov's extension theorem. Conversely, if we are given a family of candidate finite-dimensional distributions $\{\nu_{t_1,\dots,t_k}\}$, it is natural to ask: does there exist a stochastic process that realizes them? Kolmogorov's extension theorem gives a celebrated answer: if the family satisfies two natural consistency conditions (marginalization and permutation symmetry), then there exists a stochastic process $\{Y_t\}_{t\in T}$ whose finite-dimensional distributions coincide with $\{\nu_{t_1,\dots,t_k}\}$.

This result is fundamental: it allows us to construct processes like Brownian motion purely from the specification of their finite-dimensional laws, before we even worry about pathwise properties like continuity. For details and proofs, see Lamperti (1977) or Kallenberg (2002).

Independence

Two events $A, B \in \mathcal{F}$ are independent if

$$P(A \cap B) = P(A)P(B).$$

Two random variables X, Y are independent if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent. For independent X and Y, provided $E[|X|], E[|Y|] < \infty$, we have

$$E[XY] = E[X]\,E[Y].$$

Conditional Expectation

Let $H \subseteq \mathcal{F}$ be a sub-σ-algebra. For $X \in L^1(P)$, the conditional expectation $E[X|H]$ is the H-measurable random variable satisfying

$$\int_H E[X|H]\,dP = \int_H X\,dP, \quad \forall H \in H.$$

It is unique up to null sets and inherits key properties:

$$\begin{aligned}
E[aX + bY \mid H] &= aE[X|H] + bE[Y|H], \\
E[E[X|H]] &= E[X], \\
E[X|H] &= X \quad \text{if } X \text{ is } H\text{-measurable}, \\
E[X|H] &= E[X] \quad \text{if } X \text{ independent of } H, \\
E[XY|H] &= Y\,E[X|H] \quad \text{if } Y \text{ is } H\text{-measurable}.
\end{aligned}$$

Moreover, Jensen's inequality extends to this setting:

$$\varphi(E[X|H]) \le E[\varphi(X)|H], \quad \text{for convex } \varphi.$$

Filtrations and Adapted Processes

A *filtration* is an increasing family of sub-σ-algebras $\{\mathcal{F}_t\}_{t\ge 0}$ representing the information available up to time t.

A process $\{X_t\}$ is said to be adapted if X_t is $\mathcal{F}_t$-measurable for all t. This notion formalizes the idea that the process does not "look into the future."

For example, if $\{B_t\}$ is Brownian motion, then $\mathcal{F}_t^B = \sigma(B_s : 0 \le s \le t)$ is the natural filtration of the process, representing exactly the history of the motion up to time t.

2 Brownian Motion and Itô Integration

Construction of Brownian Motion

To construct a Brownian motion $\{B_t : t \ge 0\}$ using Kolmogorov's extension theorem, we must specify a consistent family of finite-dimensional distributions $\{\nu_{t_1,\dots,t_k}\}$.

For $x, y \in \mathbb{R}^n$ and $t > 0$, define the Gaussian transition kernel

$$p(t,x,y) = (2\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2t}\right), \qquad p(0,x,y) = \delta_x(y).$$

For $0 \le t_1 \le t_2 \le \cdots \le t_k$, define

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = \int_{F_1\times\cdots\times F_k} p(t_1,x,x_1)p(t_2-t_1,x_1,x_2)\cdots p(t_k-t_{k-1},x_{k-1},x_k)\,dx_1\cdots dx_k.$$

By permutation symmetry, this definition extends to any ordering of times and automatically satisfies the first consistency condition.

Remark. The second consistency condition is also satisfied: marginalizing out later coordinates leaves the earlier distribution unchanged, thanks to $\int_{\mathbb{R}^n} p(t,x,y)dy = 1$.

Theorem 2.1 (Kolmogorov Extension)**.** *There exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t : t \ge 0\}$ such that the finite-dimensional distributions of B_t coincide with $\nu_{t_1,\dots,t_k}$ constructed above. Moreover $P^x(B_0 = x) = 1$.*

Definition 2.1. *Any such process is called a (version of) n-dimensional Brownian motion starting at x.*

Remark. This construction does not produce a unique process — versions may differ in pathwise properties. We choose a version with continuous paths, justified by Kolmogorov's continuity theorem.

Kolmogorov's Continuity Theorem

Theorem 2.2 (Kolmogorov's Continuity)**.** *Suppose $\{X_t\}_{t\ge0}$ is a process such that for all $T > 0$ there exist $\alpha, \beta, D > 0$ with*

$$E\big(|X_t - X_s|^\alpha\big) \le D|t-s|^{1+\beta}, \quad 0 \le s,t \le T.$$

Then X admits a continuous version.

For Brownian motion,

$$E^x\left(|B_t - B_s|^4\right) = n(n+2)|t-s|^2,$$

so the theorem holds with $\alpha = 4$, $\beta = 1$, $D = n(n+2)$. Thus Brownian motion always has a continuous version.

Remark. If $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ is n-dimensional Brownian motion, then the coordinate processes $\{B_t^{(i)}\}$ are independent one-dimensional Brownian motions.

Basic Properties of Brownian Motion

Brownian motion $\{B_t\}_{t\ge0}$ is defined by its finite-dimensional distributions. From this definition we can immediately deduce a number of fundamental properties.

1. Gaussian process. Brownian motion is a Gaussian process. This means that for every finite set of time points

$$0 \le t_1 \le \cdots \le t_k,$$

the random vector

$$Z = (B_{t_1}, \dots, B_{t_k}) \in \mathbb{R}^{nk}$$

is multivariate normal.

The mean vector and covariance matrix are given by

$$M = \big(E[B_{t_1}], \dots, E[B_{t_k}]\big), \qquad C = [c_{jm}], \quad c_{jm} = E\big[(B_{t_j} - M_j)(B_{t_m} - M_m)\big].$$

Equivalently, the joint characteristic function has the familiar Gaussian form:

$$E^x\left[\exp\left(i\sum_{j=1}^{nk} u_j Z_j\right)\right] = \exp\Big(iu^\top M - \tfrac{1}{2}u^\top Cu\Big).$$

In particular, for one-dimensional Brownian motion starting at x,

$$E^x[B_t] = x, \qquad \mathrm{Var}^x(B_t) = E^x\big[(B_t - x)^2\big] = t.$$

In the n-dimensional case, $\mathrm{Var}^x(B_t) = nt$.

2. Covariance structure. For $0 \le s \le t$, the covariance of increments is

$$E^x\big[(B_t - B_s)^2\big] = n(t-s).$$

More generally, for $i, j \in \{1, \dots, n\}$,

$$\mathrm{Cov}(B_t^{(i)}, B_s^{(j)}) = \delta_{ij}\ \min(s,t),$$

where δ_{ij} is the Kronecker delta. This formula reflects two key features: different coordinates are independent, and covariance grows linearly with the overlap in time.

3. Independent increments. One of the defining properties of Brownian motion is that increments over disjoint intervals are independent. Specifically, for

$$0 \le t_1 < t_2 < \cdots < t_k,$$

the random variables

$$B_{t_1}, \quad B_{t_2} - B_{t_1}, \quad \dots, \quad B_{t_k} - B_{t_{k-1}}$$

are independent.

4. Stationary Gaussian increments. Not only are increments independent, they are also stationary and Gaussian. For $0 \le s < t$,

$$B_t - B_s \sim \mathcal{N}(0, (t-s)I_n).$$

Thus the distribution of an increment depends only on the length of the interval, not on its location in time.

Summary. Brownian motion is therefore:

- a Gaussian process with mean x and covariance $\min(s,t)I_n$,
- with continuous paths (after choosing the continuous version),
- and with stationary, independent, normally distributed increments.

These properties uniquely characterize Brownian motion and underlie its role as the canonical model of continuous-time noise.

Remark. If $B_t = \big(B_t^{(1)}, \dots, B_t^{(n)}\big)$ is an n-dimensional Brownian motion, then each coordinate process $\{B_t^{(j)}\}_{t\ge 0}$, $1 \le j \le n$, is itself a one-dimensional Brownian motion. Moreover, these coordinate processes are mutually independent. This follows directly from the covariance structure, since

$$\mathrm{Cov}(B_t^{(i)}, B_s^{(j)}) = \delta_{ij}\ \min(s,t).$$

Continuity and Versions of Processes

Definition 2.2 (Version / Modification)**.** *Let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, P)$. We say that $\{X_t\}$ is a version (or modification) of $\{Y_t\}$ if for every fixed $t \geq 0$,*

$$P\big(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}\big) = 1.$$

Thus, versions agree almost surely at each fixed time t, although they may differ on a null set that can vary with t. Importantly, versions have the same finite-dimensional distributions (f.d.d.'s).

Finite-dimensional distributions. Given a process $\{X_t\}_{t\geq 0}$, the collection of joint distributions of

$$(X_{t_1}, \dots, X_{t_k}) \quad \text{for all } k \geq 1,\, 0 \leq t_1 < \cdots < t_k$$

is called the system of finite-dimensional distributions. If $\{X_t\}$ and $\{Y_t\}$ are versions, then clearly

$$(X_{t_1}, \dots, X_{t_k}) \overset{d}{=} (Y_{t_1}, \dots, Y_{t_k}), \quad \forall k, t_1, \dots, t_k,$$

so they induce the same probability law on $\mathbb{R}^k$.

Why versions matter. The definition of Brownian motion via Kolmogorov's extension theorem yields only a process with the correct f.d.d.'s. However, this construction does not guarantee any regularity of sample paths: the raw object might be highly irregular, even discontinuous almost everywhere. To obtain the "classical" Brownian motion with continuous paths, one needs to show that there exists a version of the process that is continuous with probability one. This is where Kolmogorov's continuity theorem enters.

Theorem 2.3 (Kolmogorov's Continuity Theorem)**.** *Let $\{X_t : t \in T\}$ be a stochastic process such that for some $\alpha, \beta, D > 0$,*

$$\mathbb{E}\big[|X_t - X_s|^\alpha\big] \;\leq\; D\,|t-s|^{1+\beta}, \quad \forall s, t \in T.$$

Then there exists a continuous version $\{\tilde{X}_t\}$ of $\{X_t\}$ such that

$$P\big(X_t = \tilde{X}_t,\, \forall t \in T\big) = 1.$$

Moreover, the paths of $\tilde{X}_t$ are Hölder continuous of any order $\gamma < \frac{\beta}{\alpha}$.

Application to Brownian motion. For Brownian motion $B_t \in \mathbb{R}^n$, one computes

$$\mathbb{E}\big[|B_t - B_s|^4\big] = n(n+2)\,|t-s|^2.$$

Thus the continuity theorem applies with $\alpha = 4$, $\beta = 1$, and $D = n(n+2)$. Hence Brownian motion admits a continuous version with paths that are almost surely Hölder continuous of any exponent $\gamma < \frac{1}{4}$.

Key insight. Versions emphasize the distinction between:

- the *law* of a process, determined by its finite-dimensional distributions, and
- the *sample path properties*, such as continuity, differentiability, or bounded variation.

The Kolmogorov theorem is indispensable: it elevates the abstract Brownian motion from a mere collection of distributions to the concrete object with continuous trajectories that we use in stochastic calculus.

In particular, when we speak of "Brownian motion" in stochastic analysis, we always mean this continuous version.

Some Reflection on Canonical Brownian Motion

The Brownian motion defined via Kolmogorov's extension theorem is not unique. Indeed, there may exist several probability spaces $(\Omega, \mathcal{F}, P^x)$ and processes $\{B_t\}$ such that the finite-dimensional distributions satisfy the conditions of Brownian motion. However, for our purposes this non-uniqueness is not problematic: we may simply choose a convenient version to work with.

Continuous paths and identification. As established by Kolmogorov's continuity theorem, there exists a version of Brownian motion with continuous paths almost surely. Thus, for almost all $\omega \in \Omega$, we may identify ω with a continuous function

$$t \mapsto B_t(\omega), \quad t \in [0, \infty), \ B_t(\omega) \in \mathbb{R}^n.$$

Hence, we may regard Brownian motion as a probability measure P^x on the path space $C([0, \infty), \mathbb{R}^n)$. This version is called the *canonical Brownian motion*.

Why canonical? This point of view is not only intuitive but also technically advantageous. The space $C([0, \infty), \mathbb{R}^n)$ of continuous functions, equipped with the topology of uniform convergence on compact sets, is a Polish space (complete and separable metric space). This allows us to employ powerful results from measure theory and probability on Polish spaces, and is the starting point for much of the modern theory of stochastic processes (see Stroock and Varadhan (1979)).

Subtlety of measurability. At first glance, one might ask whether

$$t \mapsto B_t(\omega) \quad \text{is continuous for almost all } \omega.$$

However, the set

$$H = \{\omega \in \Omega : t \mapsto B_t(\omega) \text{ is continuous}\}$$

is not measurable with respect to the canonical product σ-algebra $\mathcal{B}(\mathbb{R}^n)^{[0,\infty)}$, since it involves uncountably many time indices.

By reformulating the construction in the canonical path space $C([0, \infty), \mathbb{R}^n)$, this measurability issue disappears: continuity is built into the path space itself. Thus, the canonical construction provides a rigorous and convenient framework in which Brownian motion is viewed as a random element of $C([0, \infty), \mathbb{R}^n)$.

From Discrete Models to Stochastic Differential Equations

We began with a deterministic growth model

$$\frac{dN(t)}{dt} = r(t)\, N(t),$$

which describes the rate of change of a population $N(t)$ with deterministic growth rate $r(t)$.

In realistic settings, however, the growth rate is subject to random environmental fluctuations. A natural idea is to add a "noise term":

$$\frac{dN(t)}{dt} = \big(r(t) + \text{noise}\big)\, N(t).$$

Generalization. For a general state process $\{X_t\}$, we may write

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{noise},$$

where

- $b(t, x)$ is the *drift* (deterministic trend),
- $\sigma(t, x)$ scales the *random fluctuations*.

Noise as Brownian motion. To make this precise, we need to model "noise" as a well-defined stochastic process. The canonical choice is one-dimensional Brownian motion $\{W_t\}$, characterized by:

1. $W_0 = 0$ almost surely,
2. Independent increments: $W_{t_2} - W_{t_1}$ is independent of the past if $t_2 > t_1$,
3. Stationary increments: $W_{t+s} - W_s \sim \mathcal{N}(0, t)$ for all $s, t \geq 0$,
4. $E[W_t] = 0$ and $\mathrm{Var}(W_t) = t$.

Thus the natural stochastic differential equation (SDE) is

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dW_t.$$

This is the standard Brownian-driven model, which forms the foundation for Itô calculus.

Motivation for the Itô Integral

To interpret an SDE rigorously, we seek processes $\{X_t\}$ satisfying

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s.$$

The first integral is classical (Riemann or Lebesgue), but the second is new: a stochastic integral with respect to Brownian motion.

The central problem. Brownian paths are almost surely continuous but nowhere differentiable and of infinite variation. Therefore, the last term cannot be understood as an ordinary Riemann–Stieltjes integral. We must instead construct a new theory of integration that is compatible with the quadratic variation of Brownian motion:

$$[B]_t = t.$$

Interpretation. The stochastic term

$$\int_0^t \sigma(s, X_s)\, dB_s$$

represents the accumulated effect of "white noise" fluctuations scaled by σ. This captures how randomness enters continuously into the dynamics of X_t.

Conclusion. The task of the next sections is to rigorously define the Itô integral, first for simple adapted processes and then extending to general square-integrable integrands. This construction will provide the foundation for the theory of SDEs.

Why the Riemann–Stieltjes Approach Fails

A natural first attempt to define the stochastic integral

$$\int_0^T f(t,\omega)\,dB_t(\omega)$$

is to view $t \mapsto B_t(\omega)$ as a continuous function for each fixed ω, and try to interpret this as a classical Riemann–Stieltjes (RS) integral of the form $\int f\,dg$.

Indeed, if g has bounded variation, the RS integral $\int f\,dg$ exists whenever f is continuous. However, Brownian paths have two pathological features:

1. They are *nowhere differentiable*, so the interpretation $\int f(t)\,dB_t = \int f(t)\,B_t' dt$ is meaningless.
2. They have *infinite total variation* almost surely:

$$V_{0,T}(B) = \sup_{\pi} \sum_{j=0}^{n-1} |B_{t_{j+1}} - B_{t_j}| = \infty \quad \text{a.s.},$$

   where the supremum runs over partitions π of $[0,T]$.

Because of (2), the RS integral cannot be applied. This failure is not only theoretical but can be demonstrated concretely by approximations.

Illustrative Example. Consider $f(t,\omega) = B_t(\omega)$. Define step-function approximations on dyadic intervals:

$$\phi_1(t,\omega) = \sum_{j\geq 0} B_{j2^{-n}}(\omega)\,1_{[j2^{-n},(j+1)2^{-n})}(t),$$

$$\phi_2(t,\omega) = \sum_{j\geq 0} B_{(j+1)2^{-n}}(\omega)\,1_{[j2^{-n},(j+1)2^{-n})}(t).$$

These correspond to left- and right-point Riemann sums.

Compute their integrals:

$$\int_0^T \phi_1(t,\omega)\,dB_t(\omega) = \sum_j B_{t_j}(\omega)\,(B_{t_{j+1}} - B_{t_j}),$$

$$\int_0^T \phi_2(t,\omega)\,dB_t(\omega) = \sum_j B_{t_{j+1}}(\omega)\,(B_{t_{j+1}} - B_{t_j}).$$

Taking expectations:

$$E\left[\int_0^T \phi_1\,dB\right] = \sum_j E[B_{t_j}(B_{t_{j+1}} - B_{t_j})] = \underline{0},$$

since increments are independent of the past.

But

$$E\left[\int_0^T \phi_2\,dB\right] = \sum_j E[B_{t_{j+1}}(B_{t_{j+1}} - B_{t_j})] = \sum_j (t_{j+1} - t_j) = \underline{T}.$$

Thus two reasonable RS-type approximations give different limits. This shows the RS definition is inconsistent for Brownian integrals.

Conceptual Reason. The breakdown arises because *RS integrals are designed for integrators of bounded variation*, while Brownian motion has variation "too large." In fact, instead of total variation, Brownian motion has a well-defined *quadratic variation*:

$$[B]_t = \lim_{|\pi|\to 0} \sum_j (B_{t_{j+1}} - B_{t_j})^2 = t,$$

which suggests that an integration theory adapted to quadratic, not total, variation is required.

Conclusion. Therefore, the Riemann–Stieltjes framework is inadequate. The correct approach, pioneered by Itô, defines the integral by:

- restricting initially to *elementary adapted processes*,
- defining the integral via increments of Brownian motion,
- and extending by L^2-limits, using the quadratic variation structure.

This leads to the Itô integral, which is consistent, linear, and satisfies the crucial *Itô isometry*.

Preliminaries: Filtration and Adaptedness

Let $\{B_t\}$ be an n-dimensional Brownian motion. Define

$$\mathcal{F}_t = \sigma(B_s^{(i)} : 0 \le s \le t,\ 1 \le i \le n),$$

the natural filtration. Intuitively, $\mathcal{F}_t$ represents the history of the process up to time t. A process $\phi(t,\omega)$ is *adapted* if $\phi(t,\cdot)$ is $\mathcal{F}_t$-measurable for each t.

Admissible Integrands

We now specify the class of processes we may integrate.

Definition 2.3. *Let $V(S,T)$ denote the class of processes $f : [S,T] \times \Omega \to \mathbb{R}$ such that:*

1. *$(t,\omega) \mapsto f(t,\omega)$ is $\mathcal{B}([S,T]) \otimes \mathcal{F}$-measurable;*
2. *$f(t,\omega)$ is adapted to $\{\mathcal{F}_t\}$;*
3. *$E\Big[\int_S^T f(t,\omega)^2 \, dt\Big] < \infty$.*

This ensures that f is square-integrable in time and adapted to the filtration.

Elementary Integrands and Definition

Definition 2.4 (Elementary Process)**.** *A function $\phi \in V(S,T)$ is elementary if it has the form*

$$\phi(t,\omega) = \sum_j e_j(\omega) \, 1_{(t_j,t_{j+1}]}(t),$$

where e_j is $\mathcal{F}_{t_j}$-measurable and square-integrable.

For such ϕ, define the stochastic integral by

$$\int_S^T \phi(t,\omega) \, dB_t(\omega) := \sum_j e_j(\omega) \, (B_{t_{j+1}} - B_{t_j}).$$

This definition parallels the Riemann sum approach, but crucially the coefficients are measurable with respect to the past, not the future.

Remark. If e_j depended on $B_{t_{j+1}}$, the integral would not be well-defined. Adaptedness ensures causality.

Itô Isometry

To extend the integral beyond elementary processes, we use the following fundamental identity.

Lemma 2.1 (Itô Isometry)**.** *If $\phi \in V(S,T)$ is elementary, then*

$$E\left[\left(\int_S^T \phi(t,\omega) \, dB_t\right)^2\right] = E\left[\int_S^T \phi(t,\omega)^2 \, dt\right].$$

Sketch. For $\phi = \sum e_j 1_{(t_j,t_{j+1}]}$, note that

$$\int_S^T \phi(t) \, dB_t = \sum_j e_j (B_{t_{j+1}} - B_{t_j}).$$

Using independence and zero mean of increments:

$$E\big[e_i e_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\big] = 0, \quad i \ne j,$$

and for $i = j$,

$$E[e_j^2 (B_{t_{j+1}} - B_{t_j})^2] = E[e_j^2](t_{j+1} - t_j).$$

Summing yields the result. □

Remark on L^2 in the Itô integral. There are *two* distinct L^2-spaces involved in the definition of the Itô integral:

- The **integrands** f are approximated by elementary processes φ_n in the Hilbert space

$$L^2([S,T]\times\Omega) := \left\{ f : [S,T]\times\Omega \to \mathbb{R} \;\middle|\; \mathbb{E}\left[\int_S^T |f(t,\omega)|^2\, dt\right] < \infty \right\}.$$

  That is,

$$\|f-\varphi_n\|^2_{L^2([S,T]\times\Omega)} = \mathbb{E}\left[\int_S^T |f(t,\omega)-\varphi_n(t,\omega)|^2\, dt\right] \;\to\; 0.$$

- The **integrals** $I_n := \int_S^T \varphi_n(t,\omega)\, dB_t$ form a Cauchy sequence in the space of square-integrable random variables

$$L^2(\Omega) := \{X : \Omega \to \mathbb{R} \mid \mathbb{E}[X^2] < \infty\},$$

  thanks to the Itô isometry

$$\mathbb{E}\left[\Big(I_n - I_m\Big)^2\right] = \|\varphi_n - \varphi_m\|^2_{L^2([S,T]\times\Omega)}.$$

  Thus $I_n \to I$ in $L^2(\Omega)$ for some limit I, which is by definition $\int_S^T f\, dB_t$.

Summary: This distinction explains why the "L^2" in the definition of the Itô integral is not the usual $L^2(\Omega)$ but rather the mixed space $L^2([S,T]\times\Omega)$:

Convergence in $L^2([0,T]\times\Omega)$ of the integrands $\iff$ Convergence in $L^2(\Omega)$ of the stochastic integrals.

Extension to General Integrands

Using the isometry, we extend the definition by density.

Construction: Itô Integral For $f \in V(S,T)$, there exists a sequence of elementary processes ϕ_n with

$$E\left[\int_S^T (f(t)-\phi_n(t))^2\, dt\right] \to 0.$$

Define

$$I(f) = \int_S^T f(t)\, dB_t := L^2 - \lim_{n\to\infty} \int_S^T \phi_n(t)\, dB_t.$$

The limit exists and is unique in $L^2(\Omega)$ by the Itô isometry.

Properties of the Itô Integral

- Linearity: $\int f + g\, dB = \int f\, dB + \int g\, dB$.
- Isometry: $E[(\int f\, dB)^2] = E[\int f^2\, dt]$.
- Martingale property: $\int_0^t f(s)\, dB_s$ is a martingale.
- Zero mean: $E[\int f\, dB] = 0$.

Remark. This construction is robust: unlike Riemann–Stieltjes, the Itô integral accommodates the roughness of Brownian motion via square-integrability and martingale structure.

Continuity and Martingale Property of the Itô Integral

An important property of the Itô integral is that it defines a martingale. Before proving this fact, we recall a classical result due to Doob.

Theorem 2.4 (Doob's Martingale Inequality)**.** *Let $(M_t)_{t\ge 0}$ be a martingale such that $t \mapsto M_t(\omega)$ is continuous for almost every ω. Then, for all $p \ge 1$, $T > 0$, and all $\lambda > 0$, we have*

$$P\left(\sup_{0\le t\le T} |M_t| \ge \lambda\right) \le \frac{1}{\lambda^p}\, \mathbb{E}[|M_T|^p]\,.$$

This inequality will be essential in controlling the supremum of martingale sequences in L^2 convergence arguments.

Theorem: Existence of a Continuous Version
Let $f \in \mathcal{V}(0,T)$, i.e., a progressively measurable process with finite energy. Then there exists a t-continuous version of the stochastic process

$$I(t) = \int_0^t f_s \, dB_s, \qquad 0 \le t \le T.$$

That is, there exists an adapted stochastic process $(J_t)_{t \in [0,T]}$ such that

$$P\left(J_t = \int_0^t f_s \, dB_s \right) = 1, \quad \forall\, t \in [0,T],$$

and $t \mapsto J_t(\omega)$ is continuous for almost every ω.

Proof. Let $\{\phi_n\} \subset \mathcal{V}(0,T)$ be a sequence of elementary (simple) processes such that

$$\mathbb{E}\left[\int_0^T |f_s - \phi_n(s)|^2 \, ds \right] \xrightarrow[n\to\infty]{} 0.$$

Define the Itô integrals

$$I_n(t) := \int_0^t \phi_n(s) \, dB_s, \qquad I(t) := \int_0^t f_s \, dB_s.$$

Then for each n, $t \mapsto I_n(t,\omega)$ is continuous for all ω, and $I_n(t)$ is an $\mathcal{F}_t$-martingale. Indeed, for $s \le t$,

$$\begin{aligned} \mathbb{E}[I_n(t) \mid \mathcal{F}_s] &= \mathbb{E}\left(\int_0^t \phi_n(u) \, dB_u \,\middle|\, \mathcal{F}_s \right) = \int_0^s \phi_n(u) \, dB_u + \mathbb{E}\left(\int_s^t \phi_n(u) \, dB_u \,\middle|\, \mathcal{F}_s \right) \\ &= I_n(s) + 0 = I_n(s), \end{aligned}$$

where the last equality follows since the future Brownian increments $(B_u - B_s)_{u \ge s}$ are independent of $\mathcal{F}_s$ and have zero mean.

Thus I_n is an $\mathcal{F}_t$-martingale with continuous paths.
Claim: The difference $I_n - I_m$ is also an $\mathcal{F}_t$-martingale. Applying Doob's martingale inequality (with $p = 2$) gives, for any $\varepsilon > 0$,

$$P\left(\sup_{0 \le t \le T} |I_n(t) - I_m(t)| > \varepsilon \right) \le \frac{1}{\varepsilon^2} \mathbb{E}\big[|I_n(T) - I_m(T)|^2 \big] .$$

By the Itô isometry,

$$\mathbb{E}\big[|I_n(T) - I_m(T)|^2 \big] = \mathbb{E}\left[\int_0^T |\phi_n(s) - \phi_m(s)|^2 ds \right] \xrightarrow[n,m\to\infty]{} 0.$$

Therefore,

$$P\left(\sup_{0 \le t \le T} |I_n(t) - I_m(t)| > \varepsilon \right) \to 0,$$

and the sequence (I_n) is Cauchy in probability in the supremum norm.
Subsequence argument. By a standard diagonal and Borel–Cantelli argument, we can extract a subsequence (I_{n_k}) such that

$$\sup_{0 \le t \le T} |I_{n_{k+1}}(t) - I_{n_k}(t)| \le 2^{-k} \quad \text{for all } k \ge k_1(\omega),$$

for almost all ω.

This shows that $(I_{n_k}(t))$ converges uniformly in t for all $t \in [0,T]$ and almost every ω. Uniform convergence preserves continuity; since each $I_{n_k}(t)$ is t-continuous, the limit

$$J_t := \lim_{k\to\infty} I_{n_k}(t)$$

is also t-continuous for almost every ω.

Finally, because $I_{n_k}(t) \to I(t)$ in $L^2(\Omega)$ for each fixed t, it follows that $J_t = I_t$ almost surely for all $t \in [0,T]$. Hence J_t is a t-continuous version of I_t. □

From now on, we assume that any Itô integral

$$\int_0^t f_s \, dB_s$$

is understood to be its *t-continuous version*.

Corollary (Martingale Property of the Itô Integral)

Let $f \in \mathcal{V}(0, T)$. For all $T > 0$, define

$$M_t := \int_0^t f_s \, dB_s, \qquad 0 \leq t \leq T.$$

Then $(M_t)_{0 \leq t \leq T}$ is an $\mathcal{F}_t$-martingale. Moreover, for any $\lambda > 0$ and $T > 0$,

$$P\left(\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right) \leq \frac{1}{\lambda^2} \mathbb{E}\left[\int_0^T f_s^2 \, ds\right].$$

Proof. The martingale property of (M_t) follows from the fact that each approximating process $I_n(t) = \int_0^t \phi_n(s) \, dB_s$ is an $\mathcal{F}_t$-martingale, and $I_n(t) \to M_t$ in $L^2(\Omega)$. The inequality is an immediate consequence of Doob's martingale inequality applied to the continuous martingale (M_t), together with Itô's isometry:

$$\mathbb{E}[|M_T|^2] = \mathbb{E}\left[\int_0^T f_s^2 \, ds\right].$$

□

Commentary. This result has two major implications:

1. The Itô integral $\int_0^t f_s \, dB_s$ is not only square-integrable in $L^2(P)$, but also admits a version whose sample paths are continuous in t almost surely.

2. The integral process is an $\mathcal{F}_t$-martingale. This property forms the analytical foundation for stochastic calculus, enabling the development of Itô's formula and stochastic differential equations.

The use of Doob's inequality ensures control of the supremum norm and validates the passage from L^2 convergence to pathwise convergence. The continuity and martingale property together justify interpreting the Itô integral as a "continuous-time martingale-valued linear operator."

Extending Itô's Integral to Accommodate the Multidimensional Case

Motivation

So far, the Itô integral has been defined only for one-dimensional Brownian motion. If we now consider two independent Brownian motions $B_t^{(1)}$ and $B_t^{(2)}$, the classical definition does *not* allow us to make sense of

$$\int_0^t B_s^{(2)}\, dB_s^{(1)},$$

since $B_s^{(2)}$ is not $\mathcal{F}_s^{(1)}$-measurable, where $\mathcal{F}_t^{(1)} := \sigma(B_u^{(1)} : 0 \le u \le t)$. This measurability issue prevents us from treating $B^{(2)}$ as an admissible integrand with respect to $B^{(1)}$.

To construct a genuinely *multidimensional* version of the Itô integral (i.e. integration with respect to a vector of Brownian motions), we must enlarge the class of admissible integrands and slightly generalize the adaptedness condition.

A New Filtration Condition

We replace the old adaptedness condition (2) in the definition of the class $\mathcal{V}$ by the following relaxed version.

$$\text{(2-new)} \quad \exists \text{ filtration } \mathcal{H} = \{\mathcal{H}_t : t \ge 0\} \text{ such that } \begin{cases} (i)\ B_t \text{ is a martingale with respect to } \mathcal{H}_t, \\ (ii)\ f_t \text{ is } \mathcal{H}_t\text{-adapted.} \end{cases}$$

Remark (Interpretation)**.**The key idea is to allow the integrand f_t to depend on a larger information set than $\mathcal{F}_t = \sigma(B_s : 0 \le s \le t)$, as long as the driving Brownian motion remains a martingale with respect to the new filtration $\mathcal{H}_t$. If (2-new) holds, then for all $t > 0$ we still have $B_t \in \mathcal{H}_t$ and

$$\mathbb{E}[B_s - B_t \mid \mathcal{H}_t] = 0, \qquad s > t,$$

which are precisely the properties required in the construction of the one-dimensional Itô integral. Thus, this modification preserves all essential martingale properties while enabling a richer class of integrands.

Remark (Broader Significance)**.**This new condition not only enables integration with respect to multidimensional Brownian motion, but also enlarges the admissible space of integrands f_t in general. It is a conceptual extension from "pathwise adaptedness" to "filtration-adapted martingale structure."

Application to the Multidimensional Case

Suppose $B_t = (B_t^{(1)}, \ldots, B_t^{(n)})$ is an n-dimensional Brownian motion. Define

$$\mathcal{F}_t^{(n)} := \sigma(B_s^{(i)} : 1 \le i \le n,\ 0 \le s \le t).$$

Each component $B_t^{(i)}$ is a martingale with respect to this joint filtration because the increments

$$B_s^{(i)} - B_t^{(i)} \quad \text{are independent of} \quad \mathcal{F}_t^{(n)} \quad (s > t).$$

Hence, we may take $\mathcal{H}_t = \mathcal{F}_t^{(n)}$ in (2-new). We then define Itô integrals of the form

$$\int_0^t f_s\, dB_s^{(i)},$$

where f_t is $\mathcal{F}_t^{(n)}$-adapted. This setup automatically includes examples such as

$$\int_0^t B_s^{(2)}\, dB_s^{(1)}, \qquad \int_0^t \sin(B_s^{(1)} + B_s^{(2)})\, dB_s^{(2)}.$$

We can now rigorously define the multidimensional Itô integral.

Definition of the Multidimensional Itô Integral

Definition 2.5. *Let $B = (B^{(1)}, \ldots, B^{(n)})$ be an n-dimensional Brownian motion. Let $\mathcal{V}_{\mathcal{H}}^{m\times n}(S,T)$ denote the set of $m \times n$ matrix-valued processes*

$$v = [v_{ij}(t,\omega)],$$

where each entry $v_{ij}(t,\omega)$ satisfies Conditions (1), (3), and (2-new) with respect to the same filtration $\mathcal{H} = \{\mathcal{H}_t : t \geq 0\}$.

If $v \in \mathcal{V}_{\mathcal{H}}^{m\times n}(S,T)$, the multidimensional Itô integral is defined by

$$\int_S^T v_t\, dB_t := \int_S^T \begin{pmatrix} v_{11}(t) & \cdots & v_{1n}(t) \\ \vdots & & \vdots \\ v_{m1}(t) & \cdots & v_{mn}(t) \end{pmatrix} \begin{pmatrix} dB_t^{(1)} \\ \vdots \\ dB_t^{(n)} \end{pmatrix}.$$

Equivalently, in component form,

$$\int_S^T v_t\, dB_t = \begin{pmatrix} \displaystyle\sum_{j=1}^n \int_S^T v_{1j}(t)\, dB_t^{(j)} \\ \vdots \\ \displaystyle\sum_{j=1}^n \int_S^T v_{mj}(t)\, dB_t^{(j)} \end{pmatrix},$$

which is an m-dimensional vector-valued Itô integral.

Remark (Interpretation)**.**Each row of the matrix v_t specifies the integrands for one component of the resulting m-dimensional Itô process. The above definition is thus a natural vectorized extension of the one-dimensional Itô integral, where integration is performed componentwise and summed over the driving Brownian components.

Remark. Simplified Notations:

- If $\mathcal{H} = \mathcal{F}^{(n)}$, we write $\mathcal{V}^{m\times n}(S,T)$.
- If $m = 1$, we write $\mathcal{V}_{\mathcal{H}}^{n}(S,T)$.
- If $\mathcal{H} = \mathcal{F}^{(n)}$ and $m = 1$, we simply write $\mathcal{V}^{n}(S,T)$.

Notation and Further Extension

We define

$$\mathcal{V}^{m\times n} := \mathcal{V}^{m\times n}(0,\infty) = \bigcap_{T>0} \mathcal{V}^{m\times n}(0,T),$$

which represents the space of all $m \times n$ integrands admissible on every finite interval.

The next step in generalization is to relax the square-integrability condition (3) in the definition of $\mathcal{V}$ to the following almost-sure finiteness condition:

$$\text{(3-new)} \qquad \mathbb{P}\left[\int_S^T f(s,\omega)^2\, ds < \infty\right] = 1.$$

This weaker assumption broadens the admissible integrands to those that are locally square-integrable almost surely, while retaining well-definedness of the stochastic integral.

Definition 2.6. *$\mathcal{W}_{\mathcal{H}}(S,T)$ denotes the class of processes satisfying conditions (1), (2-new), and (3-new). We adopt the same notational conventions for $\mathcal{W}$ as for $\mathcal{V}$. In particular,*

$$\mathcal{W}_{\mathcal{H}} = \bigcap_{T>0} \mathcal{W}_{\mathcal{H}}(0,T).$$

Remark (Summary)**.**This construction shows that the multidimensional Itô integral is not merely a collection of scalar integrals, but the natural consequence of extending the notion of adaptedness and square-integrability to vector-valued settings. The introduction of the filtration $\mathcal{H}_t$ ensures all integrands are measurable with respect to a common information structure, while maintaining the martingale property necessary for the

3 Itô Formula and Martingale Representation Formula

The 1–dimensional Itô Formula

Motivation. Like for Riemann integrals, the definition of the Itô integral as a limit is not, by itself, very helpful for computing explicit integrals. In the Riemann (deterministic) case, we rely on the fundamental theorem of calculus and the chain rule to compute integrals.

Here, we cannot hope for a fundamental theorem in the same sense, because sample paths of Brownian motion are nowhere differentiable.

However, we can establish a stochastic–calculus analogue of the chain rule: **the Itô formula**.

Some intuition

Recall that

$$\frac{1}{2}B_t^2 \;=\; \frac{t}{2} + \int_0^t B_s\,dB_s \;=\; \int_0^t \frac{1}{2}\,ds \;+\; \int_0^t B_s\,dB_s.$$

Hence, the image of the Brownian motion under the map $g(x) = \frac{x^2}{2}$ is *not* of the pure form $\int_0^t f_s\,dB_s$ for some f, but rather decomposes as

$$\int_0^t h_s\,ds \;+\; \int_0^t f_s\,dB_s.$$

It turns out that if we introduce Itô processes as sums of ds and dB_s integrals, then this class is stable under the action of sufficiently smooth maps g.

Definition 3.1 (Itô process)**.** *Let B_t be a 1–dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0},)$. A (1–dimensional) Itô process X_t is a stochastic process of the form*

$$X_t \;=\; X_0 \;+\; \int_0^t u_s\,ds \;+\; \int_0^t v_s\,dB_s,$$

where u, v are progressively measurable processes such that, for every $t > 0$,

$$\left(\int_0^t |u_s|\,ds < \infty\right) = 1, \qquad \left(\int_0^t v_s^2\,ds < \infty\right) = 1.$$

Remark (Notation)**.**Sometimes a differential form notation is used for Itô processes:

$$dX_t \;=\; u_t\,dt \;+\; v_t\,dB_t.$$

We will switch between the integral and differential views as convenient.

Statement of the Itô formula

Theorem 3.1 (1–dimensional Itô formula)**.** *Let X_t be an Itô process given by $dX_t = u_t\,dt + v_t\,dB_t$. Let $g : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be of class $C^{1,2}$ (i.e. C^1 in t and C^2 in x). Set $Y_t := g(t, X_t)$. Then Y_t is again an Itô process and formally satisfies*

$$\boxed{dY_t = \tfrac{\partial g}{\partial t}(t, X_t)\,dt + \tfrac{\partial g}{\partial x}(t, X_t)\,dX_t + \tfrac{1}{2}\,\tfrac{\partial^2 g}{\partial x^2}(t, X_t)\,(dX_t)^2.} \tag{1}$$

where $(dX_t)^2 = (dX_t)\cdot(dX_t)$ is computed according to the rules

$$dt\cdot dt = 0, \qquad dt\cdot dB_t = 0, \qquad dB_t\cdot dt = 0, \qquad dB_t\cdot dB_t = dt.$$

Equivalently, replacing dX_t by $(u_t\,dt + v_t\,dB_t)$ and $(dX_t)^2$ by $v_t^2\,dt$ in (1) yields the integral form

$$\boxed{\begin{aligned} g(t, X_t) = g(0, X_0) &+ \int_0^t \frac{\partial g}{\partial s}(s, X_s)\,ds + \int_0^t u_s\,\frac{\partial g}{\partial x}(s, X_s)\,ds \\ &+ \frac{1}{2}\int_0^t v_s^2\,\frac{\partial^2 g}{\partial x^2}(s, X_s)\,ds + \int_0^t v_s\,\frac{\partial g}{\partial x}(s, X_s)\,dB_s. \end{aligned}}$$

Proof idea and details. We first treat the case where *$g, \partial_t g, \partial_x g, \partial_{xx} g$ are bounded and where u, v are elementary (simple) adapted processes*. The general case follows by *approximation*: choose g_n with bounded derivatives such that $g_n \to g$ and the first/second partials converge uniformly on compact sets of $[0, \infty) \times \mathbb{R}$; similarly approximate u, v by elementary processes in L^1/L^2 as required.

(Telescoping sum + Taylor expansion). Fix a partition $0 = t_0 < \cdots < t_n = t$ with mesh $|\pi| = \max_j \Delta t_j$, where $\Delta t_j := t_{j+1} - t_j$. Write

$$g(t, X_t) \;=\; g(0, X_0) \;+\; \sum_{j \geq 0} \Delta g(t_j, X_{t_j}), \qquad \Delta g(t_j, X_{t_j}) := g(t_{j+1}, X_{t_{j+1}}) - g(t_j, X_{t_j}).$$

For each j, apply Taylor's theorem to g around (t_j, X_{t_j}):

$$\begin{aligned}\Delta g(t_j, X_{t_j}) = \Big[& g(t_j, X_{t_j}) + \frac{\partial g}{\partial t}(t_j, X_{t_j})\,\Delta t_j + \frac{1}{2}\,\frac{\partial^2 g}{\partial t^2}(t_j, X_{t_j})\,(\Delta t_j)^2 \\ & + \frac{\partial g}{\partial x}(t_j, X_{t_j})\,\Delta X_j + \frac{1}{2}\,\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,(\Delta X_j)^2 + \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j})\,\Delta t_j\,\Delta X_j + R_j \Big] \\ & - g(t_j, X_{t_j}),\end{aligned}$$

so that

$$\begin{aligned} g(t, X_t) = g(0, X_0) &+ \sum_{j \geq 0} \frac{\partial g}{\partial t}(t_j, X_{t_j})\,\Delta t_j + \sum_{j \geq 0} \frac{1}{2}\,\frac{\partial^2 g}{\partial t^2}(t_j, X_{t_j})\,(\Delta t_j)^2 \\ &+ \sum_{j \geq 0} \frac{\partial g}{\partial x}(t_j, X_{t_j})\,\Delta X_j + \sum_{j \geq 0} \frac{1}{2}\,\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,(\Delta X_j)^2 + \sum_{j \geq 0} \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j})\,\Delta t_j\,\Delta X_j + \sum_{j \geq 0} R_j. \end{aligned}$$

Limits of the terms as $|\pi| \to 0$. We show, term by term:

$$\sum_j \frac{\partial g}{\partial t}(t_j, X_{t_j})\,\Delta t_j \;\xrightarrow{\text{a.s.}}\; \int_0^t \frac{\partial g}{\partial s}(s, X_s)\,ds, \qquad \sum_j \frac{\partial g}{\partial x}(t_j, X_{t_j})\,\Delta X_j \;\xrightarrow{L^2}\; \int_0^t \frac{\partial g}{\partial x}(s, X_s)\,dX_s,$$

while

$$\sum_j \frac{1}{2}\,\frac{\partial^2 g}{\partial t^2}(t_j, X_{t_j})\,(\Delta t_j)^2 \;\to 0, \qquad \sum_j \frac{\partial^2 g}{\partial t \partial x}(t_j, X_{t_j})\,\Delta t_j\,\Delta X_j \;\to 0, \qquad \sum_j R_j \;\to 0.$$

The first is standard Riemann convergence (boundedness of the derivative and continuity in t suffice). The mixed term and the t^2-term vanish because $\sum (\Delta t_j)^2 \to 0$ and $|\Delta X_j| = O_P(\sqrt{\Delta t_j})$. The remainder R_j satisfies $R_j = o\big((\Delta t_j)^2 + |\Delta X_j|^2\big)$, hence $\sum_j R_j \to 0$ since $(\Delta B_j)^2 \sim \Delta t_j$ a.s. (asymptotically) and $(\Delta X_j)^2 = u(t_j)^2 (\Delta t_j)^2 + 2u(t_j)v(t_j)\Delta t_j \Delta B_j + v(t_j)^2 (\Delta B_j)^2$.

Decomposition of $\sum \frac{1}{2}\partial_{xx} g\,(\Delta X_j)^2$. When u, v are elementary (piecewise constant, adapted), we can write

$$\Delta X_j \;=\; u(t_j)\,\Delta t_j \;+\; v(t_j)\,\Delta B_j.$$

Hence

$$\begin{aligned}\sum_{j \geq 0} \frac{1}{2}\,\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,(\Delta X_j)^2 = & \sum_{j \geq 0} \frac{1}{2}\,\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,\big(u(t_j)\Delta t_j\big)^2 \\ & + \sum_{j \geq 0} \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,u(t_j)v(t_j)\,\Delta t_j\,\Delta B_j \\ & + \sum_{j \geq 0} \frac{1}{2}\,\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,v(t_j)^2\,(\Delta B_j)^2.\end{aligned}$$

The *first term* goes to 0 almost surely because

$$\left| \sum_j \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\,\big(u(t_j)\Delta t_j\big)^2 \right| \leq \frac{1}{2}\,\big(\sup_{s,x} |\partial_{xx} g|\big)\,\big(\max_j \Delta t_j\big)\,\sum_j |u(t_j)|\,\Delta t_j \;\longrightarrow\; 0.$$

The *second term* goes to 0 in L^2 as $|\pi| \to 0$ since

$$\mathbb{E}\left[\left(\sum_j \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\, u(t_j)v(t_j)\, \Delta t_j\, \Delta B_j\right)^2\right] = \sum_j \mathbb{E}\left[\left(\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\, u(t_j)v(t_j)\right)^2\right]\ (\Delta t_j)^3 \ \longrightarrow\ 0,$$

using independence and the fact that $\mathbb{E}[(\Delta B_j)^2] = \Delta t_j$. It remains to analyze the *third term*.

Key claim. With $a(t) := \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)\, v_t^2$,

$$\sum_j a(t_j)\,(\Delta B_j)^2 \ \xrightarrow[L^2(\Omega)]{}\ \int_0^t a(s)\, ds, \quad \text{as} \quad |\pi| \to 0.$$

Equivalently,

$$\sum_j a(t_j)\,\big((\Delta B_j)^2 - \Delta t_j\big) \ \xrightarrow[L^2(\Omega)]{}\ 0.$$

Indeed,

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_j a(t_j)\,\big((\Delta B_j)^2 - \Delta t_j\big)\right|^2\right] &= \sum_{i,j} \mathbb{E}\big[a(t_i)a(t_j)\,\big((\Delta B_i)^2 - \Delta t_i\big)\big((\Delta B_j)^2 - \Delta t_j\big)\big] \\
&= \sum_j \mathbb{E}\big[(a(t_j))^2\big]\ \mathbb{E}\Big[\big((\Delta B_j)^2 - \Delta t_j\big)^2\Big] \qquad \text{(cross terms vanish)} \\
&= \sum_j \mathbb{E}\big[(a(t_j))^2\big]\ \Big(\mathbb{E}[(\Delta B_j)^4] - 2\Delta t_j\, \mathbb{E}[(\Delta B_j)^2] + (\Delta t_j)^2\Big) \\
&= \sum_j \mathbb{E}\big[(a(t_j))^2\big]\ \big(3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2\big) \\
&= 2\sum_j \mathbb{E}\big[(a(t_j))^2\big]\ (\Delta t_j)^2\ \le\ 2\,(\max_j \Delta t_j)\ \sum_j \mathbb{E}\big[(a(t_j))^2\big]\, \Delta t_j \ \longrightarrow\ 0,
\end{aligned}$$

since $a \in L^2([0,t] \times \Omega)$. Therefore

$$\sum_j \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})\, v(t_j)^2\, (\Delta B_j)^2 \ \xrightarrow[L^2]{}\ \frac{1}{2}\int_0^t \frac{\partial^2 g}{\partial x^2}(s, X_s)\, v_s^2\, ds.$$

Collecting limits. Passing to the limit along partitions $|\pi| \to 0$ yields

$$g(t, X_t) = g(0, X_0) + \int_0^t \frac{\partial g}{\partial s}(s, X_s)\, ds + \int_0^t \frac{\partial g}{\partial x}(s, X_s)\, dX_s + \frac{1}{2}\int_0^t \frac{\partial^2 g}{\partial x^2}(s, X_s)\, v_s^2\, ds,$$

and since $dX_s = u_s\, ds + v_s\, dB_s$ we finally obtain

$$g(t, X_t) = g(0, X_0) + \int_0^t \Big(\frac{\partial g}{\partial s}(s, X_s) + u_s \frac{\partial g}{\partial x}(s, X_s) + \frac{1}{2}v_s^2 \frac{\partial^2 g}{\partial x^2}(s, X_s)\Big)\, ds + \int_0^t v_s \frac{\partial g}{\partial x}(s, X_s)\, dB_s.$$

This is the Itô formula in integral form, which is equivalent to the differential form (1) with the stochastic product rules

$$dt \cdot dt = 0, \quad dt \cdot dB_t = 0, \quad dB_t \cdot dt = 0, \quad dB_t \cdot dB_t = dt.$$

□

Remark (Mnemonic)**.**The convergence

$$\sum_j a(t_j)\,(\Delta B_j)^2 \ \to\ \int_0^t a(s)\, ds \quad \text{is often summarized informally as} \quad (dB_t)^2 = dt.$$

The proof above makes this precise in $L^2(\Omega)$, using the fourth moment of Gaussian increments.

Thoughts and takeaways

- *The class of Itô processes* $X_t = X_0 + \int u_s ds + \int v_s dB_s$ *is stable under smooth maps* $g \in C^{1,2}$*; the image* $g(t, X_t)$ *has drift and diffusion given by the* $\partial_t g$*,* $\partial_x g$*, and* $\partial_{xx} g$ *terms.*

- *The quadratic variation mechanism appears concretely through the limit* $\sum a(t_j)((\Delta B_j)^2 - \Delta t_j) \to 0$ *in* L^2*. This is the rigorous backbone behind the symbolic rule* $(dB_t)^2 = dt$*.*

- *In applications, the "formal differential"* $dY_t = \partial_t g\, dt + \partial_x g\, dX_t + \frac{1}{2}\partial_{xx} g\,(dX_t)^2$ *is a reliable mnemonic; the proof shows precisely why mixed terms involving* $dt\, dB_t$ *and* dt^2 *vanish.*

  Remark (Differential shorthand form of Itô's formula)**.**

$$\boxed{dg(t, X_t) = \frac{\partial g}{\partial t}(t, X_t)\, dt + \frac{\partial g}{\partial x}(t, X_t)\, dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)\,(dX_t)^2.}$$

  Substituting $dX_t = u_t\, dt + v_t\, dB_t$ and applying the Itô multiplication rules

$$(dt)^2 = 0, \qquad dt\, dB_t = 0, \qquad (dB_t)^2 = dt,$$

  we obtain

$$\boxed{dg(t, X_t) = \left(\frac{\partial g}{\partial t} + u_t \frac{\partial g}{\partial x} + \frac{1}{2} v_t^2 \frac{\partial^2 g}{\partial x^2}\right) dt + v_t \frac{\partial g}{\partial x}\, dB_t.}$$

  This expression should be viewed as a *formal shorthand* for the integral version of Itô's formula. It summarizes how the process $g(t, X_t)$ decomposes into a **drift term** (in dt) and a **martingale term** (in dB_t), while the algebraic rules $(dB_t)^2 = dt$ and $(dt)^2 = dt\, dB_t = 0$ reflect the quadratic variation of Brownian motion.

- *For full generality (unbounded derivatives, non–elementary* u, v*), justify by approximation* $g_n \to g$ *with bounded derivatives and* $u^n \to u$*,* $v^n \to v$ *in the appropriate* L^1/L^2 *senses, passing limits by dominated convergence and the Itô isometry.*

Examples: Where and How to Use Itô's Formula

Example 1: Brownian Motion with Drift and Variance Term

Let

$$X_t = \sigma B_t + \mu t, \qquad t \geq 0,$$

where σ, μ *are constants. Here we may write the dynamics explicitly as*

$$dX_t = u_t\, dt + v_t\, dB_t, \quad \text{with } u_t = \mu,\ v_t = \sigma.$$

Define

$$g(t, x) = \sigma x + \mu t.$$

Then

$$\frac{\partial g}{\partial x} = \sigma, \qquad \frac{\partial g}{\partial t} = \mu, \qquad \frac{\partial^2 g}{\partial x^2} = 0.$$

By Itô's formula,

$$dg(t, B_t) = \frac{\partial g}{\partial x}\, dB_t + \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right) dt = \sigma\, dB_t + \mu\, dt.$$

Hence,

$$\boxed{dX_t = \sigma\, dB_t + \mu\, dt.}$$

Example 2: Geometric Brownian Motion

Let

$$S_t = S_0\, e^{\sigma B_t + \mu t}, \qquad S_0 > 0.$$

We can view this as a transformation of the Brownian motion B_t *through the smooth function*

$$g(t, x) = S_0 e^{\sigma x + \mu t}.$$

The underlying process is

$$X_t = B_t,$$

which satisfies

$$dX_t = u_t\,dt + v_t\,dB_t, \qquad \text{with } u_t = 0,\ v_t = 1.$$

That is, the Brownian motion itself has no drift and unit volatility. We will apply Itô's formula to the composite process $g(t, X_t) = g(t, B_t)$.

$$dg(t, X_t) = \frac{\partial g}{\partial t}(t, X_t)\,dt + \frac{\partial g}{\partial x}(t, X_t)\,dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)\,(dX_t)^2.$$

Substituting $dX_t = u_t\,dt + v_t\,dB_t$ *and* $(dB_t)^2 = dt$, *we obtain*

$$dg(t, X_t) = \frac{\partial g}{\partial x}\,v_t\,dB_t + \Big(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\,u_t + \tfrac{1}{2}v_t^2\frac{\partial^2 g}{\partial x^2}\Big)dt.$$

Since $u_t = 0$ *and* $v_t = 1$, *this simplifies to*

$$dg(t, B_t) = \frac{\partial g}{\partial x}\,dB_t + \Big(\frac{\partial g}{\partial t} + \tfrac{1}{2}\frac{\partial^2 g}{\partial x^2}\Big)dt.$$

Compute the derivatives of g:

$$\frac{\partial g}{\partial x} = \sigma g, \qquad \frac{\partial g}{\partial t} = \mu g, \qquad \frac{\partial^2 g}{\partial x^2} = \sigma^2 g.$$

Substituting into Itô's formula gives

$$dg(t, B_t) = \sigma g\,dB_t + (\mu + \tfrac{1}{2}\sigma^2)\,g\,dt.$$

Since $g(t, B_t) = S_t$, *we have*

$$\boxed{dS_t = \sigma S_t\,dB_t + (\mu + \tfrac{1}{2}\sigma^2)S_t\,dt.}$$

Equivalently, in Itô process form

$$dS_t = u_t\,dt + v_t\,dB_t, \quad \text{where} \quad u_t = (\mu + \tfrac{1}{2}\sigma^2)S_t, \quad v_t = \sigma S_t.$$

Expectation. *Taking expectations, note that* $\log S_t = \log S_0 + \sigma B_t + \mu t$, *and since* $B_t \sim N(0, t)$,

$$\mathbb{E}[S_t] = S_0 e^{\mu t}\mathbb{E}[e^{\sigma B_t}] = S_0 e^{\mu t + \frac{1}{2}\sigma^2 t} = S_0 e^{(\mu + \frac{1}{2}\sigma^2)t}.$$

Setting $r = \mu + \frac{1}{2}\sigma^2$ *yields*

$$\boxed{\mathbb{E}[S_t] = S_0 e^{rt}.}$$

Intuitive Interpretation.

- *The process* S_t *evolves multiplicatively, meaning its rate of change is proportional to its current level.*
- *The instantaneous volatility is* σS_t, *so larger values of* S_t *lead to proportionally larger random fluctuations.*
- *The Itô correction term* $\frac{1}{2}\sigma^2$ *appears in the drift because exponential transformations of Brownian motion introduce a convexity adjustment.*
- *In financial terms,* S_t *models a stock price under continuous compounding with drift* μ *and volatility* σ.
- *Unlike linear processes,* $\log S_t$ *follows a normal distribution, hence* S_t *is lognormal.*

Example 3: Ornstein–Uhlenbeck Process

Consider the SDE

$$\boxed{dX_t = -\alpha X_t\,dt + \sigma\,dB_t, \qquad X_0 = x_0,\ \alpha, \sigma > 0.}$$

Integral Form.

$$X_t = x_0 - \alpha \int_0^t X_s \, ds + \sigma B_t.$$

Solution. *The solution can be derived using the integrating factor $e^{\alpha t}$:*

$$d(e^{\alpha t} X_t) = \alpha e^{\alpha t} X_t \, dt + e^{\alpha t} \, dX_t = e^{\alpha t} \sigma \, dB_t.$$

Integrating both sides,

$$e^{\alpha t} X_t - X_0 = \sigma \int_0^t e^{\alpha s} \, dB_s,$$

which gives

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

Intuitive Explanation. *The Ornstein–Uhlenbeck (OU) process can be viewed as a Brownian motion that is pulled back toward the origin whenever it drifts too far away. To understand this, examine the stochastic differential equation*

$$dX_t = -\alpha X_t \, dt + \sigma \, dB_t, \qquad \alpha, \sigma > 0.$$

- ***Drift term ($-\alpha X_t \, dt$):*** *This represents a restoring or mean-reverting force.*
  - *If X_t is large and positive, then $-\alpha X_t < 0$, so the drift pushes X_t downward, toward the origin.*
  - *If X_t is large and negative, then $-\alpha X_t > 0$, so the drift pushes X_t upward, again toward the origin.*
  - *The magnitude of the pull is proportional to the current distance $|X_t|$ from zero, with proportionality constant α.*
- ***Diffusion term ($\sigma \, dB_t$):*** *This adds random fluctuations, just as in standard Brownian motion. It prevents the process from ever settling exactly at the mean, keeping it in continuous motion.*
- ***Comparison with standard Brownian motion:***
  - *A standard Brownian motion B_t has no restoring drift, so it tends to wander indefinitely far from the origin with variance growing linearly in t.*
  - *In contrast, the OU process has a mean-reverting drift that counteracts this dispersion, keeping the trajectory confined to a neighborhood around zero.*
- ***Long-term behavior:*** *Because of the restoring force, the process "forgets" its initial condition exponentially fast:*

  $$E[X_t] = x_0 e^{-\alpha t} \longrightarrow 0, \qquad \text{as } t \to \infty.$$

  Eventually, the fluctuations balance the drift, and the process settles into a stationary distribution

  $$X_t \sim N\left(0, \frac{\sigma^2}{2\alpha}\right),$$

  representing equilibrium between random shocks and mean reversion.

In summary: *The Ornstein–Uhlenbeck process behaves like a "noisy spring" — a Brownian particle tethered to the origin by a linear restoring force of strength α, constantly buffeted by Gaussian noise of intensity σ.*

Moments. *Because $\int_0^t e^{-\alpha(t-s)} \, dB_s$ is an Itô integral with deterministic integrand:*

$$E[X_t] = x_0 e^{-\alpha t}.$$

Variance (using Itô isometry):

$$\mathrm{Var}(X_t) = \sigma^2 \int_0^t e^{-2\alpha(t-s)} \, ds = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}).$$

Limiting Distribution. *As $t \to \infty$:*

$$E[X_t] \to 0, \qquad \mathrm{Var}(X_t) \to \frac{\sigma^2}{2\alpha}.$$

Since the Itô integral of a deterministic function of s is Gaussian, we have

$$\boxed{X_t \xrightarrow{d} N\left(0, \frac{\sigma^2}{2\alpha}\right).}$$

Thus, the Ornstein–Uhlenbeck process is a Gaussian process with stationary limit.

The Multidimensional Itô Formula

Let

$$B_t = \left(B_t^{(1)}, \ldots, B_t^{(m)}\right)$$

be an m-dimensional Brownian motion. If each of the processes $u_t^{(i)}$ and $\sigma_t^{(ij)}$ satisfies the conditions of the definition of Itô processes for $1 \le i \le n$, $1 \le j \le m$, we can define the n-dimensional Itô process

$$\begin{cases} dX_t^{(1)} = u_t^{(1)}\,dt + \sigma_t^{(11)}\,dB_t^{(1)} + \cdots + \sigma_t^{(1m)}\,dB_t^{(m)}, \\ \vdots \\ dX_t^{(n)} = u_t^{(n)}\,dt + \sigma_t^{(n1)}\,dB_t^{(1)} + \cdots + \sigma_t^{(nm)}\,dB_t^{(m)}. \end{cases}$$

Or in matrix notation,

$$dX_t = u_t\,dt + \sigma_t\,dB_t,$$

where

$$X_t = \begin{pmatrix} X_t^{(1)} \\ \vdots \\ X_t^{(n)} \end{pmatrix}, \quad u_t = \begin{pmatrix} u_t^{(1)} \\ \vdots \\ u_t^{(n)} \end{pmatrix}, \quad \sigma_t = \begin{pmatrix} \sigma_t^{(11)} & \cdots & \sigma_t^{(1m)} \\ \vdots & \ddots & \vdots \\ \sigma_t^{(n1)} & \cdots & \sigma_t^{(nm)} \end{pmatrix}, \quad dB_t = \begin{pmatrix} dB_t^{(1)} \\ \vdots \\ dB_t^{(m)} \end{pmatrix}.$$

Such a process X_t is called an n-dimensional Itô process.

Theorem 3.2 (The general Itô formula)**.** *Let X_t be the n-dimensional Itô process defined above, and let*

$$g(t,x) = \left(g_1(t,x), \ldots, g_p(t,x)\right)$$

be a $C^{1,2}$ map from $[0,\infty) \times \mathbb{R}^n$ into $\mathbb{R}^p$. Then the process

$$Y_t = g(t, X_t)$$

is again an Itô process, whose k-th component $Y_t^{(k)}$ satisfies

$$dY_t^{(k)} = \frac{\partial g_k}{\partial t}(t, X_t)\,dt + \sum_{i=1}^{n} \frac{\partial g_k}{\partial x_i}(t, X_t)\,dX_t^{(i)} + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 g_k}{\partial x_i\,\partial x_j}(t, X_t)\,dX_t^{(i)}\,dX_t^{(j)}.$$

The differential rules are

$$dB_t^{(i)}\,dB_t^{(j)} = \delta_{ij}\,dt, \qquad dB_t^{(i)}\,dt = 0, \qquad dt\,dB_t^{(i)} = 0.$$

Proof. *The argument is analogous to the one-dimensional case. Apply Itô's lemma componentwise and use the product rules for differentials.*

The Martingale Representation Theorem

Let $(B_t)_{t\geq 0}$ be an n-dimensional Brownian motion with natural filtration $\{\mathcal{F}_t^{(n)}\}_{t\geq 0}$.

We already know that for any progressively measurable process $v \in \mathcal{V}^n(0,T)$, the process

$$X_t = X_0 + \int_0^t v_s \, dB_s$$

is a martingale with respect to $\{\mathcal{F}_t^{(n)}\}$.

*We will now prove the converse: any square-integrable $\{\mathcal{F}_t^{(n)}\}$-martingale can be represented as a constant plus an Itô integral. This is called the **Martingale Representation Theorem**, and it is fundamental in stochastic analysis and mathematical finance.*

Prerequisites and Strategy

Our goal is to show that certain families of random variables are dense in the Hilbert space $L^2(\Omega, \mathcal{F}_T, P)$. This is crucial because density, combined with the Hilbert space projection and orthogonality properties, allows us to represent any L^2-martingale as an Itô integral.

The proof builds on three key prerequisites:

1. ***Martingale convergence theorem:** if $(\mathcal{H}_n)_{n\geq 1}$ is an increasing sequence of σ-algebras with $\mathcal{H}_n \uparrow \mathcal{F}_T$, then for every $g \in L^2(\mathcal{F}_T)$,*
$$\mathbb{E}[g \mid \mathcal{H}_n] \underset{L^2}{\longrightarrow} \mathbb{E}[g \mid \mathcal{F}_T] = g.$$

2. ***Hilbert space closure identities:** For any subspace H of a Hilbert space K,*
$$(H^\perp)^\perp = \overline{H}, \qquad \text{and} \qquad H^\perp = \{0\} \implies (H^\perp)^\perp = K.$$
Hence, to show that a subspace H is dense in K, it suffices to prove that its orthogonal complement is trivial.

3. ***Inverse Fourier transform theorem:** For $\phi \in C_c^\infty(\mathbb{R}^n)$,*
$$\phi(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \widehat{\phi}(y) e^{i\langle x,y\rangle} \, dy, \quad \text{where} \quad \widehat{\phi}(y) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \phi(x) e^{-i\langle x,y\rangle} \, dx.$$
This analytic tool is used to connect the exponential martingales with the dense set of smooth cylinder functions.

With these tools, the proof proceeds in two density lemmas:

- *The first class consists of smooth cylinder functions $\phi(B_{t_1}, \dots, B_{t_n})$.*
- *The second class is the linear span of exponential martingales $\exp\{\int h \, dB - \frac{1}{2}\int h^2 dt\}$.*

Lemma 1 — Density of Cylinder Functions

Lemma 3.1 (Density of Smooth Cylinder Functions)**.** *Fix $T > 0$ and write $\mathcal{F}_T := \sigma\{B_s : 0 \leq s \leq T\}$. Let*

$$\mathcal{S} := \left\{ \phi(B_{t_1}, \dots, B_{t_n}) \ : \ n \in \mathbb{N}, \ t_i \in [0,T], \ \phi \in C_c^\infty(\mathbb{R}^n) \right\}.$$

Then
$$\overline{\mathcal{S}}^{L^2(\Omega)} = L^2(\Omega, \mathcal{F}_T, P).$$

Proof. **Finite σ-algebras and their relation to $\mathcal{F}_T$.** Fix a countable dense set $\{t_i\}_{i\geq 1} \subset [0,T]$ and define

$$\mathcal{H}_n := \sigma(B_{t_1}, \dots, B_{t_n}), \qquad n \geq 1.$$

Then $(\mathcal{H}_n)$ is an increasing sequence of σ-algebras, and we set

$$\widetilde{\mathcal{F}}_T := \sigma\Big(\bigcup_{n\geq 1} \mathcal{H}_n \Big) = \sigma\{B_t : t \in D\},$$

where $D = \{t_i\}$ is the chosen dense set.

Remark. (See Appendix 4 for a precise justification.)

By continuity of Brownian paths, for any $s \in [0,T]$ we can find $(t_i) \subset D$ with $t_i \to s$ and $B_{t_i} \to B_s$ almost surely. Hence the σ-algebra generated by $\{B_t : t \in D\}$ and the one generated by $\{B_s : s \le T\}$ coincide after completion:

$$\overline{\widetilde{\mathcal{F}}_T} = \overline{\mathcal{F}_T}.$$

Therefore, for all L^2 purposes (where random variables are identified up to P-a.s. equality), we may replace $\mathcal{F}_T$ by $\widetilde{\mathcal{F}}_T$, the smallest σ-algebra containing all $\mathcal{H}_n$.

Conditional expectations. For $g \in L^2(\Omega, \mathcal{F}_T, P)$ put

$$g_n := \mathbb{E}[g \mid \mathcal{H}_n].$$

By the martingale convergence theorem,

$$\boxed{g_n \to g \quad \text{in } L^2(\Omega)}. \tag{2}$$

Doob–Dynkin representation. Since g_n is $\mathcal{H}_n$-measurable, there exists a Borel map $\psi_n : \mathbb{R}^n \to \mathbb{R}$ with

$$g_n = \psi_n(B_{t_1}, \dots, B_{t_n}) \quad \text{a.s.} \tag{3}$$

Approximation in the Gaussian L^2 on $\mathbb{R}^n$. The vector $(B_{t_1}, \dots, B_{t_n})$ has a nondegenerate Gaussian density ρ_n on $\mathbb{R}^n$. Hence the map

$$\Phi \longmapsto \Phi(B_{t_1}, \dots, B_{t_n})$$

is an isometry from $L^2(\mathbb{R}^n, \rho_n dx)$ into $L^2(\Omega)$:

$$\mathbb{E}\big[\Phi(B_{t_1}, \dots, B_{t_n})^2\big] = \int_{\mathbb{R}^n} \Phi(x)^2 \rho_n(x)\, dx.$$

Because $C_c^\infty(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n, \rho_n dx)$, we can choose $\phi_{n,k} \in C_c^\infty(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \big|\phi_{n,k}(x) - \psi_n(x)\big|^2 \rho_n(x)\, dx \xrightarrow[k\to\infty]{} 0. \tag{4}$$

Using (3) and the isometry,

$$\begin{aligned}
\mathbb{E}\Big[\big(\phi_{n,k}(B_{t_1}, \dots, B_{t_n}) - g_n\big)^2\Big] &= \mathbb{E}\Big[\big(\phi_{n,k}(B_{t_1}, \dots, B_{t_n}) - \psi_n(B_{t_1}, \dots, B_{t_n})\big)^2\Big] \\
&= \int_{\mathbb{R}^n} \big|\phi_{n,k}(x) - \psi_n(x)\big|^2 \rho_n(x)\, dx \xrightarrow[k\to\infty]{} 0.
\end{aligned}$$

Thus for each n we can pick one index $k = k(n)$ so that

$$\boxed{\phi_{n,k(n)}(B_{t_1}, \dots, B_{t_n}) \;\to\; g_n \quad \text{in } L^2(\Omega)}. \tag{5}$$

Diagonal sequence and conclusion. Define

$$f_n := \phi_{n,k(n)}(B_{t_1}, \dots, B_{t_n}) \in \mathcal{S}.$$

By the triangle inequality together with (2) and (5),

$$\boxed{\|f_n - g\|_{L^2(\Omega)} \;\le\; \|f_n - g_n\|_{L^2(\Omega)} + \|g_n - g\|_{L^2(\Omega)} \xrightarrow[n\to\infty]{} 0}. \tag{6}$$

Hence $f_n \in \mathcal{S}$ converges to g in $L^2(\Omega)$, and the density follows. $\square$

Lemma 2 — Density of Exponential Martingale's Linear Span

Lemma 3.2 (Density of the linear span of exponential martingales)**.** *The linear span of*

$$\exp\left\{\int_0^T h(t)\,dB_t - \frac{1}{2}\int_0^T h(t)^2\,dt\right\}, \qquad h \in L^2([0,T]) \text{ deterministic,}$$

is dense in $L^2(\Omega, \mathcal{F}_T, P)$.

Proof. **Assume orthogonality.** Let $g \in L^2(\Omega, \mathcal{F}_T, P)$ satisfy

$$\mathbb{E}\left[g\exp\left\{\int_0^T h(t)\,dB_t - \tfrac{1}{2}\int_0^T h(t)^2\,dt\right\}\right] = 0 \quad \text{for all deterministic } h \in L^2([0,T]).$$

Choose h from (t,λ). Fix arbitrary $n \in \mathbb{N}$, times $t_1, \dots, t_n \in [0,T]$ and $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$. Define

$$h(t) := \sum_{k=1}^{n} \lambda_k \mathbf{1}_{[0,t_k]}(t).$$

Then

$$\int_0^T h(t)\,dB_t = \sum_{k=1}^{n} \lambda_k B_{t_k}, \qquad \int_0^T h(t)^2\,dt = \sum_{i,j=1}^{n} \lambda_i\lambda_j\,(t_i \wedge t_j) \geq 0.$$

Plugging this h into (3.1) gives

$$\mathbb{E}\left[g\exp\left\{\sum_{k=1}^{n}\lambda_k B_{t_k} - \tfrac{1}{2}\sum_{i,j=1}^{n}\lambda_i\lambda_j(t_i \wedge t_j)\right\}\right] = 0.$$

Definition of G and analytic continuation.
Multiplying by the positive constant $\exp\{\frac{1}{2}\sum_{i,j}\lambda_i\lambda_j(t_i \wedge t_j)\} > 0$ yields

$$G(\lambda) := \mathbb{E}\left[g\,e^{\sum_{k=1}^{n}\lambda_k B_{t_k}}\right] = 0 \quad \text{for all } \lambda \in \mathbb{R}^n \text{ and all } t_1, \dots, t_n \in [0,T]. \tag{†}$$

Each B_{t_k} is a centered Gaussian variable, so $e^{\sum \lambda_k B_{t_k}}$ has all exponential moments finite. Hence $G(\lambda)$ is a well-defined smooth function on $\mathbb{R}^n$.

Observe that $G(\lambda)$ is the *moment generating function (MGF)* of the random vector $(B_{t_1}, \dots, B_{t_n})$ weighted by g. In particular, $G(\lambda)$ encodes all mixed moments of g and the Brownian vector.

Since the exponential map $z \mapsto e^{zB_{t_k}}$ is entire for each B_{t_k}, we can extend the definition of G to complex arguments: for $z = (z_1, \dots, z_n) \in \mathbb{C}^n$, set

$$G(z) := \mathbb{E}\left[g\,e^{\sum_{k=1}^{n} z_k B_{t_k}}\right].$$

This G is a *holomorphic function* on $\mathbb{C}^n$, because the expectation of an entire function of Gaussian variables remains entire.

Since we already know that $G(\lambda) = 0$ for all real $\lambda \in \mathbb{R}^n$. By the **identity theorem for holomorphic functions**, the only holomorphic function that vanishes on $\mathbb{R}^n$ (which has an accumulation point) is the zero function. Hence

$$G(z) \equiv 0 \qquad \text{for all } z \in \mathbb{C}^n. \tag{$G \equiv 0$}$$

Fourier inversion and appearance of the imaginary argument.
Let $\phi \in C_c^\infty(\mathbb{R}^n)$ be an arbitrary test function. Its inverse Fourier representation is

$$\phi(x) = (2\pi)^{-n/2}\int_{\mathbb{R}^n}\widehat{\phi}(\xi)e^{i\langle x,\xi\rangle}\,d\xi, \qquad \widehat{\phi}(\xi) = (2\pi)^{-n/2}\int_{\mathbb{R}^n}\phi(x)e^{-i\langle x,\xi\rangle}\,dx.$$

Taking expectation with respect to g and the Brownian vector gives

$$\int_\Omega \phi(B_{t_1},\dots,B_{t_n})\, g\, dP = (2\pi)^{-n/2}\int_{\mathbb{R}^n} \widehat{\phi}(\xi)\mathbb{E}\big[g\, e^{i\sum_{k=1}^n \xi_k B_{t_k}}\big]\, d\xi$$
$$= (2\pi)^{-n/2}\int_{\mathbb{R}^n} \widehat{\phi}(\xi)\, G(i\xi)\, d\xi.$$

Why the argument becomes $i\xi$: in the Fourier representation the factor $e^{i\langle x,\xi\rangle}$ corresponds to evaluating G at purely imaginary vectors $z = i\xi$. That is, the Fourier transform probes the function G along the imaginary axis of its complex extension.

Since $G(z) \equiv 0$ on $\mathbb{C}^n$, in particular $G(i\xi) = 0$ for all $\xi \in \mathbb{R}^n$. Therefore,

$$\int_\Omega \phi(B_{t_1},\dots,B_{t_n})\, g\, dP = 0. \tag{$*$}$$

Orthogonality to a Dense Set and Hilbert-Space Conclusion.
Equation ($*$) shows that

$$\mathbb{E}\big[g\,\phi(B_{t_1},\dots,B_{t_n})\big] = 0, \qquad \forall\,\phi \in C_c^\infty(\mathbb{R}^n),\ t_1,\dots,t_n \in [0,T].$$

Hence, g is orthogonal to every random variable of the form $\phi(B_{t_1},\dots,B_{t_n})$ where ϕ is a smooth compactly supported function. By Lemma 3.1, the collection of all such random variables

$$\mathcal{S} = \big\{\phi(B_{t_1},\dots,B_{t_n}) : \phi \in C_c^\infty(\mathbb{R}^n),\, n\in\mathbb{N},\, t_k \in [0,T]\big\}$$

is dense in the Hilbert space $L^2(\Omega,\mathcal{F}_T,P)$. That is, the closure of $\mathcal{S}$ in the L^2 norm equals the whole space:

$$\overline{\mathcal{S}} = L^2(\Omega,\mathcal{F}_T,P).$$

Therefore, since g is orthogonal to every element of $\mathcal{S}$, it must also be orthogonal to its closure. In Hilbert-space terms, this means

$$g \in \mathcal{S}^\perp.$$

But because $\overline{\mathcal{S}} = L^2$, we have

$$\mathcal{S}^\perp = (\overline{\mathcal{S}})^\perp = (L^2)^\perp = \{0\}.$$

Hence

$$g = 0 \text{ in } L^2(\Omega,\mathcal{F}_T,P).$$

To visualize the logic:

$$g \perp \underbrace{\mathcal{S}}_{\text{dense in } L^2} \implies \mathcal{S}^\perp = (\overline{\mathcal{S}})^\perp = (L^2)^\perp = \{0\} \implies g = 0.$$

In summary,

$$\boxed{g \perp \mathcal{S} := \{\phi(B_{t_1},\dots,B_{t_n})\} \underset{\text{dense}}{\subset} L^2(\Omega,\mathcal{F}_T,P) \implies g = 0 \text{ in } L^2(\Omega,\mathcal{F}_T,P).}$$

Now, recall that at the beginning of this lemma we assumed that

$$g \perp \mathcal{H}, \qquad \text{where } \mathcal{H} = \operatorname{span}\left\{\exp\left(\int_0^T h(t)\, dB_t - \tfrac{1}{2}\int_0^T h(t)^2\, dt\right) : h \in L^2([0,T])\right\}.$$

From the above argument we have proved that this assumption forces $g = 0$. Equivalently, the only element orthogonal to $\mathcal{H}$ is the zero element. Thus,

$$\mathcal{H}^\perp = \{0\}.$$

Next, we invoke the fundamental identity of orthogonal complements in a Hilbert space:

$$(\mathcal{H}^{\perp})^{\perp} = \overline{\mathcal{H}}.$$

Combining this with $\mathcal{H}^{\perp} = \{0\}$ gives

$$\overline{\mathcal{H}} = (\mathcal{H}^{\perp})^{\perp} = \{0\}^{\perp} = L^2(\Omega, \mathcal{F}_T, P).$$

Conclusion. The closure of the linear span $\mathcal{H}$ of exponential martingale random variables is the entire space $L^2(\Omega, \mathcal{F}_T, P)$. That is,

$$\boxed{\overline{\mathcal{H}} = L^2(\Omega, \mathcal{F}_T, P).}$$

Therefore, the linear span of exponential martingales is dense in $L^2(\Omega, \mathcal{F}_T, P)$.

4 Appendix

Completion of a σ-Algebra

Given a probability space $(\Omega, \mathcal{F}, P)$, the **completion** of $\mathcal{F}$ is obtained by adding all P-null sets:

$$\overline{\mathcal{F}} := \sigma(\mathcal{F} \cup \mathcal{N}), \qquad \mathcal{N} := \{N \subset \Omega : P(N) = 0\}.$$

In words, $\overline{\mathcal{F}}$ contains every event in $\mathcal{F}$ and every subset of an event of probability zero.

This matters because random variables are considered equal in $L^2(\Omega, \mathcal{F}, P)$ if they differ only on a P-null set. Working with the completed σ-algebra ensures that such variables are automatically measurable.

Brownian Motion and the Completion of $\mathcal{F}_T$

Let $(B_t)_{t \geq 0}$ be a standard Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$.

Definition of $\mathcal{F}_T$. For a fixed $T > 0$, we define the σ-algebra generated by the Brownian path up to time T as

$$\mathcal{F}_T := \sigma\{B_s : 0 \leq s \leq T\}.$$

It represents the collection of all events whose occurrence depends only on the values of the process up to time T.

Countable dense time set. Let $D = \{t_i\}_{i \geq 1}$ be a countable dense subset of $[0, T]$ (for example, the rationals in $[0, T]$), and define

$$\widetilde{\mathcal{F}}_T := \sigma\{B_t : t \in D\}.$$

This is the σ-algebra generated by the Brownian motion evaluated at those countably many times.

Comparison of $\mathcal{F}_T$ and $\widetilde{\mathcal{F}}_T$. Clearly $\widetilde{\mathcal{F}}_T \subset \mathcal{F}_T$ because $D \subset [0, T]$. To show that these two σ-algebras are "the same for L^2 purposes," we use the fact that Brownian motion has continuous sample paths.

For every $s \in [0, T]$, we can find a sequence $(t_i) \subset D$ such that $t_i \to s$. By path continuity,

$$B_{t_i}(\omega) \to B_s(\omega) \quad \text{for all } \omega \text{ outside a } P\text{-null set.}$$

Hence B_s is the almost sure limit of $\{B_{t_i}\}_{t_i \in D}$.

Since B_{t_i} are all $\widetilde{\mathcal{F}}_T$-measurable, the limit B_s is measurable with respect to the *completion* of $\widetilde{\mathcal{F}}_T$, that is,

$$\overline{\widetilde{\mathcal{F}}_T} := \sigma(\widetilde{\mathcal{F}}_T \cup \mathcal{N}), \quad \mathcal{N} := \{N \subset \Omega : P(N) = 0\}.$$

This shows that every B_s for $s \leq T$ is measurable with respect to $\overline{\widetilde{\mathcal{F}}_T}$, and therefore

$$\mathcal{F}_T \subset \overline{\widetilde{\mathcal{F}}_T}.$$

The reverse inclusion $\widetilde{\mathcal{F}}_T \subset \mathcal{F}_T$ is immediate, and adding null sets does not change that direction, so

$$\overline{\widetilde{\mathcal{F}}_T} \subset \overline{\mathcal{F}_T}.$$

Combining the two gives the equality of completions:

$$\boxed{\overline{\widetilde{\mathcal{F}}_T} = \overline{\mathcal{F}_T}.}$$

Interpretation. The two σ-algebras may differ on a P-null set, but their completions coincide. In $L^2(\Omega, \mathcal{F}_T, P)$, random variables are identified up to null sets, so these two -algebras define the same L^2 space. Hence, for all L^2 or Itô-integral constructions, we can safely replace $\mathcal{F}_T$ by the smaller, countably generated $\widetilde{\mathcal{F}}_T$.