Stochastic analysis

A collection of definitions, theorems, and formulae

March 2, 2026

Contents

1 Stochastic analysis fundamentals

1.1 Measureable space

A *measureable space* is a pair $(\Omega, \mathcal{F})$ where

- Ω is the set of all possible outcomes of an experiment, for example $\Omega = \{H,\ T\}$ for a single two-sided coin flip. Each element $\omega \in \Omega$ is called a *sample point.*
- $\mathcal{F}$ is a σ-algebra, which is a collection of subsets of Ω with the following properties:
  (i) $\Omega \in \mathcal{F}$
  (ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, where $A^c = \Omega \setminus A$ is the complement of A in Ω
  (iii) If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

1.2 Probability measure

A *probability measure* is a function on a measureable space $(\Omega, \mathcal{F})$ where

$$P : \mathcal{F} \to [0, 1]$$

that satisfies:

(a) $P(\emptyset) = 0, \quad P(\Omega) = 1.$

(b) (σ-additivity) If $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint (i.e. $A_i \cap A_j = \varnothing$ for $i \neq j$), then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

We interpret $P(F) =$ "the probability that the event $F \in \mathcal{F}$ occurs"

1.3 Probability space

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ consisting of a set of outcomes Ω, a σ-algebra $\mathcal{F}$, and a probability measure P.

1.4 The Borel σ-algebra

If $\Omega = \mathbb{R}^n$, then one such σ-algebra supported on Ω is the Borel σ-algebra $\mathcal{B}$ which contains all open sets, all closed sets, all countable unions of closed sets, all countable intersections of such countable unions, etc.

1.5 $\mathcal{F}$-measureable functions

If $(\Omega, \mathcal{F}, P)$ is a given probability space, then a function $Y : \Omega \to \mathbb{R}^n$ is called $\mathcal{F}$-measureable if

$$Y^{-1}(U) = \{\omega \in \Omega : \; Y(\omega) \in U\}$$

for all open sets $U \subset \mathbb{R}^n$, or equivalently for all Borel sets $U \in \mathcal{B}$.

1.6 Random variables

If $(\Omega, \mathcal{F}, P)$ is a given probability space, then a *random variable* X is a $\mathcal{F}$-measureable function $X : \Omega \to \mathbb{R}^n$. Every random variable X induces a probability measure μ_X on $\mathbb{R}^n$ using the probability measure P defined on $\mathcal{F}$, given as

$$\mu_X(B) = P(X^{-1}(B))$$

and μ_X is called the *distribution of* X. The expectation is

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

and more generally if $f : \mathbb{R}^n \to \mathbb{R}$ is Borel measureable, then

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

1.7 L^p-norms and L^p-spaces of a random variable

If $X : \Omega \to \mathbb{R}^n$ is a random variable and $p \in [1, \infty)$ is a constant, then then L^p-norm of X is

$$\|X\|_p = \left(\int_\Omega |X(\omega)|^p dP(\omega)\right)^{\frac{1}{p}}.$$

Conversely, the L^p-space is the set of random variables with finite L^p-norm,

$$L^p(P) = L^p(\Omega) = \{X : \Omega \to \mathbb{R}^n \mid \|X\|_p < \infty\}.$$

In particular, for a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ writting

$$X \in L^2(\Omega) \qquad \Longleftrightarrow \qquad \mathbb{E}[\|X\|^2] < \infty$$

since X is a random variable with finite second moment $\mathbb{E}[\|X\|^2] < \infty$.

1.8 Independence of subsets

Two subsets $A, B \in \mathcal{F}$ are called *independent* if

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

1.9 Independence of random variables

Two random variables $X, Y : \Omega \to \mathbb{R}$ are independent if

$$E[XY] = E[X]E[Y]$$

1.10 Stochastic process

A *stochastic process* is a parameterized collection of random variables

$$\{X_t\}_{t \in T}$$

parameterized $X_t : \Omega \to \mathbb{R}$ on a given probability space $(\Omega, \mathcal{F}, P)$. For each fixed $t \in T$ we have a random variable

$$\omega \mapsto X_t(\omega) \qquad \omega \in \Omega.$$

On the other hand, fixing $\omega \in \Omega$ (commonly conceptualized as a particle)and considering the function

$$t \mapsto X_t(\omega) \qquad t \in T$$

produces what is called a *path of* X_t. Lastly, it is sometimes useful to write $X(t, \omega)$ and considering the function of two variables

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

which is a natural point of view since we will require that $X(t, \omega)$ be jointly measureable in t, ω.

1.11 Finite-dimensional distributions of a $\{X_t\}_{t \in T}$

A *(finite-dimensional) distribution* of the process $X = \{X_t\}_{t \in T}$ is a measure $\mu_{t_1,\ldots,t_k}$ defined on $\mathbb{R}^{nk}$, $k = 1, 2, \ldots$, by

$$\mu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = P[X_{t_1} \in F_1,\ \ldots,\ X_{t_k} \in F_k], \qquad t_i \in T.$$

Here $F_1, \ldots, F_k$ denote Borel sets in $\mathbb{R}^n$. Think of finite-dimensional distributions as anchoring the stochastic process at k points in time $t_1, \ldots, t_k$ using $X_{t_i} \in F_i$.

1.12 Kolmogorov's extension theorem

Given a family $\{\nu_{t_1,\ldots,t_k};\ k \in \mathbb{N},\ t_i \in T\}$ of probability measures on $\mathbb{R}^{nk}$, it is important to be able to construct a stochastic process $Y = \{Y_t\}_{t \in T}$ having $\nu_{t_1,\ldots,t_k}$ as its finite-dimensional distributions. One of Kolmogorov's famous theorems states that this can be done provided the family $\{\nu_{t_1,\ldots,t_k}\}$ satisfies two natural consistency conditions

Theorem 1.1 (Kolmogorov's extension theorem)**.** *For all* $t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$, *let* $\nu_{t_1,\ldots,t_k}$ *be probability measures on* $\mathbb{R}^{nk}$ *such that*

$$\nu_{t_{\sigma(1)},\ldots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\ldots,t_k}\big(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}\big), \tag{K1}$$

for all permutations σ *of* $\{1, 2, \ldots, k\}$, *and*

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}\big(F_1 \times \cdots \times F_k \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n\big), \tag{K2}$$

for all $m \in \mathbb{N}$, *where the set on the right-hand side has a total of* $k + m$ *factors.*

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ *and a stochastic process* $\{X_t\}$ *on* Ω, *with* $X_t : \Omega \to \mathbb{R}^n$, *such that*

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = P\big[X_{t_1} \in F_1,\ \ldots,\ X_{t_k} \in F_k\big],$$

for all $t_i \in T$, $k \in \mathbb{N}$, *and all Borel sets* F_i.

2 Brownian motion $\{B_t(\omega)\}_{t\in T}$

2.1 What is B_t?

Brownian motion is simply a stochastic process, where a stochastic process $\{B_t\}$ is simply a parameterized collection of random variables $B_t : \Omega \to \mathbb{R}^n$ for each time t. For a fixed time t, B_t is a random variable which maps a particle to a position: namely $\omega \in \Omega \;\mapsto\; B_t(\omega) \in \mathbb{R}^n$. On the other hand, for a fixed particle $\omega \in \Omega$, then $t \geq 0 \;\mapsto\; B_\cdot(\omega)$ is a whole path (random trajectory in $\mathbb{R}^n$).

To rigourously define Brownian motion $\{B_t\}$, we'll use Kolmogorov's extension theorem, which allows us to construct a stochastic process $\{B_t\}$ simply by defining the joint law of the random vector $(B_{t_1}, \ldots, B_{t_k})$ using the finite-dimensional distributions $\nu_{t_1,\ldots,t_k}$ for all k.

2.2 Constructing Brownian motion

To define the joint law of the random vector $(B_{t_1}, \ldots, B_{t_k})$ using the finite-dimensional distributions $\nu_{t_1,\ldots,t_k}$ for all k, first, fix $x_0 \in \mathbb{R}^n$ and define the heat kernel

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

with the convention $p(0, x, y)\, dy = \delta_x(y)$. Now fix a particle ω, any $k \in \mathbb{N}$ and set times $0 = t_0 \leq t_1 \leq \cdots \leq t_k$, and Borel sets $F_i \in \mathcal{B}(\mathbb{R}^n)$. Then define the measure $\nu_{t_1,\ldots,t_k}$ on $\mathbb{R}^{nk}$ by

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) := \int_{F_1\times\cdots\times F_k} \prod_{i=1}^{k} p(t_i - t_{i-1}, x_{i-1}, x_i)\, dx_1 \cdots dx_k.$$

These measures satisfy the consistency conditions (K1) and (K2), i.e. whenever you integrate out (marginalize) some coordinates you recover the lower-dimensional distribution for the remaining times. In particular,

$$\int_{\mathbb{R}^n} p(t, x, y)\, dy = 1 \quad \text{for all } t \geq 0,$$

which ensures the kernels define probability transitions and makes the $\nu_{t_1,\ldots,t_k}$ probability measures. Therefore, by Kolmogorov's extension theorem, there exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t\}_{t\geq 0}$ such that for every $k \geq 1$, every choice of times $0 \leq t_1 \leq \cdots \leq t_k$, and every Borel sets $F_1, \ldots, F_k \subseteq \mathbb{R}^n$,

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

where the event

$$\{B_{t_1} \in F_1, \ldots, B_{t_k} \in F_k\}$$

means that "at time t_i the position $B_{t_i}(\omega) \in \mathbb{R}^n$ falls inside the *region/set* $F_i \subseteq \mathbb{R}^n$ simultaneously for all $i = 1, \ldots, k$." Thus, Brownian motion is defined by specifying *all* these finite-dimensional distributions $\nu_{t_1,\ldots,t_k}$ at once (for all k and all time-lists $t_1, \ldots, t_k$). Therefore, we do *not* define a single ν for one fixed k only; rather, we define a *family* $\{\nu_{t_1,\ldots,t_k}\}_{k\geq 1,\ 0\leq t_1\leq\cdots\leq t_k}$, and require these measures to be mutually consistent. They are defined so that the process starts at x_0 and then makes successive "Gaussian" transitions. From Kolmogorov's extension theorem we obtain the stochastic process $\{B_t\}_{t\geq 0}$, which is called *n-dimensional Brownian motion* starting at x_0.

2.3 Filtration $\mathcal{F}_t$ of a stochastic process

The *natural filtration* generated by the process $\{B_t\}$ up to time t is

$$\mathcal{F}_t := \sigma(B_s : 0 \le s \le t).$$

This means: $\mathcal{F}_t$ is the *smallest* σ-algebra on Ω such that random variable in the stochastic process up to time t

$$B_s : \Omega \to \mathbb{R}^n, \qquad 0 \le s \le t,$$

is $\mathcal{F}_t$-measurable. Intuitively, $\mathcal{F}_t$ is exactly the information revealed by observing the path up to time t Equivalently,

$$\mathcal{F}_t = \sigma\Big(\{B_s \in A\} : 0 \le s \le t,\ A \in \mathcal{B}(\mathbb{R}^n)\Big).$$

where $\{B_s \in A\}$ is an *event* (a subset of Ω), defined by

$$\{B_s \in A\} := \{\omega \in \Omega :\ B_s(\omega) \in A\},$$

i.e. the set of outcomes (particles) for which the position at time s falls inside the region $A \subseteq \mathbb{R}^n$. Thus $\mathcal{F}_t$ is a collection of events in Ω. For example, if $t_1 \le t$ and $F_1 \in \mathcal{B}(\mathbb{R}^n)$, then

$$\{B_{t_1} \in F_1\} \in \mathcal{F}_t$$

means: the question “did the path land in F_1 at time t_1?” can be answered using only information available up to time t (since $t_1 \le t$). More generally, for $0 \le t_1 \le \cdots \le t_k \le t$,

$$\{B_{t_1} \in F_1, \dots, B_{t_k} \in F_k\} \in \mathcal{F}_t,$$

so any statement depending on finitely many observations up to time t is measurable with respect to $\mathcal{F}_t$.

2.4 Property I: Gaussian process structure

Let $\{B_t\}_{t \ge 0}$ be the n-dimensional Brownian motion constructed via its familiy of finite-dimensional distributions $\nu_{t_1,\dots,t_k}$ on the probability space $(\Omega, \mathcal{F}, P^x)$ and initialized at $x \in \mathbb{R}^n$. We now explain why Brownian motion is a Gaussian process and compute its mean and covariance.

A stochastic process is a *Gaussian process* if for every finite collection of times

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

the random vector

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

is multivariate normal. Here, a random vector Z is *multivariate normal* with mean $M \in \mathbb{R}^{nk}$ and covariance matrix $C \in \mathbb{R}^{nk \times nk}$ if its characteristic function satisfies

$$E^x\Big[e^{i\langle u, Z\rangle}\Big] = \exp\Big(-\tfrac{1}{2}u^\top C u + i\langle u, M\rangle\Big) \qquad \text{for all } u \in \mathbb{R}^{nk}.$$

For Brownian motion, the finite-dimensional distributions $\nu_{t_1,\dots,t_k}$ imply

$$E^x[B_{t_\ell}] = x \quad \text{for all } t_\ell \ge 0$$

hence

$$M = E^x[Z] = (x, x, \dots, x) \in \mathbb{R}^{nk}.$$

To compute the covariance, fix $\ell, r \in \{1, \dots, k\}$ and assume without loss of generality that $t_\ell \le t_r$. Write

$$B_{t_r} = B_{t_\ell} + (B_{t_r} - B_{t_\ell}).$$

By independent increments, $B_{t_r} - B_{t_\ell}$ is independent of B_{t_ℓ} and has mean 0, so

$$\begin{aligned}
\mathrm{Cov}(B_{t_\ell}, B_{t_r}) &= E^x\big[(B_{t_\ell} - x)(B_{t_r} - x)^\top\big] \\
&= E^x\big[(B_{t_\ell} - x)(B_{t_\ell} - x)^\top\big] + E^x\big[(B_{t_\ell} - x)(B_{t_r} - B_{t_\ell})^\top\big] \\
&= \mathrm{Var}(B_{t_\ell}) + 0 \\
&= t_\ell I_n.
\end{aligned}$$

Since the argument is symmetric in (ℓ, r), this yields

$$\mathrm{Cov}(B_{t_\ell}, B_{t_r}) = \min(t_\ell, t_r)\, I_n.$$

so covariance between any two components is determined by the earlier time. Therefore, the covariance matrix of Z has the block form

$$C = \begin{pmatrix} t_1 I_n & t_1 I_n & \cdots & t_1 I_n \\ t_1 I_n & t_2 I_n & \cdots & t_2 I_n \\ \vdots & \vdots & \ddots & \vdots \\ t_1 I_n & t_2 I_n & \cdots & t_k I_n \end{pmatrix}.$$

With this M and C from the finite dimensional distributions $\nu_{t_1,\dots,t_k}$, the characteristic function identity holds, therefore Brownian motion is a Gaussian process.

2.5 Property II: Independent increments

The process B_t has independent increments, i.e.

$$B_{t_1},\ B_{t_2} - B_{t_1},\ \dots,\ B_{t_k} - B_{t_{k-1}} \text{ are independent for all } 0 \le t_1 < t_2 < \cdots < t_k. \tag{2.2.11}$$

Since Brownian motion is a Gaussian process, independence is equivalent to being uncorrelated, which is not true for general non-gaussian random variables. Thus it suffices to show that that correlation

$$E^x\Big[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\Big] = 0 \quad \text{whenever } t_i < t_j. \tag{2.2.12}$$

Using the covariance structure $E^x[B_s B_t] = n \min(s, t)$, we compute

$$\begin{aligned}
E^x\Big[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\Big] &= E^x\Big[B_{t_i}B_{t_j} - B_{t_{i-1}}B_{t_j} - B_{t_i}B_{t_{j-1}} + B_{t_{i-1}}B_{t_{j-1}}\Big] \\
&= n\big(t_i - t_{i-1} - t_i + t_{i-1}\big) \\
&= 0.
\end{aligned}$$

Therefore, the increments are independent. In particular, for $s > t$, the increment $B_s - B_t$ is independent of the natural filtration $\mathcal{F}_t$ (path history).

2.6 Property III: Continuity from Kolmogorov's continuity theorem

A natural question is whether the sample path $t \mapsto B_t(\omega)$ is continuous for "almost every" particle ω. Stated naively, this question does not make sense because the set

$$H = \{\omega :\ t \mapsto B_t(\omega) \text{ is continuous on } [0, \infty)\}$$

need not be measurable with respect to the Borel σ-algebra $(\mathbb{R}^n)^{[0,\infty)}$. Instead, the right way to ask this question is slightly different: not whether the originally constructed stochastic process has continuous paths as written, but whether we can replace it by an equivalent stochastic process (in distribution) that does have continuous paths. To make this precise, we use the notion of a *version* (also called a modification).

Given two processes $\{X_t\}$ and $\{Y_t\}$ on the same probability space $(\Omega, \mathcal{F}, P)$, we say $\{X_t\}$ is a *version* of $\{Y_t\}$ if for every fixed t,

$$P(\{\omega : \; X_t(\omega) = Y_t(\omega)\}) = 1.$$

where we recall that P is the measure $P : \Omega \to \mathbb{R}$. That is, at each time t, the two random variables agree almost surely. If X is a version of Y, then X and Y have the same finite-dimensional distributions, so from the viewpoint of all joint laws $(X_{t_1}, \ldots, X_{t_k})$ they are indistinguishable; however, their pathwise properties (like continuity) can differ.

The continuity issue for Brownian motion is resolved using Kolmogorov's continuity theorem. It gives a practical moment condition that guarantees the existence of a continuous version: if a process $X = \{X_t\}_{t \geq 0}$ satisfies that for every $T > 0$ there exist constants $\alpha, \beta, D > 0$ such that

$$\mathbb{E}\big[|X_t - X_s|^\alpha\big] \leq D\,|t - s|^{1+\beta}, \qquad 0 \leq s, t \leq T,$$

then one can always construct another process $\widetilde{X}$ which is a version of X and whose sample paths $t \mapsto \widetilde{X}_t(\omega)$ are continuous. In particular, one applies this theorem to the process built from finite-dimensional distributions to obtain a continuous-path version, which is the usual Brownian motion used in analysis and SDEs.

3 The Ito integral

3.1 The white noise process

Suppose we have a process of the form

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{“noise”} \tag{1}$$

then a reasonable interpretation of this “noise” term is a stochastic process W_t which obeys the naturally desireable properties:

$$t_1 \neq t_2 \;\Rightarrow\; W_{t_1} \text{ and } W_{t_2} \text{ independent}, \qquad \{W_t\} \text{ stationary}, \qquad \mathbb{E}[W_t] = 0.$$

However, no such stochastic process W_t with these properties can have continuous paths. An intuitive reason is that

- if W_t had *continuous paths*, then for a fixed ω the values $W_{t_n}(\omega)$ at times $t_n \to t$ must satisfy
$$W_{t_n}(\omega) \to W_t(\omega),$$
hence nearby times force the random variables to be strongly related (they must become close on each path).
- However, the condition “W_{t_1} and W_{t_2} are independent whenever $t_1 \neq t_2$” says the opposite: even arbitrarily close times give *independent* random variables, meaning there is no relationship between W_{t_n} and W_t coming from closeness of times.

Instead, the solution to the problem above where we require continuity in addition to the natural properties of noise, is to considering a rewriting of (2)

$$dX = b(t, X_t)dt + \sigma(t, X_t)W_t dt \tag{2}$$

and set $W_t\, dt$ equal to the increment of Brownian motion:

$$W_t\, dt := dB_t.$$

3.2 Recap on Brownian motion $\{B_t\}$

The process $B = \{B_t\}_{t\geq 0}$ is a *standard Brownian motion* (also called a *Wiener process*). Concretely, $\{B_t\}$ is a stochastic process such that:

- $B_0 = 0$.
- (*Independent increments*) If $0 \leq s < t$, then the increment $B_t - B_s$ is independent of the past $\{B_u : u \leq s\}$.
- (*Stationary increments*) $B_t - B_s$ has the same distribution as B_{t-s}.
- (*Gaussian increments*) For $0 \leq s < t$,
$$\Delta B := B_t - B_s \sim \mathcal{N}(0, t - s).$$
- (*Continuous paths*) With probability 1, the map $t \mapsto B_t(\omega)$ is continuous.

A very important fact is that Brownian paths are continuous but *extremely rough*: with probability 1 they are nowhere differentiable.

3.3 SDEs definition overview

Recall that we seek to give meaning to the process

$$dX_t \;=\; b(t, X_t)\, dt \;+\; \sigma(t, X_t)\, dB_t. \tag{3}$$

However, this equation is *not* defined by treating dB_t as an ordinary derivative (since B_t is not differentiable). Instead, one defines the path $X_t = X_t(\omega)$ in an SDE as the X_t that satisfies an integral equation involving an *Itô integral*. A convenient way to remember the meaning is that the differential notation of the Itô SDE

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

is just shorthand for defining a stochastic process $X_t = X_t(\omega)$ that satisfies the integral equation

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

Hence our new goal is to define this stochastic integral above.

3.4 Natural filtration of a random variable and measurablility

For d-dimensional Brownian motion $B_t = (B_1(t), \dots, B_d(t))$, its (completed) natural filtration is

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

A random variable $h(\omega)$ is $\mathcal{F}_t$-measurable if and only if it is determined by the path $\{B_s(\omega) : s \le t\}$.

3.5 Adapted random variables

More generally, for a filtration $\{\mathcal{N}_t\}$, a process

$$g : [0, \infty) \times \Omega \to \mathbb{R}^m$$

is $\mathcal{N}_t$-*adapted* if $g(t, \cdot)$ is $\mathcal{N}_t$-measurable for each t. With $\mathcal{N}_t = \mathcal{F}_t$, adaptedness means the process does not "look into the future" of the Brownian path.

3.6 Filtered probability space and Brownian motion

Fix a filtered probability space

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \ge 0}, \mathbb{P})$$

satisfying the usual conditions. Let $\{B_t\}_{t \ge 0}$ be an $\{\mathcal{F}_t\}$-Brownian motion (possibly $\mathbb{R}^d$-valued), meaning:

- B_t is $\mathcal{F}_t$-measurable for each t (so B is adapted),
- for $0 \le s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$,
- $B_t - B_s \sim \mathcal{N}(0, (t - s) I_d)$, and B has continuous paths.

Intuitively, $\mathcal{F}_t$ represents the information revealed up to time t; increments after time t add new randomness independent of that information.

3.7 Why we cannot use ordinary Riemann–Stieltjes integrals

For a general integrand $f(t, \omega)$ we would like to define the following integral with respect to time $t \in [S, T]$

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

as a limit of Riemann-type sums

$$\sum_j f(t_j^*,\omega)\,(B_{t_{j+1}} - B_{t_j})(\omega), \qquad \Delta t = t_{j+1} - t_j = 0.$$

Unlike the classical Riemann–Stieltjes setting, the choice of sample points $t_j^* \in [t_j, t_{j+1}]$ matters here. Two conventions are fundamental:

- **Itô convention:** $t_j^* = t_j$ (left endpoints), producing $\int f\,dB$.
- **Stratonovich convention:** $t_j^* = \frac{t_j + t_{j+1}}{2}$ (midpoints), producing $\int f \circ dB$.

In these notes we focus on the Itô (left-endpoint) convention, which matches the non-anticipating requirement for f.

3.8 The class of square-integrable adapted integrands

Fix times $S < T$. We will define $\int_S^T f\,dB$ when f is adapted and square-integrable.

Definition: (Square-integrable adapted integrands) Let $\mathcal{V} = \mathcal{V}(S,T)$ be the set of all processes

$$f : [S,T] \times \Omega \to \mathbb{R}$$

such that:

(i) f is $\mathcal{B}([S,T]) \times \mathcal{F}$-measurable,
(ii) f is $\{\mathcal{F}_t\}_{t\in[S,T]}$-adapted (i.e. $f(t,\cdot)$ is $\mathcal{F}_t$-measurable for each t),
(iii) $\mathbb{E}\Big[\int_S^T f(t,\omega)^2\,dt\Big] < \infty$.

Condition (ii) is the non-anticipating requirement that aligns with the left-endpoint Itô convention.

3.9 Elementary (step) processes

(TO DO: refine this definition to match the $\phi_n \in \mathcal{V}(S,T)$ we use below. What is n here?) The Itô integral is first defined for adapted step processes, then extended by approximation to all $f \in \mathcal{V}(S,T)$. An *elementary* (or *step*) process on $[S,T]$ is a process of the form

$$\phi(t,\omega) = \sum_{j=0}^{m-1} \xi_j(\omega)\,\mathbf{1}_{(t_j,t_{j+1}]}(t), \qquad S = t_0 < t_1 < \cdots < t_m = T,$$

where each coefficient ξ_j is $\mathcal{F}_{t_j}$-measurable.

3.10 Definition (Itô integral for elementary processes).

For such ϕ, define

$$\int_S^T \phi(t,\omega)\,dB_t(\omega) := \sum_{j=0}^{m-1} \xi_j(\omega)\,(B_{t_{j+1}} - B_{t_j})(\omega).$$

This "drops" the indicator $\mathbf{1}_{(t_j,t_{j+1}]}$ because ϕ is *constant in time* on each interval $(t_j, t_{j+1}]$: the contribution of the interval is that constant value ξ_j multiplied by the corresponding Brownian increment over the same interval, $B_{t_{j+1}} - B_{t_j}$. Equivalently, this is exactly the left-endpoint Riemann-sum rule for $\int \phi\,dB$. It is well-behaved because ξ_j is measurable with respect to $\mathcal{F}_{t_j}$, while the increment $B_{t_{j+1}} - B_{t_j}$ is independent of $\mathcal{F}_{t_j}$ (by independent increments of Brownian motion).

3.11 Step functions approximate general functions

We now extend the definition from elementary processes to all $f \in \mathcal{V}(S,T)$.

Approximation principle. If $f \in \mathcal{V}(S,T)$, there exist elementary processes $\phi_n \in \mathcal{V}(S,T)$ such that

$$\mathbb{E}\left[\int_S^T |f(t,\omega) - \phi_n(t,\omega)|^2\,dt\right] \to 0 \qquad \text{as } n \to \infty.$$

The expression

$$\|f - \phi_n\|_{\mathcal{V}}^2 := \mathbb{E}\left[\int_S^T |f(t,\omega) - \phi_n(t,\omega)|^2\,dt\right]$$

is the squared L^2-distance between f and ϕ_n on $[S,T] \times \Omega$ (with respect to $dt \times \mathbb{P}$). Thus the statement says $\phi_n \to f$ in the $\mathcal{V}$-norm

$$\|g\|_{\mathcal{V}} := \Big(\mathbb{E}\int_S^T g(t,\omega)^2\,dt\Big)^{1/2}.$$

As n increases, we take "finer" approximations.

3.12 Definition of the Itô integral

Let $f \in \mathcal{V}(S,T)$ and choose any familiy of elementary approximations $\{\phi_n\}_{n\ge 1}$ satisfying

$$\mathbb{E}\left[\int_S^T (f - \phi_n)^2\,dt\right] \to 0 \qquad \text{as } n \to \infty.$$

Define

$$\int_S^T f(t,\omega)\,dB_t(\omega) := \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)\,dB_t(\omega),$$

where the limit is taken in $L^2(\mathbb{P})$. Concretely, "limit in $L^2(\mathbb{P})$" means

$$\mathbb{E}\left[\left(\int_S^T \phi_n\,dB - Y\right)^2\right] \to 0 \quad \text{for some random variable } Y,$$

and we then *define* Y to be $\int_S^T f\,dB$. The limit exists and does not depend on the particular approximating sequence.

3.13 Itô isometry and continuity of the integral map

A central identity is that the Itô integral preserves L^2 norms in a precise way.

Theorem (Itô isometry). For all $f \in \mathcal{V}(S,T)$,

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

In particular, the map

$$I : \mathcal{V}(S,T) \to L^2(\mathbb{P}), \qquad I(f) := \int_S^T f\,dB,$$

is continuous when $\mathcal{V}(S,T)$ is equipped with the norm $\|f\|_{\mathcal{V}} = \big(\mathbb{E}\int_S^T f^2 dt\big)^{1/2}$ and $L^2(\mathbb{P})$ is equipped with $\|Y\|_{L^2} = (\mathbb{E}[Y^2])^{1/2}$. Concretely, continuity here means: whenever $f_n \to f$ in

$\mathcal{V}(S,T)$ (i.e. $\|f_n - f\|_{\mathcal{V}} \to 0$), we also have

$$\left\| \int_S^T f_n \, dB - \int_S^T f \, dB \right\|_{L^2} \to 0 \qquad \text{as } n \to \infty.$$

Moreover, by the isometry,

$$\left\| \int_S^T f_n \, dB - \int_S^T f \, dB \right\|_{L^2} = \|f_n - f\|_{\mathcal{V}},$$

so I is actually an *isometry* (hence Lipschitz with constant 1) from $\mathcal{V}(S,T)$ into $L^2(\mathbb{P})$.

3.14 How this makes the SDE definition precise

Let $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ be coefficient functions, and let $B_t \in \mathbb{R}^d$ be Brownian motion. An unknown process X_t is typically required to be adapted and sufficiently integrable so that the integrals below exist (e.g. $\int_0^t \|\sigma(s, X_s)\|_F^2 ds$ has finite expectation for each t).

The Itô SDE

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

is *defined* by the requirement that for each $t \in [0,T]$,

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

where the first integral is an ordinary (pathwise) time integral and the second is the Itô integral defined above.

3.15 Key takeaway

The stochastic integral $\int f \, dB$ is constructed as an $L^2(\mathbb{P})$-limit of left-endpoint sums with adapted (non-anticipating) integrands.

3.16 Identities

Let $f, g \in \mathcal{V}(0,T)$ and let $0 \le S < U < T$. Then

1. $$\int_S^T f \, dB_t = \int_S^U f \, dB_t + \int_U^T f \, dB_t \qquad \text{for a.a. } \omega.$$

2. $$\int_S^T (cf + g) \, dB_t = c \cdot \int_S^T f \, dB_t + \int_S^T g \, dB_t \quad (c \text{ constant}) \qquad \text{for a.a. } \omega.$$

3. $$\mathbb{E}\left[\int_S^T f \, dB_t \right] = 0.$$

4. $$\int_S^T f \, dB_t \text{ is } \mathcal{F}_T\text{-measurable}.$$

3.17 Itô integrals are martingales

An important property of the Itô integral is that it is a *martingale*.

Definition (Martingales). A *filtration* (on $(\Omega, \mathcal{F})$) is a family $\mathcal{M} = \{\mathcal{M}_t\}_{t\geq 0}$ of σ-algebras $\mathcal{M}_t \subset \mathcal{F}$ such that

$$0 \leq s < t \;\Rightarrow\; \mathcal{M}_s \subset \mathcal{M}_t$$

(i.e. $\{\mathcal{M}_t\}$ is increasing). An n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, P)$ is called a *martingale* with respect to a filtration $\{\mathcal{M}_t\}_{t\geq 0}$ (and with respect to P) if

1. M_t is $\mathcal{M}_t$-measurable for all t,
2. $\mathbb{E}[\|M_t\|] < \infty$ for all t,
3. $\mathbb{E}[M_s \mid \mathcal{M}_t] = M_t \quad$ for all $s \geq t$.

Here the expectation in (2) and the conditional expectation in (3) are taken with respect to the measure $P = P^0$.

Lemma: (Ito integrals can be represented as a stochastic process) Let $f \in \mathcal{V}(0, T)$. Then there exists a t-continuous version of

$$\int_0^t f(s, \omega)\, dB_s(\omega), \qquad 0 \leq t \leq T,$$

i.e. there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, P)$ such that

$$P\left[J_t = \int_0^t f\, dB\right] = 1 \qquad \text{for all } t,\ 0 \leq t \leq T. \tag{4}$$

Theorem: (The stochastic process defined by the Ito integral is a martingale) Let $f(t, \omega) \in \mathcal{V}(0, T)$ for all T. Then

$$M_t(\omega) = \int_0^t f(s, \omega)\, dB_s$$

is a martingale w.r.t. $\mathcal{F}_t$.

4 The Ito Formula

At a high level, the Ito formula is the ito integral version of the chain rule, and is useful for evaluating ito integrals.

4.1 Ito process definition

Let B_t be 1-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$. A (1-dimensional) Itô process is a stochastic process X_t on $(\Omega, \mathcal{F}, \mathbb{P})$ of the form

$$X_t = X_0 + \int_0^t u(s,\omega)\, ds + \int_0^t v(s,\omega)\, dB_s\,. \tag{5}$$

where functions u and v are bounded and adapted. If X_t is of this form, we often use the shorthand SDE notation

$$X_t = u\, dt + v\, dBt.$$

4.2 Ito formula theorem (1-dimensional)

Theorem: Let X_t be an Itô process given by

$$dX_t = u\, dt + v\, dB_t$$

and let $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$ (i.e. g is twice continuously differentiable on $[0,\infty) \times \mathbb{R}$). Then

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

is again an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)\, dt + \frac{\partial g}{\partial x}(t, X_t)\, dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t) \cdot (dX_t)^2\,, \tag{6}$$

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

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, \qquad dB_t \cdot dB_t = dt\,. \tag{7}$$

4.3 Using the Itô formula to compute Itô integrals

A useful way to *compute* an Itô integral

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

is to manufacture an identity from Itô's formula in which this integral appears. If X_t has dynamics $dX_t = u_t\, dt + v_t\, dB_t$, then Itô's formula gives

$$d\big(g(t, X_t)\big) = \Big(g_t + u_t g_x + \tfrac{1}{2} v_t^2 g_{xx}\Big)(t, X_t)\, dt + v_t g_x(t, X_t)\, dB_t. \tag{8}$$

Thus the stochastic coefficient is always $v_t g_x(t, X_t)$. To make $\int_0^t f_s\, dB_s$ appear, the *design condition* is

$$v_t g_x(t, X_t) = f_t. \tag{9}$$

Step 1: pick X_t so that f_t becomes a function of (t, X_t). The condition (9) is easiest to satisfy when f_t can be written as

$$f_t = f(t, X_t)$$

for some explicit $f(t, X_t)$. This guides the choice of X:

- If f_t is a function of B_t (and possibly t), take $X_t = B_t$.
- If f_t involves a time-dependent shift/scale of B_t, take $X_t = \phi(t, B_t)$ (e.g. $X_t = B_t + \mu t$, $X_t = e^{-\lambda t} B_t$, etc.).
- More generally, choose an Itô process X whose dynamics you can control so that f_t can be expressed as a function of the current state X_t.

Step 2: having fixed X, choose g by integrating in x. Once $f_t = f(t, X_t)$ is achieved, a simple way to satisfy (9) is:

- choose v_t convenient (often $v_t \equiv 1$, i.e. $X_t = B_t$), and then set

$$g_x(t,x) = \frac{F(t,x)}{v_t} \quad (\text{usually } g_x(t,x) = F(t,x)),$$

- define g by integrating in x:

$$g(t,x) := \int^x \frac{F(t,y)}{v_t}\, dy,$$

  up to an additive function of t (which does not affect g_x).

After this choice, (8) becomes

$$d\big(g(t,X_t)\big) = f_t\, dB_t + \Big(g_t + u_t g_x + \tfrac{1}{2} v_t^2 g_{xx}\Big)(t, X_t)\, dt,$$

so integrating from 0 to t yields the bookkeeping identity

$$\int_0^t f_s\, dB_s = g(t,X_t) - g(0,X_0) - \int_0^t \Big(g_t + u_s g_x + \tfrac{1}{2} v_s^2 g_{xx}\Big)(s, X_s)\, ds.$$

In good examples, the remaining ds-integral is explicit (or at least simpler than the original).

Example: $\int_0^t B_s\, dB_s$. Here $f_s = B_s$ is already a function of B_s, so take $X_s = B_s$, i.e. $u_s = 0$, $v_s = 1$. We want $v_s g_x(s, X_s) = g_x(s, B_s) = B_s$, so choose $g_x(s,x) = x$, hence $g(s,x) = \frac{1}{2}x^2$. Then $Y_t = g(t, B_t) = \frac{1}{2}B_t^2$, and Itô's formula gives

$$dY_t = g_x(t,B_t)\, dB_t + \tfrac{1}{2} g_{xx}(t,B_t)\, dt = B_t\, dB_t + \tfrac{1}{2}\, dt.$$

Integrating,

$$\tfrac{1}{2}B_t^2 - \tfrac{1}{2}B_0^2 = \int_0^t B_s\, dB_s + \tfrac{1}{2}t,$$

and since typically $B_0 = 0$,

$$\int_0^t B_s\, dB_s = \tfrac{1}{2}B_t^2 - \tfrac{1}{2}t.$$

4.4 Integration by parts

Theorem: Suppose $f(s,\omega)$ is continuous and of bounded variation with respect to $s \in [0,t]$, for a.a. ω. Then

$$\int_0^t f(s)\, dB_s = f(t)B_t - \int_0^t B_s\, df_s\, .$$

4.5 Multi-dimensional Ito processes

We now turn to the situation in higher dimensions: Let $B(t,\omega) = (B_1(t,\omega), \ldots, B_m(t,\omega))$ denote m-dimensional Brownian motion. If each of the processes $u_i(t,\omega)$ and $v_{ij}(t,\omega)$ satisfies the conditions given in Definition 4.1.1 ($1 \le i \le n,\ 1 \le j \le m$) then we can form the following

n Itô processes

$$\begin{cases} dX_1 = u_1\,dt + v_{11}\,dB_1 + \cdots + v_{1m}\,dB_m, \\ \quad\vdots \\ dX_n = u_n\,dt + v_{n1}\,dB_1 + \cdots + v_{nm}\,dB_m. \end{cases} \tag{4.2.1}$$

Or, in matrix notation simply

$$dX(t) = u\,dt + v\,dB(t), \tag{4.2.2}$$

where

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad v = \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nm} \end{pmatrix}, \qquad dB(t) = \begin{pmatrix} dB_1(t) \\ \vdots \\ dB_m(t) \end{pmatrix}. \tag{4.2.3}$$

Such a process $X(t)$ is called an n**-dimensional Itô process** (or just an Itô process).

4.6 The multi-dimensional Ito formula

What is the result of applying a smooth function to the multi-dimensional stochastic process $X(t) \in \mathbb{R}^n$? The answer is given by

Theorem 4.1 (The general Itô formula)**.** *Let*

$$dX(t) = u\,dt + v\,dB(t)$$

be an n-dimensional Itô process as above. Let $g(t,x) = (g_1(t,x), \ldots, g_p(t,x))$ be a C^2 map from $[0,\infty) \times \mathbb{R}^n$ into $\mathbb{R}^p$. Then the process

$$Y(t,\omega) = g(t, X(t))$$

is again an Itô process, whose component number k, Y_k, is given by

$$dY_k = \frac{\partial g_k}{\partial t}(t,X)\,dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)\,dX_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 g_k}{\partial x_i\,\partial x_j}(t,X)\,dX_i\,dX_j,$$

where $dB_i\,dB_j = \delta_{ij}\,dt$, $dB_i\,dt = dt\,dB_i = 0$.

4.7 Martingale representation theorem

Let $B(t) = (B_1(t), \ldots, B_n(t))$ be n-dimensional. Suppose $(M_t)_{t\geq 0}$ is an $\mathcal{F}_t^{(n)}$-martingale (w.r.t. P) and that $M_t \in L^2(P)$ for all $t \geq 0$. Then there exists a unique stochastic process $g(s,\omega)$ such that $g \in \mathcal{V}^{(n)}(0,t)$ for all $t \geq 0$ and

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t g(s,\omega)\,dB(s) \qquad \text{a.s., for all } t \geq 0. \tag{10}$$

- Meaning: every square-integrable $\mathcal{F}_t^{(n)}$-martingale has no "drift" term; it is entirely generated by accumulating Brownian shocks through an adapted integrand g.
- Itô isometry intuition: the quadratic size of the martingale increment is the accumulated energy of g,

$$\mathbb{E}\Big[(M_t - \mathbb{E}[M_0])^2\Big] = \mathbb{E}\left[\left(\int_0^t g\,dB\right)^2\right] = \mathbb{E}\left[\int_0^t \|g(s,\omega)\|^2\,ds\right].$$

- Also be aware of the Ito representation theorem. While the martingale representation applies for all t (dynamic) and is used for hedging, the Itô representation is static at T and provides a payoff decomposition.

5 Solutions, existence, and uniqueness of SDEs

5.1 Existence and uniqueness for SDEs

Fix $T > 0$. Let $B(t) = (B_1(t), \dots, B_m(t))$ be m-dimensional Brownian motion and consider the n-dimensional SDE

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \qquad 0 \le t \le T, \qquad X_0 = Z, \tag{11}$$

where $b : [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times m}$ are measurable. Assume that the random variable Z – used for the initial condition $Z = X_0$ – satisfies $Z \in L^2(\Omega)$ and Z is independent of $\sigma(B_s : s \ge 0)$. Recall that $Z \in L^2(\Omega)$ implies that $\mathbb{E}[\|Z\|^2] < \infty$. Write $|\sigma|^2 := \sum_{i=1}^n \sum_{j=1}^m |\sigma_{ij}|^2$.

Assume there exist constants $C, D > 0$ such that, for all $t \in [0,T]$ and $x, y \in \mathbb{R}^n$,

$$|b(t,x)| + |\sigma(t,x)| \le C\,(1 + |x|), \qquad \text{(bounded growth rate)} \tag{12}$$

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le D\,|x - y|, \qquad \text{(global Lipschitz continuity).} \tag{13}$$

Then there exists a unique (pathwise) t-continuous $\mathcal{F}_t^Z$-adapted process $(X_t)_{0\le t\le T}$ solving (11), where $\mathcal{F}_t^Z := \sigma(Z,\, B_s : 0 \le s \le t)$, and

$$\mathbb{E}\left[\int_0^T |X_t|^2\,dt\right] < \infty \qquad \left(\text{in particular } \sup_{0\le t\le T} \mathbb{E}|X_t|^2 < \infty\right). \tag{14}$$

- Meaning: the drift b and volatility σ determine a well-posed random dynamics. Given the same Brownian path and the same initial value, there is only one possible trajectory (no branching).
- Why these conditions: (13) prevents explosions and ensures contraction in Picard iteration; (12) controls moments so the Itô integral $\int_0^t \sigma(s, X_s)\,dB_s$ is square-integrable.
- Finance/use-case intuition: once (11) is well-posed, Itô's formula applies to $g(t, X_t)$, and martingale/representation tools can be used on functionals of X without worrying that the model is ambiguous.

5.2 Strong vs. weak solutions of an SDE

Fix $T > 0$ and consider the n-dimensional SDE

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \qquad 0 \le t \le T, \qquad X_0 = Z,$$

where $b : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are given (measurable) coefficients, B is an m-dimensional Brownian motion, and Z is an $\mathbb{R}^n$-valued initial random variable.

A *strong solution* (relative to a given filtered space) is an $\mathbb{R}^n$-valued process $X = \{X_t\}_{0\le t\le T}$ defined on a *fixed* filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbb{P})$ together with a *given* $(\mathcal{F}_t)$-Brownian motion $B = \{B_t\}_{0\le t\le T}$ such that

- B is an $(\mathcal{F}_t)$-Brownian motion and Z is $\mathcal{F}_0$-measurable,
- X is $(\mathcal{F}_t)$-adapted with a.s. continuous paths,
- the integral identity holds $\mathbb{P}$-a.s. for all $t \in [0,T]$:

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

Equivalently: *given* the randomness $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and the driver B, the solution is built as an adapted functional of them.

A *weak solution* (with initial law prescribed) is a choice of *some* filtered probability space $(\hat{\Omega}, \hat{\mathcal{F}}, (\hat{\mathcal{F}}_t)_{t\geq 0}, \hat{\mathbb{P}})$ and processes $(\hat{X}, \hat{B})$ on it such that

- $\hat{B}$ is an $(\hat{\mathcal{F}}_t)$-Brownian motion and $\hat{X}$ has a.s. continuous paths,
- $\hat{X}$ is $(\hat{\mathcal{F}}_t)$-adapted and $\hat{X}_0$ has the prescribed distribution (typically $\hat{X}_0 \sim \mathrm{Law}(Z)$),
- the integral identity holds $\hat{\mathbb{P}}$-a.s. for all $t \in [0, T]$:

$$\hat{X}_t = \hat{X}_0 + \int_0^t b(s, \hat{X}_s)\, ds + \int_0^t \sigma(s, \hat{X}_s)\, d\hat{B}_s.$$

Equivalently: you only require the *existence* of a probability space and Brownian motion on which the SDE holds with the right initial law.

- Strong vs. weak is about what is fixed. Strong: the filtered space and Brownian motion are specified in advance, and X must be constructed adapted to that filtration (same sample space, same B). Weak: the underlying probability space, filtration, and Brownian motion may be chosen as part of the solution; only the coefficients and initial *distribution* are fixed.
- Consequence: strong solutions are *pathwise* objects (good for simulation/coupling and notions like pathwise uniqueness); weak solutions are *in-law* objects (often enough to identify $\mathrm{Law}(X)$ and connect to generators and PDE/Fokker–Planck).

6 The filtering problem

6.1 Filtering occurs when estimating X_t under noisy measurements

Suppose we have an (unobserved) stochastic system $X_t \in \mathbb{R}^n$, and that we continuously measure a corrupted output. The goal of filtering is to construct an estimator $\widehat{X}_t$ of X_t using only the information revealed by the observations up to time t.

Throughout this section, $(\Omega, \mathcal{F}, P)$ supports a $(p+r)$-dimensional Brownian motion (U_t, V_t) starting at 0. We assume U and V are independent, and the initial condition X_0 is independent of U and V.

6.2 The continuous-time measurement model

The unobserved stochastic process evolves according to

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dU_t \qquad t \geq 0,$$

where the drift and diffusion functions are $b : \mathbb{R}^{n+1} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^{n+1} \to \mathbb{R}^{n\times p}$. In the continuous-time observation model, we have access to a corrupted measurement

$$H_t = c(t, X_t) + \gamma(t, X_t)\,\widetilde{W}_t \qquad t \geq 0,$$

where the measurements $H_t \in \mathbb{R}^m$ are determined by the corruption functions $c : \mathbb{R}^{n+1} \to \mathbb{R}^m$, $\gamma : \mathbb{R}^{n+1} \to \mathbb{R}^{m\times r}$ and white noise $\widetilde{W}_t$ which is r-dimensional and independent of U and X_0. To obtain a mathematically well-defined model, define the integrated observation process

$$Z_t := \int_0^t H_s\,ds,$$

which yields the stochastic integral representation

$$dZ_t = c(t, X_t)\,dt + \gamma(t, X_t)\,dV_t, \qquad Z_0 = 0,$$

where V_t is an r-dimensional Brownian motion independent of U and X_0. Knowing $\{H_s\}_{0\leq s\leq t}$ is equivalent to knowing $\{Z_s\}_{0\leq s\leq t}$ since Z is the time-integral of H.

6.3 The filtering problem

Let $\mathcal{G}_t$ be the σ-algebra generated by the observation path Z_t up to time t:

$$\mathcal{G}_t := \sigma(Z_s :\ 0 \leq s \leq t)\,.$$

An estimator $\widehat{X}_t$ "based on observations" means that $\widehat{X}_t$ is $\mathcal{G}_t$-measurable.

Problem statement: the filtering problem is, among all square-integrable $\mathcal{G}_t$-measurable random variables, find the one that minimizes mean-squared error.

As a first step, define the admissible class $\mathcal{K}_t$ as the set of all $\mathcal{G}_t$-measurable random variables with finite second moment

$$\mathcal{K}_t = \mathcal{K}(Z, t) := \Big\{ Y : \Omega \to \mathbb{R}^n \ :\ Y \in L^2(P) \text{ and } Y \text{ is } \mathcal{G}_t\text{-measurable} \Big\}\,.$$

Then the best estimate $\widehat{X}_t$ is defined by

$$\widehat{X}_t = \arg\min_{Y\in\mathcal{K}_t} E\Big[\|X_t - Y\|^2\Big].$$

6.4 Conditional expectation is projection in L^2

Let $(\Omega, \mathcal{F}, P)$ be the original space that supports the stochastic process $\{X_t\}$. Recall that $\{X_t\}$ has a bounded second moment $X \in L^2(P)$. Let $\mathcal{H} \subset \mathcal{F}$ be a restricted σ-algebra and consider the closed subspace of random variables

$$\mathcal{N} := \{Y \in L^2(P) : Y \text{ is } \mathcal{H}\text{-measurable}\}.$$

Using that $L^2(P)$ is a Hilbert space with inner product $\langle X, Y\rangle = E[XY]$ (componentwise if vector-valued), there exists a unique orthogonal projection of X onto this restricted space of $L^2(P)$ random variables. We denote this orthogonal projection of X as $P_{\mathcal{N}}(X) \in \mathcal{N}$.

The key identity is

$$P_{\mathcal{N}}(X) = E[X \mid \mathcal{H}].$$

Applying this with $\mathcal{H} = \mathcal{G}_t$ and $X = X_t$ gives the filtering solution in abstract form:

$$\widehat{X}_t = P_{\mathcal{K}_t}(X_t) = E[X_t \mid \mathcal{G}_t].$$

Thus filtering is the problem of computing (or approximating) the conditional expectation $E[X_t \mid \mathcal{G}_t]$.

6.5 The Kalman–Bucy filter in the 1-dimensional linear case

Main idea: the solution to the filtering problem is the conditional expectation. Furthermore, in the linear-Gaussian setting, this conditional expectation is computable in closed form and obeys its own SDE.

Let us first recall the 1-dimensional linear system and observation model that define the setting of the filtering problem:

$$\text{(unobserved linear system)} \qquad dX_t = F(t)\,X_t\,dt + C(t)\,dU_t, \qquad F(t), C(t) \in \mathbb{R}, \tag{15}$$

$$\text{(linear observations)} \qquad dZ_t = G(t)\,X_t\,dt + D(t)\,dV_t, \qquad G(t), D(t) \in \mathbb{R}, \tag{16}$$

where $\{U_t\}, \{V_t\}$ are independent 1-dimensional Brownian motions and X_0 is independent of (U, V). Assume that the observation noise is nondegenerate $D(t) \neq 0$ for all t. Additionally, define the filtered estimate and its error variance

$$\widehat{X}_t := E[X_t \mid \mathcal{G}_t], \qquad S(t) := E\Big[(X_t - \widehat{X}_t)^2\Big]. \tag{17}$$

Then $\widehat{X}_t$ satisfies the Kalman–Bucy SDE

$$d\widehat{X}_t = \left(F(t) - \frac{G(t)^2 S(t)}{D(t)^2}\right)\widehat{X}_t\,dt + \frac{G(t)S(t)}{D(t)^2}\,dZ_t, \qquad \widehat{X}_0 = E[X_0]. \tag{18}$$

Equivalently, writing the Kalman gain

$$K(t) := \frac{G(t)S(t)}{D(t)^2},$$

one may rewrite (18) as

$$d\widehat{X}_t = F(t)\widehat{X}_t\, dt + K(t)\big(dZ_t - G(t)\widehat{X}_t\, dt\big),$$

where the term

$$dZ_t - G(t)\widehat{X}_t\, dt$$

is the observation "innovation" associated with the new (unpredicted) information in the measurement increment dZ_t relative to the prediction $G(t)\widehat{X}_t\, dt$.

The variance $S(t)$ is deterministic and solves the Riccati equation

$$\frac{dS}{dt} = 2F(t)S(t) - \frac{G(t)^2}{D(t)^2}\, S(t)^2 + C(t)^2, \qquad S(0) = E\Big[(X_0 - E[X_0])^2\Big]. \tag{19}$$

6.6 Main insight

The general filtering solution is the conditional expectation $\widehat{X}_t = E[X_t \mid \mathcal{G}_t]$, characterized as an L^2-projection. In the 1-dimensional linear-Gaussian case (15)-(16), this conditional expectation is computable in closed form and obeys its own SDE. In particular, the solution to the 1-dimensional linear-Gaussian filtering problem is characterized by the pair $(\widehat{X}_t, S(t))$, where the estimate $\widehat{X}_t$ follows (18) and the variance $S(t)$ follows (19).

7 Basic properties of diffusions

7.1 SDEs model position in moving liquid

Suppose we want to describe the motion of a small particle suspended in a moving liquid, subject to random molecular bombardments. If $b(t,x) \in \mathbb{R}^3$ is the velocity of the fluid at the point x and time t, then the following SDE is a reasonable model for a the position X_t of the particle at time t

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{7.1.2}$$

where $\sigma(t,x) \in \mathbb{R}^{3\times 3}$, B_t is 3-dimensional Brownian motion, and similarly. Thus the solution of a stochastic differential equation may be thought of as the mathematical description of the motion of a small particle in a moving fluid.

7.2 Time-homogeneous shift identity for SDE solutions

Let B_t be an m-dimensional Brownian motion and consider the time-homogeneous SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dB_t, \qquad t \geq s, \qquad X_s = x,$$

with $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ (e.g. Lipschitz continuous so the solution is well-defined and unique in law, as described in Section 5.1). Denote its solution by $X_t^{s,x}$ for $t \geq s$. For any $h \geq 0$, the solution is

$$\begin{aligned} X_{s+h}^{s,x} &= x + \int_s^{s+h} b(X_u^{s,x})\ du + \int_s^{s+h} \sigma(X_u^{s,x})\ dB_u \\ &= x + \int_0^h b(X_{s+v}^{s,x})\ dv + \int_0^h \sigma(X_{s+v}^{s,x})\ d\widetilde{B}_v, \end{aligned}$$

where we can use $\widetilde{B}_v := B_{s+v} - B_s$. On the other hand,

$$X_h^{0,x} = x + \int_0^h b\Big(X_v^{0,x}\Big)\ dv + \int_0^h \sigma\Big(X_v^{0,x}\Big)\ dB_v.$$

Since $(\widetilde{B}_v)_{v\geq 0}$ and $(B_v)_{v\geq 0}$ have the same distributions, it follows (by uniqueness in law) that

$$\{X_{s+h}^{s,x}\}_{h\geq 0} \overset{d}{=} \{X_h^{0,x}\}_{h\geq 0},$$

i.e. the family of solutions is time-homogeneous.

7.3 Girsanov's change of measure formula in the SDE setting

Fix a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}),$$

- where $(\mathcal{F}_t)_{t\geq 0}$ is a filtration (an increasing family of σ-algebras) and $\mathcal{F}$ contains $\bigvee_{t\geq 0} \mathcal{F}_t$
- recall that a random variable $X : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-*measurable* if $\{X \in B\} \in \mathcal{F}_t$ for every Borel set $B \subseteq \mathbb{R}$
- A process $(X_t)_{t\geq 0}$ is *adapted* if for every t, the random variable X_t is $\mathcal{F}_t$-measurable
- Let $(W_t)_{t\geq 0}$ be a one-dimensional $(\mathcal{F}_t)$-Brownian motion under $\mathbb{P}$

Next, fix a horizon $T > 0$ and consider an SDE under $\mathbb{P}$:

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dW_t, \qquad 0 \le t \le T,$$

with X_0 given, and measurable coefficients b, σ. Next, define the drift-shift process

$$\theta_t := \frac{b(t, X_t) - \tilde{b}(t, X_t)}{\sigma(t, X_t)}, \qquad 0 \le t \le T,$$

where $\tilde{b}$ is the drift we would *like* to have after changing measure. Note θ is adapted because X is adapted (a deterministic function of an adapted stochastic process is also adapted).

Girsanov's identity: Assume the process θ_t is square-integrable on $[0, T]$ and that the exponential process $Z_t : \Omega \to (0, \infty)$ defined as

$$Z_t := \exp\left(-\int_0^t \theta_s\,dW_s - \frac{1}{2}\int_0^t \theta_s^2\,ds\right), \qquad 0 \le t \le T,$$

is a martingale that satisifies the condition that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[Z_T] &= \int_\Omega Z_T(\omega)\mathbb{P}(\,d\omega)\\ &= 1\end{aligned}$$

(a standard sufficient condition for this is that $\mathbb{E}_{\mathbb{P}}[\exp(\frac{1}{2}\int_0^T \theta_s^2 ds)] < \infty$). Now define a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\begin{aligned}\mathbb{Q}(A) &:= \int_\Omega Z_T(\omega)\,\mathbf{1}_A(\omega)\,d\mathbb{P}(\omega) \qquad A \in \mathcal{F}_T\\ &= \mathbb{E}_{\mathbb{P}}\big[Z_T\,\mathbf{1}_A\big] \quad \text{for all } A \in \mathcal{F}_T.\end{aligned}$$

Then the process

$$W_t^{\mathbb{Q}} := W_t + \int_0^t \theta_s\,ds, \qquad 0 \le t \le T,$$

is a Brownian motion with respect to $(\mathcal{F}_t)$ *under* $\mathbb{Q}$. Equivalently,

$$dW_t = dW_t^{\mathbb{Q}} - \theta_t\,dt.$$

SDE consequence. Substituting $dW_t = dW_t^{\mathbb{Q}} - \theta_t\,dt$ into the $\mathbb{P}$-SDE gives, for $0 \le t \le T$,

$$\begin{aligned}dX_t &= b(t, X_t)\,dt + \sigma(t, X_t)\big(dW_t^{\mathbb{Q}} - \theta_t\,dt\big)\\ &= b(t, X_t)\,dt + \sigma(t, X_t)dW_t^{\mathbb{Q}} - \sigma(t, X_t)\theta_t\,dt\\ &= \big(b(t, X_t) - \sigma(t, X_t)\theta_t\big)\,dt + \sigma(t, X_t)dW_t^{\mathbb{Q}}\\ &= \tilde{b}(t, X_t)\,dt + \sigma(t, X_t)\,dW_t^{\mathbb{Q}}.\end{aligned}$$

using our definition of θ_t. So, on the *same* paths $\omega \in \Omega$, you can interpret X as solving an SDE with the same diffusion σ but a different drift $\tilde{b}$ once you switch from $\mathbb{P}$ to $\mathbb{Q}$.

Important facts / intuition.

1. Girsanov changes the *drift* by changing the measure; it does not change the diffusion coefficient σ.

2. Under $\mathbb{Q}$, the process $W^{\mathbb{Q}}$ is brownian motion. But relative to $\mathbb{P}$ it is a drifted version of W:
$$W_t^{\mathbb{Q}}(\omega) = W_t(\omega) + \int_0^t \theta_s(\omega) ds.$$

3. How to use Girsanov's change of measure: Pick an adapted process θ_t and define
$$Z_t := \exp\Big(- \int_0^t \theta_s \, dW_s - \frac{1}{2} \int_0^t \theta_s^2 \, ds \Big), \qquad 0 \le t \le T,$$
assuming (Z_t) is a true $\mathbb{P}$-martingale with $\mathbb{E}_{\mathbb{P}}[Z_T] = 1$. Define a new measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by
$$\mathbb{Q}(A) := \mathbb{E}_{\mathbb{P}}\big[Z_T \, \mathbf{1}_A\big] \qquad (A \in \mathcal{F}_T),$$
meaning that on events determined by information up to time T (i.e. events in $\mathcal{F}_T$), probabilities under $\mathbb{Q}$ are obtained by weighting probabilities under $\mathbb{P}$ by the factor Z_T.

   First the t case. For any $t \in [0, T]$ and any $\mathcal{F}_t$-measurable process $H_t(\omega) : \Omega \to \mathbb{R}$, Girsanov's gives us that
$$\mathbb{E}_{\mathbb{Q}}[H_t] = \mathbb{E}_{\mathbb{P}}[Z_t \, H_t], \qquad \text{hence} \qquad \mathbb{E}_{\mathbb{P}}[H_t] = \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{Z_t} \, H_t \right].$$
   Where the first equality is used when we seek to compute the RHS under $\mathbb{P}$, and the second equality is used when we've designed the process under $\mathbb{Q}$ to be simple, such as having zero drift. In practice H_t can be anything that depends only on the path up to t, e.g.
$$H_T = g(X_T), \qquad H = \max_{0 \le s \le T} X_s, \qquad H_T = \int_0^T h(X_s) \, ds.$$

8 Martingales