This part is based on the notes from 2016, Section 4.4.1. The convention used below for circulant matrices is slightly different from the one employed in the notes, but the two approaches are equivalent. Here we follow the convention employed here. Remark 1 and Exercise 1 were not covered during the lecture; they were included here for information purposes, but are not examinable.

Simulating Stationary Gaussian Processes

While the methods seen last week are applicable for simulating general Gaussian processes on general meshes $t_0 < t_1 < \ldots < t_{n-1}$ they are computationally expensive, since they ultimately will require $O(n^3)$ floating point operations to generate a single sample. However, in the particular case where we wish to simulate a *stationary* Gaussian process on a regular mesh $\{0, \Delta t, 2\Delta t, \ldots, (n-1)\Delta t\}$, then we can reduce the problem to a discrete Fourier transform and obtain an almost magical speedup by employing a Fast-Fourier Transform.

Indeed, suppose we wish to simulate a stationary Gaussian process $X(t)$ with mean $\mu = 0$ and covariance $C(t,s) = C(t-s)$. Then given timesteps $\{0, \Delta t, 2\Delta t, \ldots, n\Delta t\}$, the random vector $(X(t_0), X(t_2), \ldots, X(t_{n-1}))$ has a covariance matrix of the form:

$$\Sigma = \begin{pmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_1 & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ c_{n-1} & \ddots & \ddots & \ddots \end{pmatrix},$$

where the $\ddots$ indicate that the entries of Σ are constant along each diagonal: Σ is *Toeplitz matrix.*

Definition 1 (Toeplitz matrix)**.** A matrix Σ is said to be a Toeplitz matrix, if each diagonal takes a constant value, that is

$$\Sigma_{i,j} = \Sigma_{i+1,j+1},$$

for all $i, j \in \{0, \ldots, n-2\}$.

A particularly important subclass of Toeplitz matrices is that of *circulant matrices.*

Definition 2. A circulant matrix is a matrix of the form:

$$B = \begin{pmatrix} b_0 & b_1 & \cdots & b_{n-1} \\ b_{n-1} & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ b_1 & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1}$$

A circulant matrix is constructed by starting with a vector $\mathbf{b} = (b_0, \ldots, b_{n-1})$ as the first row, and obtaining the each row by a periodic right shift of the previous row. The important observation that we shall make use of is that we can embed Σ in a $N \times N$ symmetric *circulant matrix*, where

$N = 2n - 2$, as follows:

$$\Pi = \begin{pmatrix} c_0 & c_1 & \cdots & c_{n-1} & c_{n-1} & c_{n-2} & \cdots & c_1 \\ c_1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ c_{n-2} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ c_{n-1} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ c_{n-2} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ c_1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

Unfortunately, this matrix will not be nonnegative definite in general, which we require, however this will hold under some additional assumptions.

Lemma 3. *Suppose that $c_0 \geq c_1 \geq \cdots \geq c_{n-1} \geq 0$ and*

$$2c_k \leq c_{k-1} + c_{k+1},$$

for $k = 1, \ldots, n-2$, then Π is a covariance matrix, i.e. Π is nonnegative definite.

Why are we interested in this representation? At first glance this might seem like a futile exercise, however the importance of this embedding arises from the connection between circulant matrices and the discrete Fourier transform.

Definition 4. Given a vector $\mathbf{x} = (x_0, \ldots, x_{N-1})^\top \in \mathbb{C}^N$, the discrete Fourier transform of $\mathbf{x}$ is the vector

$$(\mathcal{F}\mathbf{x})_j = \sum_{k=0}^{N-1} e^{-(2\pi i/N)jk} x_k = \sum_{k=0}^{N-1} \omega^{jk} x_k,$$

for $j = 0, \ldots, N-1$ where $\omega = e^{-2\pi i/N}$.

Therefore computing discrete Fourier transform of $\mathbf{x}$ is equivalent to computing $F\mathbf{x}$, where

$$F = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{pmatrix}.$$

Exercise 1. Show that $F^{-1} = \overline{F}/N$.

Proof. This follows from the fact that the columns (or lines) of F are orthogonal in $\mathbb{C}^N$:

$$\sum_{k=0}^{N-1} F_{ki}\, \bar{F}_{kj} = \sum_{k=0}^{N-1} \omega^{ki}\, \omega^{-kj} = \sum_{k=0}^{N-1} (\omega^{(i-j)})^k = \begin{cases} N & \text{if } i = j \\ \frac{1-\omega^{(i-j)N}}{1-\omega^{(i-j)}} = 0 & \text{otherwise,} \end{cases}$$

from the definition of ω and the formula for the sum of the first n terms of a geometric series. $\square$

Normally, computing the matrix-vector produce $F\mathbf{x}$ would require $O(N^2)$ operations, however using a Fast Fourier Transform reduces this to $O(N \log N)$ steps. The connection to circulant matrices is the following. Let $\mathbf{b} = (b_0, b_1, \ldots, b_{N-1})^\top$ be a complex valued vector, and let B be the circulant matrix *generated* by $\mathbf{b}$, i.e. the same matrix as in (1) (except that here the dimension is labeled N). We then have the following fundamental result:

Lemma 5. *The circulant matrix B is diagonalized by the DFT matrix F. More specifically,*

$$B = FDF^{-1},$$

where D is a diagonal matrix containing the eigenvalues of B, $\lambda_0, \ldots, \lambda_{N-1}$,

$$\lambda_j = (F\mathbf{b})_j.$$

Proof. We calculate:

$$(BF)_{jk} = \sum_{\ell=0}^{N-1} B_{j\ell}\, F_{\ell k} = \sum_{\ell=0}^{N-1} b_{(\ell-j)\%n}\, \omega^{\ell k}.$$

Since the argument of the sum, seen as a function of ℓ, is periodic with period n, we can shift the bounds of the summation:

$$(BF)_{jk} = \sum_{\ell=j}^{N-1+j} b_{(\ell-j)\%n}\, \omega^{\ell k} = \sum_{m=0}^{N-1} b_{m\%n}\, \omega^{(m+j)k} = \omega_{jk} \sum_{m=0}^{N-1} b_m\, \omega^{mk} = F_{jk}(F\mathbf{b})_k = F_{jk}\, \lambda_k.$$

In matrix form, this is exactly $BF = FD$. □

This factorization is very commonly exploited in both numerical PDE schemes, as well as methods to perform efficient matrix-vector multiplication. Our objective is to generate a sample from the Gaussian distribution $\mathcal{N}(0, \Pi)$, which entails computing a square root of the matrix Π. If $C(\cdot)$ is such that the assumption of Lemma 3 is satisfied, then Π is a nonnegative definite matrix and Lemma 5 implies, denoting by $\mathbf{c}$ the vector associated with Π:

$$\Pi = F\,\mathrm{diag}(\boldsymbol{\lambda}/N)\,F^* = \Big(F\,\mathrm{diag}(\sqrt{\boldsymbol{\lambda}/N})\Big)\,\Big(F\,\mathrm{diag}(\sqrt{\boldsymbol{\lambda}/N})\Big)^* =: EE^*, \qquad \boldsymbol{\lambda} = F\mathbf{c}.$$

Here $\sqrt{\boldsymbol{\lambda}/N}$ is simply the vector $(\sqrt{\lambda_0/N}, \ldots, \sqrt{\lambda_{N-1}/N})^T$. Therefore, E is the square root of Π in $\mathbb{C}^{N\times N}$. The last ingredient we need is a simple extension of a result we covered in week 1:

Lemma 6. *Assume that $\Pi = EE^*$, where $\Pi \in \mathbb{R}^{N\times N}$ is a symmetric nonnegative definite matrix and $E \in \mathbb{C}^{N\times N}$ is a complex matrix, and let $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{R}^N$ be independent $\mathcal{N}(0, I_N)$ random variables. Then $\mathrm{Re}[E(\mathbf{Z}_1 + i\mathbf{Z}_2)] \sim \mathcal{N}(0, \Pi)$.*

Proof. Let us write $E = E_1 + iE_2$ for real valued matrices. Then

$$EE^* = E_1E_1^\top + E_2E_2^\top + i(E_2E_1^\top - E_1E_2^\top),$$

and since Π is real-valued, it follows that $\Pi = E_1E_1^\top + E_2E_2^\top$. Now let $\mathbf{X} = E(\mathbf{Z}_1 + i\mathbf{Z}_2)$, and notice that $\mathbf{X} = (E_1\,\mathbf{Z}_1 - E_2\,\mathbf{Z}_2) + i(E_2\,\mathbf{Z}_1 + E_1\,\mathbf{Z}_2)$. Letting $\mathbf{X}_1 = \mathrm{Re}[\mathbf{X}]$, then

$$\mathbb{E}[\mathbf{X}_1] = 0, \qquad \mathrm{Var}(\mathbf{X}_1) = \mathrm{Var}(E_1\,\mathbf{Z}_1 - E_2\,\mathbf{Z}_2) = E_1E_1^\top + E_2E_2^\top = \Pi.$$

which concludes the proof. □

This give us a method for sampling a stationary Gaussain process at equidistant intervals.

Algorithm 7 (Generating a stationary Gaussian process)**.** *Assume $\mu(t) = 0$ and that we are given covariance $C(\cdot)$ and let $t_i = i\Delta t$, $i = 0, \dots, n-1$.*

1. *Set $\mathbf{c} = (c_0, c_1, \dots, c_{n-2}, c_{n-1}, c_{n-2}, \dots, c_1)$, where $c_i = C(t_0, t_i)$, and let $N = 2n - 2$.*
2. *Set $\lambda = F\mathbf{c}$ using FFT.*
3. *Generate $\mathbf{Z} = \mathbf{Z}_1 + i\mathbf{Z}_2$, where $\mathbf{Z}_1, \mathbf{Z}_2 \sim \mathcal{N}(0, I_N)$.*
4. *Compute $\mathbf{V} = \sqrt{\operatorname{diag}(\boldsymbol{\lambda}/N)}\,\mathbf{Z}$.*
5. *Compute $\mathbf{V} = F\mathbf{Y}$ using FFT.*
6. *Output $\mathbf{X} = \operatorname{Re}(V_0, \dots, V_{n-1})^\top$.*

If we performed naive matrix-vector multiplications this algorithm would cost $O(N^2)$, however using the FFT, the algorithm is $O(N \log N)$. So what happens when the embedding circulant matrix is not nonnegative definite? Then we cannot use this approach directly. However, there are two possible approaches to generate a sample in this case:

1. Embed the symmetric Toeplitz in an even larger circulant matrix.
2. Use only the positive part of the circulant matrix.

It is typically always possible choose an large circulant matrix which is nonnegative definite. In this case, we can use the above exact scheme for generating the sample. If we must resort to option (2), then the procedure is approximate. However, one can typically quantify the error incurred in this case, so the approach is still feasible.

While this algorithm provides a perfectly adequate scheme for simulating stationary Gaussian processes, the true power of this method can be seen when using it to simulate stationary *Gaussian random fields*, i.e. a $\mathbb{R}^d$–indexed Gaussian process. Indeed, many scientific computing software libraries provide algorithms for simulating stationary GRFs based on circulant embeddings. See the lecture notes for references.

Exercise 2. Consider the stationary Gaussian process with exponential covariance $C(\tau) = e^{-|\tau|/l}$.

1. Implement a method for simulating this process in a programming language of your choice.
2. (Challenging optional exercise:) Show that the eigenvalues λ_i in this case are always positive.

Remark 1 (A word on the Cooley–Tukey algorithm)*.* The Cooley–Tukey is the most common algorithm for computing fast Fourier transforms. It relies on the observation that, assuming that N is a multiple of 2 for simplicity, and given a vector $\mathbf{x} = (x_0, \dots, x_{N-1})^\top \in \mathbb{C}^N$, the DFT of $\mathbf{x}$ can be decomposed as

$$(\mathcal{F}\mathbf{x})_j = \sum_{k=0}^{N-1} \omega^{jk} x_k = \sum_{k=0}^{N/2-1} \omega^{2kj} x_{2k} + \sum_{k=0}^{N/2-1} \omega^{(2k+1)j} x_{2k+1} =: (\mathcal{F}_1\mathbf{x} + \mathcal{F}_2\mathbf{x})_j\,, \qquad j = 0, \dots, N-1.$$

We just rewrote N sums (one for each value of j) of N terms as $2N$ sums of $N/2$ terms, so what have we gained? The gain comes from the fact that $(\mathcal{F}_1\mathbf{x})_j$ and $(\mathcal{F}_2\mathbf{x})_j$ are both periodic functions of j with period $N/2$ (why? Hint: $\omega^N = 1$), so only N sums of $N/2$ terms have to be calculated! This "splitting of the sum" procedure can then be repeated. ⊘

4. Numerical solution of stochastic differential equations.

One important class of continuous-time stochastic processes arises as a solution of Stochastic Differential Equations (SDEs). these models are important in many applications such as biology, chemistry, epidemiology, mechanics, economics and finance, etc.

In this section we will describe how to simulate/solve SDEs numerically and analyse the behaviour and performance of numerical schemes.

We consider Stochastic Differential Equations driven by Gaussian white noise: a mean-zero Gaussian process with correlation $\delta(t-s)I = C(t,s)$.
These SDEs are of the form

$$\frac{dX(t)}{dt} = b(X(t)) + \sigma(X(t)) \frac{dW(t)}{dt}$$

where $W(t)$ is a standard Brownian Motion

$b(x) \longrightarrow$ drift
$\sigma(x) \longrightarrow$ diffusion coefficient.

<u>Note that</u>: the term $\frac{dW}{dt}$ represents white noise and it is formally the derivative of a Brownian Motion. However, it is important to note that this <u>does not</u> exist in any ordinary sense of derivative

Using our intuition from ODEs, a solution of this SDE would be a stochastic process X_t satisfying

$$X_t = x_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s.$$

where $X_0 = x_0$ is a (deterministic or random) initial condition

In order to make sense of this solution, we need to understand the stochastic Integral

$$\boxed{I = \int_0^t \sigma(X_s)\, dW_s}$$

4.1. Stochastic Integrals.

Consider the following integral

$$I(t) = \int_0^t f(s)\, dW_s .$$

We want to define this integral. In order to do that, we will assume that f is a random process adapted to the filtration $\mathcal{F}_t$ generated by W_t and such that

$$\mathbb{E}\left(\int_0^T f(s)^2\, ds \right) < \infty .$$

We will define the stochastic integral using a construction similar to that of the Riemann integral. However, we need to note that now we are trying to integrate a stochastic process (or a function of a stochastic process) with respect to another stochastic process, while in the Riemann integral case we would be integrating a function with respect to time

! $\Rightarrow$ the stochastic integral I is a random variable.

We discretise the interval $[0, t]$

$$0 \quad t_1 \quad t_2 \quad \cdots \quad t_k \quad t_{k+1} \quad \cdots \quad t \qquad , \qquad \begin{aligned} t_k &= k\,\Delta t \\ t &= K \Delta t. \end{aligned}$$

and define

$$I(t) := \lim_{K \to \infty} \sum_{k=1}^{K} f(\tau_k)\left(W(t_k) - W(t_{k-1})\right).$$

Note that we evaluate f at τ_k.
The definition of the stochastic integral will depend on the choice of τ_k.

Note that: Since W_t is a standard Brownian Motion, the terms $W(t_k) - W(t_{k-1})$ are increments of a Brownian Motion $\Rightarrow$ we know that they are independent AND $W(t_k) - W(t_{k-1}) \sim \mathcal{N}(0, \Delta t)$.

This looks like a reasonable definition of the integral. However, one would expect the limit to be independent of the point τ_k where we evaluate f, and this is not the case, even if f is continuous:

example: Suppose we want to compute
$I = \int_0^T W_t \, dW_t$, We write

$$I^{(1)} = \sum_{k=1}^{K} W(t_{k-1}) \left(W(t_k) - W(t_{k-1})\right)$$

We have $\mathbb{E}(I^{(1)}) = \sum \mathbb{E}\left(W(t_{k-1}) \left(W(t_k) - W(t_{k-1})\right)\right)$

$= 0$ because $W(t_{k-1})$ and $W(t_k) - W(t_{k-1})$ are independent.

On the other hand, if we write

$$I^{(2)} = \sum_{k=1}^{K} W(t_k) \left(W(t_k) - W(t_{k-1})\right)$$

we have $\mathbb{E}(I^{(2)}) = \mathbb{E} \sum \left(W(t_k) - W(t_{k-1}) + W(t_{k-1})\right)\left(W(t_k) - W(t_{k-1})\right)$

$$= \sum_{k=1}^{K} \underbrace{\mathbb{E}\left[\left(W(t_k) - W(t_{k-1})\right)^2\right]}_{= \Delta t} + \underbrace{\mathbb{E}\left(W(t_{k-1}) \left(W(t_k) - W(t_{k-1})\right)\right)}_{= 0}$$

$\longrightarrow T$ as $K \to \infty$.

$\Rightarrow$ the integral is different depending on τ_k!
This is because the Brownian motion is a.s non-differentiable, which means it "varies too much" in the interval (t_{k-1}, t_k), which leads to this phenomenon.

Note that: For the definition of the Riemann-Stieltjes integral $\int f(x)\, dg(x)$ it is required that $g(x)$ has bounded total variation. This is not the case for BM $\rightarrow$ it has a.s. infinite total variation.

There is no way around this problem. We just need to specify each time what is our choice of τ_k when computing the integral!

The two most popular choices are:
(i) $\tau_k = t_{k-1} \longrightarrow$ Itô stochastic integral
(ii) $\tau_k = (t_k + t_{k-1})/2 \longrightarrow$ Stratonovich stochastic integral

Another alternative is the Klimontovich stochastic integral which chooses $\tau_k = t_k$.

Note that: if f is sufficiently regular, then the integrals coincide: if $\exists\ C < \infty,\ \varepsilon > 0$ such that

$$\mathbb{E}\,|f(t,\omega) - f(s,\omega)|^2 \le C\,|t-s|^{1+\varepsilon},$$

for all $s, t \in [0,t]$, then

$$\int_0^t f(s,\omega)\, dW_s = \int_0^t f(s,\omega) \circ dW_s$$

Stratonovich sense

and actually if $\tau_k = \lambda t_{k-1} + (1-\lambda) t_k$

This short section is based on Andrew Duncan's lecture notes from 2016 and on the notes by J. Michael Steele freely available here.

Construction of the Itô integral

In this section, we outline the construction of the Itô integral $\int_0^T f(t,\omega)\,dW_t$ for integrands satisfying the following assumptions:

1. $f(t,\omega)$ is $\mathcal{B}\times\mathcal{F}$-measurable where $\mathcal{B}$ is the Borel σ-algebra of $[0,T]$.
2. $f(t,\cdot)$ is $\mathcal{F}_t$ adapted for all $t \geq 0$, where $\mathcal{F}_t$ is the natural filtration associated with the Brownian motion W_t.
3. $\mathbb{E}\int_0^T f(t,\omega)^2 dt < \infty$.

We will denote by $\mathcal{J}$ the class of stochastic processes $f(t,\omega) : [0,T]\times\Omega \to \mathbb{R}$ for which the above three properties hold, and our goal will be to define the Itô integral for any $f \in \mathcal{J}$. To this end, we will first define the Itô integral on the subset $\mathcal{J}_0 \subset \mathcal{J}$ of so-called simple processes.

Definition 8. Let $\mathcal{J}_0$ be the class of stochastic processes f that admit the following representation:

$$f(\omega,t) = \sum_{i=0}^{N-1} a_i(\omega)\, I_{(t_i,t_{i+1}]}(t) \tag{2}$$

with $\mathbb{E}[|a_i|^2] < \infty$ and $a_i \in \mathcal{F}_{t_i}$, and for some $N \in \mathbf{N}$ and $0 = t_0 < t_1 < \cdots < t_n = T$.

Exercise 3. Show that $\mathcal{J}_0$ is a linear subspace of $\mathcal{J}$.

Exercise 4. Show that any $f \in \mathcal{J}_0$ admits more than one representation as a sum of the form (2).

For processes in $\mathcal{J}_0$, it is reasonable to define the Itô integral as follows, by analogy with the Riemann-Stieltjes integral:

$$I^f := \int_0^T f(\omega,t)\,\mathrm{d}W_t = \sum_{i=0}^{N-1} a_i(\omega)\,(W_{t_{i+1}} - W_{t_i}).$$

It is simple to check that this definition is unambiguous, i.e. that the value of I_f is the same for all representations of f as a sum of the form (2).

It relatively easy to show the celebrated *Itô isometry* in J_0 (not examinable).

Lemma 9 (Itô isometry in $\mathcal{J}_0$). *If $f,g \in \mathcal{J}_0$, then*

$$\mathbb{E}[I^f(\omega)\,I^g(\omega)] = \mathbb{E}\left[\int_0^T f(\omega,t)\,g(\omega,t)\,\mathrm{d}t\right] = \langle f,g\rangle_{L^2(\Omega\times[0,T])}\,.$$

Proof. The first step (left as an exercise) is to show the existence of a partition $0 = t_0 < t_1 < \ldots < t_N = T$ over which both f and g can be expressed as simple processes:

$$f(\omega,t) = \sum_{i=1}^{N-1} a_i(\omega)\, I_{(t_i,t_{i+1}]}(t), \qquad g(\omega,t) = \sum_{i=1}^{N-1} b_i(\omega)\, I_{(t_i,t_{i+1}]}(t).$$

Then, employing the definition of the Itô integral for simple processes, we obtain

$$I^f(\omega)\, I^g(\omega) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_i(\omega)\, b_j(\omega)\, (W_{t_{i+1}} - W_{t_i})\, (W_{t_{j+1}} - W_{t_j}).$$

(Note that $W_{t_i} = W_{t_i}(\omega)$ depends on ω too but, in order to alleviate the notations, we do not write this explicitly.) Taking the expectation, we have

$$\mathbb{E}[I^f(\omega)\, I^g(\omega)] = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \mathbb{E}[a_i(\omega)\, b_j(\omega)\, (W_{t_{i+1}} - W_{t_i})\, (W_{t_{j+1}} - W_{t_j})] =: \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \mathbb{E}[e_{ij}(\omega)]. \quad (3)$$

Using the basic properties of conditional expectation (read here if you don't know these properties), and employing the fact that the random variables a_i and b_i are $\mathcal{F}_{t_i}$-measurable by definition of $\mathcal{J}_0$, we obtain, in the case $i = j$:

$$\begin{aligned}
\mathbb{E}[e_{ij}(\omega)] &= \mathbb{E}[a_i(\omega)\, b_i(\omega)\, (W_{t_{i+1}} - W_{t_i})^2] \\
&= \mathbb{E}[\mathbb{E}[a_i(\omega)\, b_i(\omega)\, (W_{t_{i+1}} - W_{t_i})^2 | \mathcal{F}_{t_i}]] && \text{(Tower property)} \\
&= \mathbb{E}[a_i(\omega)\, b_i(\omega)\, \mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2 | \mathcal{F}_{t_i}]] && \text{(Pulling out known factors)} \\
&= \mathbb{E}[a_i(\omega)\, b_i(\omega)\, (t_{i+1} - t_i)] && \text{(Variance of Brownian increment)} \\
&= (t_{i+1} - t_i)\, \mathbb{E}[a_i(\omega)\, b_i(\omega)].
\end{aligned}$$

On the other hand, when $j > i$,

$$\begin{aligned}
\mathbb{E}[e_{ij}(\omega)] &= \mathbb{E}[a_i(\omega)\, b_j(\omega)\, (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})] \\
&= \mathbb{E}[\mathbb{E}[a_i(\omega)\, b_j(\omega)\, (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j}) | \mathcal{F}_{t_j}]] && \text{(Tower property)} \\
&= \mathbb{E}[a_i(\omega)\, b_i(\omega)(W_{t_{i+1}} - W_{t_i})\, \mathbb{E}[W_{t_{j+1}} - W_{t_j} | \mathcal{F}_{t_j}]] && \text{(Pulling out known factors)} \\
&= 0 && \text{(Mean of Brownian increment).}
\end{aligned}$$

An analogous calculation shows that $\mathbb{E}[e_{ij}] = 0$ when $j < i$. Substituting in (3), we deduce

$$\mathbb{E}[I^f(\omega)\, I^g(\omega)] = \sum_{i=0}^{N-1} \mathbb{E}[a_i(\omega)\, b_j(\omega)](t_{i+1} - t_i) = \mathbb{E}\left[\int_0^T f(\omega, t)\, g(\omega, t)\, \mathrm{d}t\right],$$

which concludes the proof. □

This isometry will be instrumental in extending our definition of the Itô integral over $[0, T]$ to $\mathcal{J}$. To complete this program, we need one more ingredient, which we will admit without proof.

Lemma 10 (Density of $\mathcal{J}_0$ in $\mathcal{J}$). *For any $f \in \mathcal{J}$, there exists a sequence $\{f_n\}_{n \in \mathbf{N}}$ with $f_n \in \mathcal{J}_0$ such that*

$$\|f - f_n\|_{L^2(\Omega \times [0,T])} \to 0 \qquad \text{as } n \to \infty.$$

Now, given any $f \in \mathcal{J}$, this lemma implies that there exists a sequence $\{f_n\}_{n \in \mathbf{N}}$ in $\mathcal{J}_0$ that converges to f in $L^2(\Omega \times [0, T])$. By Itô's isometry – Lemma 9 – it is clear that

$$\mathbb{E}[|I^{f_m} - I^{f_n}|^2] = \|I^{f_m} - I^{f_n}\|^2_{L^2(\Omega)} = \|f_m - f_n\|^2_{L^2(\Omega \times [0,T])},$$

and since $\{f_n\}_{n \in \mathbf{N}}$ is a Cauchy sequence in $L^2(\Omega \times [0, T])$ (because it is convergent to f and every convergent sequence is a Cauchy sequence), we deduce that $\{I^{f_n}\}_{n\in\mathbf{N}}$ is a Cauchy sequence in $L^2(\Omega)$. By completeness of $L^2(\Omega)$, this implies that $\{I^{f_n}\}_{n\in\mathbf{N}}$ converges to a limit in $L^2(\Omega)$, a limit which *we will take as the definition of the Itô integral of* f:

$$I^f := \lim_{n\to\infty} I^{f_n},$$

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

Exercise 5. Show that the Itô isometry – Lemma 9 – holds also for f, g in $\mathcal{J}$.

Proof. Let $\{f_n\}_{n\in\mathbf{N}}$ and $\{g_n\}_{n\in\mathbf{N}}$ be sequences in $\mathcal{J}_0$ converging in $L^2(\Omega\times[0,T])$ to f and g, respectively. (Again, the existence of such sequences is provided by Lemma 10.) Since the inner product of any Hilbert space $\mathcal{H}$, here denoted by $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, is a continuous function on $\mathcal{H}\times\mathcal{H}$, we have

$$\mathbb{E}[I^f\, I^g] = \left\langle I^f, I^g\right\rangle_{L^2(\Omega)} = \lim_{n\to\infty}\left\langle I^{f_n}, I^{g_n}\right\rangle_{L^2(\Omega)} = \lim_{n\to\infty}\langle f_n, g_n\rangle_{L^2(\Omega\times[0,T])} = \langle f, g\rangle_{L^2(\Omega\times[0,T])},$$

where we used Itô's isometry on $\mathcal{J}_0$. □

Link with the "Riemann sums" definition (not covered in class)

At this point, one may wonder whether the tentative definition of the Itô given at the beginning of the lecture, namely

$$\hat{I}^f = \lim_{n\to\infty}\sum_{i=0}^{N-1} f(\omega, t_i)\,(W_{t_{i+1}} - W_{t_i}), \tag{4}$$

makes sense, and whether it coincides with the formal definition of the Itô integral given above. It turns out that, under relatively weak assumptions, the limit in (4) exists in the sense of convergence in probability, and that the limit coincides with the Itô integral. For example, it is possible to show the following result (see J. Michael Steele's paper if you are interested in learning more about this):

Proposition 11. *For any continuous* $f:\mathbb{R}\to\mathbb{R}$*, and with a partition of* $[0,T]$ *given by* $t_i = iT/N$ *for* $0\le i\le N$*, it holds that*

$$\lim_{N\to\infty}\sum_{i=0}^{N-1} f(W_{t_i})\,(W_{t_{i+1}} - W_{t_i}) = \int_0^T f(W_t)\,\mathrm{d}W_t,$$

where the limit is in the sense of convergence in probability. (Note that the integral in the right-hand side is the Itô integral as defined by the extension procedure, not the tentative definition (4)*.)*

This result, of which extensions to more general integrands exist, is good news for us, because it means that we are usually allowed to view the Itô integral as the limit of a Riemann sum, which is often a more intuitive and tangible viewpoint.

Different choices of interpretation (τ_k) of the stochastic integrals are appropriate for different applications. For example, the Itô integral is the common choice in finance and biology, mostly because the result I is a martingale. The Stratonovich integral is frequently used in physics and engineering. One of the main reasons for this is that it is defined such that the chain rule holds, and that they occur as solutions of Langevin equations. The Stratonovich interpretation is also the natural choice if we are working on differentiable manifolds rather than $\mathbb{R}^d$.

4.2. Itô Stochastic Integral and Itô Formula

The Itô Stochastic integral

$$I = \int_0^T f(t,\omega)\, dW_t \approx \sum_{k=1}^{K-1} f(t_{k-1})(W_{t_k} - W_{t_{k-1}})$$

satisfies some useful properties:

(i) Additivity: $\int_0^S f\, dW_t + \int_S^T f\, dW_t = \int_0^T f\, dW_t$

(ii) Linearity: $\int_0^T (\alpha f + \beta g)\, dW_t = \alpha \int_0^T f\, dW_t + \beta \int_0^T g\, dW_t$

for every $\alpha, \beta \in \mathbb{R}$, f, g measurable, $\mathcal{F}_t$ - adapted and s.t. $\mathbb{E}\int f^2 < \infty$, $\mathbb{E}\int g^2 < \infty$.

(iii) $\mathbb{E} \int_0^T f\, dW_t = 0$

(iv) $I_T = \int_0^T f\, dW_t$ is $\mathcal{F}_T$-measurable

(v) I_T admits a continuous version

(vi) It verifies the Itô isometry:

$$\mathbb{E}\left(\int_0^T f\,dW_t\right)^2 = \mathbb{E}\int_0^T f^2\,dt.$$

or more generally,

$$\mathbb{E}\left(\int_0^t f(u)\,dW_u \int_0^s g(u)\,dW_u\right) = \mathbb{E}\int_0^{\min(t,s)} f(u)g(u)\,du$$

(vii) If f is deterministic, then I_t is a Gaussian r.v. with mean zero and variance $\int^t f^2(s)\,ds$.

(viii) $I_t = \int_0^t f(s,\omega)\,dW_s$ is a square-integrable $\mathcal{F}_t$-martingale. $(\mathbb{E}(I(t)\mid\mathcal{F}_s) = I(s),\ \forall t \geq s)$

Let us go back to the SDE

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

An Itô process X_t is a process that satisfies

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

where the integral is in the Itô interpretation.

Even though the Stratonovich integral is such that the chain rule holds, this is not the case for the Itô integral. To overcome this, we use the Itô formula.