

# 18.676: Stochastic Calculus

Lecturer: Professor Nike Sun

Notes by: Andrew Lin

Spring 2020

## Introduction

Most of the logistical information is on the class website at [1], including an official class summary and many references to relevant papers and textbooks. Here are the main points for us: there will be homework roughly once every two weeks. The first two are already posted, and they'll be due February 12 and February 24 (submitted in class). Grading is weighted 55 percent for homework, 20 percent per exam, and 5 percent for attendance. Office hours are Monday 2–4 in Professor Sun's office, 2-432.

18.675 is a prereq, so we should talk to Professor Sun if we haven't taken that class. We will be using [2] as our main textbook.

## 1 February 3, 2020

Today, we'll do an informal overview of the topics we'll do in this class. We'll start with some basic reminders: the **standard Gaussian density**

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

should be burned into our head, and the variable  $Z \sim N(0, 1)$  is distributed according to this density. We should know that if  $X, X_i$  are **symmetric random signs** ( $\pm 1$  with equal probability), and  $S_n = \sum_{i=1}^n X_i$ , then  $\frac{S_n}{\sqrt{n}}$  converges in distribution to  $N(0, 1)$ . We should know how to prove this, either using the central limit theorem or by direct combinatorial calculation (this is because  $S_n$  is a scaling of the binomial distribution).

Next, we can consider the **simple random walk** on the integers, which gives us a process  $(S_n)_{n \geq 0}$  (where  $n$  is a time index): since  $\frac{S_n}{\sqrt{n}}$  converges in distribution to a Gaussian, this means that over time  $n$ , the walk typically covers a distance on the order of  $\sqrt{n}$ . So if we rescale time by  $n$  and rescale space by  $\sqrt{n}$ , we get a process

$$X^{(n)}(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}.$$

For any fixed  $t$ , we still have the central limit theorem as before:  $X^{(n)}(t) \xrightarrow{d} N(0, t)$ . But one idea of this class is that we don't need to consider a single  $t$ : the entire process

$$X^{(n)} = \left( X^{(n)}(t) \right)_{t \geq 0}$$

converges in distribution to  $(B_t)_{t \geq 0}$ , something called a **Brownian motion**, as  $n \rightarrow \infty$ . Note that we haven't defined a Brownian motion yet, and we haven't described the topology in which this converges in distribution. We'll do everything more formally later on.

So here are some of the main goals of this class:

- Formal construction of Brownian motion
- Convergence of natural processes (like a simple random walk), also known as a “functional CLT”
- Calculations with Brownian motion (stochastic calculus).

For now, though, we’ll keep surveying some more ideas from the course: we’re going to talk a bit about Itô’s formula and give an application to the conformal invariance of planar Brownian motion.

First of all, what are some properties of our Brownian motion  $B_t$ , given our informal definition above? We should have  $B_0 = 0$ , and for any  $0 \leq s \leq t$ , we should have  $B_t - B_s \sim N(0, t - s)$ . Also, for any  $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2$ ,  $B_{t_1} - B_{s_1}$  and  $B_{t_2} - B_{s_2}$  should be independent (because they correspond to disjoint parts of the random walk). In fact, these properties actually suffice to characterize Brownian motion completely.

So now let’s try to look a bit more at Itô’s formula. Consider a process that evolves as

$$dX_t = \mu_t dt + \sigma_t dB_t.$$

An informal way to write this looks like

$$X_{t+dt} - X_t = \mu_t dt + \sigma_t \cdot N(0, dt).$$

Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a twice-differentiable function. Note that if  $X_t$  follows a deterministic smooth trajectory, then we know how  $f(X_t)$  evolves: we just have  $df(X_t) = f'(X_t)dX_t$ . But if we expand the stochastic version out, we find that

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 \\ &= f'(X_t)[\mu_t dt + \sigma_t dB_t] + \frac{f''(X_t)}{2} [\mu_t dt + \sigma_t dB_t]^2 \end{aligned}$$

and because  $dB_t$  is on the order of  $\sqrt{dt}$ , it dominates the  $\mu_t dt$  term, so the squared term is just  $\sigma_t^2 (dB_t)^2 = \sigma_t^2 dt \cdot N(0, 1)^2$ .

What Itô’s formula says is that we can actually ignore the fluctuations in the  $N(0, 1)^2$  term if we take many measurements, and so that just disappears from the expression. Thus,

$$\begin{aligned} df(X_t) &= f'(X_t)[\mu_t dt + \sigma_t dB_t] + \frac{f''(X_t)}{2} \sigma_t^2 dt \\ &= \boxed{\left[ f'(X_t)\mu_t + \frac{f''(X_t)\sigma_t^2}{2} \right] dt + f'(X_t)\sigma_t dB_t}, \end{aligned}$$

and we’ve now separated the contribution into a **drift** and a **stochastic** term.

With this, let’s do an application to planar Brownian motion – first, we’ll review a bit of complex analysis. If we have a function  $f : \mathbb{C} \rightarrow \mathbb{C}$  or  $f : D \rightarrow \mathbb{C}$  for some open  $D$ , then  $f$  is **holomorphic** or **complex differentiable** at  $z \in \mathbb{C}$  if the complex derivative

$$f'(z) = \lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h} \in \mathbb{C}$$

exists. (Being complex differentiable is much stronger than being differentiable in  $\mathbb{R}^2$  because we approach 0 in all directions in the complex plane.) If we think of our function as going from  $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ , where  $z = x + iy$  and  $f = u + iv$  (so that  $u, v$  are real-valued functions and  $x, y$  are real numbers), then  $f$  is holomorphic at  $z$  if the limits from the real axis and imaginary axis are the same: this means

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}$$

and

$$\frac{1}{i} \frac{\partial f}{\partial y} = \frac{1}{i} \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}.$$

This gives us the **Cauchy-Riemann equations**

$$u_x = v_y, \quad u_y = -v_x.$$

One useful thing to know is that the Laplacian of the real part is

$$\Delta u = u_{xx} + u_{yy} = v_{xy} - v_{yx} = 0,$$

which means that the real part of any complex differentiable function is harmonic (and so is the imaginary part by an analogous calculation).

So now, consider a two-dimensional (standard) Brownian motion  $(X_t, Y_t)$ . This just means that  $X_t$  and  $Y_t$  are independent standard one-dimensional Brownian motions; alternatively, we can take  $Z_t = X_t + iY_t$ , which is a standard Brownian motion in  $\mathbb{C}$ . Suppose we have a **conformal** map  $f : D \rightarrow D'$  (which means that  $f$  is holomorphic and has a holomorphic inverse  $f^{-1} : D' \rightarrow D$ ). Again, let  $u = \operatorname{Re} f$  and  $v = \operatorname{Im} f$ .

**Question 1.** How does  $f(Z_t)$  evolve if  $Z_t$  stops when it hits the boundary  $D$ ?

We'll need a two-dimensional version of Itô's formula for this, but the same Taylor expansion idea works:

$$du(Z_t) = \begin{bmatrix} u_x(Z_t) & u_y(Z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix} + \frac{1}{2} \begin{bmatrix} dX_t & dY_t \end{bmatrix} \begin{bmatrix} u_{xx}(Z_t) & u_{xy}(Z_t) \\ u_{yx}(Z_t) & u_{yy}(Z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}.$$

When we expand this out, we get cross terms like  $dX_t \cdot dY_t$ , which look like  $dt \cdot N(0, 1)\tilde{N}(0, 1)$ . If we add up many of these, they cancel out and become negligible, so we don't have to worry about those: this means the second term will only have the diagonal term contributions

$$\frac{1}{2} (u_{xx}(dX_t)^2 + u_{yy}(dY_t)^2),$$

but now we can replace  $(dX_t)^2$  and  $(dY_t)^2$  with  $dt$  by the same argument as above, and now  $u_{xx} + u_{yy} = 0$  because  $u$  is harmonic. So the entire second-order term actually vanishes, and we're just left with (now doing the same calculations for  $v(Z_t)$ )

$$\boxed{\begin{bmatrix} du(Z_t) \\ dv(Z_t) \end{bmatrix} = \begin{bmatrix} u_x(Z_t) & u_y(Z_t) \\ v_x(Z_t) & v_y(Z_t) \end{bmatrix} \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}}.$$

This two-by-two matrix can also be written as

$$G = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} u_x & u_y \\ -u_y & u_x \end{bmatrix} = \sqrt{u_x^2 + u_y^2} \cdot \left[ \text{(rotation matrix)} \right].$$

Now  $\sqrt{u_x^2 + u_y^2}$  is actually just the modulus of  $f'(z) = u_x + iv_x$ , and the determinant of the rotation matrix has to be 1 (because we have a conformal map). So with this, we can conclude that if  $Z_t = X_t + iY_t$  is a standard Brownian motion in  $D \subseteq \mathbb{C}$ , and  $f : D \rightarrow D'$  is conformal, then as long as  $Z_t \in D$ ,  $f(Z_t)$  evolves via

$$\boxed{\begin{bmatrix} du(Z_t) \\ dv(Z_t) \end{bmatrix} = |f'(Z_t)| O(Z_t) \begin{bmatrix} dX_t \\ dY_t \end{bmatrix}},$$

where  $O(Z_t)$  is a  $2 \times 2$  rotation matrix. Note that the standard bivariate Gaussian  $N(0, I_{2 \times 2})$  is rotationally invariant

(spherically symmetric), so it is reasonable to believe that standard Brownian motion is also rotationally invariant: if  $O$  is a  $2 \times 2$  orthogonal matrix and  $Z$  is a BM in  $\mathbb{R}^2$ , then so is  $OZ$ .

We can also consider scaling: if  $B_t$  is a Brownian motion, then  $\sigma B_t$  is equal in distribution to  $B_{\sigma^2 t}$ . This is true for any fixed  $t$  because they're both Gaussian, but the idea is that we actually can't tell two images (of the sample paths) apart. So the Brownian motion is a self-similar fractal!

**Remark 2.** *But in the formula we've derived, the scale factor and the rotation depend on the given time. (This means, for example, that when  $|f'(Z_t)|$  is big, the process runs faster.) So the process itself is not conformally invariant, but the trace (the image of the motion) is indeed conformally invariant.*

With the rest of the time today, we'll give an example of a question we can answer with this class.

### Example 3

Consider a simple random walk on a grid with  $\varepsilon$  spacing on  $(\varepsilon\mathbb{Z}) \times (\varepsilon\mathbb{Z}_{\geq 0})$ , and say that we start our walk near  $(0, y)$ . We stop when we hit the horizontal axis. What's the law of the hitting location (the  $x$ -coordinate)?

We'll think about this problem when  $\varepsilon$  is small: then we can approximate this walk with a Brownian motion in the upper-half (complex) plane.

One way to approach this is to map this problem into an easier domain:  $f(z) = \frac{i-z/y}{i+z/y}$  maps  $\mathbb{H}$  conformally into the unit disk  $\mathbb{D}$ , and our new starting point is now the origin. Since we only care about the hitting location (and not the time), and the Brownian motion is spherically symmetric, the hitting location must be uniform on  $\partial\mathbb{D}$ : that means the distribution on  $\mathbb{H}$  can be easily recovered.

So for any interval  $[a, b] \in \mathbb{R}$ , we can go ahead and calculate the probability of hitting in that interval explicitly: because  $f(x) = \frac{i-x/y}{i+x/y}$ , and the angle covered is proportional to  $\frac{|f'(x)|}{2\pi} = \frac{y}{\pi(x^2+y^2)}$ ,

$$\begin{aligned} \mathbb{P}_{\mathbb{H}}(Z_{\tau} \in [a, b]) &= \mathbb{P}_{\mathbb{D}}(\tilde{Z}_{\sigma} \in f([a, b])) \\ &= \int_a^b \frac{y \, dx}{\pi(x^2 + y^2)}. \end{aligned}$$

This last integrand is also called the **Poisson kernel** for  $\mathbb{H}$  (call it  $P_y(x)$ ), because it's closely connected to the Dirichlet problem on  $\mathbb{H}$ : if we're given a nice function  $b: \mathbb{R} \rightarrow \mathbb{R}$  and we want to know the harmonic interpolation of  $b$  to  $\mathbb{H}$ , the answer is given by

$$h(x, y) = \mathbb{E}[b(Z_{\tau}) | Z_0 = x + iy].$$

That is, we start a Brownian motion at  $x + iy$  and find the expected value of the hitting value. (We can prove this by looking at a finite graph or with direct calculation, and we'll talk about it more later.) But explicitly, this is

$$\mathbb{E}[b(Z_{\tau} + x) | Z_0 = iy] = \int_{\mathbb{R}} P_y(s) b(s + x) \, ds.$$

Since  $P_y$  is symmetric, we can replace  $s$  with  $-s$ , and this expected value becomes a convolution  $(b * P_y)(x)$  with the Poisson kernel.

## 2 February 5, 2020

Today, we'll start to formalize some of the ideas from yesterday's informal overview. Specifically, we'll be starting with Gaussian processes and Gaussian spaces. (We'll follow the textbook pretty closely for now.)

All random variables will live on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definition 4**

A  **$d$ -dimensional Gaussian vector** is an  $\mathbb{R}^d$ -valued random variable  $X$  such that  $\langle X, u \rangle$  is a one-dimensional Gaussian variable for any  $u \in \mathbb{R}^d$ .

This is somewhat fancier than other definitions and doesn't depend on a choice of basis. In addition, this definition doesn't specify that  $\langle X, u \rangle$  and  $\langle X, v \rangle$  need to be jointly Gaussian, but we'll see that it is a consequence of the definition.

**Proposition 5**

The law of  $X$  is uniquely determined by the mean vector  $\mu = \mathbb{E}X \in \mathbb{R}^d$  and the covariance matrix  $\Sigma = \mathbb{E}[(X - \mu)(X - \mu)^t] \in \mathbb{R}^{d \times d}$ .

*Proof.* Take any  $\theta \in \mathbb{R}^d$ . By definition,  $\langle X, \theta \rangle$  is Gaussian with some parameters, and now we have

$$\mathbb{E}\langle X, \theta \rangle = \langle \mathbb{E}X, \theta \rangle = \langle \mu, \theta \rangle$$

by linearity of expectation, and

$$\text{Var}(\langle X, \theta \rangle) = \text{Cov}(\langle X, \theta \rangle, \langle X, \theta \rangle) = \theta^t \Sigma \theta$$

because covariance is bilinear. This means that  $\mu$  and  $\Sigma$  tell us the distribution of  $\langle X, \theta \rangle$ , which tells us the characteristic function

$$\phi_X(\theta) = \mathbb{E} \exp(i\langle X, \theta \rangle) = \exp\left(i\langle \mu, \theta \rangle - \frac{1}{2}\theta^t \Sigma \theta\right),$$

and as we saw in 18.675, knowing all of these characteristic functions is enough to determine the law of  $X$ .  $\square$

Because of this, we'll use the notation  $X \sim N(\mu, \Sigma)$ , where  $\mu \in \mathbb{R}^d$  and  $\Sigma \in \mathbb{R}^{d \times d}$  is a symmetric positive semidefinite matrix (because the variance is nonnegative). Any such matrix has a Cholesky factorization  $\Sigma = AA^t$ , where  $A$  is a  $d \times r$  matrix. Then if we take  $Z$  to be a standard normal in  $r$  dimensions ( $r$  iid copies of a Gaussian), then

$$\mu + AZ \sim N(\mu, \Sigma).$$

(One way to check that this works is that the Fourier transform matches up.) So now if and only if  $\Sigma$  has full rank, we can take  $\Sigma = AA^t$  for an invertible square matrix  $A$ . In this case only, we can use the change of variables formula to find that  $X$  has density

$$\frac{\exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1}(x - \mu)\right)}{(2\pi)^{d/2} |\det \Sigma|^{1/2}}.$$

(If  $r < d$ , then the law of  $X$  is supported on a subspace of  $\mathbb{R}^d$  of lower dimension, so it has no density.)

One important point: for Gaussians, being independent and being uncorrelated are the same thing:

**Lemma 6**

If  $X \sim N(\mu, \Sigma)$  is a  $d$ -dimensional Gaussian vector, then  $X_i$  are mutually independent if and only if  $\Sigma$  is diagonal.

*Proof.* The forward direction is easy (independent implies uncorrelated). For the other direction, suppose  $\Sigma$  is a diagonal matrix. Then the characteristic function

$$\phi_X(\theta) = \prod_{j=1}^d \exp\left(i\theta_j - \frac{1}{2}\sigma_{jj}\theta_j^2\right)$$

has no cross-terms, so the characteristic polynomial factorizes (meaning we have independence).  $\square$

In contrast, there's the standard nonexample of  $Z \sim N(0, 1)$  and  $\varepsilon \sim \text{Unif}(\{\pm 1\})$ . Then the covariance between  $\varepsilon Z$  and  $Z$  is zero, but the two variables aren't independent because they have the same absolute value (basically, this goes wrong because the two variables aren't jointly Gaussian).

With this, we'll move on to the idea of a Gaussian space on  $(\Omega, \mathcal{F}, \mathbb{P})$ . (For the rest of today, we'll assume Gaussians are **centered**, meaning they have mean zero.) Recall that  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  is the space of all ( $\mathbb{R}$ -valued) random variables on our probability space with finite second moment – this is a Hilbert space with inner product

$$\langle X, Y \rangle_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \int_{\Omega} X(\omega)Y(\omega)d\mathbb{P}(\omega) = \mathbb{E}(XY).$$

### Definition 7

A **(centered) Gaussian space** is a closed linear subspace of  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  containing only centered Gaussian variables.

### Example 8

Take  $X \sim N(0, \Sigma)$  in  $\mathbb{R}^d$ . Then the span of the coordinate random variables  $\{X_1, \dots, X_d\}$  is a Gaussian space. Meanwhile, a non-example is the span of  $Z \sim N(0, 1)$  and  $\varepsilon Z$  (where  $\varepsilon$  is the random sign as before); this doesn't work because  $Z + \varepsilon Z$  is zero half the time.

Gaussian spaces are important because they **turn probability into geometry**. Independence of Gaussian variables will turn into orthogonality in the space:

### Theorem 9

Let  $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$  be a centered Gaussian space, and let  $(H_\alpha)_{\alpha \in I}$  be linear subspaces of  $H$ . Then the  $\sigma$ -fields  $(\sigma(H_\alpha))_{\alpha \in I}$  are independent if and only if the  $H_\alpha$  are pairwise orthogonal.

We can read the book for this – it's really a fancier version of Lemma 6. A related point is that conditional expectation among Gaussians corresponds to orthogonal projection in a Gaussian space. The first hypothesis is important here, by the way – it's important that the  $H_\alpha$  are all subspaces of a single centered Gaussian space to ensure that they are jointly Gaussian.

### Example 10

Suppose we have a bivariate normal (two-dimensional Gaussian)

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a & b \\ b & c \end{bmatrix} \right).$$

(Assume  $\Sigma$  is positive definite.) What is the law of  $Y$  conditional on  $X$ ?

The standard trick people use is to solve for  $\theta$  such that  $Y - \theta X$  is independent from  $X$ . Here,  $Y - \theta X$  will also be jointly Gaussian from  $X$  and  $Y$ , so we just need to make sure that

$$0 = \text{Cov}(Y - \theta X, X) = b - \theta a,$$

so we should set  $\theta = \frac{b}{a}$ . Then we can break  $Y$  up into a “parallel” and a “perpendicular” part:

$$Y = \frac{b}{a}X + \left(Y - \frac{b}{a}X\right),$$

where the first term is in  $\sigma(X)$  and the rest is independent of  $X$  – specifically, it’s a Gaussian with variance

$$\text{Var}\left(Y - \frac{b}{a}X\right) = \text{Cov}\left(Y - \frac{b}{a}X, Y\right) = c - \frac{b^2}{a}.$$

So altogether, we know that

$$Y|X \sim N\left(\frac{bX}{a}, c - \frac{b^2}{a}\right).$$

More generally (as an exercise), if

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}\right),$$

where  $X$  lives in a  $k$ -dimensional space and  $C$  lives in an  $\ell$ -dimensional space, we will find that

$$Y|X \sim N(B^t A^{-1} X, C - B^t A^{-1} B).$$

### Theorem 11

Let  $H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$  be a centered Gaussian space, and let  $K \subseteq H$  be a closed subspace. Then for any  $X \in H$ , we have

$$X|\sigma(K) = N(\pi_K(X), \mathbb{E}[(X - \pi_K(X))^2]).$$

(where  $\pi_K$  denotes the orthogonal projection of  $X$  onto  $K$ ).

The mean here is the “parallel” part, and the variance is the “perpendicular” part. (Also,  $\pi_K(X)$  is measurable with respect to  $\sigma(K)$ , so it is known to us already when we’re conditioning on  $\sigma(K)$ .)

Note that for a general  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ , we have

$$\mathbb{E}(X|\sigma(K)) = \pi_{L^2(\Omega, \sigma(K), \mathbb{P})}(X).$$

$\sigma(K)$  is generally very big: if  $K$  is the span of some variable  $Z$ , then  $\sigma(K)$  is the set of all measurable functions of  $Z$ . So this theorem tells us that we can project onto a smaller subspace in the Gaussian case.

### Example 12 (Kalman filter; on homework)

We have independent Gaussians  $\varepsilon_n \sim N(0, \sigma^2)$ ,  $\eta_n \sim N(0, \delta^2)$ . We have some true unknown state of a system which evolves over time:

$$0 = X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n, \quad X_{n+1} = a_n X_n + \varepsilon_{n+1}.$$

We’re given a noisy observation  $Y_i$  at each time, which satisfies  $Y_n = cX_n + \eta_n$ . (We know  $a_n$  and  $c$  and  $\sigma^2$  and  $\delta^2$ .) Our goal is to find

$$\mathbb{E}(X_n | Y_1, \dots, Y_n).$$

One idea is to think of  $H$  as the span of the  $\varepsilon_i$  and  $\eta_i$  up to some time  $n$ : this is a Gaussian space. (This is all of the noise going into the system up to time  $n$ .) From the way this is designed, all of the  $X_i$  and  $Y_i$  are linear

combinations of the  $\epsilon$ s and  $\eta$ s, so we always stay in the Gaussian space. Then if we want  $\mathbb{E}(X_n|Y_1, \dots, Y_n)$ , we just need an orthogonal projection – specifically,  $X_n$  must be a linear combination of  $Y_1$  up to  $Y_n$ .

Next, we'll talk a bit more about the Gaussian process, which is a generalization of a Gaussian vector.

**Definition 13**

Let  $I$  be an arbitrary (possibly uncountable) index set. Then a collection of random variables  $(X_t)_{t \in I}$  is a **(centered) Gaussian process** if any finite linear combination of the  $X_t$  is a one-dimensional Gaussian. The **Gaussian space generated by**  $(X_t)_{t \in I}$  is the closure of the linear span of the  $X_t$ 's. We define the **covariance function**  $\Gamma : I \times I \rightarrow \mathbb{R}$  via  $\Gamma(s, t) = \text{Cov}(X_s, X_t)$ .

$\Gamma$  is symmetric and positive semidefinite because

$$\sum_{s,t} \theta(s)\theta(t)\Gamma(s, t) \geq 0$$

for any  $\theta$  which is nonzero for **finitely many values** in  $I$ . (Otherwise, this statement may not make any sense.)

So given a  $\Gamma$ , does there necessarily exist a Gaussian process with that covariance function? The answer is yes, and this follows basically from the Kolmogorov extension theorem (which we proved in 18.675, so we won't do now). The most important example for us will be the construction of Brownian motion based on a Gaussian process with index set  $\mathbb{R}_{\geq 0}$ : that will satisfy

$$\Gamma(s, t) = \text{Cov}(B_s, B_t) = \min(s, t)$$

(because if WLOG  $s < t$ , then  $B_t = B_s + (B_t - B_s)$ , and the second term is independent of  $B_s$ ).

So why can't we just cite the Kolmogorov extension theorem right now? We want to make sure that with probability 1, our process is continuous in  $t$  – that is, for all  $\omega \in (\Omega, \mathcal{F}, \mathbb{P})$ ,  $B_t(\omega)$  is continuous in  $t$ , and that's not a guarantee we have. Specifically, it'll give us a measure on  $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$ , but that's not really the space we want to use – we want the space of continuous functions instead. So we'll come back to this a little later.

On our homework, though, there's a different construction of Brownian motion, and this one is based on the construction of "white noise," which we'll do for the rest of class. The heuristic idea is that on every "pixel" of space, we see an independent Gaussian random variable, so we get "snow on a TV screen." Here's a more formal definition:

**Definition 14**

Let  $(E, \mathcal{E})$  be a measurable space, and let  $\mu$  be a  $\sigma$ -finite measure on  $(E, \mathcal{E})$ . Then a **Gaussian white noise** on  $(E, \mathcal{E}) **with intensity  $\mu$**  is a linear isometry  $G : L^2(E, \mathcal{E}, \mu) \rightarrow H \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$  (a centered Gaussian space).$

(An **isometry** preserves the inner product, so  $\langle f, g \rangle = \langle G(f), G(g) \rangle$ ). This means that the covariance between  $G(f)$  and  $G(g)$  is the same as the inner product between  $f$  and  $g$ .)

So for any "patch" of size  $dx$  around  $x$ , we assign a Gaussian variable  $N(0, \mu(dx))$ , and we do this for all points independently. So if we take some subset  $A \subseteq E$ , we get the sum of all of the little random variables: the intensity looks like

$$G(1_A) = G(A) \sim N(0, \mu(A)).$$

This is an isometry because  $A$  and  $B$  being disjoint means  $G(A) \perp\!\!\!\perp G(B)$ :

$$0 = \langle 1_A, 1_B \rangle = \text{Cov}(G(1_A), G(1_B)).$$



And informally, we can think of

$$G(f) = \sum_{x \in E} f(x) Z_x$$

where  $Z_x = N(0, \mu(dx))$ , which helps us see that

$$\text{Cov}(G(f), G(g)) = \text{Cov}\left(\sum_x f(x) Z_x, \sum_x g(x) Z_x\right) = \sum_x f(x) g(x) \mu(dx)$$

(only the diagonal terms come out), which gives exactly the inner product between  $f$  and  $g$ . Of course, this doesn't exactly make sense, which is why we have the formal definition.

However, even when we're given the formal definition, we still need to ask whether we can construct such an object. The last question on our homework asks us to construct an explicit white noise  $G$ , and we'll get Brownian motion from that via  $B_t = G(1_{[0,t]})$ . This will be a real Brownian motion – it'll have the right covariance properties, and we'll find that  $B$  is continuous almost surely. (And this gives the historically older construction of Brownian motion.)

We'll finish by contrasting with something else we might have seen - compare this Gaussian white noise of  $N(0, \mu(dx))$  with the Poisson random measure, where for each patch of  $dx$ , we assign a Bernoulli random variable of parameter  $\mu(dx)$ . Then taking subsets  $A \subseteq E$ , we'll get a normal distribution in the white noise case with variance  $\mu(A)$ , but a Poisson distribution of parameter  $\mu(A)$  in the Poisson random measure case.

### 3 February 10, 2020

Last time, we talked about the general definition of finite-dimensional Gaussian vectors, Gaussian spaces, Gaussian processes, and Gaussian white noise. Today, we'll talk about the construction of Brownian motion, but we'll do a few things first.

Recall that if we have a covariance function  $\Gamma : I \times I \rightarrow \mathbb{R}$  which is symmetric and positive semidefinite, then there exists a Gaussian process  $(X_t)_{t \in I}$  with covariance function  $\text{Cov}(X_s, X_t) = \Gamma(s, t)$ . (We are often working with  $I = [0, \infty)$ .) We showed this with the **Kolmogorov extension theorem** – informally, the idea is that to define a measure on  $\mathbb{R}^I$ , we need to be able to write down the joint law for any finite subset of  $I$  in a consistent way, and then the Kolmogorov extension theorem gives us the measure  $(\mathbb{R}^I, \mathcal{B}_{\mathbb{R}}^{\otimes I}, \nu)$ . In particular, if we're given  $\Gamma(s, t)$ , we can just use  $\Gamma$  to get a covariance matrix for every finite collection of points in  $I$ .

#### Definition 15

In the special case where  $I = [0, \infty)$  and  $\Gamma(s, t) = \min(s, t)$ , the resulting Gaussian process  $(X_t)_{t \geq 0}$  is called a **pre-Brownian motion**.

Here's a quick connection to the material from the end of last lecture:

#### Proposition 16

Let  $(X_t)_{t \geq 0}$  be a real-valued stochastic process (any collection of random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$  indexed by  $t$ ). Then the following are equivalent:

- $X$  is a pre-Brownian motion.
- We can express  $X_t = G([0, t]) = G(1_{[0,t]})$ , where  $G$  is a Gaussian white noise on  $I = [0, \infty)$  with intensity equal to the Lebesgue measure.

*Proof.* The backwards direction follows directly from the definition of a Gaussian white noise –  $G$  is defined to be an isometry, so we get the correct covariance function. For the forward direction, we're given a pre-Brownian motion, and we need to construct an isometry. Suppose  $f$  is a step function

$$f(t) = \sum_{i=1}^n a_i 1_{(t_{i-1}, t_i]} :$$

for such a function, we define

$$G(f) = \sum_{i=1}^n a_i (X_{t_i} - X_{t_{i-1}}).$$

This defines  $G$  on the step functions, and we can check that  $G$  is an isometry on this class of functions – specifically, if  $h(t) = \sum_{i=1}^n b_i 1_{(t_{i-1}, t_i]}$  (without loss of generality they have the same break points  $t_i$ ), then the covariance of  $f$  and  $h$  satisfies

$$\langle G(f), G(h) \rangle = \sum_{i=1}^n a_i b_i (t_i - t_{i-1}) = \langle f, h \rangle$$

because the increments of  $X$  are independent by the definition of a pre-Brownian motion. This gives a isometry  $G$  from the step functions to the Gaussian space  $H$  spanned by  $X$ , and we need to define  $G$  on all of  $L^2$ . But the step functions are dense in  $L^2([0, \infty))$ , so we can extend  $G$  by finding step functions  $f_n$  that converge in  $L^2$  to a general  $f \in L^2([0, \infty))$ .

Because  $f_n$  converge to  $f$  in  $L^2$ , they form a Cauchy sequence in  $L^2$ , and  $G$  preserves distances,  $G(f_n)$  is a Cauchy sequence in the Gaussian space. Since the Gaussian space is a subspace of  $L^2$ , this means  $G(f_n)$  converges in  $L^2$  to a limit  $G(f)$ , which is the definition that we want.  $\square$

As a reminder, we're going to construct a specific white noise  $G$  which guarantees continuity of sample paths. But the generic definition we have here doesn't contain such a guarantee, because the Kolmogorov extension theorem gives us a process  $X = (X_t)_{t \geq 0}$  which is a random element of the space  $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$  for  $I = [0, \infty)$ . What events are actually in this sigma-algebra? We can ask events of the form

$$\{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\},$$

where  $A_i$  are Borel subsets of the real line, or we can ask things of the form

$$\{X_t = 0 \quad \forall t \in \mathbb{Q}\}$$

(because these are a countable intersection of the events above). Events that are **not** measurable are things like

$$\{X_t = 0 \quad \forall t \in I\},$$

because an uncountable intersection of events need not be measurable, and similarly we can't ask either of the questions

$$\{X_t \text{ continuous in } t\}, \{X_t \text{ measurable in } t\},$$

because they require us to know about  $X_t$  on uncountably many values of  $t$ . So the probability space isn't rich enough to capture properties like continuity – here is an example of something that can go wrong:

### Example 17

Let  $X$  be a Gaussian process with  $\Gamma(s, t) = \min\{s, t\}$  on  $(\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu)$ , and let's introduce some additional randomness: augment the probability space to include the uniform random variable  $U \sim \text{Unif}[0, 1]$ , independent of  $X$ . (This means we've moved to a larger space  $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^I, \mathcal{B}^{\otimes I}, \nu) \otimes ([0, 1], \mathcal{B}, \text{Leb})$ , where we draw  $X$  and then independently draw  $U$ .) Say we pick  $\omega' \in \mathbb{R}^I$  and  $u \in [0, 1]$ , and now define a new variable

$$\tilde{X}_t(\omega', u) = X_t(\omega') + 1\{t = u\}.$$

Also, define  $X_t(\omega', u)$  to just be  $X_t(\omega')$ .

Now  $X_t$  and  $\tilde{X}_t$  are closely related – for any fixed  $t$ ,  $\mathbb{P}(X_t = \tilde{X}_t)$  are equal with probability 1, because any fixed time  $t$  being equal to  $u$  has probability 0. But these are not the same process – they can't both be continuous because we add a 1 at some random time. So  $X_t$  and  $\tilde{X}_t$  are both Gaussian processes with the correct covariance, but we can't guarantee continuity even though we have the same finite-dimensional marginals here.

### Definition 18

If we have two processes  $(X_t)_{t \in I}$  and  $(\tilde{X}_t)_{t \in I}$ , then  $\tilde{X}$  is a **modification** of  $X$  if  $\mathbb{P}(X_t = \tilde{X}_t) = 1$  for any fixed time  $t \in I$ .  $X$  and  $\tilde{X}$  are **indistinguishable** if

$$\{X_t \neq \tilde{X}_t \text{ for any } t \in I\} \subseteq N$$

for some measure-zero set  $N$  (because the event may not actually be measurable).

( $X_t$  and  $\tilde{X}_t$  here are not indistinguishable, because they will be different at some time  $t = u$  with probability 1.)

The main goal of this lecture is to construct a modification of  $X$  – we'll have to change the probability space a little bit. The actual construction is very concrete, though: we'll first define Brownian motion on  $[0, 1]$ , looking at the dyadic points

$$D_n = \left\{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, \frac{2^n - 1}{2^n}\right\}.$$

These sets  $D_i$  are nested in each other ( $D_0 \subseteq D_1 \subseteq \dots$ ), and their infinite union  $D = \bigcup_{n=0}^{\infty} D_n$  is countable and dense in  $[0, 1]$ . At a very high level,  $D$  is a countable dense subset, so we only look at the process  $X$  on  $D$ . Then for any other value not in  $D$ , we'll define the process using continuity – we'll just use a limit, and we just need to show that we get a continuous process.

### Lemma 19

Let  $f : D \rightarrow \mathbb{R}$  be a function that satisfies

$$\left| f\left(\frac{i}{2^n}\right) - f\left(\frac{i-1}{2^n}\right) \right| \leq \frac{K}{2^{n\alpha}}$$

for all  $n \geq 1$  and  $1 \leq i \leq 2^n - 1$  for some  $\alpha > 0$  and  $K$  constant. Then  $f$  satisfies this type of estimate for all points in  $D$ :  $|f(s) - f(t)| \leq K'|s - t|^\alpha$  for all  $s, t \in D$  and  $K' = \frac{2K}{1-2^{-\alpha}}$  – this can be stated as  $f$  being an  **$\alpha$ -Hölder function** on  $D$ .

(This is completely deterministic – there's no probability going on here, and note that  $f$  is defined on  $D$  only.)

*Proof.* Without loss of generality, let  $s < t$ . Then we can take an integer  $p$  such that

$$\frac{1}{2^p} \leq t - s \leq \frac{1}{2^{p-1}},$$

which means  $s$  and  $t$  are either in “adjacent”  $\frac{1}{2^p}$  blocks or separated by one block. Either way, let  $s_0$  be the smallest point in  $D_p$  larger than  $s$  and  $t_0$  be the largest point smaller than  $t$ . Then, the point is that “the best way to get from  $s$  to  $t$  should use the largest jumps, because the small jumps don’t give a good estimate,” so formally we can write

$$s = s_0 - \sum_{\ell=1}^n \frac{\delta_\ell}{2^{p+\ell}}, \quad \delta_\ell \in \{0, 1\}$$

(just use each of the largest jumps if we can), and similarly

$$t = t_0 + \sum_{\ell=1}^n \frac{\eta_\ell}{2^{p+\ell}}, \quad \eta_\ell \in \{0, 1\}.$$

Then in the worst case, we take all of the individual steps, so

$$|f(s) - f(t)| \leq \frac{K}{2^{p\alpha}} + 2 \sum_{\ell \geq 1} \frac{K}{2^{(p+\ell)\alpha}} \leq \frac{K'}{2^{p\alpha}} \leq K'|s - t|^\alpha.$$

□

### Lemma 20

Suppose that  $(X_t)_{t \in [0,1]}$  is **any** stochastic process satisfying  $\mathbb{E}[|X_s - X_t|^q] \leq C|s - t|^{1+\varepsilon}$  for all  $s, t \in [0, 1]$ . Then for all  $\alpha \in \left(0, \frac{\varepsilon}{q}\right)$ , there exists a  $K_\alpha(\omega) < \infty$  such that

$$\left| X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega) \right| < \frac{K_\alpha(\omega)}{2^{n\alpha}}$$

for all  $n \geq 1$  and  $1 \leq i \leq 2^n - 1$ .

Basically, the conditions from the previous lemma hold, but now we have a random  $K$ .

*Proof.* Let  $A_n$  be the event  $\{\omega : |X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)| \geq \frac{1}{2^{n\alpha}} \text{ for any } 1 \leq i \leq 2^n - 1\}$  (the numerator of 1 here is good enough for the calculations). Ideally, this holds for all  $n$  and all  $\omega$  – instead, by union bound and Markov’s inequality, we have that

$$\mathbb{P}(A_n) \leq 2^n \cdot 2^{n\alpha q} \cdot \mathbb{E}|X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)|^q \leq 2^n 2^{n\alpha q} C \left(\frac{1}{2^n}\right)^{1+\varepsilon} = \frac{C}{2^n(\varepsilon - \alpha q)}$$

last step by assumption (the  $2^n$  is the number of  $i$ s, and then we apply Markov’s inequality). By Borel-Cantelli, because we have a geometric series,  $\mathbb{P}(A_n \text{ i.o.}) = 0$ , so if  $\omega$  lies outside this measure-zero event  $\{A_n \text{ i.o.}\}$ , we have some  $N(\omega)$  such that  $\omega \notin A_n$  for all  $n > N(\omega)$ . That means that with probability 1,

$$\sup_{n \geq 1} \left\{ \max_{1 \leq i \leq 2^n - 1} |X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)| \cdot 2^{n\alpha} \right\}$$

just has contributions of 1 for all  $n \geq N(\omega)$ , so this is

$$\leq \max \left\{ 1, \max_{n < N(\omega)} \left\{ \max_{1 \leq i \leq 2^n - 1} |X_{i/2^n}(\omega) - X_{(i-1)/2^n}(\omega)| \cdot 2^{n\alpha} \right\} \right\}$$

which defines the  $K_\alpha(\omega)$  that we want. □

**Lemma 21**

Under the same assumptions as Lemma 20, there is a modification  $\tilde{X}$  of  $X$  whose sample paths are continuous – in fact, they are  $\alpha$ -Hölder continuous for all  $\alpha \in \left(0, \frac{\varepsilon}{q}\right)$ .

*Proof.* Let  $E$  be the event that the boxed estimate of Lemma 20 holds for  $D$  –  $E$  is the complement of the event  $\{A_n \text{ i.o.}\}$  from the previous proof. Then by Lemma 19, we know that

$$|X_s(\omega) - X_t(\omega)| \leq K'_\alpha(\omega)|s - t|^\alpha$$

for all  $\omega \in E$  and all  $s, t \in D$ . Now we define  $\tilde{X}$  in the way we said we would:

$$\tilde{X}_t(\omega) = \begin{cases} \lim_{s \rightarrow t, s \in D} X_s(\omega) & \text{if } \omega \in E \\ 0 & \text{otherwise.} \end{cases}$$

Then for all  $\omega$ ,  $\tilde{X}$  is an  $\alpha$ -Hölder continuous function – it satisfies the same estimate with the same  $K'$ :

$$|\tilde{X}_s - \tilde{X}_t| \leq K'_\alpha(\omega)|s - t|^\alpha$$

now for all  $s, t \in [0, 1]$ . We do still need to check that  $\tilde{X}$  is a modification of  $X$  – in principle, it seems like we've changed a lot, because we've ignored  $X$  everywhere except on a countable set. But remember that we have the assumption

$$\mathbb{E}|X_s - X_t|^q \leq C|s - t|^{1+\varepsilon},$$

so as  $s$  converges to  $t$ , the right hand side converges to 0, meaning  $X_s$  goes to  $X_t$  in  $L^q$  and therefore also in probability. On the other hand, by definition,  $X_s$  goes almost surely to  $\tilde{X}_t$  as  $s \rightarrow t$  for  $s \in D$ . So we can use the fact that  $X_t = \tilde{X}_t$  almost surely, because these are both limits of  $X_s$  as  $s$  approaches  $t$ . □

Let's go back to the Markov bound

$$\mathbb{P}(A_n) \leq 2^n 2^{n\alpha q} \cdot C \left(\frac{1}{2^n}\right)^{1+\varepsilon}.$$

We've used the dyadic partitioning twice here: once to go from  $s$  to  $t$ , where  $s, t \in D_n$  are separated by many intervals of  $\frac{1}{2^n}$ , and once in this Markov bound to use the same bound many times. A more naive way we could have done our bound was to say that because we have  $|s - t| = \frac{L}{2^n}$  (for some  $L$ ), we might want to look at

$$\mathbb{P}\left(|X_s - X_t| \leq \left(\frac{L}{2^n}\right)^\alpha \quad \forall s \in D_n, t = s + \frac{L}{2^n}\right).$$

Doing a union bound here doesn't use the fact that our intervals are overlapping, and this would give us a bound of the form

$$\leq 2^n \cdot \frac{C \left(\frac{L}{2^n}\right)^{1+\varepsilon}}{(L/2^n)^{\alpha q}} = \frac{C \cdot L^{1-(\alpha q - \varepsilon)}}{2^n(\alpha q - \varepsilon)}.$$

The  $L^{1-(\alpha q - \varepsilon)}$  is large here (possibly close to order of  $L$ ), and  $L$  is possibly on the order of  $2^{n-1}$ , so our union bound has lost a lot from the overlapping intervals. So the point is that we should be reusing bounds that we already have! The combination of these three lemmas is called the **Kolmogorov continuity lemma**.

So let's finish this class by applying this to Brownian motion: for some  $q$ , we want to look at

$$\mathbb{E}(|X_s - X_t|^q) = \mathbb{E}[N(0, |s - t|)^q] = |s - t|^{q/2} \mathbb{E}(|Z|^q)$$

because  $X_s - X_t$  is a centered normal random variable with variance  $t - s$ . So we set  $\frac{q}{2} = 1 + \varepsilon$ , and this gives us (by the above lemmas) a modification that is  $\alpha$ -Hölder for  $\alpha < \frac{\varepsilon}{q} = \frac{q/2-1}{q}$ . This estimate is true for any positive  $q$ , so in particular if we take large  $q$ , this approaches  $\frac{1}{2}$ . So our modification is just short of  $\frac{1}{2}$ -Hölder continuous, and the obvious question is whether this is optimal. It turns out that the answer is yes, and this is the last part of our homework.

So we have a process with continuous sample paths and the correct covariance – next time, we'll talk about the probability space that is "canonical" for this Brownian motion.

## 4 February 12, 2020

As a reminder, the website for this class contains links to some useful references. Scrolling to the bottom of class contains Professor Sun's summary (which are more concise when the class follows the textbook), as well as these notes you're reading now. (These shouldn't be considered an official resource, though – they aren't being checked by the course instructors.)

Recall from last time that we defined a **pre-Brownian motion** to be a Gaussian process with covariance function  $\Gamma(s, t) = \min(s, t)$ , which allowed us to define a **Brownian motion** to be a pre-Brownian motion with continuous sample paths.

So now suppose that  $(X_t)_{t \in [0,1]}$  is a pre-Brownian motion on any probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , as long as it is rich enough to support such a process. The main content of last class was the Kolmogorov continuity lemma, which basically tells us that a **modification**  $(B_t)_{t \in [0,1]}$  of  $X$  exists with sample paths  $\alpha$ -Hölder for all  $\alpha \in (0, \frac{1}{2})$ . In other words, there exists a  $K_\alpha(\omega) < \infty$  such that

$$|B_s(\omega) - B_t(\omega)| \leq K_\alpha(\omega) \leq |s - t|^\alpha \quad \forall s, t \in [0, 1],$$

and in particular this tells us that  $B_t$  is also continuous. (We constructed this by only looking at the dyadic set, showing continuity there, and then taking limits.)

So what if we're looking at a pre-Brownian motion  $(X_t)_{t \geq 0}$  for all nonnegative  $t$  rather than just the interval  $[0, 1]$ ? The idea is that we can just apply the above results for  $[i, i + 1]$  for all integer  $i$ : this tells us that a modification  $(B_t)_{t \geq 0}$  exists which is **locally**  $\alpha$ -Hölder (on every compact interval) and therefore also continuous – this means that  $B$  satisfies the definition of a Brownian motion.

Letting  $I = [0, \infty)$ , define

$$C(I) = \{\text{continuous functions } I \rightarrow \mathbb{R}\} \subset \{\text{all functions } I \rightarrow \mathbb{R}\} = \mathbb{R}^I.$$

The sigma-algebra we can put on  $\mathbb{R}^I$  is  $\mathcal{Q} = \mathcal{B}^{\otimes I}$ , so a natural sigma-algebra to put on  $C(I)$  is

$$\mathcal{G} = \mathcal{Q}|_{C(I)} = \{C(I) \cap A : A \in \mathcal{Q}\}.$$

The Brownian motion  $B$  gives us a measurable mapping (exercise) which induces a measure on  $C(I)$ :

$$B : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (C(I), \mathcal{G}, P = B_{\#}\mathbb{P}).$$

In other words, for all  $A \in \mathcal{G}$ , we assign

$$P(A) = \mathbb{P}(B^{-1}(A)),$$

where  $B^{-1}$  denotes the pre-image of the Brownian motion. Note that  $(\Omega, \mathcal{F}, \mathbb{P})$  is not unique, because there can be all kinds of “extra randomness” in the space, but we do have some uniqueness:

**Claim 22.** *The measure  $P$ , called the **Wiener measure**, is unique.*

*Proof.* It suffices to show that the value of  $P$  is uniquely determined, and it’s enough to check this for a pi-system which generates the sigma-algebra  $\mathcal{G}$ . Consider the “simple” set

$$P(B_{t_1} \in [a_1, b_1], \dots, B_{t_n} \in [a_n, b_n]);$$

events of this type generate  $\mathcal{G}$ , and we can calculate this probability using the covariance matrix  $\Gamma(t_i, t_j)$  (where  $i, j$  run from 1 to  $n$ ):

$$= \mathbb{P}\left(N(0, \Gamma(t_i, t_j)) \in \prod [a_i, b_i]\right).$$

This is an explicit value we can find by calculating an integral, and this characterizes the value of  $P(E)$  for all events  $E$  in the pi-system generating  $\mathcal{G}$ , so we’re done.  $\square$

**Fact 23**

We’ve defined a sigma-algebra  $\mathcal{G}$  on  $C(I)$  by restriction, but here’s another way to look at it. A natural topology to put on the the space is to say that  $f_n$  converges to  $f$  if  $f_n$  **converges to  $f$  uniformly on compact sets** – this topology is metrizable, because we can define something like

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min \left\{ 1, \sup_{t \in [0, n]} |f(t) - g(t)| \right\} \leq 1.$$

Then indeed  $d(f_n, f)$  goes to 0 if and only if  $f_n \rightarrow f$  locally uniformly, so the metric captures the topology.

**Claim 24.** *The above  $\mathcal{G}$  is the same as  $\mathcal{H}$ , the Borel  $\sigma$ -algebra of  $C(I)$  in the  $d$ -topology.*

*Proof.* First, we show that  $\mathcal{G} \subseteq \mathcal{H}$ . It’s enough to show that the events in  $\mathcal{G}$  are in  $\mathcal{H}$ ; consider events of the form (which generate  $\mathcal{G}$ )

$$\{B_{t_1} \in [a_1, b_1], \dots, B_{t_n} \in [a_n, b_n]\}.$$

These are closed with respect to the  $d$ -topology (a sequence of functions in this set means its limit point will satisfy all of the  $B_{t_i} \in [a_i, b_i]$ ), and the Borel sigma-algebra contains all of the closed sets, so they are all in  $\mathcal{H}$  – this means that  $\mathcal{G} \subseteq \mathcal{H}$ .

To show the other direction, it’s enough to show that an open ball in  $\mathcal{H}$  is contained in  $\mathcal{G}$ . Consider the set

$$\{h : d(f, h) < \varepsilon\} :$$

this is contained in  $\mathcal{G}$  because we can measure  $d(f, g)$  by looking at only rational  $t$  (the definition of  $d$  takes a supremum over the whole interval  $[0, n]$ , but we can restrict to a countable set of indices by continuity). Thus, that set is also measurable in  $\mathcal{G}$ , and thus  $\mathcal{H} \subseteq \mathcal{G}$ .  $\square$

So far, we’ve been just treating Brownian motion just as a function in  $t$ , but now we want to actually make the process move forwards or backwards in time.

**Definition 25**

Define the sigma-algebras

$$\mathcal{F}_t = \sigma(B_s : s \leq t), \quad \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s.$$

$\mathcal{F}_{t+}$  gives an “infinitesimal amount of information past time  $t$ .” For example, if the process  $B$  had a right derivative

$$\lim_{h \downarrow 0} \frac{B_{t+h} - B_t}{h},$$

this would be measurable with respect to  $\mathcal{F}_{t+}$  but not  $\mathcal{F}_t$ . (And we’ll see soon that Brownian motion does not have a right derivative.)

Recall that we’ve characterized Brownian motion in a way such that it has independent increments for disjoint time intervals: this follows from the definition of the covariance function.

**Proposition 26** (Markov property, simplest version)

Suppose that  $(B_t)_{t \geq 0}$  is a Brownian motion. Then for any fixed time  $s \geq 0$ , the process  $(B_{s+t} - B_s)_{t \geq 0}$  is a Brownian motion independent of  $\mathcal{F}_s$ .

*Proof.* Let  $W_t = B_{s+t} - B_s$  be our new process. We must show that  $W_t$  is a pre-Brownian motion and that it has sample paths – the latter is true because the sample paths are just subsets of the sample paths for  $B_t$ , and the former follows because it has the correct covariance function. Thus, we just need to show independence – if Brownian motion has independent increments, then we can make the independence statement

$$(W_{t_1}, W_{t_2}, \dots, W_{t_k}) \perp\!\!\!\perp (B_{r_1}, \dots, B_{r_l})$$

for all times  $r_i \leq s$ . This tells us that  $W \perp\!\!\!\perp \mathcal{F}_s$ , as desired (by a pi-lambda argument).  $\square$

**Proposition 27** (Markov property, slight improvement)

Using the same setting as Proposition 26, we have  $W \perp\!\!\!\perp \mathcal{F}_{s+}$ .

*Proof.* Again, consider the collection of times  $(W_{t_1}, W_{t_2}, \dots, W_{t_k})$ . This is independent of  $\mathcal{F}_{s+}$  unless some of the times are 0, but because Brownian motion is continuous, we can write this as the limit

$$\lim_{\varepsilon \downarrow 0} (W_{t_1+\varepsilon}, W_{t_2+\varepsilon}, \dots, W_{t_k+\varepsilon}).$$

For any fixed  $\varepsilon > 0$ , this is independent of  $\mathcal{F}_{s+}$ , and it’s true in the limit because the limit of independent variables cannot become dependent (the limit is measurable with respect to those variables).  $\square$

This gives us the following consequence:

**Theorem 28** (Blumenthal 0-1 law)

If  $A \in \mathcal{F}_{0+}$ , then  $\mathbb{P}(A) \in \{0, 1\}$ .

A basically only sees the Brownian motion at an infinitesimal time at the beginning, and the idea is that we can’t produce something nontrivial that depends on this infinitesimal time.



*Proof.* Proposition 27 with  $s = 0$  tells us that  $\sigma(B_s : s \geq 0)$  is independent of  $\mathcal{F}_{0+}$ . But  $\mathcal{F}_{0+}$  sits inside  $\sigma(B_s : s \geq 0)$ , so  $\mathcal{F}_{0+}$  is independent of itself: thus  $\mathbb{P}(A)^2 = \mathbb{P}(A)$  and therefore  $\mathbb{P}(A) \in \{0, 1\}$  for all  $A \in \mathcal{F}_{0+}$ .  $\square$

This gives us a bit more to work with:

**Proposition 29**

Let  $B$  be a Brownian motion with  $B_0 = 0$ .

1.  $B$  will cross 0 an infinite number of times almost surely: for all  $\varepsilon > 0$ , we have (almost surely) that

$$\sup(B_s : s \in [0, \varepsilon]) > 0, \quad \inf(B_s : s \in [0, \varepsilon]) < 0.$$

2. For  $a \in \mathbb{R}$ , the hitting time

$$T_a = \inf\{t : B_t = a\}$$

is finite almost surely.

This tells us, for example, that it doesn't make sense to define "the first time  $B$  returns to 0."

*Proof.* For (1), let

$$A_\varepsilon = \{\sup(B_s : s \in [0, \varepsilon]) > 0\},$$

and define  $A = \bigcap_{\varepsilon > 0} A_\varepsilon$ . We can restrict to  $s$  rational, so  $A_\varepsilon$  is a measurable set, and similarly  $A = \bigcap_n A_{1/n}$  is also measurable.  $A$  is in  $\mathcal{F}_{0+}$ , because the events  $A_{1/n} \in \mathcal{F}_{1/n}$  are decreasing and nested. So by the zero-one law,  $\mathbb{P}(A)$  is either 0 or 1, and we want to show that it is 1: note also that  $\mathbb{P}(A_\varepsilon)$  is at least  $\frac{1}{2}$ , because Brownian motion is symmetric around 0. Since these events are decreasing, this means that

$$\mathbb{P}(A_\varepsilon) \geq \frac{1}{2} \implies \mathbb{P}(A) \geq \frac{1}{2} \implies \mathbb{P}(A) = 1.$$

(The other statement follows analogously.)

(2) is actually a consequence of (1): by part (a), we know that

$$1 = \mathbb{P}(\sup(B_s : s \in [0, 1]) > 0)$$

by taking  $\varepsilon = 1$ , but this is also equal to

$$\lim_{\delta \downarrow 0} \mathbb{P}(\sup(B_s : s \in [0, 1]) \geq \delta)$$

by continuity. Now we can rescale space by  $\frac{1}{\delta}$  and rescale time by  $\frac{1}{\delta^2}$ ; the result is still a Brownian motion, so this is

$$= \lim_{\delta \downarrow 0} \mathbb{P}\left(\sup_s \left(B_s : s \in \left[0, \frac{1}{\delta^2}\right]\right) \geq 1\right) = \mathbb{P}\left(\sup_s B_s \geq 1\right)$$

again by continuity (because  $\frac{1}{\delta^2}$  goes to infinity). This means that  $B_s$  will hit height 1 at some point with probability 1, and we can scale again to show that  $B$  hits any height  $a$  almost surely (for instance, we can multiply space by 10 and multiply time by 100).  $\square$

In particular, if we run a Brownian motion for all time, it will not converge to anything – it must hit height  $a$ , then height  $-a$ , and so on. That means it oscillates a lot, and thus Brownian motion is not particularly well-behaved.

So we can ask the question "how regular is Brownian motion?" We'll choose a few interesting results for this class that are interesting but skip the more specialized ones. This first one is important to stochastic calculus: we know

that Brownian motion is  $(\frac{1}{2} - \varepsilon)$ -Hölder, and on homework 2 we'll see that

$$\limsup_{h \downarrow 0} \frac{|B_{t+h} - B_t|}{\sqrt{2h \log(1/h)}} = 1 \text{ a.s..}$$

This tells us that the  $\frac{1}{2}$  exponent is basically correct, and this makes sense – we can do a calculation

$$|B_{t+h} - B_t| \sim \sqrt{h}N(0, 1),$$

so we should expect this to be on the order of  $\sqrt{h}$  (and the above statement shows that sometimes it is bigger).

**Definition 30**

A function  $f : [a, b] \rightarrow \mathbb{R}$  is of **bounded variation** (BV) if (taking a supremum over all sets of break points)

$$\sup_{P=\{t_i\}} \sum_{i=1}^n |f(t_i) - f(t_{i-1})| < \infty.$$

We'll denote  $\sum_{i=1}^n |f(t_i) - f(t_{i-1})|$  as  $V_p^1(f)$ . Sufficiently nice functions work – if  $f$  is  $C^1$ , then it is definitely of bounded variation, because

$$V_p^1(f) \leq \int_a^b |f'(t)| dt.$$

Meanwhile, Brownian motion is not of bounded variation – we can read this in the book or try this ourselves. (If we partition our interval  $[0, 1]$  into blocks of size  $\varepsilon$ , we have  $\frac{1}{\varepsilon}$  intervals and each increment is on the order of  $\sqrt{\varepsilon}$ , so this gives us something very large.) So is there a more suitable measure of variation we can use?

Well, consider an  $\alpha$ -Hölder function  $f$ , meaning that  $|f(s) - f(t)| \leq K|s - t|^\alpha$ . Then one way we could measure this is to consider

$$V_p^{1/\alpha}(f) = \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^{1/\alpha} \leq \sum_{i=1}^n K^{1/\alpha} |t_i - t_{i-1}| \leq K^{1/\alpha}(b - a) < \infty$$

(by the  $\alpha$ -Hölder bound). This means that Brownian motion satisfies (remembering that this is defined on a compact interval)

$$V_p^{1/\alpha}(B) < \infty \quad \forall \alpha < \frac{1}{2},$$

which means we raise the differences  $|f(t_i) - f(t_{i-1})|$  to a power higher than 2. But **this is not a good measure of variation either**: it tends to 0 as our partition gets finer. Specifically, take  $\alpha \in (0, \frac{1}{2})$ . The Brownian motion is  $\gamma$ -Hölder for  $\gamma \in (\alpha, \frac{1}{2})$ , so taking  $P$  to be the collection of points separated by  $\varepsilon$ ,

$$V_p^{1/\alpha}(B) \asymp \frac{1}{\varepsilon} (\varepsilon^\gamma)^{1/\alpha} \asymp \varepsilon^{\gamma/\alpha - 1},$$

which tends to 0. So raising the increments to any power larger than 2 doesn't work – perhaps we can try

$$V_p^2(B) = \sum_{i=1}^n (B_{t_i} - B_{t_{i-1}})^2.$$

Does this tend to a non-trivial limit as the mesh of  $P$  goes to 0? It turns out the answer is yes – this tends to the width of the interval,  $(b - a)$ .

**Proposition 31**

Suppose that  $P_n$  is a subdivision of the interval  $[0, t]$  for all  $n$ . Then

$$V_{P_n}^2(B) \rightarrow t$$

converges in  $L^2$  as the mesh of  $P_n$  goes to 0.

*Proof.* We just do a simple variance calculation: first, break up the  $t$  term into the lengths of our intervals

$$\|V_{P_n}^2(B) - t\|_2^2 = \mathbb{E} \left[ \left\{ \sum_{i=1}^n (B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1}) \right\}^2 \right].$$

Each term has mean 0, because the expected value of each increment squared is the length of the interval. In addition, the increments are independent, so if we expand the square and take expectations, the cross-terms go away and we're left with

$$\sum_{i=1}^n (t_i - t_{i-1})^2 \mathbb{E}[(Z^2 - 1)^2],$$

where  $Z$  is the standard normal. The  $\mathbb{E}$  term is just some constant, and the mesh goes to 0, so this can be bounded by

$$\leq \mathbb{E}[(Z^2 - 1)^2] \cdot \text{mesh}(P_n) \sum (t_i - t_{i-1}) = \mathbb{E}[(Z^2 - 1)^2] \cdot \text{mesh}(P_n)t,$$

which goes to 0 by assumption. □

This result is important, because the limit  $V_{P_n}^2(B)$  has to do with sum of squares of Brownian motion on small intervals, which is a good way of understanding something like

$$\int_0^t (dB_s)^2.$$

(Also, this converges in  $L^2$  but not almost surely.)

Consider the natural filtration  $\mathcal{F}_t = \sigma(B_s : s \leq t)$  as before, and define  $\mathcal{F}_\infty = \sigma(B_s : s \geq 0)$  to be the sigma-algebra of everything we know about the Brownian motion. Let  $T$  be a **stopping time** (meaning that  $\{T \leq t\} \in \mathcal{F}_t$  for all  $t$ ), and define the stopping-time sigma-algebra

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

(This captures everything we know about the process up until our random stopping time.)

**Proposition 32** (Markov property, strong version)

Let  $T$  be a stopping time. Then  $(B_{T+t} - B_T)_t$  is a Brownian motion, and it is independent of  $\mathcal{F}_T$  under the measure  $\mathbb{P}$  **where we condition on  $T$  being finite**.

We can read the proof in the book. One of the most important applications for this is the **reflection principle**: if we want to know something like

$$\mathbb{P}(B_t > a \text{ for any } t \in [0, 1]),$$

we can consider a height  $b < a$  and ask about the probability that we do exceed  $a$ , but end below  $b$  at time 1. To answer that question, we can **reflect** the Brownian motion after the stopping time when we hit  $a$ : the answer is then

equal to the probability that we end above  $2a - b$  by  $t = 1$ . In other words, if  $S_t = \sup(B_s : s \leq t)$ ,

$$\mathbb{P}(S_1 \geq a, B_1 < b) = \mathbb{P}(S_1 \geq a, B_1 \geq 2a - b) = \mathbb{P}(B_1 \geq 2a - b).$$

$B_1$  is just a standard normal, so this tells us everything we might want to know about the supremum process! We'll do this in more detail next time.

## 5 February 18, 2020

At the end of last time, we discussed the reflection principle briefly – we'll go over this in a bit more detail now. Recall the **strong Markov property** for Brownian motion, which tells us that for any stopping time  $T$ ,

$$(B_{T+t} - B_T)_{t \geq 0}$$

is a Brownian motion independent of  $\mathcal{F}_T$  (the stopping-time sigma-algebra) if we condition on  $T$  being finite. This is useful, for example, if we consider the supremum process

$$S_t = \sup\{B_s : 0 \leq s \leq t\}$$

which is a nondecreasing process.

### Theorem 33

For any  $a > 0$  and  $b \in (-\infty, a]$ , we have

$$\mathbb{P}(S_t \geq a, B_t \leq b) = \mathbb{P}(B_t \geq 2a - b).$$

The point is that this gives us a joint distribution between  $S_t$  and  $B_t$  in terms of something that is completely explicit:  $B_t$  is just a normal random variable with mean 0 and variance  $t$ .

*Proof.* Apply the strong Markov property to the stopping time  $\tau = \inf\{s \geq 0 : B_s \geq a\}$  (the first time we hit  $a$ ). Then

$$W_s = (B_{\tau+s} - B_\tau)_{s \geq 0}$$

is a Brownian motion independent of the process up to time  $\tau$ . Now we can explicitly write down

$$\mathbb{P}(S_t \geq a, B_t \leq b) = \mathbb{P}(\tau \leq t, W_{t-\tau} \leq -(a - b))$$

(because we must hit  $a$  at some point and then go down by  $(a - b)$  to get back below  $b$ ). Now we use the reflection principle:  $W$  is symmetric in law, so this is equal to

$$= \mathbb{P}(\tau \leq t, W_{t-\tau} \geq a - b) = \mathbb{P}(S_t \geq a, B_t \geq 2a - b).$$

Now  $B_t \geq 2a - b$  means that  $S_t \geq B_t \geq a$ , so we can drop the  $S_t \geq a$  assumption and we're done.  $\square$

This, in particular, tells us that the joint density of  $S_t$  and  $B_t$  is given by differentiating the given quantity:

$$= -\frac{\partial^2}{\partial a \partial b} \mathbb{P}(B_t \geq 2a - b).$$

(negative sign comes because we have  $S_t \geq a$  but  $B_t \leq b$ ), and this is supported on  $a \geq 0, b \leq a$ . And now we can take this density and integrate out the  $b$  to find the marginal law of  $S_t$ , but there's a faster way: we can decompose

via

$$\mathbb{P}(S_t \geq a) = \mathbb{P}(S_t \geq a, B_t \geq a) + \mathbb{P}(S_t \geq a, B_t \leq a) = \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq 2a - a) = \mathbb{P}(|B_t| \geq a),$$

because the law of  $B_t$  is symmetric. So this means that for a fixed  $t$ ,  $S_t$  is equally distributed as  $|B_t|$ ! But we need to be careful: the processes

$$(S_t)_{t \geq 0}, \quad (|B_t|)_{t \geq 0}$$

are not equally distributed, because  $S_t$  is increasing but  $B_t$  returns to 0 infinitely many times.

**Remark 34.** *On our homework, we'll find a bit more: it turns out that  $S_t$  and  $|B_t|$  are both equally distributed as  $S_t - B_t$  for all fixed  $t$ , and  $(S_t - B_t)$  is also equally distributed as  $|B_t|$  as a process.*

Note that this calculation also gives us the law of the first time we hit  $a$ : define  $\sigma = \sigma_a = \inf\{t \geq 0 : B_t = a\}$ . Then

$$\sigma_a \stackrel{d}{=} \inf\{t \geq 0 : B_{t/a^2} = 1\} = a^2 \sigma_1,$$

which allows us to directly find

$$\mathbb{P}(\sigma_a \geq t) = \mathbb{P}(S_t \leq a) = \mathbb{P}(|B_t| \leq a) = \mathbb{P}\left(|Z| \leq \frac{a}{\sqrt{t}}\right) = \mathbb{P}\left(\frac{a^2}{Z^2} \geq t\right).$$

where  $Z$  is a standard normal. Comparing the left and right expressions tells us that  $\sigma_a$  is equally distributed as  $\frac{a^2}{Z^2}$  – one thing we can immediately see is that  $\mathbb{E}\sigma_a$  is infinite for all  $a > 0$ .

Taking  $a = 1$ , this means that  $\sigma_1$  is distributed as  $\frac{1}{Z^2}$ . We can then calculate the density of  $\sigma$  (which is an exercise for us), which will tell us that (for large  $t$ )

$$\mathbb{P}(\sigma > t) \asymp \frac{1}{\sqrt{t}},$$

which decays quite slowly. We can compare this with the stopping time  $\tau = \inf\{t : |B_t| \geq 1\}$  (so the first time we hit either 1 or  $-1$ ). It turns out that for large  $t$ , we have very different behavior:

$$\mathbb{P}(\tau > t) \leq \exp(-\Omega(t)).$$

It's easy to see why this occurs: if we run Brownian motion for a very long time and we want it to stay confined in the interval  $[-1, 1]$ , think about time intervals of length 1. There is some positive chance it leaves in each time interval, and conditioned on staying inside  $[-1, 1]$ , we have some positive chance in the next interval. Again, we'll be more precise on the homework.

This is all we'll cover from chapter 2 of Le Gall – now we'll move on to chapter 3, which discusses continuous-time martingales. Unfortunately, it's pretty boring: under mild assumptions, all the properties from discrete-time martingales hold. So we'll go through this fairly quickly – if we took 18.675 in a different semester where we didn't cover martingales in such detail, we might have to do reading on our own.

Brownian motion  $B_t$  is an example of a continuous-time martingale, and here's another way we can construct one: let  $\zeta, \zeta_i$  be iid exponential random variables, and let

$$N_t = \max\{n : \sum_{i=1}^n \zeta_i \leq t\}.$$

Then  $N_t$  is distributed as  $\text{Pois}(t)$ , and it is an integer-valued process with right-continuous sample paths. It will turn out  $N_t - t$  is a continuous-time martingale as well, so our formalism should be able to study it. (Generally, we'll be assuming that everything we're dealing with is right-continuous.)

Throughout this chapter, everything will live on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with a **filtration**

$$(\mathcal{F}_t)_{0 \leq t \leq \infty} : \mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall s \leq t.$$

We'll call  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  a **filtered probability space**.

**Definition 35**

A process  $(X_t)_{t \geq 0}$  is **adapted** to a filtration  $\mathcal{F}_t$  if  $X_t \in \mathcal{F}_t$  for all  $t$ .

We'll also reiterate the following definition:

**Definition 36**

A random variable  $\tau : \Omega \rightarrow [0, \infty]$  is a **stopping time** if  $\{\tau \leq t\} \in \mathcal{F}_t$  for all  $t$  (stopping doesn't depend on the future). The  **$\sigma$ -field of the past up to  $\tau$**  is

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t \forall t\}.$$

So an event that only depends on time up to  $\tau$  can be rephrased as "only needing information up to  $t$  if  $\tau \leq t$ ." We should read all of the basic facts about filtrations and stopping times on our own: this is sections 3.1 and 3.2 of our book. An example of what we'll see: if  $\sigma, \tau$  are both stopping times, then their minimum  $\sigma \wedge \tau$  and maximum  $\sigma \vee \tau$  are also stopping times, and

$$\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau.$$

(And this specific fact is useful for one of the questions on our homework.)

**Definition 37**

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_+, \mathbb{P})$  be a filtered probability space, and let  $(X_t)_{t \geq 0}$  be a real-valued process. Then  $(X_t)$  is a **submartingale** if

- $X_t \in \mathcal{F}_t$  is **adapted** to the filtration.
- $\mathbb{E}|X_t|$  is finite for all  $t$  (the variable is **integrable**).
- For every  $0 \leq s \leq t$ ,

$$X_s \leq \mathbb{E}[X_t | \mathcal{F}_s].$$

$(X_t)$  is a **supermartingale** if we flip the inequality.

A martingale is both a submartingale and a supermartingale: note that  $\mathbb{E}X_t$  is constant for a martingale, nondecreasing for a submartingale, and nonincreasing for a supermartingale.

**Example 38**

Let  $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  be any integrable random variable. Then we can define

$$X_t = \mathbb{E}(Z | \mathcal{F}_t).$$

This is a martingale by basic properties of the conditional expectation, and it's bounded in  $L^1$ :

$$\mathbb{E}|X_t| = \mathbb{E}|\mathbb{E}(Z | \mathcal{F}_t)| \leq \mathbb{E}(\mathbb{E}|Z| | \mathcal{F}_t) = \mathbb{E}|Z|$$

by using Jensen's inequality.

**Example 39**

The standard Brownian motion  $B_t$  is a martingale (it satisfies all properties), but there is no random variable  $Z$  such that  $B_t = \mathbb{E}(Z|\mathcal{F}_t)$ , because  $\mathbb{E}|B_t| = t^{1/2}\mathbb{E}|N(0, 1)|$  is unbounded.

There are some other important martingales based on Brownian motion as well: we can check that  $B_t^2 - t$  is a martingale, as is  $\exp\left(\theta B_t - \frac{\theta^2 t}{2}\right)$ .

We'll use the rest of this lecture to prove some basic results about martingales.

**Proposition 40**

Let  $X_t$  be a (sub/super)martingale. Then

$$\sup \{\mathbb{E}|X_s| : 0 \leq s \leq t\} < \infty$$

for all  $t < \infty$ .

(We didn't have to prove this in the discrete case, because we only had a finite number of variables to consider between 0 and  $t$ , and we know that  $\mathbb{E}|X_s|$  is finite at any given time  $s$ .)

*Proof.* Without loss of generality, say that  $X$  is a submartingale. Then define  $(X_t)_+ = \max\{X_t, 0\}$ , which is a submartingale (because  $f(x) = \max(x, 0)$  is a convex nondecreasing function). Then  $\mathbb{E}[(X_t)_+]$  is nondecreasing, so for all  $s \leq t$ ,

$$\mathbb{E}|X_s| = \mathbb{E}(2(X_s)_+ - X_s) \leq \mathbb{E}(2(X_t)_+ - X_0).$$

This bound holds uniformly over  $s$ , so the supremum of  $\mathbb{E}|X_s|$  must indeed be finite. □

This next fact is a weak version of the optional stopping theorem for discrete-time submartingales, and we saw this in 18.675:

**Lemma 41**

Let  $X_n$  be a discrete-time submartingale, and let  $\tau$  be a bounded stopping time such that  $\tau \leq n$  almost surely. Then

$$\mathbb{E}X_0 \leq \mathbb{E}X_\tau \leq \mathbb{E}X_n.$$

*Proof.* Consider the stopped process

$$Y_k = X_{k \wedge \tau};$$

this is also a submartingale, and

$$\mathbb{E}X_0 = \mathbb{E}Y_0 \leq \mathbb{E}Y_n = \mathbb{E}X_{n \wedge \tau} = \mathbb{E}X_\tau$$

because  $\tau \leq n$  almost surely. And for the other direction,

$$\mathbb{E}X_\tau = \sum_{k=0}^n \mathbb{E}(1\{\tau = k\}X_k) \leq \sum_{k=0}^n \mathbb{E}(1\{\tau = k\}\mathbb{E}(X_n|\mathcal{F}_k))$$

by the submartingale condition, and now we can put  $1\{\tau = k\}$  inside the conditional expectation because it is

measurable with respect to  $\mathcal{F}_k$ . This gives us

$$= \sum_{k=0}^n \mathbb{E}[\mathbb{E}(\mathbb{1}\{\tau = k\}X_n | \mathcal{F}_k)] = \sum_{k=0}^n \mathbb{E}(\mathbb{1}\{\tau = k\}X_n) = \mathbb{E}X_n,$$

as desired. □

**Proposition 42** (Maximal inequality, discrete version)

Let  $Y_n$  be a discrete time (sub/super)martingale. Then for all  $\lambda \geq 0$ ,

$$\lambda \mathbb{P} \left( \max_{0 \leq k \leq n} |Y_k| \geq \lambda \right) \leq \mathbb{E}[|Y_0| + 2|Y_n|]$$

This tells us that we have control of the trajectory up to time  $n$  just by knowing something about the beginning and end time.

*Proof.* Let  $A$  be the event  $\max_{0 \leq k \leq n} |Y_k| \geq \lambda$ . (The notation  $\mathbb{E}[X; A]$  means  $\mathbb{E}[X \cdot \mathbb{1}\{X \in A\}]$ .)

Without loss of generality, assume  $Y_n$  is a supermartingale. Consider the stopping time

$$\tau = \min \{k : |Y_k| \geq \lambda \text{ or } k = n\} \in [0, n].$$

Now

$$\lambda \mathbb{P} \left( \max_{0 \leq k \leq n} |Y_k| \geq \lambda \right) \leq \mathbb{E}(|Y_\tau|; A)$$

because on the event  $A$ ,  $|Y_\tau|$  is at least  $\lambda$ . Now we can decompose this via

$$\mathbb{E}(|Y_\tau|; A) \leq \mathbb{E}(|Y_\tau|) = \mathbb{E}[Y_\tau + 2(Y_\tau)_-];$$

since  $Y_\tau$  is a supermartingale and  $(Y_\tau)_-$  is a submartingale, we can bound this via

$$\leq \mathbb{E}[Y_0 + 2(Y_n)_-],$$

which is at most the right-hand side of the result. □

One note about Proposition 42: we can get a small improvement if we assume that  $X$  is a martingale. Then

$$\lambda \mathbb{P}(\max_{k \leq n} |X_k| \geq \lambda) \leq \mathbb{E}(|X_\tau|; A) = \mathbb{E}((X_\tau)_+ + (X_\tau)_-; A)$$

and because  $X$  is a martingale,  $(X_\tau)_+$  and  $(X_\tau)_-$  are both submartingales, which gives us

$$\leq \mathbb{E}(|X_n|; A).$$

We'll generalize this to continuous-time martingales:

**Proposition 43** (Maximal inequality, continuous version)

Let  $X_t$  be a (sub/super) martingale with **right-continuous sample paths** (this is a regularity condition). Then

$$\lambda \mathbb{P} \left( \sup_{s \leq t} |X_s| \geq \lambda \right) \leq \mathbb{E}[|X_0| + 2|X_t|]$$

where  $A$  similarly refers to the event  $\sup_{s \leq t} |X_s| \geq \lambda$ .



*Proof.* Fix  $t$ . Then any sequence  $0 = t_0 < t_1 < t_2 < \dots < t_m = t$  gives a discrete-time (sub/super)martingale  $(X_{t_k})_k$ , so by Proposition 42, we have

$$\lambda \mathbb{P} \left( \max_{0 \leq k \leq m} |X_{t_k}| \geq \lambda \right) \leq \mathbb{E}(|X_0| + 2|X_t|).$$

Now take our sequences  $D_m \uparrow D$ , where  $D_1 = \{t_0 = 0, t_1 = t\}$ , such that the  $D_m$ s are nested and increase to a countable dense subset in  $[0, t]$ . This gives us

$$\lambda \mathbb{P} \left( \sup_{s \in [0, t] \cap D} |X_s| \geq \lambda \right) \leq \mathbb{E}(|X_0| + 2|X_t|),$$

and right-continuity allows us to replace  $[0, t] \cap D$  with  $[0, t]$ , and we have the desired result.  $\square$

We'll prove a few more inequalities which are weaker but easier to package and remember:

**Proposition 44** (Doob's  $L^p$  inequality, discrete version)

Let  $X_n$  be a discrete-time martingale. Then for all  $p > 1$  and finite  $n$ ,

$$\left\| \max_{0 \leq k \leq n} |X_k| \right\|_p \leq C_p \|X_n\|_p,$$

where  $C_p = \frac{p}{p-1}$ .

**Proposition 45** (Doob's  $L^p$  inequality, continuous version)

Let  $X_t$  be a martingale with right-continuous sample paths. Then for all  $p > 1$  and finite  $t$ ,

$$\left\| \sup_{0 \leq s \leq t} |X_s| \right\|_p \leq C_p \|X_t\|_p,$$

where  $C_p = \frac{p}{p-1}$ .

Here, the discrete version again implies the continuous version by the same argument – we'll just prove the discrete version here.

*Proof.* We use the above bound

$$\lambda \mathbb{P} \left( \max_{k \leq n} |X_k| \geq \lambda \right) \leq \mathbb{E}(|X_n|; A).$$

Let  $S_n = \max_{0 \leq k \leq n} |X_k|$ ; we wish to bound its  $L^p$  norm. By our usual formula,

$$\mathbb{E}((S_n)^p) = \int_0^\infty \mathbb{P}(S_n^p \geq t) dt$$

and we make a change of variables to see that this is also

$$= \int_0^\infty p y^{p-1} \mathbb{P}(S_n \geq y) dy.$$

And we know how to bound this:

$$\leq \int_0^\infty p y^{p-1} \left[ \frac{\mathbb{E}(|X_n|; S_n \geq y)}{y} \right] dy.$$

By Fubini's theorem, we can change the order of integration

$$= \mathbb{E} \left[ |X_n| \int_0^{S_n} p y^{p-2} dy \right],$$

where the indicator that  $S_n \geq y$  is equivalent to replacing the  $\infty$  with  $S_n$ . And this is just

$$= \frac{p}{p-1} \mathbb{E} [|X_n| S_n^{p-1}],$$

and by Hölder's inequality, this is

$$\leq \frac{p}{p-1} \|X_n\|_p \|S_n\|_{\frac{p}{p-1}}^{p-1} = \frac{p}{p-1} \|X_n\|_p \|S_n\|_p^{p-1}.$$

We can now divide through by  $\|S_n\|_p^{p-1}$  to get the result if the norm is finite, and otherwise we use a standard truncation argument.  $\square$

Again, it's important to note that there's no  $L^p$  inequality for  $p = 1$ : consider the simple random walk  $X_n$  on the integers starting from  $X_0 = 1$ , and let

$$\tau = \min\{n : X_n = 0\}, \quad M_n = X_{n \wedge \tau}.$$

Basically, this stops when it hits 0, which means we have a nonnegative martingale with  $\mathbb{E}M_n = 1$ . But we can't control the maximum of this process: if  $S_n = \max_{k \leq n} M_k$ , then  $\|S_n\|$  turns out to be unbounded as  $n \rightarrow \infty$ . To see this, it's enough to show that  $\|S_\infty\|$  has infinite expectation – the probability that  $S_\infty$  is at least  $a$  is the chance that a random walk hits  $a$  before 0, which is  $\frac{1}{a}$  (by optional stopping theorem for discrete-time martingales, for example). This is not summable over  $a$ , so the expectation is indeed infinite.

## 6 February 19, 2020

Yesterday, we defined what it means for  $(X_t)_{t \geq 0}$  to be a continuous-time (sub/super)martingale on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ . Last time, we showed the maximal inequality

$$\lambda \mathbb{P} \left( \sup_{0 \leq s \leq t} |X_s| \geq \lambda \right) \leq \mathbb{E} [|X_0| + 2|X_t|]$$

assuming that the process  $X$  has right-continuous sample paths. (There's a slightly stronger result for martingales, which is that  $\lambda \mathbb{P}(A) \leq \mathbb{E}[|X_t|; A]$  for an event of the type  $A = \sup_{0 \leq s \leq t} |X_s| \geq \lambda$ .) We also used this to show the  $L^p$  inequality by integrating the previous result:

$$\left\| \sup_{0 \leq s \leq t} X_s \right\|_p \leq \frac{p}{p-1} \|X_t\|_p$$

for all  $p > 1$  and  $t \in [0, \infty)$ . This requires right-continuous sample paths, and if we don't assume that fact, we actually prove that

$$\lambda \mathbb{P} \left( \sup_{s \in [0, t] \cap D} |X_s| \geq \lambda \right) \leq \mathbb{E} (|X_0| + 2|X_t|)$$

for a countable dense set  $D$ . We'll build off of this today: we'll show that under mild conditions, a (sub/super)martingale has a right-continuous modification, and it will actually do a bit better:

### Definition 46

A function  $f$  is **right continuous with left limits** (**rcll** or **càdlàg** for short) if for all  $t \geq 0$ ,  $f(t) = \lim_{s \downarrow t} f(s)$ , and for all  $t > 0$ ,  $\lim_{s \uparrow t} f(s)$  exists.

The main idea is that martingales can't oscillate too much, so we can guarantee existence of limits. We'll start with a deterministic result by controlling **upcrossing numbers**: for any subset  $I \subseteq [0, \infty)$  and any  $a < b$ , denote  $U_{a,b}^f(I)$  to be the maximum  $k$  such that there exist  $s_1 < t_1 < s_2 < t_2 < \dots < s_k < t_k$ , and  $f(s_i) \leq a$  and  $f(t_i) \geq b$  for all  $i$ . (This is the maximum number of times that we go from  $a$  to  $b$ .) Basically, if we can control upcrossing numbers, we have some regularity condition:

**Lemma 47**

Let  $D$  be a countable dense subset of  $[0, \infty)$ , and consider a function  $f : D \rightarrow \mathbb{R}$ . Say that the function is locally bounded:  $\sup\{|f(t)| : t \in [0, T] \cap D\} < \infty$  for all  $T \in D$ . Also, suppose that  $U_{a,b}^f(D \cap [0, T]) < \infty$  for all  $T \in D$  and for all rational  $a < b$  (this is to avoid issues about measurability). Then  $f$  has all of its left and right limits, and the function

$$g(t) = f(t+) = \lim_{s \downarrow t, s \in D} f(s)$$

is rcll.

*Proof.* Take any  $t \geq 0$ : suppose for the sake of contradiction that the (WLOG) right limit  $\lim_{s \downarrow t, s \in D} f(s)$  does not exist. This means that the lim sup and lim inf are different, and thus there exist rational  $a, b$  with

$$\liminf_{s \downarrow t, s \in D} f(s) < a < b < \limsup_{s \downarrow t, s \in D} f(s).$$

But this means we must cross between  $a$  and  $b$  infinitely many times, which is a contradiction with the assumption that  $U_{a,b}^f([0, T] \cap D)$  is finite. (And showing that  $g$  is rcll is a similar argument about the upcrossings.)  $\square$

This basically tells us that we need to control upcrossing numbers, and this is based on the following idea:

**Definition 48**

Let  $X_n$  be an adapted (discrete) process, so  $X_n \in \mathcal{F}_n$  for all  $n$ . In addition, say that  $H_n$  is a **previsible** process, so  $H_n \in \mathcal{F}_{n-1}$  for all  $n$ . Then the **Doob transform** is

$$(H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1}).$$

Notably, if  $X$  is a supermartingale and  $H$  is a **nonnegative bounded** previsible process, then  $H \cdot X$  is also a supermartingale. We can check this from the definition, but what it's really saying is that in a gambling system  $X$  where we can't win, with a betting strategy  $H$  (to tell us how much to bet in the next game), we can't game the expected gain  $H \cdot X$ .

**Lemma 49 (Doob's upcrossing inequality)**

Let  $X_n$  be a discrete supermartingale. Then the expected number of upcrossings

$$\mathbb{E}U_{a,b}^X([0, n]) \leq \frac{\mathbb{E}(X_n - a)_-}{b - a}.$$

*Proof.* Let  $Y$  be the value we get of betting on  $X$  only during upcrossings. More formally, let

$$Y = H \cdot X, \text{ where } H_j = 1\{j \in (\sigma_i, \tau_i]\} \text{ for some } i,$$

where  $(\sigma_i, \tau_i]$  are stopping times corresponding to being below  $a$  and above  $b$ .  $H_j$  is in  $\mathcal{F}_{j-1}$ , because

$$\{j \in (\sigma_i, \tau_i]\} = \{\sigma_i \leq j-1\} \cap \{\tau_i \leq j-1\}^c,$$

and both of these events are  $\mathcal{F}_{j-1}$ -measurable by the definition of a stopping time.

Because  $Y = H \cdot X$  is a supermartingale with  $Y_0 = 0$ , we must have  $\mathbb{E}Y_n \leq 0$ . On the other hand, we know that  $Y_n$  gets a contribution from the number of completed up-crossings, and then there's an extra term from the end where we start an up-crossing but we go very far down at the end: the worst case is just that we lose  $(X_n - a)_-$ , so

$$Y_n \geq (b-a)U_{a,b}^X([0, n]) - (X_n - a)_-$$

Taking expectations of both sides, we have

$$0 \geq \mathbb{E}Y_n \geq (b-a)\mathbb{E}[U_{a,b}^X([0, n])] - \mathbb{E}[(X_n - a)_-]$$

and rearranging gives the bound. □

### Corollary 50

Let  $X_t$  be a supermartingale, and let  $D$  be a countable dense subset of  $[0, \infty)$ . Then there exists an event  $N$  with probability zero so that for all  $\omega \notin N$ , the function  $t \mapsto X_t(\omega)$  satisfies Lemma 47.

*Proof.* The first property (locally bounded) follows by the maximal inequality

$$\lambda \mathbb{P} \left( \sup_{s \in [0, t] \cap D} |X_s| \geq \lambda \right) \leq \mathbb{E}(|X_0| + 2|X_t|) < \infty$$

and taking  $\lambda \rightarrow \infty$ . The second property (number of upcrossings is bounded) follows from Lemma 49 plus an approximation argument. Basically, for a continuous-time martingale, we know that

$$\mathbb{E} [U_{a,b}^X([0, t] \cap D)] \leq \frac{\mathbb{E}[(X_t - a)_-]}{b-a}$$

if  $X_t$  is a supermartingale. But if  $X_t$  is a continuous-time supermartingale, we can take any finite set and apply Lemma 49, and take a larger and larger set increasing to  $D$ : this gives a larger and larger number of upcrossings, but we're uniformly bounded by the right hand side, so we have the result in the limit. □

Remember that our goal is to turn our process  $X_t$  into a modification  $\tilde{X}_t$ , where  $\mathbb{P}(\tilde{X}_t = X_t) = 1$  for all  $t$ . What we've proved so far has suggested that we'll take limits from the right: we know that  $t \mapsto X_t(\omega)$  has left and right limits, but we need to check that taking limits from the right to get an rcll function is actually a modification of  $X_t$ . For example, let  $f$  be a deterministic nonincreasing function and have  $X_t = f(t)$  – this is a supermartingale, but if  $f$  is not right-continuous, there's no way for us to modify it to get right-continuous sample paths.

So from here, we'll need to use two facts: first of all, if  $X_t$  is a (sub/super)martingale, then  $\sup\{\mathbb{E}|X_s| : 0 \leq s \leq t\}$  is finite for all finite  $t$ . (We showed this last time.) Also, we'll need to talk about **backwards (sub/super)martingales**, which are indexed by  $\mathbb{Z}_{\leq 0}$  instead of  $\mathbb{Z}_{\geq 0}$ : we have  $\mathcal{F}_{-3} \subset \mathcal{F}_{-2} \subset \mathcal{F}_{-1} \subset \dots$ , and we can say things like

$$Y_{-10} \geq \mathbb{E}(Y_{-9} | \mathcal{F}_{-10})$$

for a supermartingale.

**Proposition 51**

If  $Y_n$  is a backwards (sub/super)martingale, and  $\sup_n |Y_n|_1 < \infty$ , then  $Y_n$  converges to a finite limit  $Y_{-\infty}$  almost surely and in  $L^1$ .

*Proof.* If  $(\dots, Y_{-3}, Y_{-2}, Y_{-1}, Y_0)$  is a backwards supermartingale, then we can apply the Doob upcrossing inequality for any finite  $n \in \mathbb{Z}_{\leq 0}$ :

$$\mathbb{E}U_{a,b}^Y([n, 0]) \leq \frac{\mathbb{E}(X_0 - a)}{b - a}.$$

And now we can take the limit as  $n \rightarrow -\infty$ : the total number of upcrossings is bounded by this finite quantity on the right hand side. This means that for all rational  $a, b$ ,

$$U_{a,b}^Y([-\infty, 0]) < \infty$$

almost surely, so  $Y_n$  must converge almost surely to a limit  $Y_{-\infty}$  (or else it would oscillate between two rational numbers infinitely often). The  $L^1$  convergence is a uniform integrability argument, which is trickier for supermartingales than for martingales (which we did in 18.675) – we should read this on our own.  $\square$

**Theorem 52**

Let  $X_t$  be a supermartingale and let  $D$  be a countable dense subset. Then for all  $\omega \notin N$  for some event with  $\mathbb{P}(N) = 0$ ,  $X_t(\omega)$  has left and right limits. Now define the function

$$Y_t(\omega) = \begin{cases} X_{t+}(\omega) = \lim_{s \downarrow t, s \in D} X_s(\omega) & \text{if limit exists} \\ 0 & \text{otherwise.} \end{cases}$$

This is a supermartingale with the filtration  $\mathcal{G}_t = \mathcal{F}_{t+}$ , and now note that

$$X_t \geq \mathbb{E}(Y_t | \mathcal{F}_t),$$

with equality if the map  $t \mapsto \mathbb{E}X_t$  is right continuous (this is the mild condition).

*Proof.* The first part (left and right limits) follows directly from Corollary 50 and Lemma 47. Let's check that  $Y_t$  is a supermartingale: it is clear that  $Y_t \in \mathcal{G}_t$  because of the limit definition, and for any  $s_k \downarrow t$ ,  $X_{s_k}$  is a backwards supermartingale. This backwards supermartingale is bounded in  $L^1$ , because the supremum of  $\mathbb{E}|X_s|$  is bounded on finite time intervals, so we can use Proposition 51 to show that  $X_{s_k}$  converges almost surely and in  $L^1$  to  $Y_t$ , which means that  $Y_t$  is indeed in  $L^1$  (which shows integrability). Now the supermartingale condition says

$$X_t \geq \mathbb{E}(X_{s_k} | \mathcal{F}_t),$$

and  $X_{s_k}$  converges in  $L^1$  to  $Y_t$ , so this goes to  $\mathbb{E}(Y_t | \mathcal{F}_t)$  as  $k \rightarrow \infty$ . To check the final condition (the equality case), if  $t \mapsto \mathbb{E}X_t$  is right continuous, then

$$\mathbb{E}X_t = \lim_{s \downarrow t} \mathbb{E}X_s,$$

and we can switch the limit and expectation by the  $L^1$  convergence, and this is

$$= \mathbb{E} \left( \lim_{s \downarrow t} X_s \right) = \mathbb{E}(Y_t).$$

Now,  $X_t \geq \mathbb{E}(Y_t|\mathcal{F}_t)$  and  $\mathbb{E}X_t = \mathbb{E}Y_t$ . If both of these are true, then we must have equality in the conditional statement, and indeed  $X_t = \mathbb{E}(Y_t|\mathcal{F}_t)$  almost surely, as desired.

One thing we haven't done yet is to show that  $Y_n$  is a supermartingale. Let  $s_n \downarrow s$  and  $t_n \downarrow t$ , and suppose that  $s < t, s_n < t_n$ . For any event  $A \in \mathcal{G}_s$ , note that

$$\mathbb{E}(Y_s; A) = \lim_{k \rightarrow \infty} \mathbb{E}(X_{s_k}; A)$$

by the backwards supermartingale  $L^1$  convergence, and then we can bound this

$$\geq \lim_{k \rightarrow \infty} \mathbb{E}(X_{t_k}; A) = \mathbb{E}(Y_t; A)$$

by switching the limit and expectation again. □

Now let's put everything together for the main result:

**Theorem 53**

Assume that  $\mathcal{F}_t = \mathcal{F}_{t+}$  for all  $t$  (which means we have a **right-continuous filtration**), and say that  $\mathcal{F}_t$  is complete (meaning it contains the null sets). Let  $X_t$  be a supermartingale with right-continuous mean  $\mathbb{E}[X_t]$ . Then  $X$  has a modification  $\tilde{X}$  which is also a supermartingale with respect to  $\mathcal{F}_t$  which is rcll.

*Proof.* We define

$$\tilde{X}_t(\omega) = \begin{cases} Y_t(\omega) & \text{if } \omega \notin N \\ 0 & \text{otherwise,} \end{cases}$$

where  $N$  is the set from Corollary 50. (This means that we're allowed to look into the whole future and see if things go wrong – it's just an issue with measurability.) So now  $\tilde{X}_t \in \mathcal{F}_t$  because  $\mathcal{F}_t$  is right-continuous and complete, and we need to check that it's a supermartingale – this is because  $Y_t$  is a supermartingale and  $\tilde{X}_t = Y_t$  almost surely.

So now it remains to show that  $X$  is a modification of  $\tilde{X}$ . By the last part of Theorem 52, we have (using  $\tilde{X}$  instead of  $Y$ )

$$X_t = \mathbb{E}(\tilde{X}_t | \mathcal{F}_t)$$

because the mean is right-continuous, but  $\tilde{X}_t$  is measurable with respect to  $\mathcal{F}_{t+} = \mathcal{F}_t$ , and thus

$$X_t = \tilde{X}_t \text{ a.s.}$$

Therefore,  $\tilde{X}$  is indeed a modification of  $X$ . □

Basically, with a sufficiently rich filtration and with the mild condition that the deterministic function  $\mathbb{E}X_t$  is right-continuous, we get some nice results.

**Remark 54.** *It's necessary to assume that  $\mathbb{E}X_t$  is right-continuous: otherwise, we could just consider the deterministic process  $X_t = f(t)$ . Also, assuming that the filtration is right-continuous is also necessary. For example, consider  $\Omega = \{\pm 1\}$  and let  $P$  be the uniform measure on  $\Omega$ ; define*

$$X_t(\omega) = \omega 1\{t > 1\}.$$

*$X$  is 0 and then jumps to a random bit – this is a martingale, and the filtration is generated by  $X$ .*

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

Then  $\mathcal{F}_t$  is trivial until  $t = 1$  and then jumps to the complete sigma-algebra, so the filtration is not right-continuous. And there is no modification of  $X$  that is rcll.

Next time, we'll talk about the optional stopping theorem for continuous martingales, and that will be all from chapter 3.

## 7 February 24, 2020

Recall that last time, we discussed sample path regularity for continuous-time (sub/super)martingales: it was somewhat technical, so let's review the main points. If  $D$  is a countable dense subset in  $[0, \infty)$ , we found last time that for any (sub/super)martingale  $X_t$ , then its left and right limits

$$X_{t+}(\omega) = \lim_{s \downarrow t, s \in D} X_s(\omega), \quad X_{t-}(\omega) = \lim_{s \uparrow t, s \in D} X_s(\omega)$$

exist, and if the mapping  $t \rightarrow \mathbb{E}[X_t]$  is right-continuous, we actually have

$$X_t = \mathbb{E}(X_{t+} | \mathcal{F}_t).$$

Finally, if  $\mathcal{F}_+$  is right-continuous and complete (contains the null sets), then  $X_t$  has a modification  $\tilde{X}_t = X_{t+}$  except on a null set, and this modification has sample paths which are **rcll** (right-continuous with left limits). This allows us to generally assume right-continuous sample paths in most of our discussion.

Today, the discussion will focus on optional stopping theorems, primarily for martingales. The main feature is that we can extend the statement  $X_s = \mathbb{E}(X_t | \mathcal{F}_s)$  for fixed  $s \leq t$  to **random** times: our goal is to say that if  $\sigma \leq \tau$  are both stopping times, then  $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ . There are lots of applications of this, many of which are on our homework, but here's a simple example:

### Example 55

Let  $B_t$  be a Brownian motion started from 0, and let  $\tau = \tau_a \wedge \tau_b$  for  $a < 0 < b$ .

Then an optional stopping theorem will tell us (because  $B_t$  is a martingale) that

$$0 = B_0 = \mathbb{E}B_\tau = a\mathbb{P}(\tau_a < \tau_b) + b\mathbb{P}(\tau_a > \tau_b).$$

And this allows us to explicitly calculate the probability  $p = \frac{b}{b-a}$  that we hit  $a$  before we hit  $b$ .

Note, though, that optional stopping theorems **do not hold** without further restrictions:

### Example 56

Let  $\tau$  be the first time our Brownian motion hits 1: we know  $\tau$  is finite almost surely, but  $0 = B_0 \neq \mathbb{E}B_\tau = 1$ .

So what is a reason for us to believe that the optional stopping theorem should hold? We know that if  $X_t$  is a martingale,  $\mathbb{E}X_t$  is constant (in  $t$ ), and then if  $\tau$  is a stopping time, then  $X_{t \wedge \tau}$  is a martingale. (There is actually a small caveat: we know the stopped process is a martingale in discrete time, but we haven't actually proved this for continuous time yet.) But then this means  $\mathbb{E}X_{t \wedge \tau}$  is also constant, so we know that

$$\mathbb{E}X_0 = \mathbb{E}X_{0 \wedge \tau} = \mathbb{E}X_{t \wedge \tau}$$

for all finite  $t$ . And then taking  $t \rightarrow \infty$  should give (for any finite stopping time  $\tau$ ) that

$$\mathbb{E}X_0 = \lim_{t \rightarrow \infty} \mathbb{E}(X_{t \wedge \tau}) \stackrel{?}{=} \mathbb{E}\left(\lim_{t \rightarrow \infty} X_{t \wedge \tau}\right) = \mathbb{E}X_\tau$$

and we've now reduced this to the usual  $L^1$  convergence question: "can we swap the limit and expectation?".

We can start with an almost-sure (pointwise) convergence result:

**Proposition 57**

Let  $X_t$  be a (sub/super)martingale with right-continuous sample paths, and suppose that  $\sup_t \|X_t\|_1 < \infty$  (our process is bounded in  $L^1$ ). Then  $X_t$  converges almost surely to  $X_\infty \in L^1$  (though  $L^1$  convergence may not happen).

*Proof.* Without loss of generality we can assume that  $X_t$  is a supermartingale (otherwise multiply it by  $-1$ ). Then we know by the upcrossing inequality that

$$\mathbb{E}U_{a,b}([0, T] \cap D) \leq \frac{\mathbb{E}(X_T - a)_-}{b - a},$$

and by the **monotone convergence theorem**, because the left hand side is increasing as we take  $T \rightarrow \infty$ , we know that

$$\mathbb{E}U_{a,b}(D) \leq \sup_t \frac{\mathbb{E}(X_t - a)_-}{b - a} < \infty.$$

This means that  $U_{a,b}$  is finite for all rational  $a < b$  almost surely, and take the countable union over all pairs of rationals – now  $X_t$  must converge, or else there is a rational sandwiched between the liminf and limsup. And Fatou's lemma tells us (because  $X_\infty$  is equal to the liminf of  $X_t$ ) that

$$\mathbb{E}|X_\infty| \leq \liminf_{t \rightarrow \infty} \mathbb{E}|X_t| < \infty,$$

which tells us that  $X_\infty$  is in  $L^1$ . □

Again, it's important to remember that  $X_t$  does not always converge in  $L^1$  to  $X_\infty$  under these conditions: another example is to take

$$X_t = \exp(B_t - t/2).$$

This is a nonnegative martingale with  $\mathbb{E}X_t = 1$  for all  $t$ . But as  $t \rightarrow \infty$ ,  $B_t$  is much smaller than  $\frac{t}{2}$ , so  $X_t$  converges almost surely to  $X_\infty = 0$ .

**Proposition 58**

Let  $X_t$  be a (sub/super)martingale with right-continuous sample paths, and suppose that for some  $p > 1$ , we know that

$$\sup_t \|X_t\|_p < \infty.$$

Then  $X_t \rightarrow X_\infty$  almost surely and in  $L^p$ , so it also converges in  $L^1$ .

*Proof.* This proof is the same as in the discrete case: recall Doob's  $L^p$  inequality, which tells us that

$$\left\| \sup_{s \leq t} |X_s| \right\|_p \leq \frac{p}{p-1} \|X_t\|_p.$$

Taking  $t \rightarrow \infty$ , the left side is nondecreasing in  $t$ , and the right-hand side stays bounded by the assumption, so defining the random variable  $S = \sup_{t \geq 0} |X_t|$ , this is in  $L^p$ . The conditions assumed here are strictly stronger than in



the previous proposition, so we know already that  $X_n \rightarrow X_\infty$  almost surely. And now

$$\lim_{t \rightarrow \infty} \mathbb{E}(|X_t - X_\infty|^p) \rightarrow 0$$

by the dominated convergence theorem, because  $|X_t - X_\infty|^p$  is dominated by  $(2S)^p$ , which we've shown is in  $L^1$ .  $\square$

This will help us with some but not all of the cases we're interested in, and in fact we have a precise characterization of when  $L^1$  convergence occurs. (And the proofs now will be a bit more complicated than in the discrete time case.)

### Definition 59

A collection of random variables  $\{X_i\}_{i \in I}$  is **uniformly integrable** (u.i.) if

$$\lim_{M \rightarrow \infty} \left( \sup_{i \in I} \mathbb{E}(|X_i|; |X_i| \geq M) \right) \rightarrow 0$$

As a trivial example, let  $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ , and assume that all  $|X_i| \leq Z$ . Then the  $|X_i|$ s are uniformly integrable because  $\mathbb{E}(|Z|; |Z| \geq M)$  goes to 0.

A less trivial example is to consider the collection of

$$X_{\mathcal{G}} = \mathbb{E}(Z|\mathcal{G}),$$

where  $\mathcal{G}$  is any sub- $\sigma$ -field of  $\mathcal{F}$ . Then this collection is uniformly integrable – it's a good exercise for us to work on. Finally, note that being uniformly integrable is stronger than being bounded in  $L^1$  – this is also good for us to think about.

### Theorem 60

Suppose we have a collection of random variables  $X_n$  **indexed by integer**  $n$  which converges in probability to  $X_\infty$ .

Then the following are equivalent:

1. The  $\{X_n\}$  are uniformly integrable,
2.  $X_n$  converges in  $L^1$  to  $X_\infty$ ,
3.  $\mathbb{E}|X_n|$  converges to  $\mathbb{E}|X_\infty|$ .

(We can see [3] for the proof: we also proved this last semester in 18.675.) As a word of caution, (1) implies (2) implies (3) if we have a real-indexed process  $X_t$ , but (3) does not imply (1).

### Definition 61

A martingale  $X_t$  is **closed** if there exists a  $Z \in L^1$  such that  $X_t = \mathbb{E}(Z|\mathcal{F}_t)$  for all  $t$ .

### Theorem 62

Let  $X_t$  be a right-continuous martingale. Then the following are equivalent:

1.  $X$  is closed,
2.  $\{X_t\}$  is uniformly integrable,
3.  $X_t$  converges almost surely and in  $L^1$  as  $t \rightarrow \infty$ .

Again, remember that the  $L^p$  condition was sufficient, but this is both necessary and sufficient.

*Proof.* (1) implies (2) because  $\mathbb{E}(Z|\mathcal{G})$  is a uniformly integrable family. To show that (2) implies (3), note that uniformly integrable implies

$$\sup_t \|X_t\|_1 < \infty,$$

so Proposition 57 tells us that  $X_t$  converges almost surely to  $X_\infty$ , and Theorem 60 tells us that  $X_t$  converges in  $L^1$  to  $X_\infty$  as well.

To show that (3) implies (1), note that for all  $t \leq u < \infty$ , we have

$$X_t = \mathbb{E}(X_u | \mathcal{F}_t).$$

$X_u$  converges to  $X_\infty$  in  $L^1$ , so we can pass the limit through the integral when we take  $u \rightarrow \infty$ :

$$= \mathbb{E}(X_\infty | \mathcal{F}_t).$$

Now we can just take  $Z = X_\infty$  to show that our martingale is closed. □

This result was the main step in proving the discrete-time optional stopping theorem, but the continuous-time case makes a few things more complicated. Remember that if  $X_n$  is **discrete** and adapted to  $\mathcal{F}_n$  and  $\tau$  is a stopping time, then we want to say that we have the stopping-time sigma-algebra

$$X_\tau \in \mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq n\} \in \mathcal{F}_n \forall n\}.$$

To check that  $X_\tau$  is actually measurable with respect to  $\mathcal{F}_\tau$ , we just need to check that

$$\{X_\tau \in B\} \cap \{\tau \leq n\} \stackrel{?}{\in} \mathcal{F}_n.$$

But this is just a finite union of events

$$= \{\tau \leq n\} \cap \left( \bigcup_{k=1}^n \{X_k \in B, \tau = k\} \right),$$

which is indeed in  $\mathcal{F}_n$ . But we immediately get measurability issues if we use continuous time: we then have to take an infinite union. So claiming that  $X_\tau \in \mathcal{F}_\tau$  actually requires some regularity assumptions.

**Proposition 63**

Say that  $X_t$  is adapted to  $\mathcal{F}_t$  and has right-continuous sample paths, and let  $\tau$  be a stopping time. Then  $X_\tau \in \mathcal{F}_\tau$ .

*Proof.* We'll prove that  $X_\tau$  is a composition of two maps. Fix  $t > 0$ , and consider the map

$$F : \Omega \times [0, t] \rightarrow \mathbb{R},$$

sending  $F(\omega, s) = X_s(\omega)$ . We claim that this map is measurable with respect to  $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$  – this is called being **progressive**, and it's stronger than being adapted.

To show this, we'll approximate  $F$  with something that has this property: we split our interval  $[0, t]$  into blocks of length  $\frac{t}{n}$ , and define

$$F^{(n)}(\omega, s) = X_{(\lceil s \cdot n/t \rceil)/(n/t)}(\omega)$$

(take the value of  $X$  at the right edge of the block). Each  $F^{(n)}$  is measurable, because the preimage

$$(F^{(n)})^{-1}(B) = \left[ \bigcup_{k=1}^n \left\{ \{X_{kt/n} \in B\} \times \left( \frac{(k-1)t}{n}, \frac{kt}{n} \right] \right\} \right] \cup (\{X_0 \in B\} \times \{0\})$$

which is just a collection of rectangles. And now take  $n \rightarrow \infty$ : by right-continuity,  $F^{(n)}$  converges to  $F$ , because the subscript  $(\lceil s \cdot n/t \rceil)/(n/t)$  decreases to  $s$  as  $n$  gets large. And the pointwise limit of measurable functions is measurable.

So now we want to show that  $X_\tau \in \mathcal{F}_\tau$ : we need to show that

$$\{X_{\tau \in B}\} \cap \{\tau \leq t\} \in \mathcal{F}_t$$

for all  $t$ . Fix  $t$ : let  $G(\omega) = (\omega, t \wedge \tau(\omega))$  (this is measurable), and define  $F$  as before. Then

$$(\Omega, \mathcal{F}_t) \xrightarrow{G} (\Omega \times [0, t], \mathcal{F}_t \otimes \mathcal{B}_{[0,t]}) \xrightarrow{F} (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

is a composition of two measurable maps, and notably

$$\{X_\tau \in B\} \cap \{\tau \leq t\} = \{\tau \leq t\} \cap \{\omega : F(G(\omega)) \in B\},$$

because  $t \wedge \tau(\omega)$  is just  $\tau$ . And since this composition is measurable, both terms on the right side are in  $\mathcal{F}_t$ , giving us the result.  $\square$

Let's look again at a result from discrete-time that we have to be more careful about: the stopped process  $X_{n \wedge \tau}$  can be written as the sum

$$X_0 + \sum_{k=1}^n \mathbf{1}\{k \leq \tau\} (X_k - X_{k-1}).$$

This is the same as the  $H$ -transform  $X_0 + (H \cdot X)_n$ , where  $H_k = \mathbf{1}\{\tau \geq k\}$  is  $\mathcal{F}_{k-1}$ -measurable (it's a previsible process). Since  $X_n$  is integrable by definition of a martingale,  $X_{n \wedge \tau}$  needs to be in  $L^1$  for all  $n$  – it's a sum over  $n+1$  terms and each one is in  $L^1$ . And we can check that  $X_{n \wedge \tau}$  is a martingale using the  $H$ -transform property.

But both of these are less obvious now in the continuous-time case, and we have to do a bit more.

#### Theorem 64

Let  $X_t$  be a **uniformly integrable** martingale **with right-continuous sample paths**, and let  $\sigma, \tau$  be stopping times such that  $\sigma \leq \tau$ . Then  $X_\sigma, X_\tau$  are both in  $L^1$ , and  $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ .

*Proof.* We'll sketch the proof here: **assume that we know the discrete-time result already**. Then we take approximations of  $\sigma, \tau$ : let

$$\sigma_n = \frac{\lceil \sigma \cdot 2^n \rceil}{2^n}, \tau_n = \frac{\lceil \tau \cdot 2^n \rceil}{2^n}.$$

Clearly  $\sigma_n \downarrow \sigma, \tau_n \downarrow \tau, \sigma_n \leq \tau_n$ , and the  $\sigma_n, \tau_n$  are stopping times (because they're larger than  $\sigma, \tau$  respectively). Then the discrete time optional stopping theorem tells us that

$$X_{\sigma_n} = \mathbb{E}(X_{\sigma_0} | \mathcal{F}_{\sigma_n}),$$

so the  $X_{\sigma_n}$  are uniformly integrable. Because  $X_{\sigma_n}$  converge almost surely to  $X_\sigma$  by right continuity, that means that  $X_{\sigma_n}$  also converges in  $L^1$  to  $X_\sigma$ . Similarly,  $X_{\tau_n}$  converges almost surely and in  $L^1$  to  $X_\tau$ . But

$$X_{\sigma_n} = \mathbb{E}(X_{\tau_n} | \mathcal{F}_{\sigma_n}),$$

and any event  $A \in \mathcal{F}_\sigma$  is contained in all  $\mathcal{F}_{\sigma_n}$  and therefore in their intersection. Thus

$$\mathbb{E}(X_{\sigma_n}; A) = \mathbb{E}(X_{\tau_n}; A)$$

for all events  $A \in \mathcal{F}_\sigma$ , and taking  $n \rightarrow \infty$  yields (by  $L^1$  convergence)

$$\mathbb{E}(X_\sigma; A) = \mathbb{E}(X_\tau; A).$$

This is exactly the definition of  $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ , as desired. □

We can now address the point about the stopped process  $X_{n \wedge \tau}$ :

**Corollary 65**

Let  $X_t$  be a martingale with right-continuous sample paths, and let  $\tau$  be a stopping time. Then (1)  $X_{t \wedge \tau}$  is a martingale, and (2) if  $\{X_t\}$  is uniformly integrable, then  $\{X_{t \wedge \tau}\}$  is also uniformly integrable with

$$X_{t \wedge \tau} = \mathbb{E}(X_\tau | \mathcal{F}_t).$$

*Proof.* We prove (2) first. We know that  $\tau$  and  $t \wedge \tau$  are both stopping times, so Theorem 64 tells us that  $X_\tau, X_{t \wedge \tau}$  are in  $L^1$ . Now by the optional stopping theorem,

$$X_{t \wedge \tau} = \mathbb{E}(X_\tau | \mathcal{F}_{t \wedge \tau}),$$

so we want to show that for all  $A$  in  $\mathcal{F}_t$ ,

$$\mathbb{E}(X_{t \wedge \tau}; A) = \mathbb{E}(X_\tau; A).$$

One part of this is trivial: if we consider the event  $1_A 1_{\{\tau \leq t\}}$ , this is exactly  $1_A 1_{\{\tau \leq t\}} X_\tau$ , so in the case  $\tau \leq t$  the two sides are actually identical. On the other hand, if we look at the case where  $\tau > t$ , then  $A \cap \{\tau > t\} \in \mathcal{F}_t$ , and we can check that it is in  $\mathcal{F}_\tau$  as well: this means it is in their intersection, which is  $\mathcal{F}_{t \wedge \tau}$ . So

$$\mathbb{E}(1_A 1_{\{\tau > t\}} X_{t \wedge \tau}) = \mathbb{E}(1_A 1_{\{\tau > t\}} X_\tau),$$

where we've used  $X_{t \wedge \tau} = \mathbb{E}(X_\tau | \mathcal{F}_{t \wedge \tau})$ . Adding together the cases where  $\tau \leq t$  and  $\tau > t$  yields the result.

So now we show (1) from (2):  $X_{t \wedge \tau}$  is measurable with respect to  $\mathcal{F}_{t \wedge \tau} \subseteq \mathcal{F}_t$ , so it is adapted. By Theorem 64,  $X_{t \wedge \tau}$  is in  $L^1$ , so it is integrable. We just need to show that

$$X_{s \wedge \tau} = \mathbb{E}(X_{t \wedge \tau} | \mathcal{F}_s) :$$

if we fix any finite  $t$  and define  $Y_s = X_{s \wedge t}$  (run the process only for a finite time), this is uniformly integrable (regardless of what martingale we have), because there is a single random variable  $X_t$  with conditional expectations

$$Y_s = \mathbb{E}(X_t | \mathcal{F}_{s \wedge t})$$

So now applying (2), we have that  $Y_\tau = X_{\tau \wedge t}$  is in  $L^1$ , and  $Y_s = \mathbb{E}(Y_\tau | \mathcal{F}_s)$ , so

$$X_{s \wedge \tau} = \mathbb{E}(X_{t \wedge \tau} | \mathcal{F}_s),$$

verifying the condition that we want. □

## 8 February 26, 2020

Today, we're starting with Chapter 4 of our textbook, which discusses **continuous semimartingales**: these are processes of the form

$$X_t = M_t + A_t,$$

where  $M_t$  is a **continuous local martingale** and  $A_t$  is a **finite variation process**. We haven't defined any of these terms or classified any of these processes yet, which will be the topic of the next few classes. But before that, we'll take a step back and look at where things are headed (because the last few lectures have been a bit technical).

### Example 66

Consider the standard Brownian motion  $B_t$ , which is a continuous martingale. Suppose we want to consider a function  $X_t = f(B_t)$  for a smooth function  $f$ : how does  $X_t$  evolve?

Heuristically, we can use Itô's formula, which tells us something of the form

$$dX_t = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt$$

(where the main idea is that we've replaced the  $(dB_t)^2$  term with  $dt$ ). In particular, this means  $X$  is not a martingale unless  $f$  is linear, which makes sense: a linear scale  $\mu + \sigma B_t$  should give a martingale, but otherwise we get a drift term (caused by the curvature of our function  $f''$ ). Intuitively, we can have the picture that a positively curved function  $f$  will have  $f(B_{t+dt})$  larger than our linear approximation  $f(B_t) + f'(B_t)dB_t$ .

So Itô's formula gives us a decomposition

$$X_t = \left[ \int_0^t f'(B_s)dB_s \right] + \left[ \int_0^t \frac{1}{2}f''(B_s)ds \right],$$

which turns out to exactly be the process  $M_t + A_t$  we're looking for. So this definition is really coming out of manipulations of Brownian motion!

Throughout this, it might be useful to keep the discrete-time picture in mind: say that a process  $X_n$  is adapted to  $\mathcal{F}_n$ , where  $X_0 = 0$  and  $\mathbb{E}|X_n| < \infty$  for all  $n$ . Then we know that we can decompose  $X_n$  via

$$X_n = \sum_{i=1}^n (X_i - X_{i-1} - \mathbb{E}(X_i - X_{i-1}|\mathcal{F}_{i-1})) + \sum_{i=1}^n \mathbb{E}(X_i - X_{i-1}|\mathcal{F}_{i-1})$$

Then the first term is our martingale  $M_n$ , and the second term is our finite variation process  $A_n$ . It's not entirely clear how this  $A_n$  relates to the  $A_t$  above, though: to make that more clear, imagine that our process  $X_n$  is a function  $f(S_n)$ , where

$$S_n = \varepsilon \sum_{i=1}^n \zeta_i, \quad \zeta_i \sim \text{Unif}\{\pm 1\}$$

is a random walk with step size  $\varepsilon$ . Then

$$A_{n+1} - A_n = \mathbb{E}(f(S_n + \varepsilon\zeta_{n+1}) - f(S_n)|\mathcal{F}_n),$$

and now we can Taylor expand in  $\varepsilon$  to get

$$= f'(S_n)\mathbb{E}(\zeta_{n+1}|\mathcal{F}_n) + \frac{1}{2}\varepsilon^2 f''(S_n)\mathbb{E}(\zeta_{n+1}^2|\mathcal{F}_n) + \dots$$

(taking out the constants and derivatives because they're measurable with respect to  $S_n$ ). Now we can simplify a lot

because  $\zeta$ s are symmetric random signs, and this just simplifies to

$$A_{n+1} - A_n = \frac{1}{2}\varepsilon^2 f''(S_n) + o(\varepsilon^2),$$

which now looks identical to the  $A_t$  term above. So to summarize, we want to understand this decomposition in the continuous case: chapters 4 and 5 will help us with a formal characterization of this class of processes, and then we'll prove Itô's formula, which essentially tells us that **the image of a continuous semimartingale under a smooth map is another continuous semimartingale**:

$$h(M_t + A_t) = \tilde{M}_t + \tilde{A}_t$$

for sufficiently nice  $h$ . Heuristically, the idea is that if  $X_t = M_t + A_t$ , then

$$\begin{aligned} dh(M_t + A_t) &= h'(X_t)dX_t + \frac{1}{2}h''(X_t)(dX_t)^2 \\ &= h'(X_t)(dM_t + dA_t) + \frac{1}{2}h''(X_t)((dM_t)^2 + 2dM_t dA_t + (dA_t)^2). \end{aligned}$$

A step of the martingale  $dM_t$  is "like a step of the Brownian motion with some variance," so  $dM_t$  is a Gaussian step with variance of order  $\sqrt{dt}$ , and  $dA_t$  is something deterministic of order  $dt$ . So we can toss the last two terms, and we're just left with the three terms of highest order

$$= h'(X_t)dM_t + \left( h'(X_t)dA_t + \frac{1}{2}h''(X_t)(dM_t)^2 \right)$$

(where the  $h'(X_t)$  term will be the martingale part and the rest is the finite variation part). We don't really know how to integrate any of these terms, so we have a lot to understand – we'll start with the  $h'(X_t)dA_t$  term. **What is a finite variation process, and how we take an integral against it?**

Today, we'll start by talking about this in the case where  $A_t$  is some deterministic function of time, and  $X_t$  is some deterministic process. The reason for eliminating the  $\omega$  here because  $t$  will "be our  $\omega$ ."

### Definition 67

Let  $(\Omega, \mathcal{F})$  be a measurable space (this is not the probability space of our process). A **finite signed measure** on  $(\Omega, \mathcal{F})$  is a function  $\alpha : \mathcal{F} \rightarrow \mathbb{R}$  such that  $\alpha$  is countably additive: if we have disjoint sets  $A_i \in \mathcal{F}$ , then

$$\alpha\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \alpha(A_i)$$

where the sum must be absolutely convergent.

### Example 68

If  $\alpha_+, \alpha_-$  are (actual nonnegative) measures on  $(\Omega, \mathcal{F})$ , then  $\alpha = \alpha_+ - \alpha_-$  is a signed measure. Also, if  $\mu$  is a measure on  $(\Omega, \mathcal{F})$  such that  $\int_{\Omega} |h|d\mu < \infty$ , then

$$\nu(E) = \int_E h d\mu$$

is also a signed measure – in fact,  $h = \frac{d\nu}{d\mu}$  is the Radon-Nikodym derivative.

It turns out that we can only get a signed measure by writing  $\alpha = \alpha_+ - \alpha_-$  as the difference of two measures. Let's see how to show this:

**Definition 69**

Let  $\alpha$  be a signed measure on  $(\Omega, \mathcal{F})$ . Then  $A \in \mathcal{F}$  is **positive** if  $\alpha(B) \geq 0$  for all  $B \subseteq A$  (**negative** if  $\alpha(B) \leq 0$ ), and  $A$  is a **null set** if  $\alpha(B) = 0$  for all  $B$ .

**Theorem 70** (Hahn decomposition theorem)

For any signed measure  $\alpha$  on  $(\Omega, \mathcal{F})$ , there is a bipartition  $\Omega = \Omega_+ \sqcup \Omega_-$  such that  $\Omega_+$  is positive and  $\Omega_-$  is negative. This decomposition is essentially unique – for any other decomposition  $\Omega = B_+ \sqcup B_-$ ,  $B_+ \cap \Omega_-$ ,  $B_- \cap \Omega_+$  must be null sets.

(We should refer to [3] for the proof – it's about a paragraph long.)

**Theorem 71** (Jordan decomposition theorem)

We can write any signed measure uniquely as  $\alpha = \alpha_+ - \alpha_-$  (where  $\alpha_+, \alpha_-$  are measures on  $(\Omega, \mathcal{F})$ ).

*Proof.* Just let  $\alpha_+(E) = \alpha(E \cap \Omega_+)$  and  $\alpha_-(E) = \alpha(E \cap \Omega_-)$  – we can check that this is unique. □

We'll now connect this to the idea of a finite variation process:

**Definition 72**

A continuous function  $a : [0, T] \rightarrow \mathbb{R}$  is **of finite/bounded variation** (FV or BV) if there exists a signed measure  $\alpha$  on  $([0, T], \mathcal{B}_{[0, T]})$  such that

$$a(t) = \alpha([0, t])$$

for all  $0 \leq t \leq T$ .

Because we know the Jordan decomposition, this means we can write

$$a(t) = \alpha_+([0, t]) + \alpha_-([0, t])$$

where the two terms on the right hand side are nondecreasing functions of  $t$  (because  $\alpha_+, \alpha_-$  are actual measures). We will sometimes refer to these as  $a_+(t)$  and  $a_-(t)$ .

The measure  $\mu = |\alpha| = \alpha_+ + \alpha_-$  is the **total variation measure** of  $a$ , and we define

$$v(t) = \mu([0, t]).$$

This is the “total variation of  $a$  on the interval  $[0, t]$ .” So now we want to define an integral of the form

$$\int_0^T f(s) da(s) :$$

one natural way to do this is to define it in terms of our measure

$$= \int_0^T f(s) \alpha(ds),$$

and similarly to define  $\int_0^T f(s) |da(s)| = \int_0^T f(s) \mu(ds)$ . It's important to emphasize that in both of these equations, the left hand side is new notation, while the right hand side is just a Lebesgue integral (which we know how to compute).

In order for these integrals to be well-defined, we just need to make sure that  $f$  is measurable and absolutely integrable:  $\int_0^T |f(s)|\mu(ds) < \infty$ .

Some simple properties of this integral: first of all, we have

$$\left| \int_0^T f(s) da(s) \right| \leq \int_0^T |f(s)| |da(s)|$$

by Jensen's inequality. Another result is that

$$b(t) = \int_0^t f(s) da(s)$$

is also of finite variation (which we'll need for something like Itô's formula), because we can define  $b(t) = \beta([0, t])$  where  $\beta$  is the signed measure

$$\beta(E) = \int_E f(s) \alpha(ds).$$

(This should look like the Radon-Nikodym derivative equation.) And now decomposing into positive and negative parts for both  $f$  and  $\alpha$ , the explicit formula for  $\beta$  is that

$$\beta(E) = \int_E (f_+(s)\alpha_+(ds) + f_-(s)\alpha_-(ds)) - \int_E (f_+(s)\alpha_-(ds) + f_-(s)\alpha_+(ds))$$

So to recap, we're letting  $(\Omega, \mathcal{F}) = ([0, t], \mathcal{B})$ , and we have a signed measure  $\alpha$  corresponding to a decomposition  $\Omega = \Omega_+ \sqcup \Omega_-$ . This gives a corresponding  $\alpha = \alpha_+ - \alpha_-$  and  $\mu = \alpha_+ + \alpha_-$ . Note that  $\alpha_+ \ll \mu$  (is absolutely continuous with respect to  $\mu$ ), because if a set has zero measure under  $\mu$ , it has zero measure under  $\alpha_+$ . Similarly,  $\alpha_- \ll \mu$ , and we can write down the Radon-Nikodym derivatives directly using the defining properties:

$$\frac{d\alpha_+}{d\mu} = 1_{\Omega_+} = h_+$$

(because we want  $\alpha_+(E) = \int_E \frac{d\alpha_+}{d\mu} d\mu$  to be the integral over only  $\Omega_+$ ) and similarly

$$\frac{d\alpha_-}{d\mu} = 1_{\Omega_-} = h_-.$$

Then the corresponding finite-variation process  $a : [0, t] \rightarrow \mathbb{R}$  can be written as

$$a(s) = \alpha([0, s]) = \int_0^s h(s) \mu(ds),$$

and

$$h = \frac{d\alpha}{d\mu} = h_+ - h_-.$$

This then decomposes as  $a(s) = a_+(s) - a_-(s)$ , and the total variation can be written as  $v(s) = a_+(s) + a_-(s)$ . Therefore,

$$a_+(s) = \frac{1}{2}(v(s) + a(s)), \quad a_-(s) = \frac{1}{2}(v(s) - a(s)).$$

Remember that we wanted to understand the quantity  $\int_0^t h'(X_s) dA_s$ . Once we introduce randomness back in the process, our processes  $A$  and  $X$  will usually be correlated, so we need to ask how we can calculate the integral – for example, do we need to worry about time steps? The first step is to look at a simpler quantity:



### Lemma 73

Let  $a : [0, t] \rightarrow \mathbb{R}$  be a function of finite variation. Then

$$\int_0^t |da(s)| = \sup\{V_P(a) : P \text{ subdivision of } [0, t]\},$$

where  $P = \{0 = t_0 < t_1 < \dots < t_p = t\}$  and

$$V_P(a) = \sum_{i=1}^p |a(t_i) - a(t_{i-1})|.$$

In addition, if we have an increasing set of subdivisions  $P_1 \subseteq P_2 \subseteq \dots$  and then mesh of  $P_n$  goes to zero, then  $V_{P_n}(a) \rightarrow \int_0^t |da(s)|$ .

Basically, we can approximate total variation by a discrete subdivision. This proof is similar to the Radon-Nikodym theorem, but we'll do something self-contained here:

*Proof.* We know that  $V_P(a) \leq \int_0^t |da(s)|$  by Jensen, so it suffices to show the convergence result. By scaling, we can assume without loss of generality that  $\mu([0, t]) = 1$ . This means that  $([0, t], \mathcal{B}, \mu)$  is a probability space.

Let  $\mathcal{G}_n$  be the sigma-algebra generated by the intervals of  $P_n$  – there are finitely many elements in here for each fixed  $n$ . The  $G_n$  are nondecreasing because the  $P_n$  are nondecreasing, and the sigma-algebra  $\sigma$  generated by the union of the  $G_n$  is just the Borel sigma-algebra on  $[0, t]$ .

Now let  $X(t) = h(t) = 1\{t \in \Omega_+\} - 1\{t \in \Omega_-\}$  (here we use the Hahn decomposition).  $X$  is a random variable on  $([0, t], \mathcal{B}, \mu)$ , and we have the filtration  $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \dots \subseteq \mathcal{B}$ , which gives us the **closed** martingale (meaning uniformly integrable as well)

$$X_n = \mathbb{E}(X|\mathcal{G}_n).$$

This means that  $X_n$  converges to  $X$  almost surely and in  $L^1$ , but  $X_n$  is the value of  $X$  (it's 1 or  $-$ ). Because  $\mathcal{G}_n$  is finite,  $X_n$  is **piecewise constant** on the intervals of  $P_n$ . Then (break points depend on  $n$ ) we can write

$$\mathbb{E}(X_n; [t_{i-1}^{(n)}, t_i^{(n)}]) = X_n(t_{i-1}^{(n)}) \cdot \mu([t_{i-1}^{(n)}, t_i^{(n)}])$$

(where we have cheated a tiny bit because it's possible to modify  $X$  at any point, but the point is that it's the measure of the interval times the value at any given point). But we also know that on this same interval, ( $X$  and  $h$  are the same thing)

$$\mathbb{E}(X; [t_{i-1}, t_i]) = \int_{t_{i-1}}^{t_i} h(s)\mu(ds) = \alpha([t_{i-1}, t_i]) = a(t_i) - a(t_{i-1}).$$

But these two expectations should be equal, because  $X_n = \mathbb{E}(X|\mathcal{G})$ , so **the value of  $X_n$  on  $[t_{i-1}, t_i]$  is equal to  $\frac{a(t_i) - a(t_{i-1})}{\mu([t_{i-1}, t_i])}$** .

Now because  $X_n$  converges to  $X$  in  $L^1$ ,  $\mathbb{E}|X_n|$  converges to  $\mathbb{E}|X|$ .  $X_n$  is piecewise constant, and that means

$$\mathbb{E}|X_n| = \sum_{i=1}^{|P_n|} \mu([t_{i-1}, t_i]) \cdot \left| \frac{a(t_i) - a(t_{i-1})}{\mu([t_{i-1}, t_i])} \right| = \sum_{i=1}^{|P_n|} |a(t_i) - a(t_{i-1})| = V_{P_n}(a).$$

But on the other hand,  $|X|$  is 1 almost surely (it's either 1 or  $-$ ), so

$$\mathbb{E}|X| = 1 = \mu([0, t]) = \int_0^t |da(s)|,$$

which proves the claim by taking  $n \rightarrow \infty$ .

**Corollary 74**

Let  $a$  be a function of finite variation and  $f : [0, t] \rightarrow \mathbb{R}$  be a continuous function. Then if  $P_1 \subseteq P_2 \subseteq \dots$  are subdivisions of  $[0, t]$  with mesh going to 0, then

$$\sum_{i=1}^{|P_n|} f(t_{i-1}^{(n)}) \left( a(t_i^{(n)}) - a(t_{i-1}^{(n)}) \right) \rightarrow \int_0^t f(s) da(s),$$

where again the  $t_i$  implicitly depend on  $n$ .

This is useful because we have more practice looking at things like the left hand side, so we'll be able to simplify calculations.

*Proof.* We start by trying to define the Stieltjes integral: let  $t_i^{(n)}$  be our break points corresponding to the subdivision  $P_n$ , and define

$$f^{(n)}(t) = f(t_{i-1}^{(n)}) \quad \forall t \in [t_{i-1}^{(n)}, t_i^{(n)}).$$

The  $f^{(n)}$  are bounded uniformly, because

$$\|f^{(n)}\|_\infty \leq \|f\|_\infty < \infty,$$

and also  $f^{(n)}$  converges to  $f$  pointwise (because  $f$  is continuous). Now the left hand side of the equation we're trying to show is  $\int_0^t f^{(n)} d\alpha$ , which converges to  $\int_0^t f d\alpha$  by the dominated convergence theorem. □

**Remark 75.** One last point: a function  $a : [0, \infty) \rightarrow \mathbb{R}$  has finite variation if  $a$  has finite variation on any compact interval  $[0, t]$ .

## 9 March 2, 2020

Solutions for the first two homework assignments are posted on Stellar now; the next homework is due on Monday, and we'll have office hours 3-5 on Thursday instead of today.

Recall that a function  $a : [0, T] \rightarrow \mathbb{R}$  is of **finite/bounded variation** if there exists a signed measure  $\alpha = \alpha_+ - \alpha_-$  on  $[0, T]$  such that

$$a(t) = \alpha([0, t]) = a_+(t) - a_-(t)$$

(noting that  $a(0) = 0$ ). We also mentioned that we can define the integral

$$\int_0^t f(s) ds = \int_0^t f(s) \alpha(ds),$$

which is well-defined as long as we have the absolute integrability condition

$$\int_0^t |f(s)| |da(s)| < \infty.$$

We also noted that the integral  $\int_0^t f(s) da(s)$ , as a function of  $t$ , is also of finite variation. Finally, if  $f$  is continuous, and we have a sequence of subdivisions  $P_0 \subseteq P_1 \subseteq \dots$  of  $[0, T]$  with mesh going to zero, then the discrete approximations

$$\sum_{i=1}^{|P_n|} f(t_{i-1}^{(n)}) \left( a(t_i^{(n)}) - a(t_{i-1}^{(n)}) \right)$$

converges to the integral  $\int_0^t f(s)da(s)$ .

Everything last time was deterministic, so we'll add randomness now. We'll be on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  for the rest of this lecture.

### Definition 76

An adapted process  $A_t$  (to our filtered probability space) is a **finite variation process** if all sample paths have finite variation (are FV). (We assume that we have continuous sample paths, as we will for the rest of this chapter.) If all sample paths are nondecreasing in  $t$ , then  $A_t$  is an **increasing process**.

We want our discrete approximations to converge to the integral again: recall that a process  $H_t$  is **progressive** if  $F(\omega, s) = H_s(\omega)$  is measurable with respect to the sigma-algebra  $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$ . The point here is that this is stronger than being adapted, but any adapted, continuous process is progressive.

### Proposition 77

Let  $A$  be a finite-variation process, and let  $H$  be a progressive process. Suppose that  $\int_0^t |H_s(\omega)| |dA_s(\omega)|$  is finite for all finite  $t$ . Then

$$(H \cdot A)_t = \int_0^t H_s dA_s$$

is a well-defined, finite variation process.

We'll skip over this proof for now, but the fact that this is an FV process follows from the result above that  $\int_0^t f(s)da(s)$  is FV. What we need to check is that this process is measurable with respect to  $\mathcal{F}_t$ : we should read this on our own, and this is where we use the progressive condition.

Remember that we want to study the class of continuous semi-martingales, which are of the form  $X_t = A_t + M_t$ , where  $A$  is FV and  $M$  is a (local) martingale. We'll spend some time now on the latter.

### Definition 78

A **continuous local martingale**  $M_t$  on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  is an adapted continuous process for which there exists a sequence of stopping times  $(\tau_n)$  such that

- $\tau_n(\omega) \uparrow \infty$  for all  $\omega$ ,
- The stopped processes  $(M_{t \wedge \tau_n} - M_0)$  are uniformly integrable martingales for all  $n$ . (We often say that  $\tau_n$  **reduces**  $M$ .)

### Example 79

Let  $B_t$  be a Brownian motion in  $\mathbb{R}^3$ , and define  $M_t = \frac{1}{|B_t|} = (B_{t,1}^2 + B_{t,2}^2 + B_{t,3}^2)^{-1/2}$ . This is not a true martingale, but it is a local martingale. (We'll talk more about this difference in a future lecture.)

Note that all continuous martingales are continuous local martingales, because we can just take  $\tau_n = n$ : we know that

$$X_s = \mathbb{E}(X_n | \mathcal{F}_s) \quad \forall s \leq n,$$

so for all  $n$ , the variables  $(M_{t \wedge \tau_n} - M_0)$  are just conditional expectations of  $X_n$ , so they must be uniformly integrable. Also, this implies that **we can actually leave out "uniformly integrable" from the definition** with the same conditional expectation argument.

Another note is that any stopping time  $\tau$  of a continuous local martingale  $M$  gives us a continuous local martingale  $M_{t \wedge \tau}$ . This is true because the optional stopping theorem tells us that a stopped uniformly integrable martingale is still uniformly integrable.

One set of stopping times to consider (which we can always use) is

$$\tau_n = \inf\{t : |M_t - M_0| = n\}.$$

(At any finite  $n$ , all of the stopped  $M_{t \wedge \tau_n}$  are bounded, so we have uniform integrability.)

**Proposition 80**

Let  $M$  be a nonnegative continuous local martingale with  $M_0 \in L^1$  (we need to add this condition because it's no longer true that  $M_t$  needs to be integrable in general). Then  $M$  is a true supermartingale, but not necessarily a martingale.

*Proof.* Let  $N_t = M_t - M_0$ , and let  $\tau_n$  be the reducing sequence for  $M$ . Then for all  $s \leq t$ , we have

$$N_{s \wedge \tau_n} = \mathbb{E}(N_{t \wedge \tau_n} | \mathcal{F}_s)$$

because  $\tau_n$  creates a uniformly integrable martingale, so the optional stopping theorem holds. And now adding  $M_0$  back (it's integrable),

$$M_{s \wedge \tau_n} = \mathbb{E}(M_{t \wedge \tau_n} | \mathcal{F}_s).$$

Taking limits on both sides,  $M_{s \wedge \tau_n}$  converges to  $M_s$  by continuity, and the right hand side satisfies

$$\lim_{n \rightarrow \infty} \mathbb{E}(M_{t \wedge \tau_n} | \mathcal{F}_s) \geq \mathbb{E}\left(\liminf_{n \rightarrow \infty} M_{t \wedge \tau_n} | \mathcal{F}_s\right) = \mathbb{E}(M_t | \mathcal{F}_s)$$

by Fatou's lemma and continuity. □

Again, the example to keep in mind is  $M_t = \frac{1}{|B_t|}$  for a three-dimensional Brownian motion.

**Theorem 81**

If  $X$  is both a continuous local martingale and a finite variation process, then  $X = 0$ .

(This is good, because we want to decompose  $X_t = M_t + A_t$ , and this result gives us uniqueness.)

*Proof.* Since  $X$  is finite variation, we know that

$$X_t = \int_0^t dX_s, \quad X_0 = 0.$$

Now  $X$  is a continuous local martingale, so we define

$$\tau_n = \inf\{t : \int_0^t |dX_s| = n\}$$

to be the first time the total variation exceeds  $n$ . Letting  $X_t^{(n)} = X_{t \wedge \tau_n}$ , we know that  $|X_t^{(n)}| \leq n$  for all  $t$  (because it was stopped before it could change by more than  $n$ ), which means that  $X^{(n)}$  is bounded and therefore uniformly integrable for each  $n$ .

Fix  $n$ , and let  $N = X^{(n)}$  for notation. (Note that  $N$  is a martingale.) By definition, we know that

$$\mathbb{E}[(N_{s_2} - N_{s_1})(N_{t_2} - N_{t_1})] = 0$$

for all  $s_1 \leq s_2 \leq t_1 \leq t_2$  (because the second term disappears if we condition on  $\mathcal{F}_{t_1}$  by the definition of a martingale). This means that

$$\mathbb{E}(N_t^2) = \mathbb{E} \left[ \sum_i (N_{t_i} - N_{t_{i-1}})^2 \right] = \sum_i \mathbb{E}[(N_{t_i} - N_{t_{i-1}})^2],$$

but the fact that  $\int_0^t |dX_s| = n$  means that summing squares of increments will give us a something small:

$$\mathbb{E}(N_t^2) \leq \mathbb{E} \left[ \left( \sup_i |N_{t_i} - N_{t_{i-1}}| \right) \cdot \sum_i |N_{t_i} - N_{t_{i-1}}| \right].$$

The second term is at most  $n$  by definition, and the first term goes to 0 as the mesh of  $\{t_i\}$  goes to zero. Because the whole thing inside the expectation is bounded, we know that this goes to 0 by the dominated convergence theorem. Thus  $\mathbb{E}(N_t^2) = 0$ , which means that

$$N_t = X_t^{(n)} = X_{t \wedge \tau_n}$$

is identically zero for all  $n$ , meaning that  $X = 0$ . □

From here on, we assume that  $\mathcal{F}_t$  is complete.

### Theorem 82

Let  $M$  be a continuous local martingale. Then there exists an increasing (finite variation) process  $A_t = \langle M, M \rangle_t$  such that  $M_t^2 - A_t$  is a continuous local martingale, and  $A$ , the **quadratic variation** (QV), is unique up to null sets. If  $P_1 \subseteq P_2 \subseteq \dots$  is a subdivision of  $[0, t]$ , then  $V_{P_n}^2(M)$  (the sum of squares of  $M_{t_i} - M_{t_{i-1}}$ ) converges to  $A_t$ .

Note that  $A$  is the formalization of what we labeled as  $\int_0^t (dM_s)^2$  in our earlier heuristic arguments. This is a really important quantity: as motivation for why we care about this, if  $M$  is a continuous local martingale and  $A_t = \langle M, M \rangle_t$ , then  $(M_t)$  is equal in distribution to a Brownian motion indexed by  $A_t$ . So a continuous local martingale is just a time change of Brownian motion! (As a special case, if  $\langle M, M \rangle_t = t$ , then  $M$  is just a standard Brownian motion.)

*Proof.* We'll prove this **assuming that  $M$  is a bounded continuous martingale**. (Generalizing to continuous local martingales takes very little work.) Again, we assume  $M_0 = 0$  for simplicity.

First of all, consider a discrete-time martingale  $M_n$ . We know that

$$M_n^2 = \sum_{i=1}^n (M_i^2 - M_{i-1}^2 - \mathbb{E}(M_i^2 - M_{i-1}^2 | \mathcal{F}_{i-1})) + \sum_{i=1}^n \mathbb{E}(M_i^2 - M_{i-1}^2 | \mathcal{F}_{i-1}),$$

where the first summation is a martingale. So a natural idea is to take a discrete approximation for our continuous-time martingale, and hope that the remainder converges to our process  $A$ .

It turns out we don't need the expectations: we can take

$$A^{(n)} = \sum (M_{t_i} - M_{t_{i-1}})^2.$$

More formally, in continuous time,  $M$  is a continuous bounded martingale, so we can take our increasing subdivisions of  $[0, T]$  (with mesh going to zero), and define

$$M_t^2 = \sum_{i=1}^{|P_n|} \left( M_{t \wedge t_i^{(n)}}^2 - M_{t \wedge t_{i-1}^{(n)}}^2 \right), \quad A_t^{(n)} = \sum_{i=1}^{|P_n|} \left( M_{t \wedge t_i^{(n)}} - M_{t \wedge t_{i-1}^{(n)}} \right)^2.$$

Then subtracting the two and expanding, we find that

$$N_t^{(n)} = M_t^2 - A_t^{(n)} = 2 \sum_{i=1}^{|P_n|} M_{t \wedge t_{i-1}^{(n)}} \left( M_{t \wedge t_i^{(n)}} - M_{t \wedge t_{i-1}^{(n)}} \right).$$

Notice that  $N_t^{(n)}$  is a bounded martingale.

**Lemma 83**

If we assume that  $|M_t| \leq C$  for all  $t \leq T$  and  $M_0 = 0$ , then

$$\mathbb{E} \left[ \left( \sum_i (M_{t_i} - M_{t_{i-1}})^2 \right)^2 \right] \leq 12C^4.$$

(This becomes  $48C^4$  if we don't assume  $M_0 = 0$ .)

*Proof of lemma.* Expanding out the left hand side, we get diagonal terms for  $i = j$  and nondiagonal terms otherwise:

$$= \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^4] + 2 \sum_{i < j} \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2 (M_{t_j} - M_{t_{j-1}})^2].$$

Simplify the first term by pulling out some factors, and sum over  $i$  first in the second term

$$\leq (2C)^2 \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2] + 2 \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2 (M_T - M_{t_i})^2]$$

where we've used the orthogonality of disjoint intervals when we sum over  $j > i$ . We take care of the blue term by using our bound again:

$$\leq [(2C)^2 + 2(2C)^2] \sum_i \mathbb{E}[(M_{t_i} - M_{t_{i-1}})^2] = 12C^2 \cdot \mathbb{E}[M_T^2]$$

again using orthogonality (here's where we gain the factor of 4 when  $M_0 = 0$  isn't assumed), and now this is

$$\leq 12C^2 \cdot C^2 = 12C^4.$$

□

**Lemma 84**

Again assume that  $M_t$  is bounded. Then the sequence  $(N_T^{(n)})_n$  is Cauchy in  $L^2$ .

*Proof.* Say  $m \leq n$ . Then the subdivision  $P_m$  is contained in the subdivision  $P_n$ : let  $P_n = (s_j : 1 \leq j \leq |P_n|)$  and  $P_m = (t_i : 1 \leq i \leq |P_m|)$ . Because  $P_n$  is a refinement of  $P_m$ , we can also index the  $t_i$ s as  $t_{i,k}$ , where the first index  $i$  tells us that we are in the range  $[s_{i-1}, s_i]$ . Now the expected covariance can be calculated by writing out the definition of  $N_t^{(n)}$ :

$$\frac{1}{4} \mathbb{E}[N_T^{(m)} N_T^{(n)}] \leq \mathbb{E} \left[ \sum_{i,j} M_{s_{i-1}} (M_{s_i} - M_{s_{i-1}}) M_{t_j} (M_{t_j} - M_{t_{j-1}}) \right].$$

If the time intervals are disjoint, we get zero, so we only get a contribution when the  $t$  increments are inside the  $s$  increments: we now sum over the subincrements

$$= \sum_{i,k} \mathbb{E} [M_{s_{i-1}} (M_{s_i} - M_{s_{i-1}}) M_{t_{i,k}} (M_{t_{i,k}} - M_{t_{i,k-1}})].$$

Each term here involves the times  $s_{i-1} \leq t_{i,k-1} \leq t_{i,k} \leq s_i$ , so we can further break up the increment  $M_{s_i} - M_{s_{i-1}}$  into the three components. This yields

$$\frac{1}{4} \mathbb{E}[N_T^{(m)} N_T^{(n)}] = \sum_{i,k} \mathbb{E} [M_{s_{i-1}} (M_{t_{i,k}} - M_{t_{i,k-1}}) M_{t_{i,k}} (M_{t_{i,k}} - M_{t_{i,k-1}})].$$

And now to show that this sequence is Cauchy in  $L^2$ , we want to calculate the  $L^2$  distance:

$$\frac{1}{4} \mathbb{E} \left( N_T^{(m)} - N_T^{(n)} \right)^2 = \sum_i \mathbb{E} [M_{s_{i-1}}^2 (M_{s_i} - M_{s_{i-1}})^2] - 2(\text{cross term above}) + \sum_{i,k} \mathbb{E} \left[ M_{t_{i,k-1}}^2 (M_{t_{i,k}} - M_{t_{i,k-1}})^2 \right].$$

Simplifying everything, we'll find that this is

$$= \sum_{i,k} \mathbb{E} \left[ (M_{s_i} - M_{t_{i,k-1}})^2 (M_{t_{i,k}} - M_{t_{i,k-1}})^2 \right].$$

Applying Cauchy-Schwarz, this is at most

$$\leq \mathbb{E} \left[ \sup_{i,k} |M_{s_i} - M_{t_{i,k-1}}|^4 \right]^{1/2} \mathbb{E} \left[ \left( \sum_{i,k} (M_{t_{i,k}} - M_{t_{i,k-1}})^2 \right)^2 \right]^{1/2}.$$

The second term is bounded by our previous lemma, and the first term converges to 0 again by dominated convergence theorem. This shows that we indeed have Cauchy convergence in  $L^2$ .  $\square$

This is the main point, because if we know the final time is Cauchy in  $L^2$ , we can use the maximal inequality to get convergence of the  $M$  process. We'll finish up the proof next time.  $\square$

## 10 March 4, 2020

We'll continue the proof that we started last time: the idea is that if we have a continuous local martingale  $M_t$ , we want to decompose

$$M_t^2 = (\text{continuous local martingale}) + (\text{finite variation process}).$$

We know that this is unique, because any continuous local martingale that is of finite variation must be identically zero. Continuing from last time, our goal is to show that (1) there exists a finite variation process  $A_t = \langle M, M \rangle_t$  such that (2)  $M_t^2 - A_t$  is a local martingale, and (3) if  $P_n$  is an increasing subdivision of  $[0, t]$ , we have

$$\sum_{i=1}^{|P_n|} \left( M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}} \right)^2 \xrightarrow{P} \langle M, M \rangle_t,$$

which we call the quadratic variation of  $M$ .

*Continuation of proof.* Remember that we're proving this assuming that  $M$  is bounded and that  $M_0 = 0$ . For any subdivision of  $[0, T]$ , we decompose  $M_t^2$  (for any  $t \in [0, T]$ ) as

$$M_t^2 = \sum_{i=1}^{|P_n|} \left( M_{t \wedge t_i^{(n)}}^2 - M_{t \wedge t_{i-1}^{(n)}}^2 \right),$$

which can then further be decomposed as

$$M_t^2 = \sum_{i=1}^{|P_n|} \left( M_{t \wedge t_i^{(n)}} - M_{t \wedge t_{i-1}^{(n)}} \right)^2 + 2 \sum_{i=1}^{|P_n|} M_{t \wedge t_{i-1}^{(n)}} \left( M_{t \wedge t_i^{(n)}} - M_{t \wedge t_{i-1}^{(n)}} \right).$$

The first term, which we denote  $A_t^{(n)}$ , is supposed to be our finite variation process, and we showed last time that the second term, which we denote  $N_t^{(n)}$ , is a martingale. We showed last time that  $N_T^{(n)}$  is Cauchy in  $L^2$  for any finite  $T$ , as long as  $M$  is bounded.

So now by Doob's  $L^2$  inequality,

$$\sup_{0 \leq t \leq T} \left\| N_t^{(n)} - N_t^{(m)} \right\|_2 \leq C \left\| N_T^{(n)} - N_T^{(m)} \right\|_2,$$

which goes to 0 as  $m, n \rightarrow \infty$ . Because we have a martingale, having control of the endpoint gives us control of the entire process, and now we have a kind of uniform convergence: we can find  $n_k \rightarrow \infty$  (extract a subsequence) such that

$$\sup_{0 \leq t \leq T} \left\| N_t^{(n_k)} - N_t^{(n_{k+1})} \right\|_2 \leq \frac{1}{2^k},$$

so the expected value can be bounded by the  $L^2$  norm via

$$\mathbb{E} \left[ \sum_{k=1}^{\infty} \sup_{0 \leq t \leq T} \left| N_t^{(n_k)} - N_t^{(n_{k+1})} \right| \right] < \infty.$$

Since the expectation is finite, the quantity inside the absolute value is finite almost surely. Thus the  $N^{(n_k)}$  converge uniformly on  $[0, T]$  outside of a null set  $\Omega_0$ : let  $Y_t(\omega)$  be the limit

$$Y_t(\omega) = \begin{cases} \lim_k N_t^{(n_k)}(\omega) & \omega \notin \Omega_0 \\ 0 & \text{otherwise.} \end{cases}$$

This process is adapted to  $\mathcal{F}_t$ , and we just need to check that  $Y_t$  is a martingale. Since  $N_t^{(n_k)}$  converges to  $Y_t$  almost surely and in  $L^2$ , we indeed have

$$\mathbb{E}(Y_t | \mathcal{F}_s) = Y_s$$

by taking the limit of the  $N_t^{(n_k)}$  statements and passing the integral through.

Now we have our process

$$A_t = M_t^2 - Y_t;$$

we know this can be obtained by taking the (uniform) limit (on  $[0, T]$ ) of the  $A_t^{(n)}$ s, because  $M_t^2$  is fixed and  $N_t^{(n)}$  converges to  $Y_t$  in the boxed equation above. It's not necessarily true that  $A_t^{(n)}$  is an increasing process, but  $A^{(n)}$  is nondecreasing on the set  $\{t_i^{(n)}\}$ . Taking the limit, our set of  $t_i$ s is dense, so continuity tells us that  $A$  is indeed a nondecreasing process on  $[0, T]$ .

Finally, we repeat this process for all integers  $T \geq 1$ : this gives us a collection of processes

$$M_t^2 = A_t^{(T)} + Y_t^{(T)}$$

holding for all  $t \leq T$ . We just need to show that all of these are compatible of each other – this follows from the uniqueness point we made earlier. In particular, the stopped process

$$M_{t \wedge T}^2 - A_{t \wedge T}^{(T)}$$



is a martingale, and similarly if  $T' \geq T$ , we have that

$$M_{t \wedge T}^2 - A_{t \wedge T}^{(T')}$$

is also a martingale. So if we subtract these,  $A_t^{(T)}$  and  $A_t^{(T')}$  must agree up to time  $T$ , which is what we want.

It remains to check the final point: note that

$$\sum_{i=1}^{|P_n|} \left( M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}} \right)^2 = A_T^{(n)} = M_T^2 - N_T^{(n)}.$$

This converges in  $L^2$  to  $M_T^2 - Y_T$ , which is exactly  $A_T$ , as desired.  $\square$

Everything we've done has proven the theorem in the case where  $M$  is bounded and  $M_0 = 0$ . Extending to the general case is easy, and we can read this on our own.

### Example 85

Let  $B$  be a standard Brownian motion. We've shown that  $B_t^2 - t$  is a martingale, so  $\langle B, B \rangle_t = t$ .

In a future lecture, we'll see the converse as well, which tells us that a continuous local martingale with  $\langle B, B \rangle_t$  is a Brownian motion. In fact, if  $M_t$  is both a continuous martingale and a Gaussian process, then  $M_t^2 - \mathbb{E}(M_t^2)$  is a martingale, so we have  $\langle M, M \rangle_t = \mathbb{E}(M_t^2)$ . In both of these cases, the quadratic variation is deterministic, but in general it can be random.

### Theorem 86

Suppose that  $M$  is a continuous local martingale, and  $M_0 \in L^2$ . Then the following are equivalent:

- $M$  is a true martingale bounded in  $L^2$ .
- $\mathbb{E}\langle M, M \rangle_\infty$  is finite.

If these properties hold, then  $M_t^2 - \langle M, M \rangle_t$  is also a true, uniformly integrable martingale. (Without loss of generality, we can assume  $M_0 = 0$ .)

It is natural to try saying that  $M_t^2 - \langle M, M \rangle_t$  is a martingale, so  $\mathbb{E}(M_t^2) = \mathbb{E}(\langle M, M \rangle_t)$ . The main problem is that  $M_t^2 - \langle M, M \rangle_t$  is not actually a martingale, so we need to fix that.

### Lemma 87

If  $M$  is a continuous local martingale such that  $|M_t| \leq Z \in L^1$  for all  $t$ , then  $M$  is a uniformly integrable martingale.

*Proof of lemma.* By definition, there exist stopping times  $\tau_n$  such that  $M_{t \wedge \tau_n}$  is a uniformly integrable martingale. Thus, we have

$$M_{s \wedge \tau_n} = \mathbb{E}(M_{t \wedge \tau_n} | \mathcal{F}_s),$$

and since  $M_{t \wedge \tau_n}$  is a continuous process, this converges to  $M_t$  as  $n \rightarrow \infty$ . By assumption,  $|M_t|$  is dominated by  $Z$ , so the collection  $M_{t \wedge \tau_n}$  are uniformly integrable (because they are all dominated). Thus  $M_{t \wedge \tau_n}$  converges in  $L^1$  to  $M_t$ , so we can indeed say that  $M_s = \mathbb{E}(M_t | \mathcal{F}_s)$  by the dominated convergence theorem. It now remains to show that  $M_t$  is uniformly integrable, but this is true because  $|M_t| \leq Z$ .  $\square$

We're now ready to prove the theorem above:

*Proof.* It's okay to say  $M_0 = 0$  without loss of generality. To show the forward direction, note that the Doob  $L^2$  inequality tells us that  $S = \sup_{t \geq 0} |M_t|$  is in  $L^2$ . Define the stopping time

$$\sigma_n = \inf\{t : \langle M, M \rangle_t = n\};$$

then  $Y_t = M_t^2 - \langle M, M \rangle_t$  is a local martingale, so the stopped version  $Y_{t \wedge \sigma_n}$  is also a local martingale. But

$$|Y_{t \wedge \sigma_n}| = |M_{t \wedge \sigma_n}^2 - \langle M, M \rangle_{t \wedge \sigma_n}| \leq S^2 + n \in L^1$$

(because  $S \in L^2$ ). Therefore, the stopped process  $Y_{t \wedge \sigma_n}$  is a uniformly integrable martingale by the previous lemma (dominated by  $Z = S^2 + n$ ). This means that

$$\mathbb{E}M_{t \wedge \tau_n}^2 = \mathbb{E}\langle M, M \rangle_{t \wedge \sigma_n}.$$

The left hand side is finite – it's bounded by  $\mathbb{E}(S^2)$  by assumption (because  $M$  is a true martingale). By the monotone convergence theorem, the right hand side increases to  $\mathbb{E}\langle M, M \rangle_t$ . And we have the same finite bound for all  $t$ , so taking  $t \rightarrow \infty$  finishes this direction of the proof.

On the other hand, assume the total quadratic variation is finite. We define the stopping times

$$\tau_n = \inf\{t : |M_t| = n\}.$$

Then  $Y_t = M_t^2 - \langle M, M \rangle_t$  is a local martingale, so  $Y_{t \wedge \tau_n}$  is also a local martingale. And now

$$|Y_{t \wedge \tau_n}| = |M_{t \wedge \tau_n}^2 - \langle M, M \rangle_{t \wedge \tau_n}| \leq n^2 + \langle M, M \rangle_\infty \in L^1$$

by assumption, so  $Y_{t \wedge \tau_n}$  is a uniformly integrable martingale, again by the above lemma. This means that

$$\mathbb{E}(M_{t \wedge \tau_n}^2) = \mathbb{E}\langle M, M \rangle_{t \wedge \tau_n} \leq \mathbb{E}\langle M, M \rangle_\infty < \infty,$$

so  $M$  is bounded in  $L^2$ . It remains to show that we have a martingale: we already know that

$$\mathbb{E}(M_{t \wedge \tau_n} | \mathcal{F}_s) = M_{s \wedge \tau_n},$$

For each  $t$ , the collection  $M_{t \wedge \tau_n}$  is bounded in  $L^2$ , and this means it is uniformly integrable. (We'll prove this below.) Thus,  $M_{t \wedge \tau_n}$  converges to  $M_t$ , and uniform integrability lets us pass the limit through the integral to find that

$$\mathbb{E}(M_t | \mathcal{F}_s) = M_s,$$

as desired. □

The fact that remains to be proved is the following:

**Proposition 88**

If  $\{X_i\}$  are bounded in  $L^p$  for some  $p > 1$ , then  $\{X_i\}$  is uniformly integrable.

*Proof.* Let  $q = \frac{p}{p-1}$ , so that  $\frac{1}{p} + \frac{1}{q} = 1$ . By Hölder's inequality,

$$\mathbb{E}(|X_i|; |X_i| \geq M) \leq \|X_i\|_p \|1\{|X_i| \geq M}\|_q.$$

But the  $\|X_i\|_p$  are bounded by some constant  $C$ , and the second term is just  $\mathbb{P}(|X_i| \geq M)^{1/q}$ . By Markov's inequality,

we can therefore simplify everything to

$$\leq C \left( \frac{\mathbb{E}(|X_i|^p)}{M^p} \right)^{1/q} \leq \frac{C \cdot C^{p/q}}{M^{p/q}}.$$

Taking  $M \rightarrow \infty$ , this goes to 0 uniformly over  $i$ , which means we've shown the desired uniform integrability condition.  $\square$

We'll finish by wrapping up some minor points from chapter 4.

**Definition 89**

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  be a filtered probability space. If  $M, N$  are both continuous local martingales, then

$$\langle M, N \rangle_t = \frac{1}{2} [\langle M + N, M + N \rangle_t - \langle M, M \rangle_t - \langle N, N \rangle_t].$$

(This is sometimes called the quadratic covariation.)  $M, N$  are **orthogonal** if their covariation is zero.

For example, if we take two independent Brownian motions  $B_1, B_2$ ,  $\langle B_1, B_2 \rangle = 0$ . (This is just by plugging in quantities into the definition.)

**Theorem 90 (Kunita-Watanabe)**

Let  $M, N$  be continuous local martingales, and let  $H, K$  be **measurable** (that is,  $(t, \omega) \rightarrow H_t(\omega)$  is measurable on  $\mathcal{F} \otimes \mathcal{B}_{[0, \infty)}$ ). Then

$$\int_0^\infty |H_s K_s| |d\langle M, N \rangle_s| \leq \left( \int_0^\infty H_s^2 d\langle M, M \rangle_s \right)^{1/2} \left( \int_0^\infty K_s^2 d\langle N, N \rangle_s \right)^{1/2}.$$

(An adapted continuous process is sufficient for  $H$  and  $K$ .) This is a kind of analogy to the Cauchy-Schwarz inequality. We won't go through the proof here, but the main idea is that the left hand side looks like

$$\sum_{i=1}^{|P_n|} \left| H_{t_{i-1}^{(n)}} K_{t_{i-1}^{(n)}} (M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}}) (N_{t_i^{(n)}} - N_{t_{i-1}^{(n)}}) \right|$$

as  $n \rightarrow \infty$ . And by normal Cauchy-Schwarz, this is bounded by

$$\left( \sum_i H_{t_{i-1}^{(n)}}^2 (M_{t_i^{(n)}} - M_{t_{i-1}^{(n)}})^2 \right)^{1/2} \left( \sum_j K_{t_{j-1}^{(n)}}^2 (N_{t_j^{(n)}} - N_{t_{j-1}^{(n)}})^2 \right)^{1/2}$$

which converges to the right hand side.

**Definition 91**

A process  $X_t = M_t + A_t$ , where  $M_t$  is a continuous local martingale and  $A_t$  is a finite variation process, is a **continuous semimartingale**. If we have two such processes  $X_t = M_t + A_t$  and  $Y_t = M'_t + A'_t$  on the same space, then the **covariation**  $\langle X, Y \rangle_t$  is defined to be  $\langle M, M' \rangle_t$ .

This definition makes sense, because a subdivision  $P_n$  makes the sum

$$\sum_{i=1}^n (X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}})(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}})$$

converge to  $\langle M, M' \rangle_t$ . (The finite variation doesn't contribute to the covariation or quadratic variation.)

# 11 March 9, 2020

We have an exam on Thursday evening in 2-449. (This is possibly the last thing we'll do in person before we get quarantined, but the class is not too big so this is probably okay.)

Today, we'll start discussing **stochastic integration** (from chapter 5). We're going to start seeing nice applications of the theory we've been developing, and we'll start with a bit of review. Let  $M_t$  be a local martingale: recall that if we subtract off the quadratic variation from  $M^2$  to get

$$M_t^2 - \langle M, M \rangle_t,$$

yields a local martingale, and here  $\langle M, M \rangle_t$  is an increasing (finite variation) process. We also showed that if  $M_0 \in L^2$ , then  $M$  is bounded in  $L^2$  if and only if  $\mathbb{E}(\langle M, M \rangle_\infty)$  is finite. Then in this case,  $M_t^2 - \langle M, M \rangle_t$  is a local martingale, and in fact it is also a uniformly integrable martingale because

$$\sup_t |M_t^2 - \langle M, M \rangle_t| \leq (\sup_t |M_t|^2) + \langle M, M \rangle_\infty,$$

where the first term is integral by Doob's  $L^2$  inequality and the second term is finite by assumption. (We wouldn't be asked to show Doob's  $L^2$  inequality on the exam, but we would be expected to be able to reproduce the above argument.)

At the end of last lecture, we defined the bracket

$$\langle M, N \rangle = \frac{1}{2} [\langle M + N, M + N \rangle - \langle M, M \rangle - \langle N, N \rangle].$$

In particular, this tells us that we can calculate

$$M_t N_t - \langle M, N \rangle_t = \frac{1}{2} (M_t + N_t)^2 - \langle M + N, M + N \rangle_t - \frac{1}{2} (M_t^2 - \langle M, M \rangle_t) - \frac{1}{2} (N_t^2 - \langle N, N \rangle_t)$$

where each term on the right is a local martingale, so the left side is also a local martingale. In addition, if  $M, N$  are bounded in  $L^2$ , so is  $M + N$ . Thus, all three terms on the right side are uniformly integrable martingales, so  $M_t N_t - \langle M, N \rangle_t$  is a uniformly integrable martingale as well, and in particular this means

$$\mathbb{E}[M_\infty N_\infty] = \mathbb{E}\langle M, N \rangle_\infty$$

if  $M_0 = N_0 = 0$ .

Our goal is to take a semimartingale  $X_t = A_t + M_t$  and a class of processes  $H_t$ , and to define the stochastic integral

$$(H \cdot X)_t = (H \cdot A)_t + (H \cdot M)_t = \int_0^t H_s dA_s + \int_0^t H_s dM_s.$$

We've already seen how to compute the first integral:  $A$  corresponds to a signed measure, so this is just the Lebesgue integral of  $H_s$  against that signed measure. The notation suggests that this should be a continuous-time version of the Doob transform: we defined (to prove the optional stopping theorem)

$$(H \cdot M)_n = \sum_{i=1}^n H_i (M_i - M_{i-1}).$$

In particular, in discrete time, this object is a martingale for appropriate  $H$ , and the continuous time version  $\int_0^t H_s dM_s$  will turn out to be a local martingale.

The reason we talk about everything in  $L^2$  is that this is an important case: we'll define the stochastic integral for

martingales bounded in  $L^2$  today, and extending to local martingales will be done next time.

**Definition 92**

On a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ , let  $\mathbb{H}^2$  denote the space of  $L^2$ -bounded martingales.

On this space  $\mathbb{H}^2$ , we have the useful identity from above: since  $\mathbb{E}(M_\infty N_\infty) = \mathbb{E}\langle M, N \rangle_\infty$ , we will define the **scalar product on  $\mathbb{H}^2$**  via

$$(M, N)_{\mathbb{H}^2} = \mathbb{E}(M_\infty N_\infty).$$

In particular, if  $\mathbb{E}\langle M, M \rangle = 0$ , then  $M$  is identically zero, so this is indeed a norm.  $\mathbb{H}^2$  is a **Hilbert space** – to prove this, we need to check that it's complete, which means a Cauchy sequence (with respect to this norm) converges. We won't do this in much detail because it's similar to what we've seen, but if we have

$$\lim_{m, n \rightarrow \infty} \|M^n - M^m\|_{\mathbb{H}^2} \rightarrow 0,$$

we can use the Doob  $L^2$  inequality to show uniform convergence of the  $M^n$ s.

Recall that a process  $H_t$  is **progressive** if the function  $F(\omega, t) = H_t(\omega)$  restricted to  $\Omega \times [0, t]$  is measurable with respect to  $\mathcal{F}_t \otimes \mathcal{B}_{[0, t]}$ . An equivalent characterization is to say that  $F$  is measurable with respect to  $\mathcal{P}$ , where  $A \in \mathcal{P}$  if and only if  $H_t(\omega) = 1\{(\omega, t) \in A\}$  is progressive. The point here is that there is a sigma-field  $\mathcal{P}$  that is equivalent to being progressive – it's not defined in a very useful way, though.

**Definition 93**

For an  $L^2$ -bounded martingale  $M \in \mathbb{H}^2$ , let  $L^2(M)$  be the space of all progressive processes  $H$  such that

$$\mathbb{E} \left[ \int_0^\infty H_t^2 d\langle M, M \rangle_t \right] < \infty.$$

Remember that  $\langle M, M \rangle_t$  has finite variation, so the inner integral is just a Lebesgue integral. Here,  $L^2(M)$  is equivalent to a standard  $L^2$  space

$$L^2(M) = L^2(\Omega \times [0, \infty), \mathcal{P}, \nu)$$

where the measure  $\nu$  is defined as

$$\nu(A) = \mathbb{E} \left[ \int_0^\infty 1\{(\omega, t) \in A\} d\langle M, M \rangle_t \right] < \infty.$$

This means that  $L^2(M)$  is also a Hilbert space with all of its nice properties, so we won't need to prove those. And this means we can define the inner product

$$(H, K)_{L^2(M)} = \mathbb{E} \left[ \int_0^\infty H_t K_t d\langle M, M \rangle_t \right].$$

From here, the idea is that for any  $M \in \mathbb{H}^2$ , we will define the stochastic integral with respect to  $M$  as an  $L^2$  isometry

$$\mathcal{J}^M : L^2(M) \rightarrow \mathbb{H}^2,$$

where we send a martingale

$$H \mapsto \mathcal{J}^M(H) = H \cdot M = \left( \int_0^t H_s dM_s \right)_{t \geq 0}.$$

It's helpful to write out what the isometry is directly: we should preserve scalar products, and thus we need

$$(H, K)_{L^2M} = \mathbb{E} \left[ \int_0^\infty H_t K_t d\langle M, M \rangle_t \right] = (H \cdot M, K \cdot M)_{\mathbb{H}^2} = \mathbb{E} [(H \cdot M)_\infty (K \cdot M)_\infty]$$

which can also be written as

$$= \mathbb{E} \left[ \left( \int_0^\infty H_t dM_t \right) \left( \int_0^\infty K_t dM_t \right) \right].$$

In particular, if  $H = K$ , we get what we call the **Itô isometry**.

To define such an isometry, we can define that isometry on a dense subspace of  $L^2(M)$  and extend by continuity to the entire space:

**Definition 94**

Let  $\mathcal{E}$  be the space of **elementary processes** of the form

$$H_t = \sum_{i=1}^p H_{(i)} 1\{t \in (t_i, t_{i+1}]\},$$

where each  $H_{(i)} \in \mathcal{F}_{t_i}$  is almost surely bounded by some constant.

These processes are pretty simple: we take deterministic times, and for each time interval we put a (measurable) random variable there. We definitely have  $\mathcal{E} \subseteq L^2(M)$  because of the boundedness condition, and note that  $\mathcal{E}$  makes no reference to  $M$  at all.

**Proposition 95**

$\mathcal{E}$  is dense in  $L^2(M)$  (with respect to the scalar product on  $L^2(M)$ ).

*Proof.* We've seen that simple functions are dense in  $L^2$ ; this is a bit more complicated because of the sigma-algebra. It's enough to verify that if  $K \in L^2(M)$  with  $K \perp \mathcal{E}$ , then  $K = 0$ . (Then the usual Hilbert space theory gives us denseness.)

If  $K$  is orthogonal to  $\mathcal{E}$ , this means that for all  $H \in \mathcal{E}$ ,

$$0 = (H, K)_{L^2(M)} = \mathbb{E} \left[ \int_0^\infty H_t K_t d\langle M, M \rangle_t \right].$$

We will deduce that  $X_t = \int_0^t K_s d\langle M, M \rangle_s$  is identically zero. We claim this variable  $X$  is well-defined as a finite-variation process. For this, we just need absolute integrability: we need to consider the quantity

$$\mathbb{E} \left[ \int_0^\infty |K_s| d\langle M, M \rangle_s \right],$$

and by Cauchy-Schwarz, this is at most

$$\leq \sqrt{\mathbb{E} \left[ \int_0^\infty |K_t|^2 d\langle M, M \rangle_t \right] \mathbb{E} [1 \cdot d\langle M, M \rangle_t]} < \infty,$$

where we know the first term is finite because of the inner product on  $L^2(M)$ , and second term because  $M$  is in  $\mathbb{H}^2$ . And this also tells us that  $X_t$  is in  $L^1$  for all  $t$ , because the left hand side here upper bounds  $|X_t|$ . Pick the process

$$H_r(\omega) = F_{(s)} 1\{r \in (s, t]\},$$

where  $F_{(s)}$  is a bounded random variable that is measurable with respect to  $\mathcal{F}_s$ . We've assumed that  $K \perp \mathcal{E}$ , so

$$0 = (H, K)_{L^2(M)} = \mathbb{E} \left[ F_{(s)} \int_s^t K_r d\langle M, M \rangle_r \right].$$

But now the integral is the same as the difference between  $X_t$  and  $X_s$ , so this is

$$= \mathbb{E}[F_{(s)}(X_t - X_s)].$$

We know that  $X_t \in L^1$ , so this last calculation tells us that  $X$  is a martingale. Since it is also a finite variation process, this means  $X = 0$ .

And now  $K = 0$  **as an element of**  $L^2(M)$ : in other words, knowing  $X_t$  is identically zero means we only care about  $K$  being zero except on a set of measure zero with respect to  $L^2(M)$ .  $\square$

### Theorem 96

Let  $M \in \mathbb{H}^2$  be an  $L^2$ -bounded martingale. For  $H \in \mathcal{E}$ , where  $H_t = \sum_{i=1}^p H_{(i)} 1\{t \in (t_i, t_{i+1}]\}$ , define

$$\mathcal{J}^M(H)_t = \sum_{i=1}^p H_{(i)} (M_{t \wedge t_{i+1}} - M_{t \wedge t_i}).$$

Then  $\mathcal{J}^M$  defines an isometry from  $\mathcal{E}$  (with the  $L^2(M)$  scalar product) into  $\mathbb{H}^2$ . Therefore,  $\mathcal{J}^M$  extends to an isometry from  $L^2(M)$  into  $\mathbb{H}^2$ .

(The idea here is that integrating  $1dM_t$  should give us back the original martingale, so integrating elementary processes gives us increments of the martingale.)

*Proof.* We know that  $\mathcal{J}^M(H)$  is in  $\mathbb{H}^2$  (because  $H_{(i)}$  are bounded and the stopped  $M$ s are in  $\mathbb{H}^2$ ), and the quadratic variation process

$$\langle \mathcal{J}^M(H), \mathcal{J}^M(H) \rangle_t = \sum_{i=1}^p H_{(i)}^2 (\langle M, M \rangle_{t \wedge t_{i+1}} - \langle M, M \rangle_{t \wedge t_i}).$$

(This is only easy because  $H$  takes on a very simple form.) But this sum is also equal to

$$= \int_0^t H_s^2 d\langle M, M \rangle_s$$

by the definition of the Lebesgue integral for this simple-like function. So now we can check the isometry property on  $\mathcal{E}$ :

$$(\mathcal{J}^M(H), \mathcal{J}^M(H))_{\mathbb{H}^2} = \mathbb{E} \left[ \sum_i H_{(i)}^2 (\langle M, M \rangle_{t_{i+1}} - \langle M, M \rangle_{t_i}) \right]$$

by the definition of the  $\mathbb{H}^2$  scalar product, but by our observation above, this is also equal to (setting  $t = \infty$ )

$$= \mathbb{E} \left[ \int_0^\infty H_s^2 d\langle M, M \rangle_s \right] = (H, H)_{L^2(M)}.$$

So  $\mathcal{J}^M$  is an isometry on  $\mathcal{E}$ , and we can extend this to  $L^2(M)$  by continuity.  $\square$

**Remark 97.** We should have checked at some point that  $\mathcal{J}^M(H)$  doesn't depend on the representation of  $H$ : currently,  $H$  is written down in an explicit form, but there might be different ways to do so. But if there were two different representations, we could consider the identity

$$\|\mathcal{J}^M(H - H')\|_{\mathbb{H}^2} = \|H - H'\|_{L^2(M)},$$

and this is 0 if  $H$  and  $H'$  are really the same function.

We'll spend the rest of the time on a useful result for Itô's formula: "the stochastic integral commutes with the bracket."

**Proposition 98**

Let  $M \in \mathbb{H}^2$ , and let  $H \in L^2(M)$ . Then  $H \cdot M$  is the unique element of  $\mathbb{H}^2$  such that  $\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$  for all  $N \in \mathbb{H}^2$ .

In particular, if we apply this twice, we find that

$$\langle H \cdot M, K \cdot N \rangle = HK \cdot \langle M, N \rangle$$

as long as everything is well-defined, meaning  $M, N \in \mathbb{H}^2$ ,  $H \in L^2(M)$ ,  $K \in L^2(N)$ . This can be written as

$$\left\langle \int_0^t H_s dM_s, \int_0^t K_s dN_s \right\rangle = \int_0^t H_s K_s d\langle M, N \rangle_s,$$

which we'll see soon.

*Proof.* The key step is to show the identity at  $t = \infty$  for an elementary process  $H \in \mathcal{E}$ . For this case, we know that

$$(H \cdot M)_t = \sum_i H_{(i)} (M_{t \wedge t_{i+1}} - M_{t \wedge t_i}),$$

where each term is a martingale. Then the covariation of this with  $N$  is just

$$\langle H \cdot M, N \rangle_\infty = \sum_i H_{(i)} (\langle M, N \rangle_{t_{i+1}} - \langle M, N \rangle_{t_i}) = \int_0^\infty H_t d\langle M, N \rangle_t = (H \cdot \langle M, N \rangle)_\infty.$$

Now we want to approximate

$$\langle H \cdot M, N \rangle = \lim_{n \rightarrow \infty} \langle H^n \cdot M, N \rangle = \lim_{n \rightarrow \infty} (H^n \cdot \langle M, N \rangle)_\infty = H \cdot \langle M, N \rangle,$$

where  $H^n \in \mathcal{E}$  – we just need to justify the limits on the left and right. We proved the Kunita-Watanabe inequality last time, which says that

$$\left| \int H_s K_s d\langle M, N \rangle_s \right| \leq \left| \int H_s^2 d\langle M, M \rangle_s \right|^{1/2} \left| \int K_s^2 d\langle N, N \rangle_s \right|^{1/2}.$$

This means that

$$\mathbb{E}[\langle X, N \rangle] \leq \mathbb{E}[\langle X, N \rangle \langle X, N \rangle]^{1/2} \leq \mathbb{E}[\langle X, X \rangle]^{1/2} \mathbb{E}[\langle N, N \rangle]^{1/2}$$

using normal Cauchy-Schwarz, and this is  $\|X\|_{\mathbb{H}^2} \|N\|_{\mathbb{H}^2}$ . This means that taking  $X$  and taking its quadratic covariation with  $N$  is continuous with respect to the  $\mathbb{H}^2$  norm, so if we let  $X = (H - H^n) \cdot M$ , we know that  $H^n \cdot M$  converges to  $H \cdot M$  in the  $\mathbb{H}^2$  metric, and thus we've justified the left limit. **(Here,  $X$  is in  $\mathbb{H}^2$ .)**

For the right limit, consider

$$\mathbb{E}[(X \cdot \langle M, N \rangle)_\infty].$$

Apply Kunita-Watanabe and Cauchy-Schwarz again with  $H = X$  and  $K = 1$  to find

$$\leq \|X\|_{L^2(M)} \|N\|_{\mathbb{H}^2}.$$

Similarly, we take  $X = H - H^n$ , which goes to 0 in  $L^2$ . **(Here,  $X$  is in  $L^2(M)$ .)** □



So all of the properties are easy to prove when  $H$  is an elementary processes – we'll generalize next time!

## 12 March 11, 2020

This will be the last class we have in person – we'll continue after spring break on videoconference. If we don't have efficiently high-speed internet, we should let Professor Sun know by email.

We started talking about stochastic integration last time: if  $\mathbb{H}^2$  denotes the (Hilbert) space of  $L^2$ -bounded martingales, we defined a scalar product

$$(M, N)_{\mathbb{H}^2} = \mathbb{E}(M_{\infty}N_{\infty}) = \mathbb{E}\langle M, N \rangle_{\infty}.$$

We also defined the (Hilbert) spaces

$$L^2(M) = \left\{ \text{progressive } H : \|H\|_{L^2(M)} = \mathbb{E} \left[ \int_0^{\infty} H_t^2 d\langle M, M \rangle_t \right] < \infty \right\}.$$

Then for any  $M \in \mathbb{H}^2$ , we can define the stochastic integral, which is an isometry  $\mathcal{J}^M$  from  $L^2(M) \rightarrow \mathbb{H}^2$ , sending  $H$  to  $H \cdot M$ . One property of  $H \cdot M$  is that it is the unique element in  $\mathbb{H}^2$  such that

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle \quad \forall N \in \mathbb{H}^2. \quad (1)$$

Today, and from now on, we'll use the notation

$$\langle M \rangle = \langle M, M \rangle.$$

Our goal now is to extend stochastic integration to local martingales (and therefore to semimartingales). We'll start with a few remarks: if  $M$  is a local martingale and  $\tau$  is a stopping time, then the stopped process  $M_t^{\tau} = M_{t \wedge \tau}$  is also a local martingale (this follows straightforwardly from the definitions). Then  $M_t^{\tau} - \langle M \rangle_t$  is a local martingale, so plugging in the stopped process, we see that

$$(M^{\tau})_t^2 - \langle M \rangle_{t \wedge \tau}$$

is also a local martingale. But by definition, there is a unique process that yields a local martingale when we subtract off from  $(M^{\tau})$ , so we've now found that

$$\langle M^{\tau} \rangle = \langle M \rangle_{t \wedge \tau} \implies \langle M^{\tau} \rangle = \langle M \rangle^{\tau}.$$

Also, if  $M, N$  are both local martingales, then

$$\langle M^{\tau}, N^{\tau} \rangle = \langle M, N \rangle^{\tau} = \langle M, N \rangle^{\tau}. \quad (2)$$

(This last fact is a bit harder to prove, but we can read it on our own.)

### Lemma 99

If  $M \in \mathbb{H}^2$  and  $H \in L^2(M)$ , then for all stopping times  $\tau$ , we have

$$(H \cdot M)^{\tau} = (H \cdot 1_{[0, \tau]}) \cdot M = H \cdot (M^{\tau}).$$

*Proof.* First, apply (Eq. (2)): note that for any  $N \in \mathbb{H}^2$ , we have

$$\langle (H \cdot M)^{\tau}, N \rangle_t = \langle H \cdot M, N \rangle_{t \wedge \tau} = \langle H \cdot \langle M, N \rangle \rangle_{t \wedge \tau}.$$

(last equality by Eq. (1)). But this is just the integral of  $H$  against a finite-variation process, stopped at some time  $\tau$ , so we understand this: it is indeed just

$$= ((H \cdot 1_{[0,\tau]}) \cdot \langle M, N \rangle)_t.$$

Thus,  $(H \cdot M)^\tau$  satisfies the characterizing property Eq. (1) for  $(H \cdot 1_{[0,\tau]}) \cdot M$ , so we've shown the first identity.

For the second equality, we note that

$$\langle H \cdot (M^\tau), N \rangle = H \cdot \langle M^\tau, N \rangle = H \cdot (\langle M, N \rangle^\tau).$$

Again, this is the integral of  $H$  against a finite-variation process, which is just

$$((H \cdot 1_{[0,\tau]}) \cdot \langle M, N \rangle)_t,$$

which is what we want (we've shown the characterizing property again). □

To extend our definition, we'll need to be a bit more general:

**Definition 100**

Let  $M$  be a local martingale, and keep the definition  $L^2(M)$  from before. Then define the (larger) space

$$L^2_{\text{loc}}(M) = \left\{ \text{progressive } H : \int_0^t H_s^2 d\langle M \rangle_s < \infty \forall \text{ finite } t, \text{ a.s.} \right\}.$$

For  $L^2(M)$ , we integrate from 0 to  $\infty$  and require the expectation to be finite, but here we only need the integral to be almost surely finite. The main goal of today is to prove the following result:

**Theorem 101**

Let  $M$  be a local martingale and let  $H \in L^2_{\text{loc}}(M)$ . Then there exists a unique local martingale  $H \cdot M$  with initial value 0, such that

$$\langle H \cdot M, N \rangle = H \cdot \langle M \cdot N \rangle$$

for all local martingales  $N$ .

(Then it makes sense to denote  $(H \cdot M)_t = \left( \int_0^t H_s dM_s \right)_t$ .)

*Proof.* Let  $M_0 = 0$  without loss of generality – nothing depends on the initial value. We'll start by talking about how to construct this process: we want to extend our definition from the  $L^2$ -bounded case, so we define the stopping times

$$\tau_n = \inf \left\{ t \geq 0 : \int_0^t (1 + H_s^2) d\langle M \rangle_s \geq n \right\}.$$

We can check that  $\tau_n$  goes to  $\infty$  almost surely –  $\langle M \rangle_s$  is a local martingale, we assume  $H^2$  is in  $L^2_{\text{loc}}(M)$ , and  $M$  is a local martingale, so neither term in the integrand gets large too fast. Furthermore, we know that  $M^{\tau_n} \in \mathbb{H}^2$ , because the total quadratic variation by time  $\infty$  will be at most  $n$  almost surely. Thus, it has finite expectation and will definitely be in  $\mathbb{H}^2$ . Similarly,  $H \in L^2(M^{\tau_n})$ , because by time  $\tau_n$ , the integral  $\int H_s^2 d\langle M \rangle_s$  is bounded by  $n$ .

This implies immediately that

$$X^n = H \cdot (M^{\tau_n})$$

is well-defined and is an element of  $\mathbb{H}^2$ . Clearly, we want to send  $n \rightarrow \infty$ , and thus for  $m > n$ , we define the variables

$$(X^m)^{\tau_n} = (H \cdot M^{\tau_m})^{\tau_n}$$

to ensure that  $X^m$  and  $X^n$  are **consistently** defined. By Lemma 99, this is the same as  $H \cdot ((M^{\tau_m})^{\tau_n})$ , but we know  $\tau_n < \tau_m$ , so this is just  $H \cdot M^{\tau_n}$ . So the various  $X^n$  are consistent, meaning that there is a single  $X$  such that  $X^{\tau_n} = X^n$  for all  $n$ . This  $X$  is a local martingale, because  $X^{\tau_n}$  is the same as  $X^n$ , which is an  $L^2$ -bounded martingale, which means it is uniformly integrable.

Next, we need to check that this  $X = H \cdot M$  satisfies the desired property  $\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$ . Assume without loss of generality that  $N_0 = 0$ . We'll basically use the fact that the desired property holds at our stopping times – let

$$\sigma_n = \inf\{t \geq 0 : |N_t| = n\}$$

be a sequence of stopping times for us to control the behavior of the  $N$ s, and let  $\gamma_n = \sigma_n \wedge \tau_n$ . Notice that  $N^{\sigma_n}$  is a local martingale bounded by  $n$ , so it is a true martingale in  $\mathbb{H}^2$ . Then Eq. (2) tells us that

$$\langle H \cdot M, N \rangle^{\gamma_n} = \langle (H \cdot M)^{\tau_n}, N^{\sigma_n} \rangle,$$

and we can simplify this by Lemma 99 to

$$= \langle H \cdot (M^{\tau_n}), N^{\sigma_n} \rangle = H \cdot \langle M^{\tau_n}, N^{\sigma_n} \rangle$$

by Eq. (1), and this is just

$$= H \cdot (\langle M, N \rangle)^{\gamma_n} = (H \cdot \langle M, N \rangle)^{\gamma_n},$$

last step because we have a finite variation process  $\langle M, N \rangle$ . And now taking  $\gamma_n \rightarrow \infty$  yields the desired property for  $X = H \cdot M$ .

Finally, we show that  $X$  is indeed unique: if

$$\langle \tilde{X}, N \rangle = H \cdot \langle M, N \rangle = \langle X, N \rangle,$$

then we know that  $\langle X - \tilde{X}, N \rangle = 0$ , and take  $N = \langle X - \tilde{X} \rangle$  to show that we have a local martingale  $X - \tilde{X}$  with quadratic variation 0, so  $X = \tilde{X}$ .  $\square$

Note now that Lemma 99 also extends to the case where  $M$  is a local martingale and  $H$  is in  $L^2_{\text{loc}}(M)$ , because we've proved the analogous arguments for local martingales now.

We'll now discuss something useful for our next homework: we should recall that if  $M \in \mathbb{H}^2$  and  $H \in L^2(M)$ , we have two nice properties:

$$\mathbb{E} \left[ \int_0^t H_s dM_s \right] = 0,$$

because  $\int_0^t M_s dM_s$  is an  $L^2$ -bounded martingale and thus has expectation 0. We can also calculate the second moment here using the Itô isometry: by definition, we have

$$\mathbb{E} \left[ \int_0^t H_s d\langle M \rangle_s \right] = \|H \cdot \mathbf{1}_{[0,t]}\|_{L^2(M)}^2,$$

and then the isometry tells us that this is also

$$= \|(H \cdot \mathbf{1}_{[0,t]}) \cdot M\|_{\mathbb{H}^2}^2 = \|(H \cdot M)^t\|_{\mathbb{H}^2}^2.$$

by Lemma 99. And remembering that the  $\mathbb{H}^2$  norm is just the expectation of the final value, we find that this is

$$= \mathbb{E}[\langle (H \cdot M)_t \rangle] = \mathbb{E} \left[ \left( \int_0^t H_s dM_s \right)^2 \right].$$

This means we have the first and second moment of the stochastic integral, but **these don't necessarily hold if we're in the general situation where  $M$  is a local martingale**. But they do hold under restricted conditions: if we have

$$\mathbb{E} \left[ \int_0^t H_s^2 d\langle M \rangle_s \right] < \infty,$$

then the statement does hold. This is because  $X = (H \cdot M)^t$  has total quadratic variation

$$\langle X \rangle_\infty = \mathbb{E} \left[ \int_0^t H_s^2 d\langle M \rangle_s \right] < \infty$$

by assumption, so  $X \in \mathbb{H}^2$  as long as we stop the process at time  $t$ .

In general, if we're interested in the second moment of  $\int_0^t H_s dM_s$ , we have the **upper bound**

$$\mathbb{E} \left[ \left( \int_0^t H_s dM_s \right)^2 \right] \leq \mathbb{E} \left[ \int_0^t H_s^2 d\langle M \rangle_s \right]. \quad (3)$$

(If the right hand side is finite, then we have equality; otherwise, this is a vacuous inequality anyway.)

With this, we can also make a definition for semimartingales: if we have such a process  $X = A + M$ , we want to define

$$H \cdot X = H \cdot A + H \cdot M.$$

We want to do this for some reasonably large class of  $H$ -processes – both stochastic integrals must exist.

### Definition 102

A progressive process  $H$  is **locally bounded** if

$$\sup_{s \leq t} |H_s| < \infty$$

for all finite  $t$ .

For any such process, we know that for any  $t$ ,

$$\int_0^t |H_s| |dA_s| \leq \left( \sup_{s \leq t} |H_s| \right) \cdot \int_0^t |dA_s| < \infty,$$

so  $H \cdot A$  is well-defined. Similarly, note that

$$\int_0^t H_s^2 d\langle M \rangle_s < \infty$$

for the same reason, so  $H \in L_{\text{loc}}^2(M)$  for any local martingale  $M$ . Thus,  $H \cdot M$  will be well-defined as well.

This stochastic integral  $H \cdot X$  is now also a semimartingale – it's already written in its canonical decomposition. Specifically,  $H \cdot A$  is the finite variation part, and  $H \cdot M$  is the local martingale part.

We'll use the remaining time to prove a useful convergence property: we'll definitely need a dominated convergence type result for stochastic integrals.

### Proposition 103

Let  $X = A + M$  be a semimartingale, and suppose that  $H^n, H, K$  are locally bounded with  $K \geq 0$ . Suppose that  $H_s^n \rightarrow H_s$  for all  $s \leq t$ , and  $|H_s^n| \leq K_s$  ( $K$  dominates everything), where

$$\int_0^t K_s |dA_s| < \infty, \quad \int_0^t K_s^2 d\langle M \rangle_s < \infty$$

almost surely. Then  $(H^n \cdot X)_t$  converges in probability to  $(H \cdot X)_t$ .

*Proof.* First of all, we have

$$\int_0^t H_s^n dA_s \xrightarrow{\text{a.s.}} \int_0^t H_s dA_s$$

by the usual dominated convergence theorem, because the finite-variation integral is just a Lebesgue integral and everything's dominated by  $K$ . To show the other part, let

$$\tau_k = t \wedge \inf \left\{ r \geq 0 : \int_0^r K_s^2 d\langle M \rangle_s \geq k \right\}.$$

Now Eq. (3) implies that

$$\mathbb{E} \left[ \left( \int_0^{\tau_k} (H_s^n - H_s) dM_s \right)^2 \right] \leq \mathbb{E} \left[ \int_0^{\tau_k} (H_s^n - H_s)^2 d\langle M \rangle_s \right] \rightarrow 0$$

by two applications of the dominated convergence theorem: inside the expectation, the  $H_s$  are both dominated by  $K$ , so we have the integrand dominated by  $4K^2$ . Integrating up to  $\tau_k$  keeps this finite, so the thing inside the integrand goes to 0 almost surely for the integral  $d\langle M \rangle_s$ . Then we want the expectation to converge, and it's dominated by  $4k^2$  because we stop at  $\tau_k$ . So again this expectation will go to 0.

To get the conclusion, note that

$$\mathbb{P} \left( \left| \int_0^t (H_s^n - H_s) dM_s \right| \geq \varepsilon \right) \leq \mathbb{P} \left( \left| \int_0^{\tau_k} (H_s^n - H_s) dM_s \right| \geq \varepsilon \right) + \mathbb{P}(\tau_k \neq t)$$

and this makes sense because  $\tau_k$  is equal to  $t$  for large enough  $k$ . We know the integral in the first term converges in  $L^2$  to 0, so the probability in the first term goes to 0 as  $n \rightarrow \infty$ . And similarly the second term disappears as  $k$  goes to  $\infty$ .  $\square$

### Corollary 104

Let  $X$  be a semimartingale, and let  $H$  be an adapted, continuous process (in particular, this means it will be locally bounded and also progressive). Then given a subdivision  $P_n = (t_i^{(n)})$  subdividing  $[0, t]$  with mesh going to 0, we have

$$\sum_{i=1}^{|P_n|} H_{t_{i-1}^{(n)}} \left( X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) \xrightarrow{P} \int_0^t H_s dX_s.$$

What's important is that if  $X$  were a finite-variation process, it wouldn't matter whether we take  $H_{t_{i-1}^{(n)}}$  or  $H_{t_i^{(n)}}$  in the sum on the left-hand side. But it matters here in the local martingale case, and we can check that directly: one example is the case where  $H = X$ . Then the corollary tells us that

$$\sum_{i=1}^{|P_n|} X_{t_{i-1}^{(n)}} \left( X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} \right) \xrightarrow{P} \int_0^t X_s dX_s,$$

but if we replace  $(i - 1)$  with  $i$ , we have

$$\sum_{i=1}^{|P_n|} X_{t_i^{(n)}} (X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}) = \sum_{i=1}^{|P_n|} X_{t_{i-1}^{(n)}} (X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}) + \sum_{i=1}^{|P_n|} (X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}})^2 \xrightarrow{P} \int_0^t X_s dX_s + \langle X \rangle_t.$$

Finally, if we add the two of these together, we get

$$\sum_{i=1}^{|P_n|} (X_{t_i^{(n)}} + X_{t_{i-1}^{(n)}}) (X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}) = \sum_{i=1}^{|P_n|} (X_{t_i^{(n)}}^2 - X_{t_{i-1}^{(n)}}^2) = X_t^2 - X_0^2.$$

Thus, putting everything together tells us that

$$X_t^2 - X_0^2 = 2 \int_0^t X_s dX_s + \langle X \rangle_t.$$

A special case of Itô's formula is that for sufficiently smooth  $f$ , we have

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s.$$

We can understand this statement now: if  $f$  is twice continuously differentiable,  $f'$  is a continuous, adapted process, so the first term on the right hand side is a semimartingale (it's an integral of an adapted process against a semimartingale), and the second term is a finite variation term.

We'll end here for now – all of the cool applications will be done via the internet, unfortunately. Because we have two canceled lectures, we'll be asked (if possible) to read the section about Itô's formula.

As a reminder, the midterm tomorrow is canceled, and next class resumes after spring break. From 3:30-5:00, Professor Sun will be in room 2-175 for general advising hours.

## 13 March 30, 2020

The first midterm will be on Thursday in the usual timeslot – if this doesn't work for us, we should email Professor Sun. There will be no lecture on Wednesday due to the exam, so we'll have office hours during the lecture timeslot. (The system puts us in a virtual waiting room, and Professor Sun will "admit" us into the internet office.)

First of all, for a quick review of stochastic integration, we should scroll down to the bottom of the course webpage and read the "Review of stochastic integration" section. Here are the main points: we can define stochastic integration as an isometry in the  $L^2$  bounded case by approximating with elementary processes, and then in general we can use the characterization

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$$

for all  $N \in \mathbb{H}^2$ .

We'll start today with Itô's formula, and we're going to state it in multiple dimensions here:

**Theorem 105** (Itô's formula)

Let  $X_t = (X_t^1, \dots, X_t^p)$  be a process that evolves in  $\mathbb{R}^p$  such that each  $X_t^i$  is a continuous semimartingale. Then given a twice continuously differentiable function  $F : \mathbb{R}^p \rightarrow \mathbb{R}$ ,

$$F(X_t) - F(X_0) = \sum_{i=1}^p \int_0^t \partial_i F(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^p \int_0^t \partial_{ij} F(X_s) d\langle X^i, X^j \rangle_s$$

is also a continuous semimartingale with the above decomposition (where the first term is a local martingale and the second term is a finite variation process).

Remember that the first term is integration with respect to a semimartingale, and the second is essentially a Lebesgue integral, so we now understand all of the individual components of this statement. We won't prove this during class – we should read the proof in Le Gall on our own.

**Fact 106**

This formula holds if  $F$  is defined only on an open set  $U \subseteq \mathbb{R}^p$ , as long as  $t \leq \tau_\varepsilon = \inf\{t : \text{dist}(X_t, U^c) \leq \varepsilon\}$ . In other words, we need to stay at least  $\varepsilon$  away from the boundary so that we can define a function  $\tilde{F}$  consistent with  $F$ .

**Proposition 107**

If  $M$  is a local martingale, then

$$\mathcal{E}(\lambda M)_t = \exp\left(\lambda M_t - \frac{\lambda^2}{2} \langle M \rangle_t\right)$$

is also a local martingale for all  $\lambda \in \mathbb{C}$ .

*Proof.* By Itô's formula with  $X_t = \lambda M_t - \frac{\lambda^2}{2} \langle M \rangle_t$  (so that  $\mathcal{E}(\lambda M)_t = \exp(X_t)$ ) and  $F$  just being the exponential function, we find that

$$d\mathcal{E}_t = \exp(X_t) dX_t + \frac{1}{2} \exp(X_t) d\langle X \rangle_t,$$

because the derivatives of the exponential are just the exponential itself. Plugging in the values that we know, this is

$$= \exp(X_t) \left[ \lambda dM_t - \frac{\lambda^2}{2} d\langle M \rangle_t + \frac{1}{2} \lambda^2 \langle M \rangle_t \right],$$

because  $X_t$  gets its quadratic variation only from the local martingale part ( $\lambda M_t$ ). The last two terms cancel, and now we're just left with the local martingale term, meaning that  $\mathcal{E}(\lambda M)_t$  is indeed a local martingale.  $\square$

**Theorem 108** (Lévy's characterization of Brownian motion)

If  $X$  is a continuous adapted process on  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  taking values in  $\mathbb{R}^d$ , then the following are equivalent:

- $X$  is a Brownian motion with respect to  $\mathcal{F}_t$ .
- The  $X^i$ 's are continuous local martingales with

$$\langle X^i, X^j \rangle_t = t \cdot 1\{j = i\}.$$

*Proof.* We know the forward direction already, because Brownian motion is a continuous local martingale, and a Brownian motion in  $\mathbb{R}^d$  is just  $d$  independent Brownian motions.

For the other direction, we'll want to use the Fourier transform, so it's natural to consider the exponential martingale

$$\mathcal{E}(i\theta \cdot X)_t,$$

where  $\theta \in \mathbb{R}^d$  and  $X$  is our  $\mathbb{R}^d$ -valued process. This can also be written out explicitly as

$$= \exp\left(i\theta \cdot X_t - \frac{1}{2}i^2|\theta|^2 t\right),$$

since we're assuming (by the second condition) that the quadratic variation  $\langle \theta \cdot X \rangle_t = |\theta|^2 t$ . The exponential of the first part ( $i\theta \cdot X_t$ ) is bounded because it's only varying on the unit circle, and the exponential of the second part is bounded **on any finite interval**. Therefore,  $\mathcal{E}(i\theta \cdot X)_t$  is actually a uniformly integrable martingale up to any finite time, so we can apply the optional stopping theorem:

$$\mathbb{E}[\mathcal{E}(i\theta \cdot X_t) | \mathcal{F}_s] = \mathcal{E}(i\theta \cdot X_s)$$

for all  $0 \leq s \leq t < \infty$ . Plugging in the above expression and rearranging, we find that

$$\mathbb{E}[\exp(i\theta \cdot (X_t - X_s)) | \mathcal{F}_s] = \exp\left(-\frac{|\theta|^2(t-s)}{2}\right).$$

But now the left hand side is the characteristic function (or Fourier transform) of  $X_t - X_s$ , given  $\mathcal{F}_s$ , and the right hand side is the characteristic function of  $N(0, (t-s)I_d)$ . So conditioned on  $\mathcal{F}_s$ ,  $X_t - X_s$  has the correct normal distribution, which means that  $X$  has the same finite dimensional distributions as the Brownian motion in  $\mathbb{R}^d$ , which means that  $X$  is a Brownian motion (because we assumed continuous sample paths).  $\square$

We've mentioned the next result before in previous classes:

**Theorem 109 (Dambis-Dubins-Schwarz)**

Let  $M$  be a continuous local martingale, and suppose that  $\langle M \rangle_\infty = \infty$  almost surely (just for simplicity). Then there exists a Brownian motion  $B$  such that almost surely, we have  $M_t = B_{\langle M \rangle_t}$  for all  $t \geq 0$ .

(If the quadratic variation is not infinite, then we just have a Brownian motion parameterized up to some time.)

*Proof.* We're going to construct a  $B$  first and then show that it satisfies the characterization. Without loss of generality, we can assume that  $M_0 = 0$ , and we'll denote  $A_t = \langle M \rangle_t$ . Since  $A$  is a nondecreasing process, we can define an "inverse" by defining

$$\tau_r = \inf\{t \geq 0 : A_t \geq r\}.$$

We now define  $B_r = M_{\tau_r}$ : here,  $B$  will be a Brownian motion with respect to the filtration  $\mathcal{G}_r = \mathcal{F}_{\tau_r}$ . We just need to check that this satisfies the above Lévy characterization. Note that  $M^{\tau_r}$  and  $(M^{\tau_r})^2 - \langle M \rangle^{\tau_r}$  are both uniformly integrable martingales, because the quadratic variation is bounded up to time  $\tau_r$ :  $M^{\tau_r}$  accumulates a total variation of  $r$  by definition, and then  $(M^{\tau_r})^2 - \langle M \rangle^{\tau_r}$  is therefore also a uniformly integrable martingale. So now

$$B_s = M_{\tau_s} = \mathbb{E}[M_{\tau_r} | \mathcal{F}_{\tau_s}] \quad \forall r \geq s$$

by the optional stopping theorem, and plugging in our definitions, this is just  $\mathbb{E}[B_r | \mathcal{G}_s]$ . We also know that

$$B_s^2 - s = (M_{\tau_s})^2 - \langle M \rangle_{\tau_s},$$



and this is a uniformly integrable martingale, so we can again apply the optional stopping theorem to say that this is

$$= \mathbb{E}[(M_{\tau_r})^2 - \langle M \rangle_{\tau_r} \mid \mathcal{F}_{\tau_s}] = \mathbb{E}[B_r^2 - r \mid \mathcal{G}_s]$$

for all  $r \geq s$ . So  $B$  is a local martingale with quadratic variation  $\langle B \rangle_s = s$ , and now we just need to make sure  $B$  is continuous (so that we can apply Lévy's characterization). If  $A$  were strictly increasing,  $\tau$  would be a continuous function, and then  $B_r = M_{\tau_r}$  is a composition of two continuous functions, so it must be continuous. So our problem is that  $A_t$  might be flat (constant) on some interval  $[\tau_r, \tau_{r+}]$ , where

$$\tau_{r+} = \inf\{t : A_t > r\}.$$

If this interval is nontrivial, then  $\tau_r < \tau_{r+}$ , and we need to check if  $B$  is continuous at  $r$ . But if  $A$  is flat on this interval,  $M$  **must also be constant**, because any local martingale with constant quadratic variation does not evolve with time (this is a lemma that we need to check, but we should read the book for details). Thus  $B$  is continuous, and thus it is indeed a Brownian motion.  $\square$

### Example 110

Recall the example from our first lecture, where we consider a holomorphic function  $f : \mathbb{C} \rightarrow \mathbb{C}$ . (This means that if we write  $f = u + iv$ , then the Cauchy-Riemann equations tell us that  $f' = u_x + iv_x = v_y - iu_y$ .)

We can now check that  $f$  applied to a Brownian motion yields another Brownian motion:

$$f(B_t) = \beta_{A_t}, \quad A_t = \int_0^t |f'(B_s)|^2 ds,$$

where  $\beta$  is a Brownian motion.

**Remark 111.** *We're going to skip two topics in the book for now: the Burkholder–Davis–Gundy inequality, as well as the stochastic integral representation for martingales.*

For the rest of today, we'll discuss Girsanov's theorem – this will give us some practice working with all of these objects so far. We did a lot of exercises involving change of measure, and we'll start with a simple example: an **exponential change of measure**. Suppose we have a random variable  $X$  with moment generating function  $m(\theta) = \mathbb{E}(e^{\theta X}) < \infty$ . We can then define a new "tilted" probability measure with the Radon–Nikodym derivative:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{e^{\theta X}}{m(\theta)}.$$

In other words, the probability of an event  $A$  is

$$\mathbb{P}_\theta(A) = \mathbb{E} \left[ 1_A \frac{d\mathbb{P}_\theta}{d\mathbb{P}} \right],$$

and this now explains why we normalize by  $m(\theta)$ : we want  $\mathbb{P}_\theta(\Omega) = 1$ . In particular, under this new measure, notice that

$$\mathbb{E}_\theta X = \mathbb{E} \left[ X \frac{d\mathbb{P}_\theta}{d\mathbb{P}} \right] = \frac{\mathbb{E}[X e^{\theta X}]}{\mathbb{E}[e^{\theta X}]} = \frac{m'(\theta)}{m(\theta)},$$

so shifting the distribution by  $\mathbb{P}_\theta$  changes the mean.

### Example 112

As an explicit example, we'll exponentially tilt the standard normal  $X \sim N(0, 1)$ , and take  $\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \frac{e^{\theta X}}{\exp(\theta^2/2)}$ .

Then the density of  $X$  under  $\mathbb{P}_\theta$  is

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \exp\left(\theta x - \frac{\theta^2}{2}\right),$$

where the first term is the density of  $X$  under  $\mathbb{P}$  and the second term is the Radon-Nikodym derivative. This factors nicely as

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \theta)^2\right),$$

so now  $X$  is distributed as  $N(\theta, 1)$ . In other words, this exponential tilt just moves the center of our distribution in this case.

**Example 113**

Suppose  $\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}\right)$  is bivariate normal, and we do a change of measure  $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\theta Y - \frac{\theta^2}{2}\right)$ .

We can repeat the calculation above, but another way to work through this is to use the characteristic functions: it suffices to calculate (for all real numbers  $t$ )

$$\mathbb{E}_{\mathbb{Q}} [e^{itX}] = \mathbb{E} [e^{itX} e^{\theta Y} e^{-\theta^2/2}].$$

We'll replace  $Y$  with  $aX + \sqrt{1-a^2}W$ , where  $W$  is some independent normal (this is consistent with the covariance between  $X$  and  $Y$ ). Plugging this in, we find that

$$\mathbb{E}_{\mathbb{Q}} [e^{itX}] = \exp\left(\frac{1}{2}(it + \theta a)^2 + \frac{\theta^2 t^2 (1-a)^2}{2} - \frac{\theta^2}{2}\right),$$

where the first term comes from the characteristic function of  $X$  and the second term comes from the characteristic function of  $W$ . The  $\theta^2$  terms cancel, and we're just left with

$$= \exp\left[-\frac{t^2}{2} + it\theta a\right].$$

Therefore,  $X$  is distributed normally as  $N(\theta a, 1)$  under  $\mathbb{Q}$ .

In both of these cases, we just have a finite number of random variables, and now let's go to a sequence  $X_1, \dots, X_n$  of iid random variables under  $\mathbb{P}$ . We'll now define the change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{i=1}^n \frac{e^{\theta X_i}}{\mathbb{E}[e^{\theta X}]}$$

by changing the measure for each  $X_i$ . Then the  $X_i$  are iid under  $\mathbb{Q}$  as well, and now the process

$$S_k = \sum_{i=1}^k X_i$$

is a random walk under both  $\mathbb{P}$  and  $\mathbb{Q}$  (just with different jump distributions). Notice that we can also write this change of measure as

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp(\theta S_n - n \log m(\theta))$$

for any finite  $n$  (though if we take  $n \rightarrow \infty$ ,  $\mathbb{Q}$  may not be absolutely continuous with respect to  $\mathbb{P}$ ). If we take the two

examples above, we can consider a sequence of independent random variables

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a_i \\ a_i & 1 \end{bmatrix} \right).$$

Then we can consider the two processes

$$M_k = \sum_{i=1}^k \sigma_i X_i, \quad L_k = \sum_{i=1}^k \tau_i Y_i,$$

where the coefficients  $\sigma_i, \tau_i \in \mathcal{F}_{i-1}$  can also be random. (In other words, we add a Gaussian increment times some random number which is measurable with respect to the past.) Note that  $M$  and  $L$  are both martingales, and now we can define the “exponential martingale for  $L$ ”

$$D_k = \exp \left( L_k - \frac{1}{2} \sum_{i=1}^k \tau_i^2 \right).$$

We can check that  $D_k$  is also a martingale – this should look very similar to the continuous exponential martingale we talked about earlier in class – and for any finite time, we can now define

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = D_n.$$

Then under  $\mathbb{Q}$ , the law of  $X_i$  looks like

$$X_i \sim N(a_i \tau_i, 1)$$

by applying our above argument, and thus  $M_k$  under  $\mathbb{Q}$  behaves like  $M_k$  under  $\mathbb{P}$ , plus an extra drift term  $\sum_{i=1}^k \sigma_i a_i \tau_i$ . Another way to say this is that  $M_k - \sum_{i=1}^k \sigma_i a_i \tau_i$  is a martingale under  $\mathbb{Q}$  (while  $M_k$  itself is a martingale under  $\mathbb{P}$ ). To say a little bit more about the drift term, this is a **measure of the covariation of  $M$  with  $L$** : since  $M$  is the sum of  $\sigma_i X_i$  and  $L$  is the sum of  $\tau_i Y_i$ , it makes sense that there is a covariation of  $a_i \sigma_i \tau_i$ .

The point of Girsanov’s theorem is to give a continuous-time version of this:

**Theorem 114** (Girsanov’s theorem, informal)

Let  $M$  and  $L$  be local martingales under  $\mathbb{P}$ , and define a change of measure  $\frac{d\mathbb{Q}}{d\mathbb{P}} = D_\infty = \mathcal{E}(L)_\infty$ . Then  $M - \langle M, L \rangle$  is a martingale under  $\mathbb{Q}$ .

(The resemblance between this result and the discrete case should be clear.) We should be a bit careful here: it’s not always true that  $\mathcal{E}(L)_\infty$  is a valid Radon–Nikodym derivative because of **absolute continuity**, and let’s see how that can fail in the discrete case.

Suppose we have probability measures  $\mu, \nu$  on  $(\Omega, \mathcal{F})$  with  $\mathcal{F}_n \uparrow \mathcal{F}$ , and suppose that  $\nu_n = \nu|_{\mathcal{F}_n}$  are absolutely continuous with respect to  $\mu_n = \mu|_{\mathcal{F}_n}$ . Then we know that  $D_n = \frac{d\nu_n}{d\mu_n}$  is a martingale under  $\mu$ , but it may not need to converge under  $\nu$  – thus, we define

$$D_\infty = \limsup D_n.$$

Recall that we can then decompose our measure into a continuous and singular part:

$$\nu(A) = \int_A D_\infty d\mu + \nu(A \cap \{D_\infty = \infty\}),$$

which just tells us that  $D_\infty$  might not be absolutely continuous with respect to  $\nu(A)$ , as long as there is a positive component where  $D_n$  diverges in the limit.

### Example 115

Let  $\Omega = \prod_{i=1}^{\infty} \{0, 1\}$ , and let  $\mathcal{F}_n = \{A \times \prod_{i=n+1}^{\infty} \{0, 1\}, A \subseteq \{0, 1\}^n\}$  be contained in the set of events that only depend on the first  $n$  variables. Suppose that

$$\mu = \bigotimes_{i=1}^{\infty} \text{Ber}(p), \quad \nu = \bigotimes_{i=1}^{\infty} \text{Ber}(q)$$

with  $0 < p < q < 1$ .

It's clear that  $\mu$  and  $\nu$  are not absolutely continuous with respect to each other: a sample  $X \sim \mu$  looks like  $(X_1, X_2, \dots)$  where the  $X_i$  are iid Bernoulli with parameter  $p$ , and similarly  $X \sim \nu$  looks like iid Bernoullis with parameter  $q$ . But  $\frac{1}{n} \sum_{i=1}^n X_i$  converges to  $p$  under one measure and  $q$  under the other by the law of large numbers, so  $\mu$  and  $\nu$  are not absolutely continuous.

But in the discrete case, we know that  $\mu_n$  is just the law of  $(X_1, \dots, X_n)$  under  $\mu$ . Even if  $\mu$  and  $\nu$  are singular with respect to each other, we do have  $\mu_n \ll \nu_n \ll \mu_n$  for all finite  $n$  – “seeing  $n$  bits doesn't tell us for sure whether it comes from an iid Bernoulli  $p$  or iid Bernoulli  $q$ .” Indeed, we can calculate explicitly that

$$D_n = \frac{d\nu_n}{d\mu_n} = \prod_{i=1}^n \frac{e^{\theta X_i}}{m(\theta)},$$

where we pick  $\theta$  correctly to move the mean from  $p$  to  $q$ : we need

$$\mathbb{E}_{\theta} X_i = q = \frac{\mathbb{E}[X_i e^{\theta X_i}]}{\mathbb{E}[e^{\theta X_i}]} = \frac{pe^{\theta}}{pe^{\theta} + 1 - p}.$$

Doing all of the algebra, we find that  $D_n$  concentrates around a specific point:

$$D_n = \frac{\left(\frac{q(1-p)}{p(1-q)}\right)^{S_n}}{\left(\frac{1-p}{1-q}\right)^n} \approx \left(\frac{q}{p}\right)^{nq} \left(\frac{1-q}{1-p}\right)^{n(1-q)} e^{\theta S_n} = \exp(nH(q|p)),$$

where  $H(q|p)$  is the binary relative entropy – it is zero if and only if  $q = p$  and strictly positive otherwise. This goes to  $\infty$  as  $n \rightarrow \infty$ , so indeed  $\mu$  and  $\nu$  are mutually singular.

Basically, we should remember that Girsanov's theorem doesn't work for arbitrary  $L$ , so we need to understand which  $L$  produce a valid change of measure in the continuous case. We'll discuss this more next time.

## 14 April 6, 2020

We'll finish the discussion of Girsanov's theorem today – we'll start by recalling last week's calculation. Suppose we're on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and we have independent Gaussians of the bivariate distribution

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & a_i \\ a_i & 1 \end{bmatrix}\right),$$

such that  $\mathcal{F}_n$  is the sigma-algebra generated by these variables. If  $\sigma_i, \tau_i$  are bounded random variables that are  $\mathcal{F}_{i-1}$ -measurable, we can define

$$M_k = \sum_{i=1}^k \sigma_i X_i, \quad L_k = \sum_{i=1}^k \tau_i Y_i.$$

These are martingales under  $\mathbb{P}$ , and now we can consider the process up to some finite time  $n$ : then we can calculate the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = D_n = \prod_{i=1}^n \exp\left(\tau_i Y_i - \frac{\tau_i^2}{2}\right) = \exp\left(L_n - \sum_{i=1}^n \frac{\tau_i^2}{2}\right).$$

Then by the martingale property,  $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_k} = D_k$  for all  $k \leq n$ , and now  $D_k$  is a discrete-time version of the exponential martingale

$$\mathcal{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L \rangle_t\right).$$

(Remember that this gives us a strictly-positive continuous local martingale.) This  $D$  then helps us define a **change of measure**: recall that we found last time that  $X_i \sim N(a_i \tau_i, 1)$  is a shifted Gaussian under  $\mathbb{Q}$  (conditioned on  $\mathcal{F}_{i-1}$ , so that we know the value of  $\tau_i$ ), so this means that  $M_k$  under  $\mathbb{Q}$  looks like  $M_k$  under  $\mathbb{P}$ , except with a drift term  $\sum_{i=1}^k \sigma_i \tau_i a_i$ . This is similarly a discrete-time version of the “covariation of  $M$  with  $L$ .”

At the end of last lecture, we stated the informal version of Girsanov’s theorem: we said that if  $M$  and  $L$  are local martingales under  $\mathbb{P}$ , we can define  $\frac{d\mathbb{Q}}{d\mathbb{P}} = D_\infty = \mathcal{E}(L)_\infty$ . Then  $M - \langle M, L \rangle$  is a local martingale under  $\mathbb{Q}$  – we’ll formalize this today.

We’ll start with some setup: we are working on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ , where our filtration  $\mathcal{F}_t$  is right-continuous and complete.

**Proposition 116**

Suppose  $\mathbb{Q} \ll \mathbb{P}$ . Then  $D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$  is a uniformly integrable martingale, so it has an rcll modification.

We’ll skip the proof of this – the fact that it’s a martingale is easy to check from the definition of conditional expectation, and the rcll modification is from results of Chapter 3. The main point is that we’ll be working with rcll modifications from now on.

**Lemma 117**

Suppose  $D_t$  is a continuous local martingale with  $D_0 = 1$ , such that  $D_t > 0$  for all  $t$ . Then we can write

$$D_t = \mathcal{E}(L)_t$$

for a continuous local martingale  $L$ .

*Proof.* Apply Itô’s formula to  $Y_t = \log D_t$ . Then

$$dY_t = \frac{1}{D_t} dD_t - \frac{1}{2D_t^2} d\langle D \rangle_t.$$

If we take  $L_t$  such that  $dL_t = dY_t + \frac{1}{2D_t^2} d\langle D \rangle_t$  (to get rid of the finite variation term), we know that  $L$  is a local martingale, and its quadratic variation comes from the local martingale term  $\frac{1}{D_t} dD_t$ : thus

$$d\langle L \rangle_t = \frac{1}{D_t^2} d\langle D \rangle_t,$$

and now

$$dY_t = dL_t - \frac{1}{2} d\langle L \rangle_t.$$

Integrating this, we find that

$$Y_t = \log D_t = L_t - \frac{1}{2} \langle L \rangle_t,$$

and exponentiating both sides tells us that  $D_t = \mathcal{E}(L)_t$ , as desired. We can also write the above  $L$  explicitly:

$$L_t = \int_0^t \frac{1}{D_s} dD_s.$$

□

**Theorem 118** (Girsanov)

Assume that  $\mathbb{Q} \ll \mathbb{P}$ , and  $D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \mathcal{E}(L)_t$ . Assume that  $\mathcal{F}_0$  is trivial, so  $D_0 = 1$ . If  $M$  is a continuous local martingale under  $\mathbb{P}$ , then  $M - \langle M, L \rangle$  is a continuous local martingale under  $\mathbb{Q}$ .

This theorem essentially tells us that the class of martingales is only changing by the drift term  $\langle M, L \rangle$ : in particular, the quadratic variation under  $\mathbb{P}$  and under  $\mathbb{Q}$  are the same.

*Proof.* Let  $X$  be any adapted process. Note that if  $D \cdot X$  (the product, not the stochastic integral) is a continuous martingale under  $\mathbb{P}$ , then  $X$  is a continuous martingale under  $\mathbb{Q}$ . To check this, first we make sure  $X$  is in  $L^1$ :

$$\mathbb{E}_{\mathbb{Q}}|X_t| = \mathbb{E}_{\mathbb{P}}(D_t|X_t|),$$

and because  $D$  is positive, this is  $\mathbb{E}_{\mathbb{P}}(|D_t X_t|)$ , which is finite by assumption of  $D \cdot X$  being a martingale. Now we check the martingale property: for any  $s \leq t$  and any event  $A \in \mathcal{F}_s$ ,

$$\mathbb{E}_{\mathbb{Q}}[X_t 1_A] = \mathbb{E}_{\mathbb{P}}[D_t X_t 1_A] = \mathbb{E}_{\mathbb{P}}[D_s X_s 1_A] = \mathbb{E}_{\mathbb{Q}}[X_s 1_A],$$

where we've used a change of measure in the first and third equalities and the martingale property in the second. Thus  $\mathbb{E}_{\mathbb{Q}}[X_t | \mathcal{F}_s] = X_s$ , as we want. Similarly, we can show that if  $D \cdot X$  is a continuous **local** martingale under  $\mathbb{P}$ ,  $X$  is a continuous **local** martingale under  $\mathbb{Q}$ .

We'll apply this with  $X = M - \langle M, L \rangle$ : we want to show that this is a martingale under  $\mathbb{Q}$ , so it suffices to show that  $D \cdot X$  is a martingale under  $\mathbb{P}$ . Remember that  $D$  evolves as

$$dD_t = \mathcal{E}(L)_t dL_t = D_t dL_t$$

(see the explicit expression from Lemma 117), and now using Itô's formula,

$$d(D \cdot X)_t = D_t dX_t + X_t dD_t + d\langle D, X \rangle_t$$

(we only have the mixed partials in the Hessian, and the factor of 2 cancels).  $X_t dD_t$  is already a martingale, so we don't need to expand it out further. By the definition of  $X_t$ , we can rewrite this as

$$X_t dD_t + D_t(dM_t - d\langle M, L \rangle_t) + d\langle D, X \rangle_t,$$

and now  $X$  is  $M$  plus a finite variation term, so  $d\langle D, X \rangle_t = d\langle D, M \rangle_t$ . But  $d\langle M, L \rangle_t = \frac{1}{D_t} d\langle M, D \rangle_t$ , so the last two terms cancel, and we're just left with

$$d(D \cdot X)_t = X_t dD_t + D_t dM_t.$$

This indeed shows that  $D \cdot X$  is a local martingale. □

Note that this is not the typical way that we apply Girsanov's theorem: instead, we start with a continuous local martingale  $L$  such that  $L_0 = 0$  and  $\langle L \rangle_{\infty} < \infty$  almost surely. We know that this means  $L_t$  converges almost surely to a limit  $L_{\infty}$ , so we know that  $\mathcal{E}(L)_t$  is a continuous local martingale. In particular, because it is nonnegative, it is a

supermartingale, and thus it converges almost surely to  $\mathcal{E}(L)_\infty$  such that  $\mathbb{E}\mathcal{E}(L)_\infty \leq 1$  by Fatou's lemma. And if the expectation is equal to 1,  $\mathcal{E}(L)_t = D_t$  is a uniformly integrable martingale (see our homework), and thus we can define

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$$

and apply Girsanov with this  $L$ . So we need to make sure  $L$  has this condition  $\mathbb{E}\mathcal{E}(L)_\infty = 1$  to make sure all of this is valid.

**Fact 119**

Theorem 5.23 in Le Gall gives a few criteria for this condition being satisfied. If  $L$  is a continuous local martingale with  $L_0 = 0$ , then **Novikov's condition**

$$\mathbb{E} \exp\left(\frac{1}{2}\langle L \rangle_\infty\right) < \infty$$

implies that  $L$  is a uniformly integrable martingale with  $\mathbb{E} \exp\left(\frac{1}{2}\langle L \rangle_\infty\right) < \infty$  (**Kazamaki's criterion**), which tells us that  $\mathcal{E}(L)$  is a uniformly integrable martingale.

We can read the proof on our own – we'll instead focus on applications during this class.

We'll start by constructing a solution for a stochastic differential equation.

**Example 120**

Suppose we want to solve the differential equation

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

where  $b$  is a measurable function with  $|b(t, x)| \leq g(t)$  and  $\int_0^\infty g(t)^2 dt < \infty$ .

*Solution.* Let  $X$  be a Brownian motion under  $\mathbb{P}$ , and let  $L_t = \int_0^t b(s, X_s)dX_s$ . Since  $X$  is a Brownian motion under  $\mathbb{P}$ ,

$$\langle L \rangle_\infty = \int_0^\infty b(t, X_t)^2 dt$$

(since  $d\langle X \rangle_t = dt$ ), and we know that this is bounded by  $\int_0^\infty g(t)^2 dt < \infty$ , so it itself is finite. Thus, Novikov's condition is satisfied, which means that we can define a new measure  $\mathbb{Q}$  such that  $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$ . Applying Girsanov's theorem now tells us that  $B = X - \langle X, L \rangle$  is a local martingale under  $\mathbb{Q}$  – this process  $B$  is a Brownian motion under  $\mathbb{Q}$ , because  $B$  is  $X$  minus a finite variation process (here we're using Lévy's characterization). We have that  $X = \langle X, L \rangle + B$  under  $\mathbb{Q}$  – in differential form, this says that

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

by plugging in the definition of  $L$ , and this is exactly what we wanted. Notice that the only assumption we used is that  $b(t, x)$  is measurable and bounded by an  $L^2$  function  $g(t)$ . □

Our next application is the **Cameron-Martin formula**:

**Example 121**

Let  $L_t = \int_0^t g(s)dB_s$ , where we have a deterministic function  $g(s)$ , and again define  $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(L)_\infty$ .

Then

$$\tilde{B} = B_t - \langle B, L \rangle_t = B_t - \int_0^t g(s)ds$$

will be a Brownian motion under  $\mathbb{Q}$ , as long as  $\mathbb{E}\mathcal{E}(L)_\infty = 1$ . So this gives an **explicit change of measure** between a Brownian motion and a Brownian motion plus a deterministic function  $h(t)$ , where  $h = -\int_0^t g(s)ds$ . And it makes sense that we can't use every possible  $h$ : for example, the law of  $B_t + ct$  is not absolutely continuous with respect to the law of  $B_t$ , because  $\frac{B_t}{t}$  goes to 0 almost surely, while  $\frac{B_t+ct}{t}$  goes to  $c$  almost surely – the two measures are mutually singular.

To proceed, notice that if we define

$$A_t = \langle L \rangle_t = \int_0^t g(s)^2 ds,$$

then  $L_t$  behaves as  $\beta_{A_t}$ , where  $\beta$  is a Brownian motion. Since  $A_t$  is a **deterministic** time change, this means that  $L_t$  is distributed normally as  $N(0, A_t)$ , and thus **if**  $\int_0^\infty g(t)^2 dt = \langle L \rangle_\infty$  is finite, then  $L_\infty$  is just distributed as  $N(0, \langle L \rangle_\infty)$ , and indeed  $\mathbb{E}\mathcal{E}(L)_\infty = 1$ . So here, we don't need Novikov's condition to see that this last condition holds, because we can calculate the law directly.

In other words, this means that the law of  $(B_t + h(t))$  is absolutely continuous with respect to the law of  $B_t$  if and only if we can write  $h(t) = \int_0^t g(s)ds$  such that  $\int_0^\infty g(t)^2 dt < \infty$ . Such functions  $h$  form the **Cameron-Martin (CM) space**.

We'll spend the remainder of this class discussing an application to the large deviations principle – this often goes under the name of **Schilder's theorem**. First, we can recall Cramér's theorem, which tells us about large deviations for the empirical mean  $\frac{1}{n} \sum_{i=1}^n X_i$  of a random variable. Suppose that  $m(\theta) = \mathbb{E}(e^{\theta X_i})$  is finite for all  $\theta \in \mathbb{R}$ : we can then define the cumulant generating function  $\kappa(\theta) = \log m(\theta)$ . Cramér's theorem then tells us that for any  $a > \mathbb{E}[X]$ ,

$$\frac{1}{n} \log \mathbb{P}\left(\frac{S_n}{n} \geq a\right) \rightarrow -I(a),$$

where  $I(a) = \sup_\theta(\theta a - \kappa(\theta))$ . In other words, the probability is exponentially decaying with rate given by this function  $I$ .

We proved the upper bound by using Markov's inequality – to show the lower bound, we used a change of measure – choose a  $\theta$  such that

$$\mathbb{E}_\theta X_i = a + \varepsilon.$$

Then we can let  $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{\exp(\theta S_n)}{\exp(n\kappa(\theta))}$ , and the idea is that the mean of  $X_i$  is slightly larger than  $a$ , so the event  $\{\frac{S_n}{n} \geq a\}$  is now a typical event – the sum of  $n$  iid terms with mean slightly larger than  $a$  is likely to give something larger than  $a$ . So

$$1 \approx \mathbb{Q}\left(\frac{S_n}{n} \geq a\right) \geq \mathbb{Q}\left(a \leq \frac{S_n}{n} \leq a + 2\varepsilon\right).$$

On this event, the Radon-Nikodym derivative is roughly constant, so this is approximately

$$\approx \mathbb{E}_\mathbb{P}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \mathbf{1}\left\{a \leq \frac{S_n}{n} \leq a + 2\varepsilon\right\}\right) \approx \exp(\theta na - n\kappa(\theta)) \mathbb{P}\left(\frac{S_n}{n} \in [a, a + \varepsilon]\right).$$

We're going to do something similar now, but with Brownian motion sample paths instead.



### Theorem 122 (Schilder)

For simplicity, consider Brownian motion on the interval  $[0, T]$ : consider the continuous functions  $C[0, T]$  with the sup-norm  $\|\cdot\|_\infty$ , and let  $W[0, T]$  be the space of functions in  $C[0, T]$  that start at 0. Then for any  $A \subseteq W[0, T]$ ,

$$-\Lambda(A^\circ) \leq \liminf_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon}B \in A) \leq \limsup_{\varepsilon \downarrow 0} \varepsilon \log \mathbb{P}(\sqrt{\varepsilon}B \in A) \leq -\Lambda(\bar{A}),$$

where  $A^\circ$  is the interior of  $A$ ,  $\bar{A}$  is the closure of  $A$ , and  $\Lambda(A) = \inf_{h \in A} I(h)$ , where  $I(h) = \frac{1}{2} \int_0^T h'(t)^2 dt$  is analogous to the  $I(a)$  of Cramér's theorem.

(The liminf and limsup is not very particular to this new setting – it's the function  $I(h)$  that is relevant. If we cared in Cramér's theorem about a general Borel set  $A$ , we'd have the liminf and limsup there, too.)

*Proof sketch for lower bound.* Consider the set

$$A = \{f : \|f - h\|_\infty < \delta\}.$$

We want to do the analogous proof as in Cramér: we want to make  $A$  into a typical event, so we set  $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\frac{1}{\sqrt{\varepsilon}} \int_0^T h'(t) dB_t - \frac{1}{2\varepsilon} \int_0^T h'(t)^2 dt\right)$ . The Cameron-Martin formula now tells us that  $A$  is approximately a regular Brownian motion but with a drift term  $\frac{h}{\sqrt{\varepsilon}}$ , which is exactly what we want for  $B$ : now

$$1 \approx \mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} 1_A\right),$$

and now we can approximately evaluate the Radon-Nikodym derivative by setting  $B = \frac{h}{\sqrt{\varepsilon}}$  here to find that

$$1 \approx \exp\left(\frac{1}{2\varepsilon} \int_0^T h'(t)^2 dt\right) \mathbb{P}(A) \implies \varepsilon \log \mathbb{P}(A) \approx -\frac{1}{2} \int_0^T h'(t)^2 dt = -I(h).$$

□

If we want to extend this from a finite time interval to  $[0, \infty)$ , we should refer to the book of Deuschel and Stroock.

## 15 April 8, 2020

We haven't covered everything from chapter 5, but we'll hold off on stochastic differential equations for now. We'll spend the next few lectures on continuous-time Markov processes, because the theory goes beyond Brownian-type processes. (Most of this comes from Le Gall chapter 6, but we'll go a bit beyond that as well.)

We'll start with a review of the discrete-time Markov chains:

### Definition 123

A **discrete-time Markov process** or **Markov chain** on a finite state space  $E = \{1, \dots, k\}$  is a discrete process  $(X_n)_{n \geq 0}$ , where each  $X_n \in E$ , where there is a transition matrix  $P \in \mathbb{R}^{k \times k}$  with entries  $p_{xy} = \mathbb{P}(X_{n+1} = y | X_n = x)$ .

This matrix  $P$  maps  $\mathbb{C}^k \rightarrow \mathbb{C}^k$ : basically, we have that

$$(Pf)(x) = \sum_y p_{x,y} f(y) = \mathbb{E}(f(X_{n+1}) | X_n = x),$$

which means that the transition matrix is a linear operator. (We'll take this view in the continuous-time case as well.) Here,  $P$  is a **stochastic matrix**: since the rows sum to 1, the constant vector  $P\mathbf{1} = \mathbf{1}$  is a right eigenvector with eigenvalue 1. And if we have  $n$  steps of the chain (that is, if we only observe every  $n$ th step), the transition matrix is just  $P^n$ . We'll call these equations (for all  $n$ ) the **Chapman-Kolmogorov equations**.

Recall the following result:

**Theorem 124 (Perron-Frobenius)**

If a Markov chain is irreducible (there is a sequence of steps that leads from any state to any other state) and aperiodic (the gcd of the directed cycle lengths is 1), then 1 is a simple eigenvalue (it has multiplicity 1), and all other eigenvalues have modulus strictly less than 1. Then the associated left eigenvector  $\pi^*$  for the eigenvalue 1 (such that  $\pi^*P = \pi^*$ ) has all positive entries, and it is the stationary distribution of the chain.

This can be thought of as a theorem about matrices with real positive entries. The idea is that if we diagonalize  $P = UDU^{-1}$ , the column vectors of  $U$  are the right eigenvectors  $u_i$ , and the rows of  $U^{-1}$  are the left eigenvectors  $v_i$ : we notice that

$$P^n = U D^n U^{-1} = \sum_{i=1}^k (d_i)^n u_i v_i^*.$$

Since 1 is a simple eigenvalue and all of the other eigenvalues have modulus less than 1,  $P^n \pi^*$  will approach  $\mathbf{1} \pi^*$ .

What's interesting about this is first generalizing the state space and then making this into a continuous process. We'll begin by looking at a **measurable** state space  $(E, \mathcal{E})$ :

**Definition 125**

A **Markov transition kernel** on a space  $(E, \mathcal{E})$  is a map  $Q : E \times \mathcal{E} \rightarrow [0, 1]$ , such that

- For all  $x \in E$ , the function  $Q(x, \cdot)$  is a probability measure on  $(E, \mathcal{E})$ .
- For all  $A \in \mathcal{E}$ , the function  $Q(\cdot, A)$  is measurable.

Here,  $Q(x, \cdot)$  represents the law of the process at some future time, given that it is at  $x$  right now. (This corresponds to the row vector of  $x$  in the transition matrix  $P$ .) The second condition about  $Q(\cdot, A)$  is more technical: since we have the result

$$Pf(x) = (\text{row vector of } x)(\text{column vector } f) = \mathbb{E}(f(X_{n+1}) | X_n = x),$$

we can generalize that in this case: if we have a measurable function  $f : E \rightarrow \mathbb{C}$ , we will say that

$$Qf(x) = \int_E f(y)Q(x, dy),$$

which is the Lebesgue integral of  $f$  against the measure  $Q(x, \cdot)$ . This is completely analogous to the discrete case, and now for any measurable bounded function  $f$ ,  $Qf$  will also be measurable and bounded, and this is where we use the second condition on  $Q(\cdot, A)$ . ( $Qf$  being measurable follows directly for indicator functions, and then we can approximate with indicators in general.)

We'll denote

$$B(E) = \{\text{bounded measurable functions on } E\},$$

and we'll give it the sup-norm  $\|f\| = \|f\|_\infty$ . Then any Markov transition kernel  $Q$  maps  $B(E)$  to itself, and  $Q$  here is a **contractive operator**:  $\|Qf\| \leq \|f\|$ , because  $Qf$  is an expectation of the function  $f$ , so it is uniformly bounded by the sup-norm.

Here,  $E$  has no conditions other than being a measurable space, but moving forward will require further regularity conditions. (We'll point those out as we need them.)

**Definition 126**

A **transition semigroup** on a state space  $(E, \mathcal{E})$  is a collection of transition kernels  $(Q_t)_{t \geq 0}$  which satisfy the following conditions:

- For all  $x \in E$ ,  $Q_0(x, \cdot)$  (the law of where we go in zero time) is the Dirac measure  $\delta_x$ .
- For all  $s, t \geq 0$ ,  $Q_{s+t} = Q_s Q_t$ .
- For all measurable  $A \in \mathcal{E}$ , the map  $(t, x) \rightarrow Q_t(x, A)$  is measurable with respect to  $\mathcal{B}_{[0, \infty)} \otimes \mathcal{E}$ .

The second condition here is the Chapman-Kolmogorov condition, and it's less trivial because we're making sure that all of our  $Q_t$ s are consistent. Note that we're defining the composition  $Q_s Q_t$  in terms of how they map  $B(E)$  to itself.

**Definition 127**

A **Markov process with transition semigroup  $Q_t$**  on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  is an adapted process  $(X_t)_{t \geq 0}$  such that

$$\mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] = (Q_t f)(X_s)$$

for all bounded measurable functions  $f \in B(E)$  and all  $s, t \geq 0$ .

We're defining such a process with restrictions, so we should make sure it exists. If  $E$  is nice enough (for example, if we have a Polish space), then existence of such a Markov process with the transition semigroup  $(Q_t)_{t \geq 0}$  comes from the Kolmogorov extension theorem. Here, our index set is  $[0, \infty)$  (which is uncountable), and we just need to specify the consistent finite-dimensional distributions, which just come from the  $Q_t$ s. We won't check the Kolmogorov extension theorem itself, but we just need to make sure  $E$  is nice.

However, just like with Brownian motion and martingales, there is no guarantee of sample path regularity. We may talk a bit about this later on, but we'll focus on the aspects that are different from what we've already seen.

Going back to the discrete-time case, we can also define a semigroup  $(Q_n)_{n \geq 0}$ , but the object is less useful because it's just  $(I, P, P^2, P^3, \dots)$  – it's specified by just one transition matrix  $P$ . So the first mystery is whether there is a “basic building block” analogous to  $P$  for Markov processes in continuous time which encodes information of the entire semigroup  $(Q_t)_{t \geq 0}$ . The answer is “generally yes,” but the story is more complicated – this is only true for **Feller processes**, which we'll define later.

To make that analogy, we'll need a few more concepts:

**Definition 128**

The  **$\lambda$ -resolvent** of a semigroup  $(Q_t)_{t \geq 0}$  (for some  $\lambda > 0$ ) is the operator  $R_\lambda : B(E) \rightarrow B(E)$  such that

$$(R_\lambda f)(x) = \int_0^\infty e^{-\lambda t} Q_t f(x) dt.$$

We should think of this as the Laplace transform (in the time-coordinate) of the semigroup. Remember that  $Q_t f$  is the expectation of the state at time  $t$ , given that we're currently at  $x$ : recall that the exponential distribution  $\text{Exp}(\lambda)$

has density  $\lambda e^{-\lambda t}$ , so what this really says is that

$$(R_\lambda f)(x) = \frac{1}{\lambda} \mathbb{E}(f(X_{\tau_\lambda}) | X_0 = x),$$

where  $\tau_\lambda$  is a random time distributed according to  $\text{Exp}(\lambda)$ . So when  $\lambda$  is large, we emphasize times close to 0, and vice versa.

**Lemma 129** (Resolvent equation)

For any  $\lambda, \mu > 0$ ,

$$R_\lambda - R_\mu + (\lambda - \mu)R_\lambda R_\mu = 0.$$

*Proof.* It's enough to prove this for  $\lambda \neq \mu$  (otherwise this is clearly 0). The composition of the two resolvents satisfies

$$(R_\lambda(R_\mu f))(x) = \int_0^\infty e^{-\lambda s} Q_s(R_\mu(f(x))) ds = \int_0^\infty e^{-\lambda s} Q_s \left( \int_0^\infty e^{-\mu t} Q_t f dt \right) (x) ds.$$

Expanding out the definition of  $Q_s$ , this is

$$\int_0^\infty e^{-\lambda s} \int_E \left( \int_0^\infty e^{-\mu t} Q_t f(y) dt \right) Q_s(x, dy) ds,$$

and now we can apply Fubini's theorem (because there are no integrability issues when we have bounded functions).

This evaluates to

$$\int_0^\infty e^{-\lambda s} \int_0^\infty e^{-\mu t} \int_E Q_t f(y) Q_s(x, dy) dt ds,$$

and the inner integral  $\int_E Q_t f(y) Q_s(x, dy)$  means that we start at  $x$  and evolve for time  $s$ , ending up at  $y$ , where we evaluate  $Q_t f$ . This means we are at state  $y$  and evolve at time  $t$ , and see what the value of  $f$  looks like there, so this inner integral all just evaluates to  $Q_{s+t} f(x)$ . Thus, we end up with

$$\int_0^\infty e^{-\lambda s} e^{\mu s} \int_0^\infty e^{-\mu s} e^{-\mu t} Q_{s+t} f(x) dt ds,$$

where we've slipped in an  $e^{\mu s - \mu s}$  to separate the two integrals, and now the inner integral is all in terms of  $s + t$ : this is equivalent to (setting  $r = s + t$ )

$$\int_0^\infty e^{-\lambda s} e^{\mu s} \int_s^\infty e^{-\mu r} Q_r f(x) dr ds.$$

This is a double integral over pairs  $(r, s)$ , and we can again change the order of integration to

$$= \int_0^\infty e^{-\mu r} Q_r f(x) \int_0^r e^{-\lambda s} e^{\mu s} ds dr.$$

This just evaluates to

$$= \int_0^\infty Q_r f(x) \frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} dr,$$

which is exactly  $\frac{R(\mu) - R(\lambda)}{\lambda - \mu} f(x)$ , as desired.  $\square$

This is mostly an algebraic manipulation – we won't use it today, but we will in the next few lectures. The main idea is to become more familiar with idea of composition of operators: the key idea is “composing” the operators using Chapman-Kolmogorov.

We're now moving on, and we're going to need some more regularity for the next step: we'll assume that our space  $E$  is metrizable, locally compact (around any point, we can find a compact set that contains a neighborhood of the point), and  $\sigma$ -compact ( $E$  is a countable union of compact sets). In particular, this implies that  $E$  is a Polish space –

examples of such spaces  $E$  include open subsets of  $\mathbb{R}^d$ , as well as much more general spaces.

Let  $\mathcal{E}$  be the Borel  $\sigma$ -field of  $E$ , and also denote  $E = \bigcup_{n=1}^{\infty} K_n$  (where the  $K_n$  are **compact and nested** – this exists by assumption of being  $\sigma$ -compact).

**Definition 130**

A function  $f : E \rightarrow \mathbb{R}$  **tends to zero at infinity** if

$$\lim_{n \rightarrow \infty} \sup_{x \in E \setminus K_n} |f(x)| = 0.$$

Remember that a Polish space is defined to be separable (containing a countable dense subset) and completely metrizable (topologically homeomorphic to a complete metric space). For example, if we take  $E = (0, 1)$ , this is locally compact and  $\sigma$ -compact (it's the countable union of the sets  $[\frac{1}{n}, 1 - \frac{1}{n}]$ ), so this is indeed a Polish space. But it's not a complete metric space, because the point  $\frac{1}{n}$  doesn't converge to anything in the space. It **is** instead completely metrizable, because  $(0, 1)$  is topologically homeomorphic to  $\mathbb{R}$ . So being completely metrizable is a topological property – we don't need to put a complete metric on the space. In particular, “tending to zero at infinity” for the interval  $(0, 1)$  means that we tend to 0 at the endpoints 0 or 1.

Now denote

$$C_0(E) = \{\text{continuous real functions on } E \text{ tending to zero at infinity}\}.$$

This is a subspace of  $B(E)$ , which is a Banach space with the sup norm – we'll also look at  $C_0(E)$  with the sup norm. This now gives us enough of the topological setup to define the class of processes that we want:

**Definition 131**

On a space  $E$  satisfying the above conditions, a **Feller semigroup** is a transition semigroup  $(Q_t)_{t \geq 0}$  such that

- $Q_t$  maps  $C_0(E)$  into itself for all  $t \geq 0$ ,
- For all  $f \in C_0(E)$ ,  $\|Q_t f - f\| \rightarrow 0$  as  $t \rightarrow 0$ .

A **Feller process** is a Markov process with a Feller semigroup.

Remember that  $Q_t f(x)$  is the expected value of  $f$  at time  $t$ , given that we're at  $x$  at time 0. So the Feller property tells us that the process doesn't make a large jump in a small amount of time:  $Q_t f$  is very close to the value of  $f$ . However, discontinuous jumps are still allowed, and many natural Feller processes do have jumps – we're just not likely to make a jump immediately at any time, so our process might evolve continuously for some interval and then make a jump.

**Definition 132**

Given a Feller semigroup  $(Q_t)_{t \geq 0}$ , set  $D(L) = \text{domain}(L)$  to be the set

$$\left\{ f \in C_0(E) : \frac{Q_t f - f}{t} \text{ converges in } C_0(E) \text{ as } t \downarrow 0 \right\}.$$

(We require convergence in the sup-norm.) Then for any  $f \in D(L)$ , define

$$Lf = \lim_{t \downarrow 0} \frac{Q_t f - f}{t}.$$

$L$  is the **(infinitesimal) generator** of  $(Q_t)_{t \geq 0}$ , and it is an operator from  $D(L)$  to  $C_0(E)$ .

We'll spend the rest of the lecture on a heuristic preview of material for next week. Here,  $L$  is the "derivative" of  $Q_t$  at time  $t = 0$ . Notice that at a general time,

$$\frac{d}{dt}Q_t = \lim_{s \rightarrow 0} \frac{Q_{t+s} - Q_t}{s},$$

and Chapman-Kolmogorov tells us that this is

$$= \lim_{s \downarrow 0} \frac{Q_s Q_t - Q_t}{s} = LQ_t.$$

If we just have a real-valued function  $q : [0, \infty) \rightarrow \mathbb{R}$  which solves  $q' = \ell q$  with  $q(0) = 1$ , we know that the unique solution is  $q(t) = e^{\ell t}$ . The Laplace transform of such a function is

$$r(\lambda) = \int_0^{\infty} e^{-\lambda t} q(t) dt = \frac{1}{\lambda - \ell}$$

for  $\lambda > \ell$ . By analogy, we might guess that  $Q_t$  is similarly an exponential of the form  $Q_t = e^{tL} = \sum_{k \geq 0} \frac{(tL)^k}{k!}$ , and that the resolvent is  $(\lambda - L)^{-1}$ . This isn't a rigorous argument, but it's our best guess – our goal for next week will be to explore how valid this analogy is.

### Example 133

Consider Brownian motion on  $\mathbb{R}^d$ .

$Q_t(x, \cdot)$  is then the distribution  $N(x, tI_{d \times d})$  (this is how the process evolves starting at  $x$  in time  $t$ ), and then the generator in Brownian motion looks like

$$Lf = \lim_{t \downarrow 0} \frac{\mathbb{E}f(x + B_t) - f(x)}{t}.$$

We don't actually have all of the tools needed to evaluate this rigorously, but if  $f$  is nice enough and we let our Brownian motion go for a small amount of time,  $B_t$  is small, so we should be able to Taylor expand. This yields

$$\lim_{t \downarrow 0} \frac{1}{t} \mathbb{E} \left[ \nabla f(x) \cdot B_t + \frac{1}{2} B_t^t (\text{Hess } f) B_t \right],$$

and  $B_t$  has mean 0 so the first term goes away in expectation. The Hessian of  $f$  is a  $d \times d$  matrix here, and we're only picking up the diagonal terms here because the Brownian motion  $B_t$  has independent entries in the  $d$  dimensions. Thus, we'll get the trace of the Hessian, which is the Laplacian:

$$Lf = \frac{1}{2} \Delta f(x).$$

Again, this highlights the connection between Dirichlet theory and Brownian motion, which we'll explore more soon! We can also calculate  $R_\lambda$  in this case, but the integral is more complicated: it turns out to be related to the Green kernel (the inverse of the Laplacian) when we take  $\lambda \downarrow 0$ , since we're saying that the resolvent is the inverse of  $L$ .

## 16 April 13, 2020

From survey responses that we filled out, it was mentioned that some of us are having trouble seeing the slides during class – as a reminder, there's a Dropbox link at the bottom of the course webpage, which has slides basically synchronized with lecture.

We'll be talking more about Feller processes today, as well as a special case of these processes that are particularly simple. For review, say we have a Markov process  $X_t$  with transition semigroup  $(Q_t)_{t \geq 0}$ : here, remember that

$$Q_t(x, \cdot) = \mathbb{P}(X_{s+t} \in \cdot | X_s = x, \mathcal{F}_s)$$

for all  $s, t \geq 0$  and  $x \in E$ . As discussed last time, we can view  $Q_t$  as an operator on  $B(E)$ : given a bounded function  $f$ , we have

$$(Q_t f)(x) = \int f(y) Q_t(x, dy) = \mathbb{E}[f(X_{s+t}) | C_s = x].$$

The **Chapman-Kolmogorov equations** tell us that  $Q_{s+t} = Q_s Q_t$ , and we can define a Laplace transform

$$R_\lambda = \int_0^\infty e^{-\lambda t} Q_t dt.$$

Since  $\lambda e^{-\lambda t}$  integrates to 1,  $\lambda R_\lambda$  is a Markov kernel, meaning that multiplying by  $\lambda$  makes it properly normalized. We showed (with some algebra) the resolvent equation

$$R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0,$$

which shows that  $R_\lambda$  and  $R_\mu$  commute.

Last time, we defined Feller processes to be those such that

$$\|Q_t f - f\| \rightarrow 0 \text{ as } t \downarrow 0$$

for all  $f \in C_0(E)$ . We also defined the space  $D(L)$  (the "domain" of  $L$ ) to be the space of functions  $f \in C_0(E)$  such that  $\lim_{t \downarrow 0} \frac{Q_t f - f}{t}$  **exists in**  $C_0(E)$  (with respect to the sup-norm topology). Our goal here is to understand how  $L$  determines the semigroup  $Q_t$ .

**Proposition 134**

$Q_t$  and  $L$  commute with each other on  $D(L)$  (where they are both defined) for all  $t$ .

*Proof.* We know that

$$Q_t L f = Q_t \left( \lim_{s \downarrow 0} \frac{Q_s f - f}{s} \right),$$

and now the limit is in the sup-norm, while  $Q_t$  is an operator which is contractive (the norm of  $Q f$  is at most the norm of  $f$ ). Thus the limit converging in the sup norm means that we can bring in  $Q_t$  and it will converge as well, so this is equal to

$$= \lim_{s \downarrow 0} \frac{Q_t Q_s f - Q_t f}{s} = \lim_{s \downarrow 0} \frac{Q_s(Q_t f) - (Q_t f)}{s} = L Q_t f,$$

with the middle equality by Chapman-Kolmogorov. □

The next result we'll prove is a differential relation:

**Proposition 135**

For all  $f \in D(L)$  and for all  $t \geq 0$ ,

$$\int_0^t Q_s L f ds = Q_t f - f = \int_0^t L Q_s f ds.$$

In other words, the time-derivative of  $Q_t f$  is given by  $Q_t L f$ . The two results were proven at different times – they're known as the Kolmogorov forward and backwards equations, respectively.

*Proof.* From the previous result, we just need to prove one of the two equalities, and we can also fix a point  $x \in E$ . Let  $h(t) = Q_t f(x)$ , and take the right-derivative of  $h$ : this is equal to

$$\lim_{s \downarrow 0} \frac{1}{s} [Q_{s+t} f(x) - Q_t f(x)] = LQ_t f(x),$$

and now since  $Q_t$  is contractive, this is uniform convergence in  $t$ , so  $h'(t) = LQ_t f(x)$ . □

To explain the last argument a bit more, note that

$$\left\| Q_t \left( \frac{1}{s} (Q_s - I) f \right) \right\| \leq \left\| \frac{1}{s} (Q_s - I) f \right\|,$$

and because the right side converges uniformly, so does the left side.

We'll explain the names a bit more: for a large class of (**diffusive**) processes,  $L$  is a differential operator. For example,  $L = \frac{1}{2} \Delta$  for a Brownian motion (as we discussed earlier).

- For the Kolmogorov forward equation, suppose that we know  $X_0 \sim \mu$  is initially distributed in some way, and we want to know how the forward evolution affects the law of the process at some final time. We know that  $X_t \sim \mu Q_t(dy) = \int_{x \in E} \mu(dx) Q_t(x, dy)$ , and what we know now is that

$$\frac{d}{dt} (\mu Q_t) = (\mu Q_t) L,$$

which gives us a partial differential equation: if  $\mu Q_t$  has a density  $p(t, x)$  in the  $x$ -coordinate at time  $t$ , then

$$\partial_t p(t, x) = L_x^* p(t, x).$$

Then solving this PDE forward in time from the initial conditions  $p(0, x)$  just comes from the initial density  $\mu$ .

- Meanwhile, for the Kolmogorov backward equation, we care about the final state of the system: suppose we want to know something like

$$\mathbb{E}(f(X_T) | X_0 = x) = (Q_T f)(x).$$

Then we know that

$$\frac{d}{dt} Q_t f(x) = LQ_t f(x),$$

and now if  $f(t, x) = Q_t(f, x)$ , we get the PDE

$$\partial_t f(t, x) = L_x f(t, x),$$

and we can solve this backward from the final condition  $f(T, x) = f(x)$ .

We'll now return to the connection from last time between the **resolvent** and the **generator**:

### Proposition 136

For a semigroup  $Q_t$ , let  $\mathcal{R}$  be the range of  $R_\lambda$ :

$$\mathcal{R} = \{R_\lambda f : f \in C_0(E)\}.$$

This doesn't depend on  $\lambda$ , and it's a dense subset of  $C_0(E)$  (in the sup-norm topology).

*Proof.* Rearranging the resolvent equation,

$$R_\mu = R_\lambda (I + (\lambda - \mu)R_\mu),$$



so the range of  $R_\mu$  is contained in the range of  $R_\lambda$ , and this is interchangeable. To show that the range is dense, we can consider  $\lambda R_\lambda f$  for any function  $f \in C_0(E)$ : this is

$$= \int_0^\infty \lambda e^{-\lambda t} Q_t f dt,$$

which is saying we wait an (exponential random variable) amount of time and evolve  $f$  by that amount of time. Doing a change of variables, this is the same as

$$= \int_0^\infty e^{-t} Q_{t/\lambda} f dt,$$

and now taking  $\lambda \rightarrow \infty$ ,  $Q_{t/\lambda} f$  converges uniformly to  $f$  by the Feller property, while  $e^{-t}$  is integrable, so the integral converges uniformly to  $\int_0^\infty e^{-t} f dt = f$  by the dominated convergence theorem. Thus we can indeed approximate any  $f$ .  $\square$

### Theorem 137

For any Feller semigroup,  $D(L) = \mathcal{R}$ , and the two functions  $R_\lambda : C_0(E) \rightarrow \mathcal{R}$  and  $\lambda - L : D(L) \rightarrow C_0(E)$  are inverses of each other.

*Proof.* It suffices to show that (1)  $(\lambda - L)R_\lambda g = g$  for all  $g \in C_0(E)$ , and (2)  $R_\lambda(\lambda - L)f = f$  for all  $f \in D(L)$ .

For (1), we know that

$$LR_\lambda g = \lim_{s \downarrow 0} \frac{1}{s} (Q_s R_\lambda g - R_\lambda g) = \lim_{s \downarrow 0} \frac{1}{s} \left( Q_s \left( \int_0^\infty e^{-\lambda t} Q_t g dt \right) - \left( \int_0^\infty e^{-\lambda t} Q_t g dt \right) \right),$$

and now we can move  $Q_s$  inside the integral by Fubini's theorem:

$$= \lim_{s \downarrow 0} \frac{1}{s} \left( e^{\lambda s} \int_0^\infty e^{-\lambda s} e^{-\lambda t} Q_{s+t} g dt - \int_0^\infty e^{-\lambda t} Q_t g dt \right),$$

and a change of variables means this is the same as

$$= \lim_{s \downarrow 0} \frac{1}{s} \left( (e^{\lambda s} - 1) \int_0^\infty e^{-\lambda t} Q_t g dt - e^{\lambda s} \int_0^s e^{-\lambda t} Q_t g dt \right).$$

Taking  $s$  down to 0, we just end up with

$$= \lambda R_\lambda g - g,$$

which shows the desired result.

For (2), take any  $f \in D(L)$  to get

$$\lambda R_\lambda f = \lambda \int_0^\infty e^{-\lambda t} Q_t f dt$$

and by the Kolmogorov forward equation, we can rewrite  $Q_t f$ :

$$= \int_0^\infty \lambda e^{-\lambda t} \left( f + \int_0^t Q_s Lf ds \right) dt.$$

The first term integrates to  $f$ , and the second term can be written as well:

$$= f + \int_0^\infty Q_s Lf \int_s^\infty \lambda e^{-\lambda t} dt ds = f + \int_0^\infty e^{-\lambda s} Q_s Lf ds = f + R_\lambda Lf.$$

$\square$

### Corollary 138

A Feller semigroup  $Q_t$  is uniquely determined by its generator  $L$  (though we need to specify the domain  $D(L)$  for which the limit is well-defined).

*Proof.* Take  $g \in C_0(E)$ . If we know  $L$ , we know  $R_\lambda g = (\lambda - L)^{-1}g$ , and if  $g \geq 0$ , then  $Q_t g \geq 0$  for all  $t$ . Since we know that

$$(R_\lambda g) = \int_0^\infty e^{-\lambda t} (Q_t g)(x) dt,$$

(fixing  $x$  gives us a nonnegative function  $(Q_t g)(x)$  of  $t$ ), so we know the Laplace transform of  $Q_t g(x)$ , meaning we know the nonnegative function  $Q_t g(x)$ . This characterizes  $Q_t$  for all  $t$  for nonnegative functions, which gives us the characterization on  $C_0(E)$ .  $\square$

Our expression here is less explicit than in the discrete case, in which we just said that  $Q_n = P^n$ . So it's natural, for example, to ask if  $Q_t = \exp(tL)$ , and that's what we'll discuss next.

First, though, we can do an example:

### Example 139

Consider a Brownian motion in  $\mathbb{R}^d$ . We can check the generator  $L = \frac{1}{2}\Delta$ .

The semigroup can be explicitly written down: we have

$$Q_t(x, dy) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right) dy,$$

because this is just the probability to go from  $x$  to  $y$ , which can be written in terms of a Gaussian density. Then

$$R_\lambda(x, dy) = \int_0^\infty e^{-\lambda t} Q_t(x, dy) = \int_0^\infty e^{-\lambda t} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right) dt dy,$$

and we showed earlier that this is equal to  $(\lambda - L)^{-1}$ . The inverse of the Laplacian is the **Green kernel**, so it's natural to ask what happens when we take  $\lambda = 0$ : generally, the resolvent will not always be defined, since we can't always evaluate the Laplace transform at 0, but in this case the integral converges as long as  $d > 2$  (since  $t$  getting larger makes the exponential to go 1). We can then just compute directly, and we'll find that

$$R_\lambda(x, dy) = 2G(x, dy).$$

The reason this argument doesn't work well when  $d = 1, 2$  is that the Brownian motion returns to each state infinitely often – then it is possible to get the classical Green kernel back, but we need to do some renormalization.

We'll now turn our attention to the earlier question of whether  $Q_t = \exp(tL)$ : suppose we have a space  $(E, \mathcal{E})$ , and let  $Y_n$  be a discrete-time Markov chain with a transition kernel  $P(x, dy)$ . Basically, at every integer time, we just move according to  $P$ .

### Definition 140

The canonical way to make a discrete-time Markov chain into a continuous time process is to let  $N_t$  be a Poisson process of some rate  $c$ : in particular,  $N_t$  is an integer distributed according to  $\text{Pois}(ct)$ . Then we let  $X_t = Y_{N_t}$ , and this is known as a **pseudo-Poisson process**.

Here,  $X$  and  $Y$  have the same trajectory up to a time change, and the only difference is that  $Y$  jumps at integer times while  $X$  jumps at random times  $\tau_n$ : specifically, the increment  $\tau_{n+1} - \tau_n$  are iid exponential random variables with rate  $c$ .

The generator of such a process is related to the chance of making a jump in a time  $t$ , where  $t$  is small. This is the probability that our exponential clock will ring, so it's  $1 - e^{-ct} \approx ct$ . Therefore,

$$Lf(x) = \lim_{t \downarrow 0} \mathbb{E} \left[ \frac{f(X_t) - f(x)}{t} \middle| X_0 = x \right] = c \int (f(y) - f(x)) P_x(dy).$$

In operator notation, this says that  $L = c(P - I)$ . So given the jump rate and transition kernel of a discrete Markov chain, we can find  $L$ , and this is a bounded operator:  $\|Lf\| \leq 2c\|f\|$ . This means that we can define  $\exp(tL)$  for this specific  $L$ : we just set

$$\exp(tL)f = \sum_{k=0}^{\infty} \frac{(tL)^k f}{k!}.$$

We want to claim that this is the same as the semigroup: note that

$$e^{tL} = e^{tc(P-I)} = e^{-ctI} e^{ctP} = e^{-ct} \sum_{k=0}^{\infty} \frac{(ctP)^k}{k!} = \sum_{k=0}^{\infty} \left( \frac{e^{-ct}(ct)^k}{k!} \right) P^k.$$

And now this parenthetical term is related to the Poisson distribution:

$$= \sum_{k=0}^{\infty} \mathbb{P}(N_t = k) P^k,$$

and this is exactly how the process should evolve in time  $t$ : we figure out how many times we'll make a change, and then we evolve by  $P$  that many times.

We'll finish by discussing the Yosida approximation theorem. Suppose we're back in the general case with a Feller semigroup  $(Q_t)_{t \geq 0}$  and a generator  $L$ : we have no control on the boundedness of  $L$ , so we can't always define  $\exp(tL)$ . But we can define the  $\lambda$ -approximation

$$L^{(\lambda)} = \lambda L R_\lambda = \lambda (\lambda - (\lambda - L)) R_\lambda = \boxed{\lambda(\lambda R_\lambda - I)},$$

where we've used the fact that  $\lambda - L$  is the inverse of  $R_\lambda$ , and then we can show that this last expression is also equal to  $\lambda R_\lambda L$  by the same argument. But here, the boxed expression looks a lot like  $L = c(P - I)$ , so  $L^{(\lambda)}$  is the generator of a pseudo-Poisson process with transition kernel  $\lambda R_\lambda$ . In other words, there is a discrete Markov chain that jumps according to  $\lambda R_\lambda$  with rate  $\lambda$ . Here,  $L$  is pretty general, while  $L^{(\lambda)}$ . Denote the semigroup with this generator to be

$$Q_t^{(\lambda)} = \exp(tL^{(\lambda)}).$$

**Theorem 141** (Yosida approximation theorem)

For all  $f \in D(L)$ , we have  $\|L^{(\lambda)}f - Lf\|$  converging to 0 in the sup-norm as  $\lambda \rightarrow \infty$ . In addition,

$$\|Q_t^{(\lambda)}f - Q_t f\| \leq t \|L^{(\lambda)}f - Lf\|,$$

so the left-hand side converges on any bounded interval. Finally, for all  $f \in C_0(E)$  (a larger domain),  $\|Q_t^{(\lambda)}f - Q_t f\|$  converges to 0 (the semigroup converges to the actual semigroup) on any bounded time interval.

This is one way to approximate a Markov process: we can use simple processes of the form  $Q_t^{(\lambda)}$  and take  $\lambda \rightarrow \infty$ . But at the end of the day, we're interested in the process, not the semigroup, so we're curious about whether the

processes  $(Q_t^{(\lambda)})_{t \geq 0}$  converge in law to  $Q_t$  as well. It turns out that this holds somewhat generally, and we might talk about this later on.

## 17 April 15, 2020

We'll cover three separate topics today, concluding our discussion of Markov processes. The first will be **sample path regularity** for Markov processes, which will provide some nice connections with martingales. Last time, we discussed that the resolvent and generator are related for a Feller process. The other reason that this resolvent is important is that we can construct a supermartingale out of our process:

### Lemma 142

Let  $X_t$  be a Markov process with semigroup  $(Q_t)_{t \geq 0}$  and resolvent  $R_\lambda$ . For any bounded nonnegative function  $h \in B(E)$ ,  $h \geq 0$ , the process  $S_t = e^{-\lambda t} R_\lambda h(X_t)$  is a supermartingale for all  $\lambda > 0$ .

(Here, we're using the function  $h$  to go from an abstract space  $E$  to the reals, so that we can do things like addition and subtraction.)

*Proof.*  $S_t$  is a nonnegative process – we can check that  $S_t \in L^1$  for all  $t$ , because  $\lambda R_\lambda$  is a bounded (Markov transition) operator:  $\|\lambda R_\lambda h\| \leq \|h\|$ , so there are no integrability issues at any finite time  $t$ . Then we can check the martingale property: we have

$$\mathbb{E}[S_{s+t} | \mathcal{F}_s] = e^{-\lambda(s+t)} \mathbb{E}[R_\lambda h(X_{s+t}) | \mathcal{F}_s] = e^{-\lambda(s+t)} \mathbb{E}\left[\int_0^\infty e^{-\lambda r} Q_r dr h(X_{s+t}) | \mathcal{F}_s\right].$$

By Fubini's theorem, we can move the expectation inside to get

$$= e^{-\lambda(s+t)} \int_0^\infty e^{-\lambda r} \mathbb{E}[Q_r h(X_{s+t}) | \mathcal{F}_s] dr.$$

But the expectation can be directly evaluated with the semigroup, leaving

$$= e^{-\lambda(s+t)} \int_0^\infty e^{-\lambda r} Q_r Q_s h(X_s) dr.$$

Chapman-Kolmogorov shows that this is equal to

$$e^{-\lambda s} \int_0^\infty e^{-\lambda(t+r)} Q_{t+r} h(X_s) dr,$$

and a change of variables means that this is equivalent to

$$= e^{-\lambda s} \int_t^\infty e^{-\lambda r} Q_r h(X_t) dr.$$

$h$  is nonnegative, and thus the integrand is nonnegative: we can then bound this by

$$\leq e^{-\lambda s} \int_0^\infty e^{-\lambda r} Q_r h(X_t) dr = e^{-\lambda s} R_\lambda h(X_s) = S_s,$$

as desired. □

Earlier in the class (chapter 3), we proved that when our filtration is right-continuous and complete and  $X_t$  is a supermartingale such that  $t \rightarrow \mathbb{E}X_t$  is right-continuous,  $X$  has an **rcll** modification  $X$  which is also a supermartingale. We'll take advantage of this to prove that Markov processes have a similar property:

### Theorem 143

Suppose  $\mathcal{F}_t$  is right-continuous and complete on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ , and  $X_t$  is a Feller process with semigroup  $Q_t$ . Then  $X$  has a modification  $\tilde{X}$  which is also a Markov process with semigroup  $Q_t$  and rcll sample paths.

Proving this directly is not so straightforward – the crucial thing that we used to prove the original result about martingales is Doob's upcrossing inequality (which gives us control over right and left limits). Markov processes don't necessarily have upcrossings in generic spaces  $E$ , but this result is showing that nice enough (Feller) processes give us enough control with martingales.

*Proof sketch.* We'll assume for simplicity that  $E$  is compact, so that  $C_0(E) = C(E)$  (the semigroup is defined on all continuous functions). As a (topological) exercise, there exists a countable subset  $\{f_n\} \subseteq C(E)$  which **separate** the points of  $E$ : in other words, for all  $x \neq y$ , there exists an  $f_n$  such that  $f_n(x) \neq f_n(y)$ .

Consider the set

$$\mathcal{H} = \{R_p f_n : \text{integers } p \geq 1, n \geq 1\},$$

which is a countable set. We showed last time that  $\lambda R_\lambda$  converges to the identity as  $\lambda \rightarrow \infty$ , and thus  $\mathcal{H}$  also separates the points of  $E$  (for any  $x, y$  that are distinct, we can find an  $f_n$  such that  $f_n(x) \neq f_n(y)$ , and then we can pick sufficiently large  $p$  so that  $R_p f_n(x) \neq R_p f_n(y)$ ). Now for any  $h = R_p f_n \in \mathcal{H}$ , define the process

$$S_t^h = e^{-pt} h(X_t) = e^{-pt} R_p f_n(X_t).$$

This is a supermartingale by Lemma 142, and we just need to check that the expected value  $t \rightarrow \mathbb{E}S_t^h$  is right-continuous: this is true because of the Feller property. And now chapter 3 tells us that there is a modification of  $S_t^h$  which is rcll, and now we can simultaneously define the countably many modifications  $\tilde{S}_t^h$  for all  $h \in \mathcal{H}$ . To finish, we take a countable dense subset  $D \subseteq [0, \infty)$  and take the limits

$$\lim_{s \downarrow t, s \in D} X_s(\omega), \quad \lim_{s \uparrow t, s \in D} X_s(\omega).$$

We know that the  $S_t^h$ s are real-valued, while  $X_s$  are  $E$ -valued – the claim we're making is that these limits above exist in  $E$ . And this is because the limit failing to exist would violate the rcll property for some supermartingale  $\tilde{S}_t^h$ . Basically, given two sequences  $s_k \downarrow t$  and  $\tilde{s}_k \downarrow t$ , where all sequences are in  $D$ , where  $X_{s_k}(\omega) \rightarrow x$  and  $X_{\tilde{s}_k}(\omega) \rightarrow y$ , we can take the  $h$  that separates  $x$  and  $y$ . But then  $\tilde{S}_t^h$  does have a limit along both of these sequences, which is a contradiction. Thus,  $X$  has an rcll modification, as desired.

In general, if  $E$  is not a compact space, we can use a one-point compactification when we have a Feller process. Applying the above argument to  $E \cup \{\Delta\}$  gives an rcll modification  $\tilde{X}$  on the larger state space, and we just need to show that  $\tilde{X}$  does not visit this extra point  $\Delta$ . This is basically because we have a Markov semigroup, so all of the probability is supported on  $E$ .  $\square$

### Theorem 144

Take the same setting as the above result. Then  $\tilde{X}$  satisfies the strong Markov property (that is, the Markov property with a stopping time).

(We won't say anything more about this here; the main argument is that it's a Markov process, so it satisfies the simple Markov property, and then we can approximate random times with random times with a discrete set of

possibilities, where being a Feller process helps with continuity.)

We'll now move on to our next topic, **Lévy processes**. These processes don't have to be continuous, so they aren't covered as much in our textbook, but there's a large field of research about them: we can see [4] for more information. In short, this is a specific class of Feller processes (including Brownian motion and Poisson processes) which are "spatially homogeneous."

**Definition 145**

A **Lévy process** is a real-valued process  $X_t$  with stationary and independent increments (for all  $s \leq t$ ,  $X_t - X_s$  is independent of  $\mathcal{F}_s$  and  $X_t - X_s$  is equidistributed as  $X_{t-s}$ ), such that  $X_t$  converges in probability to  $X_0 = 0$  as  $t \rightarrow 0$ .

To understand the spatial homogeneity point, remember that

$$Q_t(x, dy) = \mathbb{P}(X_{s+t} \in dy | X_s = x) = \mathbb{P}(X_{s+t} - X_s \in d(y - X_s) | X_s = x),$$

and the conditioning goes away. So  $Q_t(x, dy)$  depends only on the difference between  $x$  and  $y$ :

$$Q_t(x, dy) = Q_t(d(y - x)).$$

In other words, this is a Markov process such that deciding where we'll be at the next step is independent of our current position.

There are a lot of known facts about these processes, and we'll talk today about the characterizing features of a Lévy process. We know that the characteristic function for  $X_t$  looks like

$$\mathbb{E} [e^{i\theta X_t}] = \mathbb{E} [e^{i\theta(X_{t/2} + (X_t - X_{t/2}))}],$$

and the increments  $X_{t/2}$  and  $X_t - X_{t/2}$  are iid, so we can factor the function. In fact, we can break this up into smaller and smaller chunks, and the idea is that we end up with a characteristic function that looks like

$$\mathbb{E} [e^{i\theta X_t}] = e^{t\psi(\theta)}.$$

Such a process does not have a lot of degrees of freedom: for example,  $X_t$  is **infinitely divisible**, meaning

$$X_t = X_{t/k} + (X_{2t/k} - X_{t/k}) + \dots + (X_t - X_{t-t/k}),$$

where all  $k$  terms on the right side are iid increments. The class of infinitely divisible processes is indeed somewhat restricted:

**Theorem 146**

A Lévy process is characterized by three numbers  $(a, \sigma^2, \nu)$  ( $a$  a drift term,  $\sigma^2$  a rate of diffusion, and  $\nu$  a jump measure), where  $\nu$  is a possibly infinite signed measure on  $\mathbb{R} \setminus \{0\}$  such that  $\int \min\{1, x^2\} \nu(dx) < \infty$ .

*Proof sketch.* First of all, Lévy processes are Feller processes (we can see our textbook for this). Thus, we can assume that we're working with the rcll modification  $X_t$ . An rcll process  $X_t$  has countably many discontinuities, which are the points where

$$\Delta X_t = X_t - X_{t-} = 0.$$

Take the empirical measure of the jumps

$$\eta = \sum_t \delta_{(t, \Delta X_t)}$$

(we can think of this as drawing points in the  $t, X_t$  plane corresponding to the jumps), and  $\eta$  is then a random measure on  $[0, \infty) \times (\mathbb{R} \setminus \{0\})$ . Because our process has stationary and independent increments, two blocks of time must have independent  $\eta$ s, meaning that  $\eta$  must be a Poisson random measure with intensity  $\mathbb{E}\eta = (\text{Leb}) \otimes \nu$ . So  $\nu$  is the distribution of the jumps, and that “takes care of the discontinuity part.”

From here, we want to remove the jumps and end up with a continuous process, and the idea is that what we end up with is basically a Brownian motion. Ideally, we’d subtract off  $\sum_{s \leq t} \Delta X_s$  – there’s only countably many jumps – but we don’t know if this sum is convergent. Instead, we can subtract off all jumps that are large: consider

$$J_t = \sum_{s \leq t} \Delta X_s \mathbf{1}\{|\Delta X_s| > 1\},$$

because there are finitely many such large jumps by the rcll property. Now let  $\varepsilon \in (0, 1]$ , and define

$$M_t^\varepsilon = \sum_{s \leq t} (\Delta X_s - \mathbb{E}\Delta X_s) \mathbf{1}\{|\Delta X_s| \in (\varepsilon, 1]\}.$$

This is also well-defined for any positive  $\varepsilon$ , and on our homework we’ll show that  $M_t^\varepsilon$  has a well-defined limit  $M_t$  as  $\varepsilon \rightarrow 0$ . (We need the second-moment property on  $\nu$ , and then we need the martingale  $L^2$  inequality.)

So now the process  $Y = X - M - J$  is a continuous Lévy process, and this will indeed be related to a Brownian motion via  $at + \sigma B_t$ . The issue is that we’ve **assumed nothing about integrability** in the definition of a Lévy process, and we’ll see how to fill in the details on the homework.  $\square$

Again, the significant point here is that Lévy processes only do two things: the continuous part evolves as a Brownian motion, and the discontinuous part has jumps evolving at a Poissonian rate.

### Theorem 147

Let  $\xi$  be a real-valued random variable. The following are equivalent:

- There exists a Lévy process such that  $X_1$  is equidistributed as  $\xi$ .
- The law of  $\xi$  is infinitely divisible.
- There exists a triangular array  $\xi_{i,j}$  such that each row is an iid sequence of length  $m_n$  and

$$\sum_{j=1}^{m_n} \xi_{n,j} \xrightarrow{d} \xi.$$

This is an important result, because one idea from last semester is that such a triangular array with mild conditions (Lindeberg-Feller) means  $\xi$  must be a normal random variable. But we also know that there are non-Gaussian infinitely divisible random variables (such as the Cauchy, gamma, and Poisson distribution), so we must violate the Lindeberg-Feller conditions. In particular, this forces the random variables in our array to be “heavy-tailed,” and the theorem above is interesting because this theorem covers **all** iid triangular arrays with row sums converging to a limit (in contrast with Lindeberg-Feller).

We’ll close with a brief note about our final topic, **approximation of Markov chains**. Recall that last time, we had a Feller semigroup  $Q_t$  with generator  $L$ , and we mentioned that

$$L^{(\lambda)} = \lambda(\lambda R_\lambda - I)$$

generates a pseudo-Poisson process, with an associated semigroup  $Q_t^{(\lambda)} = \exp(tL^{(\lambda)})$ . The Yosida approximation theorem from last time stated that  $Q_t^{(\lambda)} \rightarrow Q_t$  as  $\lambda \rightarrow \infty$ , and now if  $X_t^{(\lambda)}$  is such a realization – a process with semigroup  $Q_t^{(\lambda)}$  – we want to know whether  $X^{(\lambda)}$  converges in distribution to  $X$ .

One case this can help us understand is whether a simple random walk converges in distribution to Brownian motion. As another example, the average of  $n$  Cauchy random variables is Cauchy (we can use the characteristic function here), and we can ask whether the process  $X_t^{(n)} = S_{\lceil nt \rceil} / n$  converges as a distribution to a continuous time analog (the Lévy process with  $X_1$  distributed according to the Cauchy distribution), given that  $X_t^{(n)}$  converges to  $t \cdot (\text{Cauchy})$  for any fixed  $t$ .

The first issue we need to worry about is the topology for convergence in distribution: if we have an rcll random variable that can also have jumps, the **sup-norm topology is not a good choice** anymore. For example, the process  $X^{(n)}(t)$  which is 0 for some time  $1 + \frac{1}{n}$  and then takes on the value 1 after that should converge to the process which jumps up to 1. But this isn't true in the sup-norm topology: instead, we must use the **Skorohod topology**, which allows us to slightly reparameterize the time. And from here, the idea is that semigroup convergence is equivalent to weak convergence in the Skorohod topology for Feller processes. It turns out that semigroup convergence is easier to show in these kinds of situations (we can check each  $Q_t$  on its own), and this means it's easier for us than showing convergence in law directly.

## 18 April 22, 2020

We'll quickly finish discussing Markov processes today (chapter 6) and move on to some preparation for potential theory (chapter 7). Recall that we started by discussing Markov chains in discrete time in a finite state space: in such a case, our chain is completely specified by a transition matrix  $P$ , and under mild conditions (irreducible and aperiodic), the Perron-Frobenius theorem guarantees convergence in law to a unique stationary distribution  $\pi^*$  (that is,  $\pi^*P = \pi^*$ ).

We've been discussing how to extend this to continuous time: the dynamics of the system are now specified by a semigroup  $Q_t$ , and in a particularly nice class of processes known as Feller processes, we have a generator  $L$  that determines the semigroup, and we know that the  $\lambda$ -resolvent satisfies  $R_\lambda = (\lambda - L)^{-1}$  here. A further subclass of these Feller processes is the space of pseudo-Poisson processes, where the generator  $L$  just looks like  $c(P - I)$ : in such a case,  $Q_t = \exp(tL)$ , and the Yosida approximation tells us that a Feller process can be approximated by these pseudo-Poisson processes.

Finally, a different subclass of the Feller processes is the Lévy processes, which include standard Brownian motion  $B_t$ , standard Poisson process  $N_t$ , and the first passage time process  $T_a$  (the infimum  $t$  such that  $B_t > a$ ). Then  $(T_a)_{a \geq 0}$  is a Lévy process because it has independent increments, and we can completely characterize its behavior:

### Proposition 148

We can represent  $T_a$  via

$$T_a = \int_{[0, a] \times [0, \infty]} x \eta(ds dx),$$

where  $\eta$  is a Poisson random measure with intensity  $\text{Leb} \otimes \frac{1_{\{x > 0\}}}{\sqrt{2\pi x^{3/2}}}$ .

This is a special case of the processes on our homework: it's good to see the details of this result worked out if we haven't seen something like this before, and we can see this on the online notes.

With this, we'll move on to potential theory – we'll first cover the subject in a discrete setting because there's



enough going on already. The central idea of what we'll be doing is **relating Markov chains to Dirichlet problems**. Suppose we have a discrete state space  $V$ , and there is a Markov chain  $Y_n$  evolving in  $V$ .

**Definition 149**

A Markov chain  $Y_n$  is **reversible** if there exists a symmetric function  $c : V \times V \rightarrow [0, \infty)$  such that

$$p(x, y) = \frac{c(x, y)}{c(x)},$$

where  $c(x) = \sum_y c(x, y)$ .

We'll only discuss reversible Markov chains here: any such chain can be described with a weighted graph  $G = (V, E, c)$ , where the vertices are the elements of the state space and the (undirected) edges are of the form  $(xy)$ , where  $c(x, y) > 0$ . (For any edge  $e$ ,  $c(e)$  denotes the **conductance**.)

We can also define the **weighted adjacency matrix**  $A = \{c(x, y)\}$ : we'll assume there are no self-loops, so the matrix has zeros on the diagonal. If we let  $D$  be the diagonal matrix such that  $D(x, x) = c(x)$  (corresponds to the degree of the vertex  $x$ ), we know that the transition matrix of the chain must be  $P = D^{-1}A$ . For a function  $u : V \rightarrow \mathbb{R}$ , consider the function

$$Lu(x) = \mathbb{E}_x(u(Y_1) - u(Y_0)) :$$

this is a discrete-time version of the generator, and it measures the expected change in one step of the chain, given that we started the chain at  $x$ . But this is just

$$= (Pu)(x) - u(x) = (P - I)u(x),$$

and here  $L = P - I$  is the **weighted graph Laplacian**: we can write it as  $L = P - I = D^{-1}A - I = D^{-1}(A - D)$ .

**Remark 150.** *This is called the graph Laplacian, because we can consider the case where we have a random walk on  $G = \mathbb{Z}^d$ . Then  $x$  is a point in  $\mathbb{Z}^d$ , and there are  $2d$  possibilities for where we can go:*

$$Lu(x) = \frac{1}{2d} \sum_{i=1}^d (u(x + e_i) - u(x)) - (u(x) - u(x - e_i)).$$

*This is a difference of first derivatives, so this behaves like a second derivative in the  $i$ th coordinate (summed over  $i$ ). So this is approximately*

$$\approx \frac{1}{2d} \sum_i \partial_i^2.$$

From here on out, we'll assume that  $G$  is a finite connected graph: we'll be interested in **potential functions** (also called **harmonic functions**), which are essentially functions  $u$  such that  $Lu = 0$ . But if we require that everywhere, this means  $u$  **must be a constant**, because we can consider  $x \in \operatorname{argmax}(u)$ . Since  $u$  is harmonic,

$$Lu(x) = 0 \implies u(x) = \mathbb{E}_x(u(Y_1)).$$

But  $u(x)$  is the largest possible value of  $u$ , and the right hand side is a weighted average of  $u$ , so this means  $u(y) = u(x)$  for all  $y$  adjacent to  $x$ . Continuing throughout the connected graph, this means that  $u$  is constant: this is known as the **maximum principle**.

So to make things interesting, we'll take some subset  $B$  of the vertices  $V$ , which we call the **boundary**  $B$ , and we won't require the function  $u$  to be harmonic on  $B$ . Then we get a more general maximum principle:

### Lemma 151

Let  $G$  be a finite connected graph with boundary  $B$  and interior  $U = V \setminus B$ , and let  $u : V \rightarrow \mathbb{R}$  be a function such that  $(Lu)|_U = 0$  and  $u|_B = 0$  (harmonic on the interior, zero on the boundary). Then  $u = 0$ .

*Proof.* Consider  $x \in \operatorname{argmax}_{x \in U}(u(x))$ : again,  $u(x)$  is the weighted average of the  $u$ s around it, which implies that  $u(y) = u(x)$  for all  $y \sim x$ . If we continue in this way, we eventually reach the boundary (where the value is zero), and this tells us that everything must be zero.  $\square$

### Definition 152

A **Dirichlet boundary value problem** consists of finding a harmonic function  $u : V \rightarrow \mathbb{R}$  such that  $(Lu)|_U = 0$  and  $u|_B = f$  for some boundary condition  $f$ . If such a  $u$  exists, the solution is called the **harmonic interpolation** of  $f$ .

Note that if we have two solutions  $u', u''$  with the same boundary data  $f$ , then  $u' - u''$  is still a harmonic function, and now this function is zero on the boundary, so Lemma 151 tells us that  $u' - u'' = 0$ . Thus, the harmonic interpolation must be unique.

To show existence, we can just write down a solution directly: we just define

$$u(x) = \mathbb{E}_x f(Y_\tau),$$

where  $\tau$  is the first hitting time of the boundary. This has the correct boundary conditions, because  $\tau = 0$  for any  $x \in B$ , meaning that  $u(x) = f(x)$ , and we can check for ourselves that  $Lu(x) = 0$  for all  $x \in U$ . (This means there is always a unique harmonic interpolation from the boundary to the rest of the graph, as long as both the boundary and the interior are nonempty. And the boundary does not need to correspond in any visual sense to an actual boundary.)

The reason these are also known as potential functions is that we can view  $G = (V, E, c)$  as a wiring diagram of an **electrical network**. Basically, we take some nodes and connect them with wires, where each wire has a number  $c(e)$  attached to it, representing the **electrical conductance** of  $e$  (then  $r(e) = \frac{1}{c(e)}$  is the **electrical resistance** we might have seen in 8.02).

There are two main things we should know about electrical networks:

- Ohm's law: if we hold two vertices  $x$  and  $y$  at fixed voltages  $v(x)$  and  $v(y)$  (for instance, the two ends of a battery), then that imposed voltage difference causes an electrical current

$$i(x, y) = \frac{v(x) - v(y)}{r(xy)}.$$

(This is how we define the current  $i$ , and we'll take the convention that everything "flows downhill" from positive to negative voltage.) We can also write this as

$$i(x, y) = c(x) \frac{c(xy)}{c(x)} (v(x) - v(y)) = c(x) p(x, y) (v(x) - v(y)).$$

Now if we have a **voltage function**  $v : V \rightarrow \mathbb{R}$  which tells us that voltage at each vertex, then the **net current** into the node  $x$  from its neighbors  $y$  is

$$(\operatorname{div} i)(x) = \sum_{y \sim x} i(y, x).$$

Substituting the previous expression in, this is equal to

$$c(x) \sum_{y \sim x} p(x, y)(v(y) - v(x)) = c(x)Lv(x).$$

- Kirchoff's node law: if  $x$  is not connected to an external electrical source or sink, then the current into  $x$  is the same as the current out of  $x$ . (If electrons flow in, they also need to flow out.) In other words, this means that  $Lv(x) = 0$ : the voltage function should be harmonic everywhere other than the sources and sinks.

This means that the Dirichlet boundary value problem with boundary condition  $f : B \rightarrow \mathbb{R}$  will take on the value of the voltage function  $v : V \rightarrow \mathbb{R}$  if we impose voltages  $f$  on  $B$ . This is a nice view, because it's easy to calculate properties of electrical networks: for example, having two edges in parallel with conductance  $c_1$  and  $c_2$  is equivalence to a single edge with conductance  $c_1 + c_2$ , and having two edges in series with resistance  $r_1$  and  $r_2$  yields a single edge with resistance  $r_1 + r_2$ . So there are various rules for reducing a network, and sometimes this reduction allows us to take a complicated network and reduce to a single wire with an effective conductance (if we, for example, impose a voltage of 1 on one vertex and a voltage of 0 on another). We'll make this more formal:

**Definition 153**

Suppose our boundary  $B \subset V$  is further partitioned into  $A \sqcup Z$ , where  $A$  is the set of **sources** and  $Z$  is the set of **sinks**. If we have the boundary condition  $f = 1_A$  (connect the positive end of a battery to  $A$  and the negative end to  $Z$ ), then there exists a **total current**

$$I(A \rightarrow Z) = - \sum_{x \in A} (\text{div } i)(x) = \sum_{x \in Z} (\text{div } i)(x).$$

Then we define the **effective resistance**  $R_{\text{eff}}(A \rightarrow Z) = \frac{1}{I(A \rightarrow Z)} = \frac{1}{C_{\text{eff}}(A \rightarrow Z)}$ .

We can check that the effective resistance is the same if we switch  $A$  and  $Z$ , so we often represent the effective resistance with a double arrow  $R_{\text{eff}}(A \leftrightarrow Z)$ . This is interesting from a probabilistic point of view, because we can write things in terms of these conductances and resistances. For simplicity, we can take  $A = \{a\}$  to be a single point, and we can calculate the **escape probability** from  $a$  to  $Z$

$$\mathbb{P}(a \rightarrow Z) = \mathbb{P}_a(\inf\{n \geq 1 : Y_n \in Z\} < \inf\{n \geq 1 : Y_n = a\})$$

Basically, this tracks the probability that we escape to  $Z$  before returning to  $a$ . To calculate this, we can condition on the first step of the chain:

$$\mathbb{P}(a \rightarrow Z) = \sum_{y \sim a} \mathbb{P}_a(Y_1 = y) \mathbb{P}_y(\inf\{n \geq 0 : Y_n \in Z\} < \inf\{n \geq 0 : Y_n = a\}).$$

But now the second probability is  $\mathbb{E}_y(1_Z(Y_\tau))$ , where  $\tau$  is the first hitting time of  $a \cup Z$ . And now

$$\mathbb{E}_y(1_Z(Y_\tau)) = \mathbb{E}_y(1 - 1\{Y_\tau = a\}),$$

and the expectation here is just the voltage function: thus

$$\mathbb{P}(a \rightarrow Z) = \sum_{y \sim a} p(a, y)(1 - v(y)),$$

where  $v$  is the voltage function where  $v(a) = 1$  and  $v(Z) = 0$ . This may look familiar if we write it as

$$= \sum_{y \sim a} p(a, y)(v(a) - v(y)) = -Lv(a) :$$

as we derived earlier, this is therefore equal to

$$= -\frac{(\operatorname{div} i)(a)}{c(a)} = \frac{I(a \rightarrow Z)}{c(a)} = \frac{C_{\text{eff}}(a \leftrightarrow Z)}{c(a)}.$$

So there is a nice equivalence between a probabilistic quantity and an electrical one here!

We'll spent the time remaining to talk about the discrete Green kernel. Suppose again that we have a boundary  $B \subset V$  and interior  $U = V \setminus B$  – everything we define here depends on our choice of  $U$ , but we'll omit the subscript. The **Green kernel**

$$G_x(y) = \mathbb{E}_x \left( \sum_{n=0}^{\tau-1} 1\{Y_n = y\} \right),$$

the **expected number of times to hit**  $y$ , depends on  $\tau$ , which is the first time we leave  $U$ :

$$\tau = \inf\{n \geq 0 : Y_n \in B\}.$$

$G_x(y)$  can also be rewritten as

$$= \sum_{n=0}^{\infty} \mathbb{E}_x (1\{Y_n = y\}; \tau > n) = \sum_{n=0}^{\infty} p_n(x, y),$$

where  $p_n$  is the probability that we hit  $y$  after  $n$  steps, assuming we haven't left  $U$ . By reversibility, we know that

$$c(x)p_n(x, y) = c(y)p_n(y, x),$$

so

$$g_y(x) = \frac{G_x(y)}{c(y)} = \frac{G_y(x)}{c(x)}$$

is symmetric in  $x$  and  $y$ . If we let  $G$  be the matrix with  $G(x, y) = G_x(y)$ , then

$$(D^{-1}G)(x, y) = g_x(y)$$

is a symmetric matrix. The key identity we should keep in mind is that for any  $x \in U$ , we can condition on the first step of the chain:

$$g_x(x) = \frac{G_x(x)}{c(x)} = \frac{1}{c(x)} (1 + \mathbb{E}_x(G_{Y_1}(x))) = \frac{1}{c(x)} \left( 1 + \frac{\mathbb{E}_x(G_x(Y_1))c(x)}{c(Y_1)} \right) = \frac{1}{c(x)} + \mathbb{E}_x g_x(Y_1).$$

Comparing the left and right sides,

$$Lg_x(x) = -\frac{1}{c(x)},$$

and we can check that for any other point  $y \in U \setminus x$ , we have  $(Lg_x)(y) = 0$ . (The only contribution comes from the initial  $+1$  that we get from the starting point.) This means that  $Lg_x = -\frac{1}{c(x)}$ , from which it follows that we have the matrix identity

$$LG = -I$$

on  $U$ . So we can define  $G$  in the probabilistic way (expected number of visits), but it turns out to also be equal to the matrix inverse of the Laplacian. We should be a bit careful: remember that the definition of  $L$  does not depend on the choice of  $U$ , but in the identity above, we're restricting  $L$  to only contain the rows and columns for the interior  $U$ .

# 19 April 27, 2020

To give a bit more time to think about the homework, the deadline is pushed to Thursday. To finish the class, there will be a test next Thursday, May 7, and a final problem set on Tuesday, May 12.

Last time, we started discussing potential theory for discrete space and time: we let  $G = (V, E, c)$  be a weighted graph, where the weights determine the transition of the Markov chain. We then defined the weighted adjacency matrix

$$A(i, j) = c(ij),$$

and letting  $D = \text{diag}(c(x) = \sum_y c(x, y))$  be the matrix which tracks the outdegree from each vertex, we know that  $P = D^{-1}A$  is the transition matrix, and  $L = P - I = D^{-1}(A - D)$ . This was helpful for solving the Dirichlet problem: we have a vertex set  $V = B \sqcup U$ , and we want to find a function  $u : V \rightarrow \mathbb{R}$  such that  $Lu = 0$  on  $U$  and  $u = f$  on  $B$ . Such a function exists and is unique, and the answer is given by

$$u(x) = \mathbb{E}_x f(Y_\tau),$$

where  $\tau$  is the first hitting time of the boundary  $B$ . And we can write this in another way using the Green kernel:

$$G_U(x, y) = \mathbb{E}_x \left( \sum_{n=0}^{\tau-1} 1\{Y_n = y\} \right)$$

tracks the total number of visits to  $y$ , and we found that  $G_U = -(L_U)^{-1}$  can be written in terms of the Laplacian matrix (only taking the rows and columns from  $U$ ). And we can use this to rewrite the solution  $u$ : writing the Laplacian in block form,

$$\begin{bmatrix} L_U & L_{UB} \\ L_{BU} & L_B \end{bmatrix} \begin{bmatrix} u_U \\ u_B \end{bmatrix} = \begin{bmatrix} 0 \\ * \end{bmatrix},$$

so we must have

$$L_U u_U + L_{UB} u_B = 0 \implies u_U = -(L_U)^{-1} L_{UB} f = G_U L_{UB} f.$$

Here,  $K_U = G_U L_{UB}$  is also known as the discrete Poisson kernel: we can write it as

$$K_U(x, z) = \sum_y G_U(x, y) L_{UB}(y, z),$$

where the only nonzero terms here come from  $x, y \in U, z \in B$ , so this can also be written as

$$\sum_{y \in U} (G_U(x, y) - G_U(x, z)) p(y, z)$$

$G_U(x, z)$  is just zero here, but there will have a continuous analog which is important: the difference will become a derivative.

All of this has an analog in the continuous setting, which is what we'll talk about today. We won't talk about things in full generality – we started with a general weighted graph in the discrete case, and it's possible to similarly use a general Feller process with infinitesimal generator  $L$ , which takes the place of the discrete Laplacian  $L$ . But in our case, we'll just discuss Brownian motion in  $\mathbb{R}^d$ , so that  $L$  is just  $\frac{1}{2}\Delta$ .

### Definition 154

Let  $U \subseteq \mathbb{R}^d$  be an open subset. A function  $u \in L^1_{\text{loc}}(U)$  (locally integrable – bounded on compact sets is strong enough) satisfies the **mean value property** on  $U$  if for all  $x \in U$  and  $r > 0$  such that  $\overline{B_r(x)} \subset U$ ,

$$u(x) = \frac{1}{|B_r(x)|} \int_{B_r(x)} u(y) dy = \frac{1}{|S_r(x)|} \int_{S_r(x)} u(y) dy$$

where  $|B_r(x)|, |S_r(x)|$  denotes the volume of the ball  $B_r(x)$  and sphere  $\partial B_r(x)$ .

We'll use the following analysis fact:

### Theorem 155

A function  $f \in L^1_{\text{loc}}(U)$  satisfies the mean value property on  $U$  (in particular,  $f$  is twice continuously differentiable and in fact smooth) if and only if  $f$  is harmonic on  $U$  – that is,  $\Delta f = \sum_{i=1}^n \partial_i^2 f = 0$ .

We'll now move to the continuous Dirichlet problem: we'll assume for simplicity that  $U$  is a bounded domain, so the boundary condition is a continuous function  $f : \partial U \rightarrow \mathbb{R}$ . Our goal is then to find a continuous harmonic function  $u : \overline{U} \rightarrow \mathbb{R}$  such that  $\Delta u = 0$  on  $U$  and  $u = f$  on  $\partial U$ .

It turns out that the solution looks similar to the discrete case as long as we have some regularity condition:

### Theorem 156

If  $U$  satisfies the exterior cone condition, then the solution to the Dirichlet problem is

$$u(x) = \mathbb{E}_x(f(B_\tau)),$$

where  $\tau$  is the hitting time of the boundary ( $\tau = \inf\{t : B_t \notin U\}$ ).

We won't actually deal with domains that don't satisfy the exterior cone condition in this class, so we won't worry too much about that detail.

*Proof sketch.* Because  $U$  is a bounded domain,  $\partial U$  is compact. Since  $f$  is continuous and defined on a compact domain, it is bounded, and thus the function  $u$  we define above is bounded (in particular, it's definitely in  $L^1_{\text{loc}}$ ). We must show  $u$  satisfies the boundary condition (though we use the exterior cone condition here), and we want to check the mean value property. If we consider a ball  $B_r(x)$  and let  $\sigma$  be the hitting time of the boundary  $S_r(x)$  when we start from  $x$ , then the strong Markov property tells us that

$$\mathbb{E}_x(f(B_\tau) | \mathcal{F}_\sigma) = \mathbb{E}_{B_\sigma}(f(B_\tau)) = u(B_\sigma),$$

while the law of iterated expectations tells us that

$$u(x) = \mathbb{E}(\mathbb{E}(f(B_\tau) | \mathcal{F}_\sigma) | B_0 = x) = \mathbb{E}(u(B_\sigma) | B_0 = x) = \frac{1}{|S_r(x)|} \int_{S_r(x)} u(y) dy.$$

because we're going to uniformly hit a point on the sphere. Setting these equal, we indeed find that the mean-value property is satisfied.  $\square$

We'll now examine the continuous Green kernel: we can't exactly define a "number of visits" in  $\mathbb{R}^d$ , but we can just use a density instead.

**Definition 157**

Let  $U \subseteq \mathbb{R}^d$ , and let  $p_t(x, y; U)$  be the transition kernel of Brownian motion that is killed upon exiting  $U$  (that is,  $\mathbb{P}_x(B_t \in A; \tau_U > t) = \int_A p_t(x, y, U) dy$  for all  $A$ ). The **Green kernel** on  $U$  is defined to be

$$G_U(x, y) = \int_0^\infty p_t(x, y; U) dt.$$

In other words, the total time that we expect to spend in a set  $A$  is

$$\mathbb{E}_x \left[ \int_0^\infty 1\{B_t \in A\} dt \right] = \int_A G_U(x, y) dy.$$

This integral will be finite except maybe at  $y = x$  because we have a bounded domain, but we won't worry too much about those concerns. We showed in the discrete case that  $G_U = -(L_U)^{-1}$ , which means that the function  $G_U(x, \cdot)$  is harmonic on  $U \setminus \{x\}$  and  $(L G_U)(x) = -1$ . The first statement is still true in the continuous case, but the second doesn't work because our function  $G_U$  is singular at  $x$  – we'll need to restate the "inverse" condition.

Let  $P_{t,U}$  be the operator such that

$$P_{t,U}f(x) = \int_U p_t(x, y; U) f(y) dy.$$

(This can also be rewritten as  $\mathbb{E}_x(f(B_t); \tau_U > t)$ .) Also, let  $G_U$  be the operator such that

$$G_U(f(x)) = \int_U G_U(x, y) f(y) = \int_0^\infty P_{t,U}f(x) dt.$$

**Theorem 158**

The Green kernel inverts the Laplacian: for any smooth function  $f$  with compact support in  $U$ ,  $-\frac{1}{2}\Delta_U G_U f = f$ .

*Proof sketch.* If we consider the quantity

$$\frac{1}{t} \int_0^t P_{s,U}f(x) ds = \frac{1}{t} \left( \int_0^\infty P_{s,U}f(x) ds - \int_t^\infty P_{s,U}P_{s,U}f(x) ds \right),$$

the first term is  $G_U f(x)$ , while the second part can be rewritten as  $P_{t,U}G_U f(x)$  by Chapman-Kolmogorov. Thus, this can also be written as

$$= -\frac{1}{t}(P_{t,U} - I)G_U(f(x)).$$

Taking  $t \rightarrow 0$ , the original expression will converge to  $f$ , and the  $\frac{P_{t,U} - I}{t}$  in the final expression will converge to the generator of the process, which is  $\frac{1}{2}\Delta_U$ .  $\square$

Everything we've been discussing so far has used probabilistic quantities to say things about harmonic functions, but we can also work in reverse: in  $\mathbb{R}^d$ , we have the standard Green kernel

$$\Gamma(x, y) = \frac{|x - y|^{2-d}}{(d-2)|S^{d-1}|}$$

where  $|S^{d-1}|$  is the volume of the standard sphere, whenever  $d \neq 2$ , and

$$\Gamma(x, y) = \frac{1}{|S^1|} \log \frac{1}{|x - y|}$$

for  $d = 2$ . Note that  $\Gamma(x, \cdot)$  is always harmonic on  $\mathbb{R}^d \setminus x$ , which we can check by directly taking the derivative. This helps us calculate probabilities for Brownian motion: if we have two balls of radius  $\varepsilon$  and  $R$  centered at the origin, and we start the process at some  $x$  in the annulus  $U$  between the balls, we can stop when we hit either boundary and consider the quantity  $\mathbb{P}_x(\tau_\varepsilon < \tau_R)$ : this is also

$$\mathbb{E}_x f(B_\tau),$$

where  $f : \partial U \rightarrow \mathbb{R}$  is defined so that it is 1 on the inner boundary  $\partial B_\varepsilon$  and 0 on the outer boundary  $\partial B_R$ . So we've constructed the harmonic interpolation

$$u(x) = \mathbb{P}_x(\tau_\varepsilon < \tau_R) = \begin{cases} \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - \varepsilon^{2-d}} & d \neq 2 \\ \frac{\log R - \log |x|}{\log R - \log \varepsilon} & d = 2 \end{cases}$$

(using the fact that  $\Gamma(x, \cdot)$  is harmonic, and verifying that the boundary conditions are satisfied). So this tells us an exact probability for the Brownian motion hitting distance  $\varepsilon$  before distance  $R$ , and now if we take  $R \rightarrow \infty$ , the chance that  $\tau_\varepsilon < \tau_R$  goes to 1 for  $d = 1, 2$ , but it goes to  $\left(\frac{\varepsilon}{|x|}\right)^{d-2} < 1$  for  $d \geq 3$ . So Brownian motion starting at  $x$  hits any ball not containing  $x$  in dimensions 1 or 2, but not necessarily in larger dimensions. (Thus, Brownian motion is **recurrent** for  $d = 1$  or 2 but **transient** otherwise.)

Now, we can compare our  $G_U(x, y)$  (defined for a bounded  $U$ ) to the classical  $\Gamma(x, y)$  (defined for the whole space). In  $\mathbb{R}^d$ ,  $\Gamma$  inverts the Laplacian, meaning that integrating against a smooth, compactly supported test function  $f$  yields

$$\int_{\mathbb{R}^d} \Gamma(x, y) \Delta_y f(y) dy = f(x).$$

It turns out that we can take  $U = \mathbb{R}^d$  in the definition of  $G_U(x, y) = \int_0^\infty p_t(x, y; U) dt$ : we can check that

$$\int_0^\infty p_t(x, y; U) dt = 2\Gamma(x, y)$$

for  $d \geq 3$  but not for  $d = 1, 2$  (because the integral diverges). Nevertheless, we can still get a relation between the Green kernel (with occupation densities) and the classical kernel: we recenter by taking any fixed vector  $w$  of norm 1, and then we have

$$\int_0^\infty (p_t(x, y; \mathbb{R}^d) - p_t(x, x + w; \mathbb{R}^d)) dt = 2\Gamma(x, y).$$

We'll finish by discussing the Feynman–Kac formula: we'll discuss a discrete-time version of it to give some intuition (for the continuous-time version on our homework). As before, let  $G = (V, E, c)$  define a reversible Markov chain  $Y_n$ . Let  $f : V \rightarrow \mathbb{R}$  be a function, and let  $w : V \rightarrow [0, \infty)$  be another nonnegative function: define

$$u(n, x) = \mathbb{E}_x \left[ f(Y_n) \prod_{k=0}^{n-1} \frac{1}{1 + w(Y_k)} \right].$$

Then we can calculate  $u(n+1, x)$  by conditioning on the first visit of the chain:

$$u(n+1, x) = \sum_y p(x, y) \frac{1}{1 + w(x)} u(n, y) = \frac{1}{1 + w(x)} \sum_y (p(x, y)(u(n, y) - u(n, x))) + u(n, x),$$

and the sum over  $y$  in the right expression is just the discrete Laplacian  $L_x u(n, x)$ , meaning that

$$u(n+1, x) - u(n, x) = L_x u(n, x) - w(x)u(n, x).$$

This is a discrete PDE – our initial condition here is that  $u(0, x) = f(x)$ , and  $u(n, x)$  gives us the final solution. In the continuous version of this result, we're similarly given two functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $w : \mathbb{R}^d \rightarrow [0, \infty)$ : then similarly



defining

$$u(t, x) = \mathbb{E}_x \left[ f(B_t) \exp \left( - \int_0^t w(B_s) ds \right) \right],$$

we can find (see homework) that  $u$  solves the partial differential equation

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2} \Delta_x u(t, x) - w(x)u(t, x)$$

such that  $u(0, x) = f(x)$ . A special case, where  $w = 0$ , just gives us the **heat equation** (heat diffuses like Brownian motion), and the solution in that case is  $u(t, x) = \mathbb{E}_x f(B_t)$ . At equilibrium, we know that  $\frac{\partial u}{\partial t}$  should be zero, so the only way for the function to not be constant is if we have a nonconstant boundary condition.

## 20 April 29, 2020

We'll start with Chapter 8 of Le Gall today, discussing stochastic differential equations, existence and uniqueness of solutions, and the case where we have Lipschitz coefficients.

### Definition 159

Let  $\sigma, b : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$  be real-valued functions of time and space that are locally bounded and measurable. A **(weak) solution** of the stochastic differential equation (SDE)

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

(this is the usual informal notation) consists of

- a filtered probability space  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$  (we'll assume the filtration  $\mathcal{F}_t$  is complete),
- an  $\mathcal{F}_t$  Brownian motion  $B_t$ , and
- an  $\mathcal{F}_t$ -adapted process  $X_t$  with continuous sample paths, such that

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

We'll define the right-hand side of this equation to be  $\Phi(X)_t$ : notice that  $\Phi$  then implicitly depends on  $\Omega, \mathcal{F}, \mathcal{F}_t$ , and  $B$ . Because  $\sigma$  and  $b$  are locally bounded,  $\int_0^t \sigma(s, X_s)dB_s$  is a local martingale  $M(X)_t$  and  $\int_0^t b(s, X_s)ds$  a finite variation process  $A(X)_t$  (as we're used to). In addition, if  $X_0 = x \in \mathbb{R}$ , then we say  $x$  is a **solution for  $E_x(\sigma, b)$** .

We'll focus on the one-dimensional case, though many results generalize to the multi-dimensional case. The way to think about this SDE is that it is the system governed by the ODE

$$\frac{df}{dt} = b(t, f(t))dt,$$

except with some noise  $\sigma dB_t$ . Note that we have existence and uniqueness of solutions for this ODE under mild conditions, so we'd like to establish an analogous idea for SDEs, but things are a bit more complicated. In particular, there are a few notions of what a "solution" means here:

### Definition 160

Taking the definition of  $E(\sigma, b)$  above, we know what a **weak solution** of the SDE looks like. A **strong solution** satisfies the additional condition that  $X_t$  is adapted to the Brownian filtration  $\sigma(B_s : s \leq t) \subseteq \mathcal{F}_t$ .

The idea is that  $X_t$  appears on both sides of the weak solution, so we may want to solve for an  $X_t$  where all of the randomness comes from the Brownian motion randomness alone.

**Definition 161**

Again, taking the definition of  $E(\sigma, b)$ , we have **weak uniqueness** of solution if all solutions of  $E_x(\sigma, b)$  have the same law. We have **pathwise uniqueness** if given  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$ , any two solutions  $X$  and  $Y$  with  $X_0 = Y_0$  almost surely are indistinguishable.

Assuming that we have weak **existence** (so that we do have a weak solution), pathwise uniqueness is stronger than weak uniqueness. This isn't an obvious fact – we can read the book for an example where an SDE has weak uniqueness but not pathwise uniqueness (we can construct a probability space so that they are not indistinguishable). Showing that pathwise uniqueness implies weak uniqueness is the **Yamada-Watanabe theorem**. The idea is that indistinguishability requires us to look on a single probability space: if we are given  $\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B$ , and a starting point  $x$ , we have a unique solution  $X$ . But then if we have a different probability space and are given  $\Omega', \mathcal{F}', \mathcal{F}'_t, \mathbb{P}', B', x$ , we will also have a unique solution  $X'$ , and the theorem tells us that  $X$  and  $X'$  will have the same law.

First, we'll note a technical result:

**Fact 162 (Gronwell's lemma)**

Let  $g$  be a nonnegative bounded function on  $[0, t]$ , such that there exist  $a, b \geq 0$  such that

$$g(t) \leq a + b \int_0^t g(s) ds$$

for all  $t \in [0, T]$ . Then  $g(t) \leq ae^{bt}$  for all  $t \in [0, T]$ .

This is a fact about deterministic functions – it's basically a calculus fact, so we'll omit the proof. Note that if we have equality, this is easy:  $g'(t) = bg(t)$ , so  $g$  is exponential and satisfies the initial condition  $g(0) = a$ .

We'll be considering a class of processes where we can prove all of the things that we want: specifically, we'll assume our coefficient functions  $\sigma, b : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$  are continuous (jointly as a function of space and time) and  $K$ -Lipschitz in the space coordinate, meaning that

$$|\sigma(t, x) - \sigma(t, y)| \leq K|x - y|, \quad |b(t, x) - b(t, y)| \leq K|x - y|$$

for all  $x, y \in \mathbb{R}$ .

**Theorem 163**

If  $\sigma, b$  are continuous and  $K$ -Lipschitz, then for all  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$  and all  $x \in \mathbb{R}$ , there exists a **strong solution**  $X$  for  $E_x(\sigma, b)$  on the the probability space, and we have **pathwise uniqueness** of solutions (meaning any other solution  $Y$  is indistinguishable from  $X$ , so all solutions are strong).

*Proof.* Let's first show pathwise uniqueness. We are already given  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$ : suppose that  $X, Y$  are both solutions with  $X_0 = Y_0$  almost surely. (For this part, the starting point does not need to be fixed.) Our goal is to show that  $X$  and  $Y$  are indistinguishable.

Let  $\tau$  be the stopping time

$$\tau = \inf\{t \geq 0 : |X_t - X_0| \geq M \text{ or } |Y_t - Y_0| \geq M\}.$$

Using the trivial inequality  $(u + v)^2 \leq 2(u^2 + v^2)$  (for all  $u, v \in \mathbb{R}$ ), we can consider the function

$$h(t) = \mathbb{E} \left[ (X_{t \wedge \tau} - Y_{t \wedge \tau})^2 \right].$$

Because we stop our process before we're more than  $M$  away from the starting point, and  $X$  and  $Y$  have the same starting point, this function is bounded by  $(2M)^2$ . Using that above inequality by writing  $X$  and  $Y$  as a sum of the local martingale and FV parts, we can also write this out as

$$h(t) \leq 2\mathbb{E} \left[ \left( \int_0^{t \wedge \tau} (\sigma(s, X_s) - \sigma(s, Y_s)) dB_s \right)^2 \right] + 2 \left[ \left( \int_0^{t \wedge \tau} (b(s, X_s) - b(s, Y_s)) ds \right)^2 \right].$$

The first term is the expectation of the square of a stochastic integral, and we have the result from earlier study that

$$\mathbb{E} \left[ \left( \int_0^t H_s dM_s \right)^2 \right] \leq \mathbb{E} \left[ \int_0^t H_s^2 d\langle M \rangle_s \right].$$

Using that result and also using Cauchy-Schwarz on the second term, we can bound this as

$$\leq 2\mathbb{E} \left[ \int_0^{t \wedge \tau} (\sigma(s, X_s) - \sigma(s, Y_s))^2 ds \right] + 2t\mathbb{E} \left[ \int_0^{t \wedge \tau} (b(s, X_s) - b(s, Y_s))^2 ds \right].$$

Applying the Lipschitz condition, this is bounded by

$$\leq 2K^2(1+t)\mathbb{E} \left[ \int_0^{t \wedge \tau} (X_s - Y_s)^2 ds \right] \leq 2K^2(1+t) \int_0^t h(s) ds \leq 2K^2(1+T) \int_0^t h(s) ds$$

for all  $t \leq T$ . Now  $h$  is a bounded nonnegative function, and Gronwall tells us that  $h = 0$  on  $[0, T]$  (because the constant term is 0), so  $X_{t \wedge \tau} = Y_{t \wedge \tau}$  almost surely for all  $t \in [0, T]$ : now take  $M \rightarrow \infty$  and  $T \rightarrow \infty$  to show that  $X_t = Y_t$  almost surely for all  $t$ . This doesn't mean (by definition) that they are indistinguishable, but we have continuity, so we do have indistinguishability in this particular case (by assumption of continuity).

Now we'll show existence of a strong solution, and the calculation here will be fairly similar to what we've just done. Remember that we're working with a given  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$ , and a weak solution is just a solution of the fixed point equation  $X = \Phi(X)$ . This motivates the idea of iterating  $\Phi$ : let  $X^0$  be the constant process such that  $(X^0)_t = x$ , and define

$$X^n = \Phi^n(X^0).$$

We want to show that  $X^n$  converges to the fixed point solution  $X = \Phi(X)$ , and we'll do this by bounding the difference between  $X^n$  and  $X^{n+1}$ . Define

$$G_n(t) = \mathbb{E} \left[ \sup_{s \leq t} (X_s^{n+1} - X_s^n)^2 \right].$$

By the same calculation as before, this is bounded by

$$\leq 2\mathbb{E} \left[ \sup_{s \leq t} \left( \int_0^s (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1})) dB_s \right)^2 \right] + 2\mathbb{E} \left[ \sup_{s \leq t} \left( \int_0^s (b(s, X_s^n) - b(s, X_s^{n-1})) ds \right)^2 \right]$$

(where we've gone from  $n+1$  to  $n$  and  $n$  to  $n-1$  on the right-hand side by plugging in the definition of  $X^n$ ). The first term can be controlled because it is the supremum of a local martingale, so we can apply Doob's  $L^2$  inequality, and then we bound that result with the same strategy as above. Then the second is bounded by Cauchy-Schwarz again:

$$\leq 8\mathbb{E} \left[ \int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1}))^2 ds \right] + 2t\mathbb{E} \left[ \int_0^t (b(s, X_s^n) - b(s, X_s^{n-1}))^2 ds \right].$$

Applying the Lipschitz assumption again, this is bounded by

$$\leq 2K^2(4+t)\mathbb{E}\left[\int_0^t (X_s^n - X_s^{n-1})^2 ds\right] \leq \boxed{2K^2(4+T)\int_0^t g_{n-1}(s)ds}.$$

This means we have a bound for  $g_n$  in terms of  $g_{n-1}$ . Note that we have control over

$$g_1(t) = \mathbb{E}\left[\sup_{s \leq t} (X_s^1 - x)^2\right],$$

where  $X_s^1 = \Phi(X^0)_s = x + \int_0^s \sigma(r, x)dB_r + \int_0^s b(r, x)dr$ . Because  $\sigma, b$  are both locally bounded, this function  $g_1(t)$  is bounded by some constant  $c(T)$  for all  $t \in [0, T]$  (notice that we're bounding the **expectation** of the second moment of  $(X_s^1 - x)$ , not the function itself). Inductively integrating the boxed bound above, we know that

$$g_{n+1}(t) \leq c(T)(2K^2(4+T))^n \frac{t^n}{n!}.$$

But this decays quickly because of the  $n!$  in the denominator, which means that summing the sup-norm differences

$$\mathbb{E}\left[\sum_n \sup_{t \leq T} |X_t^{n+1} - X_t^n|\right] \leq \sum_n \sqrt{g(T)} < \infty.$$

Thus,  $\sum_n \sup_{t \leq T} |X_t^{n+1} - X_t^n|$  is almost surely finite for any  $T$ , which means that  $X^n$  converges uniformly to  $X$  on  $[0, T]$  (and generally any compact time interval). We know that  $X_s^1$  is adapted to the Brownian filtration, and in general integrating against a Brownian motion still keeps things adapted. so  $X^n$  is adapted to the Brownian filtration, and thus the limit process  $X$  is adapted as well. So we just need to check that this is actually a (weak) solution to  $X = \Phi(X)$ , but we showed that

$$X^n = \Phi(X^{n-1}),$$

so taking  $n \rightarrow \infty$  makes both  $X^n$  and  $X^{n-1}$  in the above equation converge to  $X$ . So we just need to check that  $\Phi(X^n)$  converges to  $\Phi(X)$ , where

$$\Phi(X^n) = x + \int_0^t \sigma(s, X_s^{n-1})dB_s + \int_0^t b(s, X_s^{n-1})ds.$$

It's clear that the second term converges to  $\int_0^t b(s, X_s)ds$  because  $b$  is continuous and  $X_s^{n-1}$  converges uniformly to  $X^n$  (so we can apply dominated convergence theorem). To argue that the first term converges to what we want, we use the dominated convergence theorem on the difference  $\int_0^t (\sigma(s, X_s) - \sigma(s, X_s^{n-1}))dB_s$  for stochastic integrals, where the dominating process is

$$D_s = K \left[ \sum_n \sup_{r \leq s} |X_r^n - X_r^{n-1}| \right],$$

and this finishes the construction. □

### Theorem 164

Consider the space of functions  $C([0, \infty], \mathbb{R})$ . Let  $W$  be the Wiener measure (the law of Brownian motion started from 0),  $\mathcal{B}_C$  be the Borel sigma-algebra on  $C([0, \infty], \mathbb{R})$ , and let  $\mathcal{G}$  be the sigma-algebra  $\sigma(\mathcal{B}_C, N)$ , where  $N$  is the set of  $W$ -negligible sets. If  $\sigma, b$  are continuous and  $K$ -Lipschitz, then for all  $x \in \mathbb{R}$  there exists a measurable function

$$F_x : (C([0, \infty], \mathbb{R}), \mathcal{G}) \rightarrow (C([0, \infty], \mathbb{R}), \mathcal{B}_C)$$

such that

- for all  $t$ ,  $F_x(w)_t$  coincides almost surely with a measurable function of  $(w(s) : s \leq t)$ ,
- For all  $w$ , the map  $x \rightarrow F_x(w)$  is continuous as a map from  $\mathbb{R} \rightarrow C([0, \infty], \mathbb{R})$ .
- For all  $x$  and for all  $\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B$ ,  $F_x(B)$  is the unique solution of  $E_x(\sigma, b)$ , and this is also true if we replace  $x$  with a random starting point  $U \in \mathcal{F}_0$ .

The second point above tells us that if we start from two points  $x, y \in \mathbb{R}$  that are close to each other, and we're using the same Brownian motion for both, our paths will look similar when  $\sigma, b$  are bounded and Lipschitz. Note that the way to get our function  $F_x(w)$  is to apply the previous theorem with the filtered probability space

$$(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B) = (C([0, \infty], \mathbb{R}), \mathcal{G}, \mathcal{G}_t = \sigma(w(s) : s \leq t, N), W, w)$$

where  $w$  is the canonical Brownian motion. Then the solution we get out of the theorem will be  $F_x(w)$  – this is adapted to the Brownian motion, so we get the first condition above. We just need to show that the mapping is continuous and that this is true on any probability space, and we'll do this next time.

## 21 May 4, 2020

Recall that we've been looking at the stochastic differential equation

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

and we're assuming that  $\sigma$  and  $b$  are  $K$ -Lipschitz in the  $x$  coordinate. Last time, we showed that we can find a strong solution  $X_t^x$  adapted to the Brownian motion and started at  $x$ , given any  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$ , and this solution is unique by **pathwise** uniqueness.

**Remark 165.** *The Yamada-Watanabe theorem, which states that pathwise uniqueness implies weak uniqueness, can be applied here, but we can also prove weak uniqueness directly in this  $K$ -Lipschitz case.*

We'll now start with a proof of last time's result, which stated that we have a measurable mapping

$$F_x : (C([0, \infty), \mathbb{R}), \mathcal{G}) \rightarrow (C([0, \infty), \mathbb{R}), \mathcal{B})$$

such that  $F_x(w)_t$  is measurable of  $(w_s)_{s \leq t}$ , the map  $x \rightarrow F_x(w)$  is continuous in  $x$ , and  $F_x(B)$  solves the stochastic differential equation  $E_x(\sigma, b)$ .

*Proof.* We already showed there exists a strong solution  $X_t^x$  for any  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B, x)$  – we'll apply that here with the space

$$(C([0, \infty), \mathbb{R}), \mathcal{G}, \mathcal{G}_t, W, w, x).$$

We want to show continuity, and showing this is similar to the **Kolmogorov continuity lemma** – recall that if we have any stochastic process  $F_t$  which takes values in a complete separable metric space  $(S, d)$  with the bound

$$\mathbb{E}[d(F_s, F_t)^q] \leq C|s - t|^{1+\varepsilon},$$

then there exists a modification  $\tilde{F}_t$  that is  $\alpha$ -Hölder continuous for all  $\alpha \in (0, \frac{\varepsilon}{q})$ . We applied this to Brownian motion earlier in the class, and remember that one part of the proof was to do a union bound (for the interval  $[0, 1]$ )

$$\mathbb{P}\left(|X_{i/2^n} - X_{(i-1)/2^n}| \leq \left(\frac{1}{2^n}\right)^\alpha \quad \forall i\right) \leq 2^{n(1+\alpha q)}/2^n(1 + \varepsilon).$$

Notice we don't require independent increments here! So we'll apply the Kolmogorov continuity lemma to  $F_x(\omega)$  (where we index by **position**  $x \in \mathbb{R}$  instead of by time), where this process takes values in  $(S, d)$  for

$$S = C([0, \infty), \mathbb{R})$$

and the metric (between functions)

$$d(f, g) = \sum_{n \geq 1} \frac{1}{2^n} \min \left\{ 1, \sup_{t \leq n} |f(t) - g(t)| \right\}.$$

In order to apply this, we need the estimate boxed above – this is a calculation we can read on our own. But once we verify this, we've indeed checked that we have continuity.

The last thing we will check is that  $F_x(B)$  solves our SDE given any  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, B)$  – we know it's true on the Wiener space, but we need to check that it's true if we're given any Brownian motion  $B$ . In other words, we know that  $F_x(w)$  solves  $E_x(\sigma, b)$ , meaning

$$F_x(w)_t - \left( x + \int_0^t \sigma(s, w(s))dw(s) + \int_0^t b(s, w(s))ds \right) = 0,$$

where  $w$  is our Brownian motion. Call this left hand side  $\Psi(w)_t$ : we also know that

$$\int_w |\Psi(w)|dW(w) = 0$$

(integration with respect to the Wiener measure). We know that any Brownian motion  $B$  has the same law as  $W$ , so we must also have

$$\int |\Psi(B)|d\mathbb{P}(B) = 0.$$

As a technical sidenote, we do need to make sure  $\Psi$  is measurable as a function of  $w$  – the main difficulty here is showing that the stochastic integral  $\int_0^t \sigma(s, w(s))dw(s)$  is a measurable function of  $w$ , but this follows from the approximation

$$\int_0^t \sigma(s, w(s))dw(s) = \lim_{n \rightarrow \infty} \sum_{i=1}^{2^n} \sigma\left(\frac{(i-1)t}{2^n}, w\left(\frac{(i-1)t}{2^n}\right)\right) \left( w\left(\frac{it}{2^n}\right) - w\left(\frac{(i-1)t}{2^n}\right) \right).$$

□

Note that this theorem we've just proved implies weak uniqueness – the solution comes from applying the same map  $F_x$  to our Brownian motion, no matter what probability space we're on, so the law of  $X^x$  is given by

$$\mathbb{P}(X^x \in A) = \mathbb{P}(F_x(B) \in A) = \mathbb{P}(B \in (F_x)^{-1}A)$$

(for any event  $A$ ), and because the law of Brownian motion is given by the Wiener measure, this is just  $W((F_x)^{-1}A)$ . So the law is just  $(F_x)_\#W$ , and we don't need to use the Yamada-Watanabe theorem in this case.

**Remark 166.** If we replace our starting point  $X_0 = X$  with a random variable  $X_0 = U \in \mathcal{F}_0$ , we can still get a solution to our SDE with  $F_U(B)$ . We can read the book for more details here as well.

We will now make a connection to Markov processes: suppose for this part of the lecture that  $\sigma$  and  $b$  don't depend on time and that they are still  $K$ -Lipschitz in  $x$ .

**Theorem 167**

If  $\sigma(t, x) = \sigma(x)$  and  $b(t, x) = b(x)$  are  $K$ -Lipschitz, let  $X$  be a solution of the SDE  $E(\sigma, b)$  on any probability space  $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ . Then  $X$  is a Markov process, and the semigroup can be described via

$$Q_t f(x) = \mathbb{E}_x f(X_t) = \int f(F_x(w)_t) dW(w).$$

(Remember that we showed the existence and uniqueness of a strong solution  $X$  already.)

*Proof.* Recall that  $f$  must be bounded and measurable here. We'll first show that

$$\mathbb{E}(f(X_{s+t} | \mathcal{F}_s) = Q_t f(X_s).$$

Define the shifted process

$$\hat{X}_t = X_{s+t} = X_s + \int_s^{s+t} \sigma(X_r) dB_r + \int_s^{s+t} b(X_r) dr,$$

and we can rewrite this with changed time as

$$= \hat{X}_0 = \int_0^t \sigma(\hat{X}_r) d\hat{B}_r + \int_0^t b(\hat{X}_r) dr.$$

Thus,  $\hat{X}_t$  solves the SDE  $E_{\hat{X}_0}(\sigma, b)$  on the probability space  $(\Omega, \mathcal{F}, \hat{\mathcal{F}}_t = \mathcal{F}_{s+t}, \mathbb{P})$  with the shifted Brownian motion  $\hat{B}_t = B_{s+t} - B_s$ . Therefore,  $F_{X_s}(\hat{B})$  must solve our differential equation (by the result we proved above), meaning we have

$$\hat{X}_t = F_{X_s}(\hat{B}) \implies \mathbb{E}[f(X_{s+t}) | \mathcal{F}_s] = \mathbb{E}[f(F_{X_s}(\hat{B})_t) | \mathcal{F}_s]$$

and the right side  $\hat{B}$  is independent of  $\mathcal{F}_s$ , so we can just write this as

$$= \int f(F_{X_s}(w)_t) dW(w) = Q_t f(X_s),$$

as desired.

To finish, we need to show that  $Q_t$  is a valid semigroup, which means that (1)  $Q_0(x, \cdot)$  is the Dirac measure  $\delta_x$  concentrated at  $x$ , (2) the Chapman-Kolmogorov equations  $Q_{s+t} = Q_s Q_t$  are satisfied, and (3) the map  $(t, x) \rightarrow Q_t(x, A)$  must be measurable – this last point follows because  $Q_t$  is continuous in  $t$  and  $x$ . The rest is more straightforward.  $\square$

**Theorem 168**

In the same setting as the theorem above,  $Q_t$  is a Feller semigroup. The space of functions  $A = C^{2,\text{cpt}}(\mathbb{R})$  that are compactly supported and twice differentiable is contained in the domain  $D(L)$ , and we can write

$$Lf(x) = b(x)f'(x) + \sigma^2 \frac{f''(x)}{2}$$

for any  $f \in A$ .

*Proof sketch.* We'll omit the proof of the Feller property, which is showing that (1) whenever  $f \in C_0(\mathbb{R})$ , we have  $Q_t f \in C_0(\mathbb{R})$ , and (2)  $Q_t f \rightarrow f$  converges in the sup-norm as  $t \downarrow 0$ . Then we can apply Itô's formula to  $f(x)$ , which tells us that

$$df(X_t) = f'(X_t)dX_t + \frac{f''(X_t)}{2}\sigma(X_t)^2 dt,$$

and the SDE tells us that this is

$$= f'(X_t)(\sigma(X_t)dB_t + b(X_t)dt) + \frac{f''(X_t)}{2}\sigma(X_t)^2 dt.$$

If we then subtract off the drift term and define

$$M_t = f(X_t) - f(x) - \int_0^t f'(X_s)b(X_s) + \frac{f''(X_s)\sigma(X_s)^2}{2} ds,$$

then  $M_t = \int_0^t \sigma(X_s)dB_s$  is a local martingale. Suppose  $Gf(X_s)$  is the integrand in the boxed expression for  $M_t$  – we want to show that  $G$  is the generator we're looking for. For simplicity, let's assume  $\sigma, b$  are bounded. Then  $M$  is a true martingale, meaning that

$$0 = \mathbb{E}M_t = \mathbb{E}_x f(X_t) - f(x) - \int_0^t \mathbb{E}_x Gf(X_s) ds.$$

Defining through by  $t$  and taking a limit, we find that

$$Lf(x) = \lim_{t \downarrow 0} \frac{Q_t f(x) - f(x)}{t} = \lim_{t \downarrow 0} \frac{1}{t} \int_0^t \mathbb{E}_x Gf(X_s) ds,$$

and we can write this out as

$$= \lim_{t \downarrow 0} \frac{1}{t} \int_0^t Q_s Gf(x) ds.$$

But we know that  $Q_s(Gf)$  converges to  $Gf$  by the Feller property, so this last expression is indeed  $Gf(x)$ , as desired.  $\square$

So **time-independent SDEs with Lipschitz coefficients correspond to Feller processes.**

**Example 169** (Ornstein-Uhlenbeck process)

Consider the SDE

$$dX_t = dB_t - \lambda X_t dt.$$

Consider the process  $M_t = e^{\lambda t} X_t$ : the reason for this is that applying Itô's formula tells us that

$$d(e^{\lambda t} X_t) = e^{\lambda t}(dB_t - \lambda X_t dt) + \lambda e^{\lambda t} X_t dt,$$

and the drift terms cancel out. This means that

$$d(e^{\lambda t} X_t) = e^{\lambda t} dB_t,$$

meaning that  $M_t$  is a local martingale with increments given by integrating a deterministic function  $e^{\lambda t}$  against a Brownian motion. So if we take  $M_0$  to be some integrable function,  $M_t$  will be a true martingale, and

$$e^{\lambda X_t} - X_0 = M_t - M_0 = \int_0^t e^{\lambda s} dB_s,$$

so rearranging yields

$$X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dB_s,$$



and we've found our solution. And now if  $X_0$  is deterministic or normal and independent of  $B$ , then  $X_t$  is actually a **Gaussian process**, where we can calculate that

$$\text{Var}(X_t) = \frac{\text{Var}(X_0)}{e^{2\lambda t}} + \int_0^t \frac{ds}{e^{2\lambda(t-s)}} = \frac{\text{Var}(X_0)}{e^{2\lambda t}} + \frac{1}{2\lambda} \left(1 - \frac{1}{e^{2\lambda t}}\right).$$

So now we can pick the variance of  $X_0$  to be  $\frac{1}{2\lambda}$ , so that  $\text{Var} X_t = \frac{1}{2\lambda}$  is constant for all time  $t \geq 0$ . We can also check that  $\text{Cov}(X_s, X_t) = \frac{1}{2\lambda \exp(\lambda|t-s|)}$ , which only depends on the difference between  $s$  and  $t$ . This means the centered Gaussian process  $X$  is also **stationary**, and such a process is called the Ornstein-Uhlenbeck process.

**Example 170 (Geometric Brownian motion)**

Consider the SDE

$$dX_t = \sigma X_t dB_t + r X_t dt.$$

This is the crudest possible model we can have for the stock market (we have some rate of appreciation, as well as some volatility). The idea here is to apply Itô's formula to  $\log X_t$  (as long as  $X$  stays positive), which yields

$$d(\log X_t) = \frac{1}{X_t} (\sigma X_t dB_t + r X_t dt) - \frac{1}{2X_t^2} \sigma^2 X_t^2 dt.$$

The  $X_t$  drift terms cancel, and we're left with

$$= \sigma dB_t + \left(r - \frac{\sigma^2}{2}\right) dt,$$

so this easily integrates to

$$X_t = X_0 \exp\left(\sigma B_t - \left(r - \frac{\sigma^2}{2}\right) t\right).$$

## 22 May 11, 2020

Today, we'll discuss a few results that are applications of what we've learned in this class, centered around the **Dyson Brownian motion in random matrices**.

**Definition 171**

Let  $(B_{ij})_{i \leq j}$  be a family of iid Brownian motions. Then the **symmetric matrix Brownian motion** is the symmetric matrix  $H$  such that

$$H(t)_{ij} = \begin{cases} \sqrt{2} B_{ij} & i = j \\ B_{ij} & i < j \\ B_{ji} & j < i. \end{cases}$$

In other words, all entries evolve according to independent Brownian motions, except that we want the matrix to be symmetric. We include the  $\sqrt{2}$  factor here because another way that we can obtain this matrix is via

$$H(t) = \frac{X(t) + X(t)^*}{\sqrt{2}},$$

where  $X$  is a standard Brownian motion in  $\mathbb{R}^{n \times n}$  (meaning **all** entries are independent), and then we're just symmetrizing and rescaling. Then all off-diagonal entries evolve like standard Brownian motions, but the diagonal terms will have a larger variance.

### Definition 172

The **Hermitian matrix Brownian motion** is similarly defined as

$$H(t) = \frac{X(t) + X(t)^*}{\sqrt{2}},$$

where  $X$  is a standard Brownian motion in  $\mathbb{C}^{n \times n}$ .

We say that  $\frac{H(t)}{\sqrt{t}}$  for a symmetric BM  $H$  is a sample from the **Gaussian orthogonal ensemble** or **GOE**, and similarly  $\frac{H(t)}{\sqrt{t}}$  for a Hermitian BM  $H$  is a sample from the **Gaussian unitary ensemble** or **GUE**.

We can show that  $H$  sampled from either GOE or GUE will always have  $n$  distinct eigenvalues almost surely, which are real because we have a Hermitian matrix: we'll order them via

$$\lambda_1 < \dots < \lambda_n.$$

In fact, it is true that the ordered eigenvalue process

$$\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))$$

is such that the eigenvalues never collide for all time almost surely: this eigenvalue process does not leave the **Weyl chamber**

$$W_n = \{z \in \mathbb{R}^n : z_1 < \dots < z_n\}.$$

There are two main results we'll be covering today:

### Theorem 173

If  $H$  is a symmetric or Hermitian matrix Brownian motion, then the eigenvalue process solves the  **$\beta$ -Dyson SDE**

$$d\lambda_i(t) = \beta \sum_{\ell \neq i} \frac{1}{\lambda_i(t) - \lambda_\ell(t)} dt + \sqrt{2} dB_i(t),$$

where  $B_i$  are independent Brownian motions, where we must have  $\beta = 1$  in the symmetric case and  $\beta = 2$  in the Hermitian case.

Notice that all eigenvalues  $\lambda_\ell < \lambda_i$  give a positive contribution to the drift term, and all eigenvalues  $\lambda_\ell > \lambda_i$  give a negative contribution. Essentially, the eigenvalues will “repel” each other, because this contribution gets larger as the eigenvalues grow closer – this is related to why the eigenvalues are distinct almost surely for the GOE and GUE.

We won't focus too much on the proof of existence and uniqueness for now – it turns out that for any  $x \in W_n$  in the Weyl chamber, there is a unique strong solution of the  $\beta$ -Dyson SDE started from  $x$  for all  $\beta \geq 1$  which stays inside the chamber for all time.

We'll also cover the following result:

### Theorem 174

For any  $x \in W_n$ , the  $\beta = 2$  Dyson process started from  $x$  is equidistributed as an  $n$ -dimensional Brownian motion started from  $x$ , conditioned to stay inside the Weyl chamber.

(Note that  $n$ -dimensional Brownian motion will almost surely exit the Weyl chamber, because even two Brownian motions will intersect almost surely. So we'll need to be more precise about this statement to work with it.)

We'll first discuss the main ideas of the first result – most of the work is calculation:

*Proof sketch of Theorem 173 for the symmetric matrix case.* Call the entries of our symmetric matrix  $H_{jk}$ . We can calculate the first and second derivative of the  $i$ th eigenvalue with respect to each matrix entry  $H_{jk}$  (for all  $j \leq k$ ). Then we can apply Itô's formula: we must break into separate terms because the Brownian motions are different on the diagonals, so we have

$$d\lambda_i(t) = \left[ \sum_{k=1}^n \frac{\partial \lambda_i}{\partial H_{kk}} \sqrt{2} dB_{kk} + \sum_{j < k} \frac{\partial \lambda_i}{\partial H_{jk}} dB_{jk}(t) \right] + \left[ \frac{1}{2} \sum_{k=1}^n 2 \frac{\partial^2 \lambda_i}{\partial H_{kk}^2} + \frac{1}{2} \sum_{j < k} \frac{\partial^2 \lambda_i}{\partial H_{kk}^2} \right] dt.$$

Call the continuous local martingale term  $\sqrt{2}dB_i(t)$ : we can use the Lévy characterization to check that  $B_i$  is indeed a Brownian motion. And then we can also check that the calculation works out for the finite variation term.

To calculate the needed derivatives, note that

$$\frac{\partial \lambda_i}{\partial H_{jk}} = \frac{d}{dt} \lambda_i(H(t)),$$

where

$$H(t) = H + t(E_{jk} + E_{kj})$$

and  $E_{jk}$  and  $E_{kj}$  are the matrices that are 0 except with a 1 in  $(j, k)$  and  $(k, j)$  respectively. We know that  $Hu_i = \lambda_i u_i$ , and then we can use implicit differentiation.  $\square$

(And the method of proof is the same for the Hermitian case.)

We'll need a technical lemma for the second result:

**Lemma 175 (Andréief integration formula)**

For "nice" functions  $f_i$  and  $g_i$ , the expressions

$$\frac{1}{n!} \int_{\mathbb{R}^n} \det \begin{bmatrix} f_1(x_1) & \cdots & f_1(x_n) \\ \vdots & \ddots & \vdots \\ f_n(x_1) & \cdots & f_n(x_n) \end{bmatrix} \begin{bmatrix} g_1(x_1) & \cdots & g_1(x_n) \\ \vdots & \ddots & \vdots \\ g_n(x_1) & \cdots & g_n(x_n) \end{bmatrix} dx,$$

and

$$\det \begin{bmatrix} \int_{\mathbb{R}} f_1 g_1 & \cdots & \int_{\mathbb{R}} f_1 g_n \\ \vdots & \ddots & \vdots \\ \int_{\mathbb{R}} f_n g_1 & \cdots & \int_{\mathbb{R}} f_n g_n \end{bmatrix}$$

are equal.

*Proof.* Write out both determinants: the left hand side becomes

$$\frac{1}{n!} \int_{\mathbb{R}^n} \left( \sum_{\sigma} \text{sgn}(\sigma) \prod_{i=1}^n f_{\sigma(i)}(x_i) \right) \left( \sum_{\tau} \text{sgn}(\tau) \prod_{j=1}^n g_{\tau(j)}(x_j) \right).$$

This can be rewritten as

$$= \frac{1}{n!} \sum_{\sigma, \tau} \text{sgn}(\sigma\tau) \prod_{i=1}^n \left( \int f_{\sigma(i)}(x_i) g_{\tau(i)}(x_i) dx_i \right)$$

where we're saying that the integral over the product is the product over the integral of the individual independent  $x_i$ s, and then we can sum over all permutations  $\rho = \sigma\tau$ , but we'll count each one  $n!$  times. This yields the formula of the determinant on the right side.  $\square$

**Lemma 176** (Karlin-McGregor formula)

Let  $B(t)$  be a Brownian motion in  $\mathbb{R}^n$ , and let  $T$  be the first exit time of  $B$  from the Weyl chamber. Then for any  $x \in W_n$  and measurable subset  $A \subseteq W_n$ ,

$$\mathbb{P}_x(B(t) \in A; T > t) = \int_A \det \begin{bmatrix} p_t(x_1, y_1) & \cdots & p_t(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p_t(x_n, y_1) & \cdots & p_t(x_n, y_n) \end{bmatrix} dy,$$

where  $p_t$  is the transition density of a usual one-dimensional Brownian motion.

If the BMs were all independent and we didn't care whether they collided or not, the transition kernel would just be  $p_t(x_1, y_1) \cdots p_t(x_n, y_n)$ . So this result is saying that we need to use the determinant of  $p_t$ s instead if we require our BMs not to collide.

Here, we're **restricting** on the left side to the event that we haven't left in time  $t$  (not conditioning).

*Proof.* Let  $T_j$  be the collision time  $\inf\{t : B_i(t) = B_{i+1}(t)\}$ , meaning that  $T$  will be the minimum of  $\{T_1, \dots, T_{n-1}\}$ . Again expanding out the determinant and writing the transition  $p_t$  functions in terms of a Brownian motion yields

$$q_t(x, y) dy = \sum_{\sigma} \text{sgn}(\sigma) \mathbb{E}_x \left( \prod_{i=1}^n 1\{B_i(t) \in dy_{\sigma(i)}\} \left[ 1\{T \geq t\} + \sum_{j=1}^{n-1} 1\{T = T_j < t\} \right] \right)$$

where the last bracket term is just 1. When we look at the contribution  $1\{T \geq t\}$  term, meaning we haven't left the chamber, the only permutation that is relevant is that where the  $y$ 's haven't gone out of order from the  $x$ 's, so this contributes

$$\mathbb{E}_x \left( \prod_{i=1}^n 1\{B_i(t) \in dy_i\}; T \geq t \right).$$

We can make a switching argument to show that the contribution from the other part is zero (if  $T_j$  happens before  $t$ , we can take the two Brownian motions that cross and switch them, changing the sign of the difference). So the density is indeed given by what we have written above, which is the claimed result.  $\square$

In order to condition on "not colliding," we'll want to construct a martingale.

**Corollary 177**

For any  $x \in \mathbb{R}^n$ , let

$$v(x) = \prod_{i < j} (x_j - x_i) = \det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}$$

(this is the classic **Vandermonde determinant**). Then if  $B$  is a Brownian motion in  $\mathbb{R}^n$  and  $T$  is the exit time from the Weyl chamber, then

$$M_t = v(B_{t \wedge T})$$

is a **nonnegative martingale**.

*Proof.*  $M_t$  is indeed nonnegative because each term  $(x_j - x_i)$  is nonnegative while in the Weyl chamber. Consider  $\mathbb{E}_x(v(B_{t \wedge T}))$ : if  $T \leq t$ , then  $M_t$  is already zero, so there is no contribution to the expectation, meaning that we can

just calculate (taking  $(y_1, \dots, y_n)$  to be the location of the Brownian motion at time  $t$ )

$$\mathbb{E}_x(v(B_{t \wedge T}); T > t) = \int_{W_n} \det \begin{bmatrix} 1 & y_1 & \cdots & y_1^{n-1} \\ 1 & y_2 & \cdots & y_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_n & \cdots & y_n^{n-1} \end{bmatrix} \det \begin{bmatrix} p_t(x_1, y_1) & \cdots & p_t(x_1, y_n) \\ \vdots & \ddots & \vdots \\ p_t(x_n, y_1) & \cdots & p_t(x_n, y_n) \end{bmatrix} dy$$

where we've used the the Karlin-MacGregor formula. But notice that symmetry means that we can integrate over  $\mathbb{R}^n$  instead of  $W_n$  and add a factor of  $\frac{1}{n!}$ , and now using the Andréief integration formula yields

$$= \det \begin{bmatrix} 1 & x_1 & \cdots & \mathbb{E}[(x_1 + \sqrt{t}Z)^{n-1}] \\ 1 & x_2 & \cdots & \mathbb{E}[(x_2 + \sqrt{t}Z)^{n-1}] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & \mathbb{E}[(x_n + \sqrt{t}Z)^{n-1}] \end{bmatrix}$$

(because integrating 1 against  $p_t(x_1, y_1)$  will just yield  $x_1$  by the properties of the transition kernel). We claim this is just equal to the simpler Vandermonde matrix

$$\det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} = v(x),$$

because we can expand out the rightmost columns by the binomial theorem and use row operations to subtract off lower powers. And this is basically the martingale identity we want once we use the Markov property.  $\square$

### Proposition 178

Let  $B(t)$  be a Brownian motion started from  $x \in W_n$ . Then there exists a unique measure  $\mathbb{Q}$  such that for all stopping times  $S < \infty$ ,

$$\frac{d\mathbb{Q}|_{\mathcal{F}_S}}{d\mathbb{P}|_{\mathcal{F}_S}} = \frac{M_S}{M_0} = \frac{v(B(S \wedge T))}{v(x)}.$$

Then the law of  $B$  under  $\mathbb{Q}$  can be thought of as Brownian motion conditioned not to exit  $W_n$ : it is a Feller process with generator given by

$$Lf(x) = \langle \nabla \log v(x), \nabla f(x) \rangle + \frac{1}{2} \Delta f(x).$$

*Proof.* Let  $R_i$  be the first time that  $M_t \geq i$ :  $M^{R_i}$  is now a bounded martingale, so the optional stopping theorem tells us that the exit time satisfies

$$\mathbb{P}(T > R_i) = \frac{v(x)}{i}.$$

Define the measure

$$\frac{d\mathbb{Q}_i}{d\mathbb{P}_i} \Big|_{\mathcal{F}_{R_i}} = \lim_{t \rightarrow \infty} \frac{M_{t \wedge R_i}}{v(x)} = \frac{i \mathbf{1}\{T > R_i\}}{v(x)}$$

(we'll either exit the chamber in finite time and get zero, or we'll stay inside  $W_n$ ). But because  $M$  is a martingale, the  $\mathbb{Q}_i$  are consistent with each other, so there exists a  $\mathbb{Q}$  such that its restrictions to  $\mathcal{F}_{R_i}$  are consistent with the  $\mathbb{Q}_i$ s,

with the desired value of  $\frac{d\mathbb{Q}|_{\mathcal{F}_S}}{d\mathbb{P}|_{\mathcal{F}_S}}$ . And notice that conditioning on  $T$  being large enough yields

$$\mathbb{P}(A|T > R_i) = \frac{\mathbb{P}(A \cdot \mathbf{1}\{T > R_i\})}{v(x)/i} = \mathbb{E}_P \left( \mathbf{1}_A \frac{i \mathbf{1}\{T > R_i\}}{v(x)} \right),$$

which is the same Radon-Nikodym derivative as before, so this expression is  $\mathbb{Q}(A)$ . And we can think of the first term as conditioning on the BM never exiting, because it takes arbitrarily long time to travel arbitrarily large distances (take  $i$  to infinity). Then we can find the generator by noting that

$$Lf(x) = \lim_{t \downarrow 0} \frac{1}{t} [\mathbb{E}_{\mathbb{Q}}(f(B(t))|B(0) = x) - f(x)],$$

and the change of measure makes this

$$= \lim_{t \downarrow 0} \frac{1}{t} \left[ \mathbb{E}_{\mathbb{P}} \left( f(B(t)) \frac{v(B_{t \wedge T})}{v(x)} | B(0) = x \right) - f(x) \right].$$

Now  $v(x)$  is a constant, and Itô's formula tells us that

$$d(f(B(t))M(t)) = f(B(t))dM_t + f'(B(t))M_t dB_t + \frac{1}{2}M_t f''(B(t))dt + f'(B(t))d\langle B, M \rangle_t.$$

We can ignore the martingale term because we're taking expectations;  $M_t$  has mean  $v(x)$ , and  $d\langle B, M \rangle = \nabla v$ . Evaluating everything indeed yields

$$Lf(x) = \frac{1}{2}\Delta f(x) + \left\langle \nabla f(x), \frac{\nabla v(x)}{v(x)} = \nabla \log v(x) \right\rangle,$$

and we're done. □

*Proof of Theorem 174.* Again, the  $\beta = 2$  Dyson SDE is

$$d\lambda_i(t) = 2 \sum_{\ell \neq i} \frac{1}{\lambda_i(t) - \lambda_\ell(t)} dt + \sqrt{2} dB_i(t).$$

Apply a rescaling  $\theta = \frac{\lambda}{\sqrt{2}}$ : then

$$d\theta_i(t) = \sum_{\ell \neq i} \frac{dt}{\theta_i - \theta_\ell} + dB_i(t).$$

Then the SDE gives us a generator

$$L = \sum_i \sum_{\ell \neq i} \frac{1}{x_i - x_\ell} \frac{\partial f}{\partial x_i} + \frac{1}{2}\Delta f,$$

and we can check this is indeed  $\langle \nabla \log v(x), \nabla f(x) \rangle + \frac{1}{2}\Delta f$ , which is the generator for the non-colliding Brownian motion. Since the generators are the same, the two processes are equal in law, as desired. □

As a final note, the reason we care about this identification is that when we consider some symmetric random matrix  $X$ , we may want it to have spectral statistics like those of the GOE (this has to do with **universality theorems** for random matrices). To show this, we take a flow on the matrices, meaning  $H(0) = X$  and  $H(t)$  evolves via an Ornstein-Uhlenbeck process (it's like a Brownian motion, but we want it to stay stationary). And  $H(\infty)$  then looks like the stationary distribution for Ornstein-Uhlenbeck, which is the GOE, and now we can consider

$$|\mathbb{E}f(x) - \mathbb{E}f(H)| \leq |\mathbb{E}f(H(0)) - \mathbb{E}f(H(t))| + |\mathbb{E}f(H(t)) - \mathbb{E}f(H(\infty))|$$

for a GOE matrix  $H$ . If we take  $t$  very small, the first term is small by perturbation theory, but the surprising fact is that the Dyson process mixes very quickly, so we can also control the second term (this is called the **fast mixing of**

the **Dyson process**). Many references for further study can be found on the official course website.

## References

- [1] N. Sun. *18.676. Stochastic Calculus*. URL: <https://math.mit.edu/~nsun/s20-18676.html>.
- [2] J.F.L. Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Graduate Texts in Mathematics. Springer International Publishing, 2013.
- [3] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, 2019.
- [4] O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer New York, 2002.