# An introduction to SPDE: Lecture Notes

Sam Punshon-Smith*

November 15, 2019

## Contents



*Division of Applied Mathematics, Brown University, Providence, RI, 02906, MD 20742 punshs@brown.edu

# 1 Introduction

The goal of this course (and these lecture notes) is to give a concise relatively self-contained introduction to the field of stochastic partial differential equations (SPDE's) aimed at graduate students. We will mainly be taking the approach of Da Prato and Zabczyk [DPZ14] and study a large class of SPDE which can be view as abstract stochastic evolution equation in a separable Hilbert space $H$ of the form

$$\mathrm{d}X = (AX + B(X))\mathrm{d}t + Q(X)\mathrm{d}W, \tag{1.1}$$

where $W$ is a certain generalization of a Wiener process in $H$, $A$ is a certain dissipative operator and $B(x)$ and $Q(x)$ are appropriate nonlinear functions on $H$ (to be described in more detail later). Abstractly we may view a solution $t \mapsto X(t)$ to (1.1) as a random trajectory in a $H$ with certain statistical properties (such as the Markov property, and certain path-wise regularity properties). Such equations include a broad class of interesting Stochastic PDE that have some form of regularizing or smoothing effect, typically parabolic in nature. While there are many interesting examples of hyperbolic SPDE (such as stochastic wave equations, stochastic continuity equations or stochastic conservation laws), these can be tricky to deal with in generality (due to issues with regularity) and can go well beyond the scope of this course. We will also stay in the regime where the nonlinearities are well behaved on the space that the solutions live in (i.e. Lipschitz continuous) and will typically avoid the case when the regularity of the solution is $X(t)$ is not enough to make sense of nonlinearities $F(X)$ (as is the case of singular SPDE with space-time white noise). Additionally, if time permits we will cover some of the basic ergodic theory and limit theorems of Markov processes associated to (1.1) such as existence and uniqueness of stationary measures, geometric ergodicity, the law of large numbers and the central limit theorem.

Much of the material in these notes will draw upon material in Da Prato/Zabczyk's book [DPZ14] as well as the lecture notes by Martin Hairer `http://www.hairer.org/notes/SPDEs.pdf`. However I will have my own take on the topic and will introduce topics not-contained in either book. It will be assumed that the students are familiar with and have taken graduate level courses in Analysis, Functional Analysis, and Probability. However, I will include a brief refresher for students who need a reminder.

## 1.1 Examples

**Example 1.1** (White Noise)**.** The most fundamental object that we will study in this course is the Gaussian white noise $\xi(t,x)$ (or cylindrical Wiener process). In the context of SPDE on $\mathbb{R}^n$ we will view $\xi(t,x)$ for each $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^n$ as a centered Gaussian random variable completely decorrelated in space and time i.e.

$$\mathbf{E}\,\xi(t,x)\xi(s,y) = \delta(t-s)\delta(x-y).$$

Alternatively we can can interpret such an expression in the sense of distribution. That is if we define for each $f \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^n)$, $\langle f, \xi \rangle$ the pairing of $f$ and $\xi$ on $\mathbb{R}_+ \times \mathbb{R}^n$ we have for test functions $f, g \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^n)$

$$\mathbf{E}\langle f, \xi\rangle\langle g, \xi\rangle = \langle f, g\rangle.$$

Note that since $\xi$ is Gaussian, specifying it's mean and covariance should characterize it completely. However in order to make this more precise, we will need the machinery of Gaussian measures Hilbert (and Banach spaces). This will be covered in great detail later.

**Example 1.2** (Stochastic heat equation)**.** The stochastic heat equation is one of the simplest non-trivial examples of a linear stochastic PDE. On $\mathbb{R}^n$ it is given by

$$\partial_t f - \Delta f = \xi$$

$f(t,x)$ is a scalar quantity (modeling a temperature of concentration) and $\xi(t,x)$ is a homogeneous (shift invariant) mean zero Gaussian noise, white in time and *colored in space* satisfying

$$\mathbf{E}\xi(t,x)\xi(s,y) = \delta(t-s)C(x-y).$$

where $C(x)$ is some covariance kernel describing the range of spatial correlations. This equation actually arises quite naturally as many particle limits of certain stochastic particle systems, especially when looking at fluctuations about the limits of certain empirical densities. We will see that depending on the regularity of $C(x)$ and the dimensionality of the problem, this equation may not have function valued solutions. We will return to a formal analysis of this equation in the next section.

**Example 1.3** (Stochastic reaction diffusion equation)**.** The stochastic reaction diffusion equation on $\mathbb{R}^n$ can be seen as a nonlinear perturbation of the stochastic heat equation given by

$$\partial_t f + F(f) - \Delta f = \xi.$$

here $F(f)$ is a polynomial nonlinearity acting pointwise on $f$ and $\xi$ is as in the previous example. The nonlinearity models a "reaction" effect, causing the solution to grow or decay depending on it's size. This effect competes with the "diffusion" effect that wants to spread everything out. The resulting behavior is rather complicated and can lead traveling fronts with very intricate patterns. As a stochastic PDE, the nonlinearity poses certain problems when $C(x)$ is rough since nonlinear function $F(f)$ are not well-defined on distributions or measures. In general we will only consider equations which have nonlinearities $F(f)$ which are locally Lipschitz continuous with respect to the topology that we expect the solution $u$ to be continuously evolving in.

**Example 1.4** (Stochastic Navier-Stokes equations)**.** The stochastic Navier-Stokes equations on $\mathbb{R}^n$ (here typically $n = 2, 3$) is one of the most widely studied nonlinear stochastic PDE. It governs the motion of an incompressible velocity field $u(t,x) \in \mathbb{R}^n$ given by

$$\partial_t u + u \cdot \nabla u + \nabla p - \Delta u = \xi, \quad \operatorname{div} u = 0.$$

Here $p$ is the pressure, which enforces the divergence-free conditions (can think of it as a Lagrange multiplier). In dimension $n = 2$ the equations above are globally well-posed for suitably regular noise. However, in dimension $n = 3$ well-posedness is an open problem (with or with out noise). At best, one has global existence of weak solutions (distributional solution) in this case. For the stochastic problem we will see that lack of uniqueness for the deterministic problem brings a whole host of other problems on the stochastic side. In the stochastic case one only has weak martingale solutions (probabilistically weak) which bring another level of complexity.

The dynamics of the Navier-Stokes equations (when well-defined) can be extremely chaotic, even with the presence of the Laplacian (this is the essence of the turbulence problem). The addition of noise can help "tame" this chaos by understanding it in a statistical sense and allowing one to put probability measures on the dynamics and study certain stationary measures associated with the long-time behavior.

## 1.2 Formal Analysis of the stochastic heat equation

Lets study the stochastic heat equation

$$\partial_t f - \Delta f = \xi$$

more carefully. It is well known that when $\xi = 0$, the initial value problem for this equation can be solved by convolution of the initial data with the heat kernel

$$u(t,x) = \int_{\mathbb{R}^n} K(t, x-y) u_0(y) \mathrm{d}y.$$

where

$$K(t,x) := \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{|x|^2}{4t}\right).$$

This action of the heat kernel on a function forms a semi-group $S(t)$ of operators (we will spend more time on this later) defined by

$$(S(t)f)(x) := (K(t,\cdot) \star f)(x) = \int_{\mathbb{R}^n} K(t, x-y) f(y) \mathrm{d}y.$$

Proceeding formally and presuming that $\xi(t,x)$ is a nice function (which it's not), we can write the the solution $u(t,x)$ to such an equation can be written using Duhamel's formula (or variation of constants) as

$$u(t) = S(t)u_0 + \int_0^t S(t-s)\xi(s,\cdot)\mathrm{d}s.$$

The quantity $\int_0^t S(t-s)\xi(s,\cdot)\mathrm{d}s$ is referred to as a *stochastic convolution* and is a fundamental object in the study of such SPDE (we will study this object more carefully in later sections). Assuming the initial data $u_0 = 0$ for now. The solution $u(t,x)$ can be written as a space time convolution

$$u(t,x) = \int_0^t \int_{\mathbb{R}^n} K(t-s, x-y)\xi(s,y)\mathrm{d}y\mathrm{d}s. \tag{1.2}$$

Of course such a formula needs justification due to the potentially singular nature of $\xi$ in time (it is not a well-defined function). It is reasonable to expect though, due to the presence of the convolution, that such a solution is significantly smoother than the noise forcing it.

The formula (1.2) tells us that the quantity $u(t,x)$ is a mean zero Gaussian variable (being a sum of mean zero Gaussians). However, it's covariance is more complicated. Additionally, due the the stationarity of $\xi(s,x)$ in space (invariance of the statistics under space shifts) $u$ is also stationary in space. Indeed lets compute the variance

$$\begin{aligned}
\mathbf{E}|u(t,x)|^2 &= \int_0^t \int_0^t \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K(t-s, x-y)K(t-s', x-y')\mathbf{E}\xi(s,y)\xi(s',y')\mathrm{d}y\mathrm{d}y'\mathrm{d}s\mathrm{d}s' \\
&= \int_0^t \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K(t-s, x-y)K(t-s, x-y')C(y-y')\mathrm{d}y\mathrm{d}y'\mathrm{d}s \\
&= \int_0^t \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K(s,y)K(s,y')C(y-y')\mathrm{d}y\mathrm{d}y'\mathrm{d}s \\
&= \frac{1}{(4\pi)^n} \int_0^t \frac{\Phi(s)}{s^n}\mathrm{d}s,
\end{aligned}$$

where we have defined

$$\Phi(s) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \exp\left(-(|y|^2 + |y'|^2)/4s\right) C(y-y')\mathrm{d}y\mathrm{d}y',$$

and we have used change of variables $x-y \to y$, $x-y' \to y'$ and $t-s \to s$. Therefore we see that $u(x,t)$ only has finite variance if $s \mapsto \Phi(s)/s^n$ is integrable on $[0,t]$ for each $t$. When such a quantity is not integrable, we do not expect $u(t,x)$ to have well-defined pointwise values. Indeed, when noise is spatially white $C(y-y') = \delta(y-y')$, then it is easy to see that

$$\Phi(s) = (2\pi s)^{n/2}$$

and therefore $\Phi(s)/s^n \approx s^{-n/2}$, which is only integrable near $s = 0$ in dimension $n = 1$. Consequently, $u(t, x)$ only makes sense pointwise (or as a function) in one dimension. Additionally if $C(y)$ is a locally integrable function which is homogeneous of degree $-\alpha$ (i.e. for any $\lambda > 0$, $C(\lambda x) = \lambda^{-\alpha} C(x)$), then by change of variables

$$\Phi(s) = s^{n-\alpha/2}\Phi(1)$$

and it follows that $\Phi(s)/s^n \approx s^{-\alpha/2}$, which is only integrable when $\alpha < 2$. Consequently if $C(x)$ is too singular at 0 (even if is it locally integrable) then $u(t, x)$ is not a well defined function either (for instance in dimension $n \geq 3$ if $n - 1 < \alpha < n$ and $C(y) = |y|^{-\alpha}$).

However, it is possible to make sense of $u(t, x)$ as a distribution in any dimension. Indeed, take $\varphi(x)$ to be a smooth, compactly supported test function and study the pairing

$$\langle u(t), \varphi\rangle := \int_{\mathbb{R}^n} u(t, x)\varphi(x)\mathrm{d}x = \int_0^t \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} K(t - s, x - y)\xi(s, y)\varphi(x)\mathrm{d}y\mathrm{d}x.$$

Changing the order of integration, we can write such a solution as

$$\langle u(t), \varphi\rangle = \int_0^t \int_{\mathbb{R}^n} v(t - s, y)\xi(s, y)\mathrm{d}s\mathrm{d}y$$

where $v(s, y)$ is a solution of the heat equation

$$\partial_t v - \Delta v = 0$$

with initial data $v(0, x) = \varphi(x)$. Again, computing the variance as we did previously, we find

$$\mathbf{E}|\langle u(t), \varphi\rangle|^2 = \int_0^t \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} v(s, y)v(s, y')C(y - y')\mathrm{d}y\mathrm{d}y'\mathrm{d}s$$

Since $v(t, x)$ is a solution to the heat equation with smooth initial data $\varphi(x)$. Consequently an application of Young's convolution inequality (see here) implies

$$\mathbf{E}|\langle u(t), \varphi\rangle|^2 \leq \left(\int_0^t \int_{\mathbb{R}^n} v(s, y)^2\mathrm{d}y\mathrm{d}s\right)\left(\int_{\mathbb{R}^n} C(y)\mathrm{d}y\right).$$

Since $\|v(t)\|_{L^2} \leq \|\varphi\|_{L^2}$, such a quantity is finite if $C(y)$ is integrable (including a delta function), meaning

$$\mathbf{E}|\langle u(t), \varphi\rangle|^2 \lesssim \|\varphi\|_{L^2}^2. \tag{1.3}$$

This suggest that when $n \geq 2$ then $u(t, x)$ can be made sense of as a distribution (i.e. when paired against a suitably regular test function).

**Remark 1.5.** It may be tempting to conclude that (1.3) would imply that $u(t)$ is actually $L^2$-valued when $\xi$ is a space-time white noise. However, this is not true since such a statement would require proving something like

$$\mathbf{E}\left(\sup_{\|\varphi\|_{L^2}=1} |\langle u(t), \varphi\rangle|^2\right) \lesssim 1,$$

which cannot be obtained from the analysis above (and is in fact not true).

## 2 Preliminaries

The goal of this section is to recall some preliminaries and notation from measure theory and probability. Much of this is standard, and some of the proofs will be omitted for brevity. A student well-versed in measure theory and probability can safely skip this section and return to it when needed.

In what follows we will denote $(\Omega, \mathscr{F})$ a measurable space, that is $\Omega$ is a set (*any* set) and $\mathscr{F}$ is a sigma-algebra of subsets of $\Omega$. Recall a sigma-algebra $\mathscr{F}$ is a family of subsets that is closed under taking countable unions and complements. All subsets If one takes the view the $\Omega$ is a configuration space of all possible "events", then one can think of $\mathscr{F}$ as the set of all well-formed logic statements about events in $\Omega$.

### 2.1 Random variables in Banach spaces

Let $(\mathcal{U}, \mathscr{G})$ be another measurable space, then a $\mathcal{U}$-valued random variable $X$ is a *measurable* mapping from $\Omega$ to $\mathcal{U}$ in the sense that for each $A \in \mathscr{G}$

$$X^{-1}(A) = \{\omega \in \Omega \,:\, X(\omega) \in A\} \in \mathscr{F}.$$

In other words, the question of whether $X$ belongs to a certain set $A \in \mathscr{G}$ must be well-formed statement belonging to $\mathscr{F}$.

**Remark 2.1.** The issue with measurability is not just a technical one. Indeed, if $\mathcal{U} = \mathbb{R}$, then for any $A \notin \mathscr{F}$, the indicator function $\mathbb{1}_A$ defined by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

is not measurable. It may seem feasible that one could simply add $A$ to $\mathscr{F}$ and then "close" $\mathscr{F}$ with respect to that new set. However, there may always be more non-measurable sets due to the fact that $\mathscr{F}$ must be closed under *countable* unions and complements. Of course, one could start with the power set $2^\Omega$, which is also a sigma-algebra which contains *all* subsets of $\Omega$. However, it is well-known that unless $\Omega$ is a countable space, it is "impossible" to define a probability measure $\mathbf{P}$ defined on all of $2^\Omega$, making the measurable space $(\Omega, 2^\Omega)$ rather useless in probability theory. An example of a non-measurable set on $[0, 1]$ with the Lebesgue sigma-algebra was first produced by Vitali using the axiom of choice. The construction of such a set is standard in most analysis text books (see for instance [Roy88, Fol99]).

In general, we will require a little more structure on the space $(\mathcal{U}, \mathscr{G})$. Namely, we would like a notion of convergence of one random variable to another. We will achieve this by taking $\mathcal{U}$ to a be a separable *Banach space*, namely a complete normed linear space, with norm $\|\cdot\|$, and a countable dense subset $\mathcal{U}_0 = \{u_k\}_{k\in\mathbb{N}}$. We will also take $\mathscr{G} = \mathscr{B}(\mathcal{U})$ to be the Borel $\sigma$-algebra of $\mathcal{U}$, defined to be the smallest sigma-algebra that contains the balls

$$B_R(u_0) = \{u \in \mathcal{U} \,:\, \|u - u_0\| \le r\},$$

for each $u_0 \in \mathcal{U}$ and $r \ge 0$. Denote the dual of $\mathcal{U}$ by $\mathcal{U}^*$, defined to be the Banach space consisting of all continuous linear functionals on $\mathcal{U}$ and let $\|\cdot\|_*$ be it's norm defined by

$$\|\ell\|_* = \sup_{\|u\|=1} \ell(u).$$

**Remark 2.2.** Note that we do not assume that $\mathcal{U}$ is *reflexive* (i.e. that $\mathcal{U}^{**} = \mathcal{U}$), since in general it is not needed and is not true in many interesting examples.

**Example 2.3.** Let us briefly recall some examples of Banach spaces and their properties.

- The spaces $L^p(\mathcal{X}, \mu)$ for any countably generated measure space $(\mathcal{X}, \mu)$ are separable reflexive Banach spaces for $p \in (1, \infty)$.
- The space $L^1(\mathcal{X}, \mu)$ (in the above setting) is separable , but not reflexive (it's dual is $L^\infty$).
- The space $L^\infty(\mathcal{X}, \mu)$(in the above setting) is neither separable, nor reflexive (it's dual is the space of finitely additive set functions).
- The Sobolev space $W^{s,p}(\mathbb{R}^n)$ is a separable, reflexive Banach space for $p \in (0, \infty)$.
- The space $C_b(\mathcal{X})$ of bounded continuous functions on a compact topological space $\mathcal{X}$ is separable, but not reflexive (it's dual is $\mathcal{M}_b(\mathcal{X})$)
- The space $\mathcal{M}_b(\mathcal{X})$ of bounded Radon measures on a compact topological space $\mathcal{X}$ is separable if $\mathcal{X}$ is, but not reflexive.
- The space $C_b(\mathcal{X})$ of continuous functions on a non-compact (but locally compact) space $\mathcal{X}$ is neither separable nor reflexive.
- Every Hilbert space $\mathcal{H}$ is reflexive by the Riesz representation theorem.

Recall that the weak topology on a Banach space $\mathcal{U}$ is the weakest topology such that all linear functionals $\ell \in \mathcal{U}^*$ are continuous. We denote $[\mathcal{U}]_w$ the space $\mathcal{U}$ equipped with it's weak topology. The Borel sigma-algebra $\mathscr{B}([\mathcal{U}]_w)$ is then defined to be the smallest sigma-algebra generated by sets of the form

$$\{u \in \mathcal{U} \,:\, \ell(u) \leq r\}, \tag{2.1}$$

where $\ell \in \mathcal{U}^*$ and $r \geq 0$. In general the sigma-algebras $\mathscr{B}(\mathcal{U})$ and $\mathscr{B}([\mathcal{U}]_w)$ may be different, with the obvious inclusion $\mathscr{B}([\mathcal{U}]_w) \subseteq \mathscr{B}(\mathcal{U})$ (since the weak topology is strictly smaller than the strong topology). However, when $\mathcal{U}$ is separable, the two sigma-algebras are the same.

**Proposition 2.4.** *Let $\mathcal{U}$ be separable. Then $\mathscr{B}(\mathcal{U}) = \mathscr{B}([\mathcal{U}]_w)$.*

*Proof.* It is a straightforward consequence of separability (exercise 2.1) that there exists a sequence $\{\ell_k\}_{k\in\mathbb{N}} \subset \mathcal{U}^*$ with $\|\ell_k\|_* = 1$ such that for each $u \in \mathcal{U}$

$$\|u\| = \sup_k \ell_k(u).$$

It follows by a simple check that this implies the identity

$$\{u \in \mathcal{U} \,:\, \|u - u_0\| \leq r\} = \bigcap_{j\geq 0} \bigcup_{k\in\mathbb{N}} \{u \in \mathcal{U} \,:\, \ell_k(u - u_0) \leq r + 1/j\}.$$

Since every ball is a countable union and intersection of sets of the form (2.1), it follows that $\mathscr{B}(\mathcal{U}) \subseteq \mathscr{B}([\mathcal{U}]_w)$. □

**Exercise 2.1.** Show that if $\mathcal{U}$ is a separable Banach space then there exists a sequence $\{\ell_n\}_{n\in\mathbb{N}} \subset \mathcal{U}^*$ such that

$$\|u\| = \sup_{n\in\mathbb{N}} \ell_n(u).$$

**Remark 2.5.** Note that the result of Proposition implies that if $\mathcal{U}$ is separable, then $X$ is a $\mathcal{U}$ valued random variable if and only if for each $\ell \in \mathcal{U}^*$, $\ell(X)$ is a real-valued random variable (in fact it suffices to check against a certain countable set $\{\ell_n\}_{n\in\mathbb{N}}$). Intuitively this means that a $\mathcal{U}$ valued random variable is determined by linear "measurements" on it. This idea is very powerful, and will serve for the basis of our definition of a Gaussian measure on a separable Banach space later. It is also one of the main reasons why separability is assumed.

It it a standard fact that linear combinations of random variables remain random variables (in terms of measurability). Additionally measurable functions of random variables obviously remain random variables. In general it is convenient to have a strategy to determine whether a given function $X : \Omega \to \mathcal{U}$ is actually measurable (apart from Proposition 2.4). Indeed taking pointwise limits of random variables gives a very useful approach to determining measurability.

**Proposition 2.6.** *Let $\mathcal{U}$ be a separable Banach space and let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathcal{U}$-valued random variables defined on $(\Omega, \mathscr{F})$ (measurable with respect to $\mathscr{B}(\mathcal{U})$). If $X : \Omega \to \mathcal{U}$ is a function such that for each $\omega \in \Omega$ and $\ell \in \mathcal{B}^*$*

$$\lim_{n\to\infty} |\ell(X_n(\omega) - X(\omega))| = 0,$$

*then $X$ is a random variable (i.e. measurable with respect to $\mathscr{B}(\mathcal{U})$).*

*Proof.* The proof is fairly standard. First, note that in light of Proposition 2.4 it is sufficient to show that for each $\ell \in \mathcal{U}^*$ and $r > 0$ the set $\{\omega \in \Omega : \ell(X(\omega)) \leq r\}$ belongs to $\mathscr{F}$. In light of this, and that fact that

$$\ell(X) = \lim_{n\to\infty} \ell(X_n) = \inf_n \sup_{k\geq n} \ell(X_k),$$

it is not hard to show that

$$\{\omega \in \Omega : \ell(X(\omega)) \leq r\} = \bigcap_{j\geq 0} \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} \{\omega \in \Omega : \ell(X_k(\omega)) \leq r + 1/j\}.$$

By assumption $\{\omega \in \Omega : \ell(X_k(\omega)) \leq r\}$ belongs to $\mathscr{F}$, therefore we are done because we have expressed $\{\omega \in \Omega : \ell(X(\omega)) \leq r\}$ as a countable union and intersection of sets in $\mathscr{F}$. □

**Remark 2.7.** Note that the above approximation is pointwise and *not* pointwise almost everywhere. The subtle difference between the two modes of convergence depends on whether the underlying probability space $(\Omega, \mathscr{F}, \mathbf{P})$ is complete. Namely that all subsets of measure zero sets are measurable. Currently we hace made no assumptions of completeness of the probability space.

It is also useful to approximate any $\mathcal{U}$-valued random variable by simple ones. Recall, a *simple random variable* is a random variable that takes countably many values. Indeed any simple random variable $X$ can be represented as

$$X(\omega) = \sum_{k\in\mathbb{N}} u_k \mathbb{1}_{A_k}(\omega),$$

where $\{u_k\}_{k\in\mathbb{N}} \subset \mathcal{U}$ and $\{A_k\}_{k\in\mathbb{N}} \subset \mathscr{F}$. Then we have the following approximation theorem.

**Proposition 2.8.** *Let $\mathcal{U}$ be a separable Banach space and let $X$ be a $\mathcal{U}$-valued random variable. Then there exists a sequence $\{X_n\}_{n\in\mathbb{N}}$ of simple random variables such that*

$$\|X_n(\omega) - X(\omega)\| \to 0, \quad \text{as } n \to \infty \quad \text{monotonically.}$$

*Proof.* Again the proof is elementary. Let $\{u_k\}_{k\in\mathbb{N}}$ be a separable subset and for each $n$, $k \le n$, define the set

$$V_k^n = \{u \in \mathcal{U} \,:\, \|u - u_k\| \le \min_{j\le n} \|u - u_j\|\}.$$

Note that $\{V_k^n\}_{k\le n}$ forms a "Voronoi" partition of $\mathcal{U}$ and satisfies $\bigcup_{k\le n} V_k^n = \mathcal{U}$. We then define the simple functions $\{X_n\}_{n\in\mathbb{N}}$ by

$$X_n(\omega) = \sum_{k\le n} u_k \mathbb{1}_{\{X(\omega)\in V_k^n\}}.$$

The proof now follows from the fact that

$$\|X_n(\omega) - X(\omega)\| \le \sum_{k\le n} \|X(\omega) - u_k\| \mathbb{1}_{\{X(\omega)\in V_k^n\}} = \min_{j\le n} \|X(\omega) - u_j\|,$$

which goes to 0 monotonically as $n \to \infty$ by separability. □

Finally, we introduce another powerful approach developed by Dynkin for determining whether a set or mapping is measurable, as well as when two probability measures are equal. We say a collection of subsets $\mathscr{K} \subset \Omega$ form a $\pi$-system if $\emptyset \in \mathscr{K}$ and if $A, B \in \mathscr{K}$, then $A \cap B \in \mathscr{K}$. The following propositon will come in very useful.

**Proposition 2.9.** *Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be probability measures on $(\Omega, \mathscr{F})$ and let $\mathscr{K}$ be a $\pi$-system such that $\mathscr{F}$ is the smallest sigma-algebra containing $\mathscr{K}$. If $\mathbf{P}_1 = \mathbf{P}_2$ on $\mathscr{K}$ then $\mathbf{P}_1 = \mathbf{P}_2$ on $\mathscr{F}$.*

*Proof.* The proof can be found in [DPZ14] Prop 1.5. □

## 2.2 Bochner integral

Recall a *probability measure* $\mathbf{P}$ on $(\Omega, \mathscr{F})$ is a non-negative set function on $\mathscr{F}$ which is sigma-additive (countably additive on disjoint sets) and has unit mass $\mathbf{P}(\Omega) = 1$. The triple $(\Omega, \mathscr{F}, \mathbf{P})$ is usually referred to as an *abstract probability space*.

Let $\mathcal{U}$ be a separable Banach space, and let $X$ be a $\mathcal{U}$-valued random variable. Our goal is to define the notion of an integral of $X$ (or expectation) with respect to $\mathbf{P}$, denoted by

$$\mathbf{E}X = \int_\Omega X \mathrm{d}\mathbf{P} = \int_\Omega X(\omega)\mathbf{P}(\mathrm{d}\omega).$$

The above integral of a Banach space valued function is referred to as the *Bochner integral* and is a generalization of the Lebesgue integral to vector valued quantities. Naturally we can define it on simple functions

$$X = \sum_{k\le n} u_k \mathbb{1}_{A_k}$$

by the identity

$$\int_\Omega X(\omega)\mathbf{P}(\mathrm{d}\omega) = \sum_{k\le n} u_k \mathbf{P}(A_k).$$

Obviously the properties of linearity and additivity hold here. Additionally, the triangle inequality easily implies that

$$\left\| \int_\Omega X \mathrm{d}\mathbf{P} \right\| \le \int_\Omega \|X\| \mathrm{d}\mathbf{P}, \tag{2.2}$$

which directly relates the Bochner integral of $X$ to the Lebesgue integral on $\|X\|$.

In general, we would like to extend this integral to general $\mathcal{U}$-valued random variables. In light of (2.2) we say that a $\mathcal{U}$-valued random variable $X$ is *Bochner integrable* if $\|X\|$ is Lebesgue integrable with respect to $\mathbf{P}$. By Proposition it readily follows that we can approximate $X$ by simple functions $\{X_n\}_{n\in\mathbb{N}}$ and therefore

$$\left\|\int_\Omega X_n \mathrm{d}\mathbf{P} - \int_\Omega X_m \mathrm{d}\mathbf{P}\right\| \leq \int_\Omega \|X_n - X_m\| \mathrm{d}\mathbf{P} \to 0$$

as $n, m \to \infty$ by the dominated convergence theorem for the Lebesgue integral. Consequently $\left\{\int_\Omega X_n \mathrm{d}P\right\}_{n\in\mathbb{N}}$ is a Cauchy sequence and therefore has a limit. This implies that we can define the Bochner integral of $X$ as

$$\int_\Omega X \mathrm{d}\mathbf{P} := \lim_{n\to\infty} \int_\Omega X_n \mathrm{d}\mathbf{P}. \tag{2.3}$$

It is simple to check that this definition is independent of the approximation chosen, and that linearity, additivity, and the triangle inequality (2.2) still hold.

**Remark 2.10.** Many important theorems for the Lebesgue integral carry over to the Bochner integral. In particular, Lebesgue dominated convergence and Fubinis Theorem also hold for the Bochner integral. (see [Yos95] for more details)

**Exercise 2.2.** Show that the definition (2.3) is independent of the approximation chosen, as long as $\int_\Omega \|X - X_n\| \mathrm{d}\mathbf{P} \to 0$, and that linearity, additivity, and the triangle inequality (2.2) still hold.

In general, given $\mathbf{P}$, each $\mathcal{U}$-valued random variable $X$ gives rise to a measure on $\mathcal{U}$ called it's *law* defined by

$$\mathrm{Law}(X)(A) := \mathbf{P}\{\omega \in \Omega : X(\omega) \in A\},$$

for each $A \in \mathscr{B}(\mathcal{U})$.

**Proposition 2.11** (Change of variables)**.** *Let $X$ be a random variable with values in a separable Banach space $\mathcal{U}$ and denote it's law by $\mu = \mathrm{Law}(X)$. Additionally, let $F$ be continuous mapping between separable Banach spaces $\mathcal{U}$ and $\mathcal{V}$ (equipped with their Borel sigma-algebra) which is Bochner integrable with respect to $\mu$. Then $F(X)$ is Bochner integrable with respect to $\mathbf{P}$ and the following formula holds*

$$\int_\Omega F(X) \mathrm{d}\mathbf{P} = \int_{\mathcal{U}} F(u)\mu(\mathrm{d}u),$$

*where both integrals are interpreted as Bochner integrals*

*Proof.* The proof follows again by simple function approximation using the approximation from the proof of Proposition 2.8,

$$X_n(\omega) = \sum_{k\leq n} u_k \mathbb{1}_{\{X(\omega)\in V_k^n\}}.$$

to see that

$$\int_\Omega F(X_n) \mathrm{d}\mathbf{P} = \sum_{k\leq n} F(u_k)\mathbf{P}\{X(\omega) \in V_k^n\} = \sum_{k\leq n} F(u_k)\mu(V_k^n) = \int_{\mathcal{U}} F_n(u)\mu(\mathrm{d}u).$$

where $F_n$ is a simple function approximation of $F(u)$ defined by

$$F_n(u) = \sum_{k\leq n} F(u_k)\mathbb{1}_{V_k^n}(u).$$

The rest of the argument is follows as in the construction of the Bochner integral after realizing that

$$F(X_n) = \sum_{k \leq n} F(u_k) \mathbb{1}_{X \in V_k^n}$$

is also a monotonic simple function approximation of $F(X)$ since

$$\|F(X_n) - F(X)\| \leq \sum_{k \leq n} \|F(u_k) - F(X)\| \mathbb{1}_{X \in V_k^n} \leq \sup_{k \leq n} \sup_{u \in V_k^n} |F(u_k) - F(u)|,$$

which goes to zero monotonically by separability and the continuity of $F$. Passing the limit as $n \to \infty$ in the simple function approximations gives the result. ☐

Finally Bochner integrals behave well with respect to action under linear functionals.

**Proposition 2.12.** *Let $X$ be a Bochner integrable random variable with values in a separable Banach space $\mathcal{U}$ and let $\ell \in \mathcal{U}^*$. Then*

$$\ell \left( \int_\Omega X \mathrm{d}\mathbf{P} \right) = \int_\Omega \ell(X) \mathrm{d}\mathbf{P}.$$

*Proof.* Again use simple function approximation and note that

$$\ell \left( \int_\Omega X_n \mathrm{d}\mathbf{P} \right) = \int_\Omega \ell(X_n) \mathrm{d}\mathbf{P}.$$

and that $\ell(X_n)$ is also a simple function approximation of $\ell(X)$ which converges monotonically $|\ell(X_n) - \ell(X)| \to 0$. We can then pass the limit as $n \to \infty$ on both sides to conclude. ☐

## 2.3 Conditional expectation

Just as with the real valued case, it is possible to define the Bochner integral of real valued random variables, it is possible to define the conditional expectation for Bochner integrable random variables taking values in a separable Banach space $\mathcal{U}$.

**Proposition 2.13.** *Let $X$ be a Bochner integrable random variable on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ with values in a separable Banach space $\mathcal{U}$, and let $\mathscr{G} \subset \mathscr{F}$ be a sub sigma-algebra. Then there exists a $\mathbf{P}$ almost surely unique random variable*

$$\mathbf{E}[X|\mathscr{G}]$$

*defined on $(\Omega, \mathscr{G}, \mathbf{P})$ called the* conditional expectation of $X$ with respect to $\mathscr{G}$ *such that for all $A \in \mathscr{G}$*

$$\int_A X \mathrm{d}\mathbf{P} = \int_A \mathbf{E}[X|\mathscr{G}] \mathrm{d}\mathbf{P}. \tag{2.4}$$

*Moreover we have*

$$\|\mathbf{E}[X|\mathscr{G}]\| \leq \mathbf{E}[\|X\||\mathscr{G}].$$

*Proof.* The uniqueness is left as an exercise. Consider the simple random variable

$$X_n = \sum_{k \leq n} u_k \mathbb{1}_{A_k}$$

Then we define

$$\mathbf{E}[X_n|\mathscr{G}] = \sum_{k \leq n} u_k \mathbf{E}[1_{A_k}|\mathscr{G}].$$

It is clear that $\mathbf{E}[X_n|\mathscr{G}]$ is $\mathscr{G}$ measurable and satisfies (2.4) and

$$\|\mathbf{E}[X_n|\mathscr{G}]\| \le \sum_{k\le n} \|u_k\|\mathbf{E}[1_{A_k}|\mathscr{G}] = \mathbf{E}[\|X_n\||\mathscr{G}].$$

In light of the fact that

$$\int_\Omega \|\mathbf{E}[X_n|\mathscr{G}] - \mathbf{E}[X_m|\mathscr{G}]\|\mathrm{d}\mathbf{P} \le \int_\Omega \|X_n - X_m\|\mathrm{d}\mathbf{P}.$$

We see that convergence of $X_n$ in $L^1(\Omega;\mathcal{U})$ implies that $\{\mathbf{E}[X_n|\mathscr{G}]\}$ converges in $L^1(\Omega;\mathcal{U})$ and therefore we define

$$\mathbf{E}[X|\mathscr{G}] = \lim_{n\to\infty} \mathbf{E}[X_n|\mathscr{G}].$$

This implies that

$$\int_\Omega \mathbf{E}[X|\mathscr{G}]\mathrm{d}\mathbf{P} = \lim_{n\to\infty}\int_\Omega \mathbf{E}[X_n|\mathscr{G}]\mathrm{d}\mathbf{P} = \lim_{n\to\infty}\int_\Omega X_n\mathrm{d}P = \int_\Omega X\mathrm{d}\mathbf{P},$$

and the following limits holds in $L^1(\Omega)$

$$\|\mathbf{E}[X|\mathscr{G}]\| = \lim_{n\to\infty}\|\mathbf{E}[X_n|\mathscr{G}]\| \le \lim_{n\to\infty}\mathbf{E}[\|X_n\||\mathscr{G}] = \mathbf{E}[\|X\||\mathscr{G}].$$

□

**Exercise 2.3.** Show the $\mathbf{P}$ almost sure uniqueness of the conditional expectation.

**Exercise 2.4.** Using the properties of the Bochner integral and the uniqueness of the conditional expectation, show that for each $\ell \in \mathcal{U}^*$,

$$\ell(\mathbf{E}[X|\mathscr{G}]) = \mathbf{E}[\ell(X)|\mathscr{G}].$$

## 2.4 Probability measures on Banach spaces

Since a Banach space valued random variable on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ gives rise to a probability measure on $\mathcal{U}$ via $\mathrm{Law}(X)$. It is important to study to properties of probability measures on Banach spaces. The extra structure on $\mathcal{U}$ allows for a study of the space of probability measures on $\mathcal{U}$ and the convergence of sequences of measures. For the remainder of this section, we will assume that $\mathcal{U}$ is a separable Banach space and denote $\mathscr{P}(\mathcal{U})$ is the space of probability measures on $(\mathcal{U}, \mathscr{B}(\mathcal{U}))$.

**Remark 2.14.** In general much of what we prove here can be achieved with general topological spaces or polish spaces (complete metrizable spaces). The assumption of separability can also be replaced by more general conditions on the existence of a countable family of separating (potentially non-linear) functionals on the space. This can be very useful when working in stochastic PDE that require working with a Banach space equipped with it's weak topology, which cannot even be metrized and potentially doesn't have a countable dense subset. This general framework is due to Jakubowski [Jak97] and can be used to prove general versions of Prokhorov's theorem and the Skorohod representation theorem, which we cover later.

To begin, we prove an analogue of Proposition 2.6, showing that random variables in separable Banach spaces are determined by all linear "measurements" on them. For each $\ell \in \mathcal{U}^*$ and $\mu \in \mathscr{P}(\mathcal{U})$ we denote $\ell_*\mu$ the *push-forward* of $\mu$ under $\ell$ defined to be a probability measure on $\mathbb{R}$ given by

$$\ell_*\mu(A) = \mu(\ell^{-1}(A)), \quad A \in \mathscr{B}(\mathbb{R}).$$

**Proposition 2.15.** *Let $\mu, \nu \in \mathscr{P}(\mathcal{U})$. If*

$$\ell_* \mu = \ell_* \nu \quad \textit{for all} \quad \ell \in \mathcal{U}^*,$$

*then $\mu = \nu$.*

*Proof.* Note that $\ell_* \mu = \ell_* \nu$ implies that $\mu(A_r) = \nu(A_r)$, for all sets $A_r = \{u \in \mathcal{U} : \ell(u) \leq r\}$. Note that the sets $\mathscr{K} = \{A_r\}_{r \in \mathbb{R}}$ form a $\pi$-system and by Proposition 2.4, the sigma-algebra generated by $\mathscr{K}$ is $\mathscr{B}(\mathcal{U})$. It then follows from Proposition 2.9 that $\mu = \nu$ on $\mathscr{B}(\mathcal{U})$. □

We will also find it useful to take the Fourier transform of a probability measure $\mu \in \mathscr{P}(\mathcal{U})$ defined as a functional $\hat{\mu}$ on $\mathcal{U}^*$ for each $\ell \in \mathcal{U}^*$ by

$$\hat{\mu}(\ell) := \int_{\mathcal{U}} e^{i\ell(u)} \mu(\mathrm{d}u).$$

The function $\hat{\mu}(\ell)$ is often called the *characteristic function* of $\mu$. It is not hard to show that $\hat{\mu}(\ell)$ uniquely determines a probability measure.

**Proposition 2.16.** *Let $\mu, \nu \in \mathscr{P}(\mathcal{U})$. If $\hat{\mu}(\ell) = \hat{\nu}(\ell)$ for all $\ell \in \mathcal{U}^*$, then $\mu = \nu$.*

*Proof.* First note that the assumption of the theorem is equivalent to $\hat{\mu}(\lambda \ell) = \hat{\nu}(\lambda \ell)$ for all $\lambda \in \mathbb{R}$ and $\ell \in \mathcal{U}^*$. By Proposition 2.11 this is equivalent to

$$\int_{\mathbb{R}} e^{i\lambda x} (\ell_* \mu)(\mathrm{d}x) = \int_{R} e^{i\lambda x} (\ell_* \nu)(\mathrm{d}x).$$

Since one dimensional Fouerier transforms of measures are uniquely determined, we deduce that $\ell_* \mu = \ell_* \nu$. Applying Proposition 2.15 completes the proof. □

### 2.4.1 Weak convergence

In general we are interested in convergence properties of measures (or more generally compactness properties of the space $\mathscr{P}(\mathcal{U})$). To begin, we introduce a notion of convergence on $\mathscr{P}(\mathcal{U})$. A sequence of probability measures $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathscr{P}(\mathcal{U})$ is said to *converge weakly* to a probability measure $\mu \in \mathscr{P}(\mathcal{U})$ if for every bounded continuous function $\varphi \in C_b(\mathcal{U})$ we have

$$\lim_{n \to \infty} \int_{\mathcal{U}} \varphi \, \mathrm{d}\mu_n = \int_{\mathcal{U}} \varphi \, \mathrm{d}\mu.$$

This convergence induces a topology on $\mathscr{P}(\mathcal{U})$ known as the *narrow topology* (sometime also called weak topology), defined to be the weakest topology such that for each $\varphi \in C_b(\mathcal{U})$ the mapping

$$\mu \mapsto \int_{\mathcal{U}} \varphi \mathrm{d}\mu$$

is continuous.

**Remark 2.17.** The integration of $\varphi \in C_b(\mathcal{U})$ against $\mu$ is often viewed as a linear functional on $C_b(\mathcal{U})$. Indeed it is well know that the space of all finite measures on $\mathcal{U}$, $\mathscr{M}(\mathcal{U})$ is the dual of $C_0(\mathcal{U})$ the space of continuous functions that vanish at $\infty$. Consequently the pairing between a probability measure $\mu \in \mathscr{P}(\mathcal{U})$ and a function $\varphi \in C_b(\mathcal{U})$ is often denoted by

$$\mu(\varphi) := \int_{\mathcal{U}} \varphi \mathrm{d}\mu.$$

It is important to note that integration against all functions in $C_b(\mathcal{U})$ completely determines probability measures.

**Proposition 2.18.** *Let $\mu, \nu \in \mathscr{P}(\mathcal{U})$. If*

$$\mu(\varphi) = \nu(\varphi),$$

*for all $\varphi \in C_b(\mathcal{U})$ then $\mu = \nu$.*

*Proof.* For any closed set $F$, one can find a sequence of $\varphi_n \geq 0$ such that $\varphi_n \to \mathbb{1}_F$ boundedly. Passing the limit in the integral implies that

$$\mu(F) = \nu(F)$$

for all closed sets $F$. Since the closed sets form a $\pi$-system, Proposition 2.9 implies that $\mu = \nu$. □

An important theorem the characterizes the narrow topology is known as the *portmanteau theorem*

**Theorem 2.19** (Portmanteau)**.** *Let $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathscr{P}(\mathcal{U})$ and $\mu \in \mathscr{P}(\mathcal{U})$. The following are equivalent.*

- $\{\mu_n\} \to \mu$ *weakly.*
- $\mu_n(\varphi) \to \mu(\varphi)$ *for all continuous bounded $\varphi$ on $\mathcal{U}$.*
- $\mu_n(\varphi) \to \mu(\varphi)$ *for all Lipschitz bounded $\varphi$ on $\mathcal{U}$.*
- $\limsup_n \mu_n(F) \leq \mu(F)$ *for all closed $F \subset \mathcal{U}$.*
- $\liminf_n \mu_n(O) \geq \mu(O)$ *for all open sets $O \subset \mathcal{U}$.*
- $\lim_n \mu_n(A) = \mu(A)$ *for all Borel $A$ with $\mu(\partial A) = 0$.*

It is important to recognize that *all* probability measures on $\mathcal{U}$ defined on the Borel sigma-algebra are actually regular measures in the sense that they can be inner approximated by compact sets.

**Proposition 2.20** (Regularity)**.** *A probability measure $\mu \in \mathscr{P}(\mathcal{U})$ is a regular measure. That is, for every $A \in \mathscr{B}(\mathcal{U})$ and $\epsilon > 0$, there is a compact set $K_\epsilon$ such that*

$$\mu(A \backslash K_\epsilon) \leq \epsilon.$$

*Proof.* Let $\epsilon > 0$. By separability, for each $n \in \mathbb{N}$ we can cover $\mathcal{U}$ by a countable family of closed balls $\{B_k^n\}_{k\in\mathbb{N}}$ of radius $1/n$. These sets can be made into a disjoint covering $\{\tilde{B}_j^n\}$ such that for each $N \geq 0$

$$\bigcup_{k\leq N} B_k^n = \bigcup_{k\leq N} \tilde{B}_k^n.$$

By countable additivity and finiteness of the measure $\mu$, there exists an $N_n > 0$ such that

$$\mu\left(\mathcal{U}\backslash \bigcup_{k\leq N_n} B_k^n\right) = \sum_{k\geq N_n} \mu(\tilde{B}_k^n) < \epsilon 2^{-n}$$

Defining

$$K = \bigcap_{n\in\mathbb{N}} \bigcup_{k\leq N_n} B_k^n,$$

we see that $K$ is a compact set since it is closed and totally bounded (i.e for each $\delta > 0$ it can be covered by finite many balls of radius $\delta$). Moreover, we have

$$\mu(\mathcal{U} \backslash K) \leq \sum_{n \in \mathbb{N}} \mu \left( \mathcal{U} \backslash \bigcup_{k \leq N_n} B_k^n \right) < \epsilon.$$

□

Our aim is to deduce a relative compactness criterion for subsets of $\mathscr{P}(\mathcal{U})$ in the narrow topology. We say subset $M \subset \mathscr{P}(\mathcal{U})$ is *relatively (sequentially) compact* if for every subsequence $\{\mu_n\} \subset M$, there exists a subsequence $\{\mu_{n_k}\}$ converging to $\mu \in \mathscr{P}(\mathcal{U})$ weakly.

Naturally, the idea is to use the fact that for each $\varphi \in C_b(\mathcal{U})$, the real-valued sequence $\{\int \varphi \mathrm{d}\mu_n\}$ has a convergence subsequence. Now if $C_b(\mathcal{U})$ were separable, then a standard diagonalization argument (similar to the proof of Arzela-Ascoli) would allow us to extract a subsequence $\{\mu_{n_k}\}$ such that $\lim_{k\to\infty} \int \varphi \mathrm{d}\mu_{n_k}$ exists for $\varphi$ in a dense subset of $C_b(\mathcal{U})$. This limit easily extends to define a continuous functional $I(\varphi)$ on $C_b(\mathcal{U})$ which can be represented uniquely by a probability measure $\mu$ by the Riesz-Markov-Kakutani representation theorem.

The problem here is that (as remarked in Example 2.3) $C_b(\mathcal{U})$ is not separable and therefore the above diagonalization procedure does not work. The key obstruction here is due to the fact that $\mathcal{U}$ is not compact. Indeed if $\mathcal{U}$ were a compact space then $C_b(\mathcal{U})$ would be separable, and the above procedure would work.

**Remark 2.21.** In fact the space $\mathscr{P}(\mathcal{U})$ is not compact in the narrow topology since a sequence of probability measures can lose mass at infinity. For example if $\{u_n\}_{n\in\mathbb{N}}$ is a sequence such that $\|u_k\| \to \infty$ (and therefore has no convergence subsequence), then the sequence of delta measures $\{\delta_{u_k}\}$ has no convergent subsequence in the narrow topology, since

$$\int_{\mathcal{U}} \varphi \, \mathrm{d}\delta_{u_k} = \varphi(u_k),$$

which does not have a convergent subsequence for all $\varphi \in C_b(\mathcal{U})$. Recall $\delta_u$ is the measure defined by $\delta_u(A) = \mathbb{1}_A(u)$ for each $A \in \mathscr{B}(\mathcal{U})$.

To understand how to get around this obstacle, we introduce the following extension of the regularity property introduced in Proposition 2.20 to subsets of $\mathscr{P}(\mathcal{U})$. We say family $M \subset \mathscr{P}(\mathcal{U})$ of probability measures is called *tight* (or *uniformly tight*) if for each $\epsilon > 0$ there exists compact set $K_\epsilon \subset \mathcal{U}$ such that

$$\sup_{\mu \in M} \mu(\mathcal{U} \backslash K_\epsilon) < \epsilon.$$

Such a condition prevents the measures $\{\mu_n\}$ from sending mass off to infinity, since they must keep most of their mass concentrated on a compact set. The following theorem is known as Prokhorov's theorem relates tightness of measures to relative compactness in $\mathscr{P}(\mathcal{U})$.

**Theorem 2.22** (Prokhorov)**.** *A family $M \subseteq \mathscr{P}(\mathcal{U})$ is tight if and only if $M$ is relatively compact in $\mathscr{P}(\mathcal{U})$.*

*Proof.* Let $\{\mu_n\}$ be a sequence in $M$. By the tightness condition, for each $j \geq 0$, let $K_j$ be a compact set such that for all $n$

$$\mu_n(\mathcal{U} \backslash K_j) \leq 1/j,$$

We may assume without loss of generality that $\{K_j\}$ is increasing $K_1 \subset K_2 \subset \dots$. Following the diagonalization argument above, for each $j$, we can produce a subsequence of measures that when restricted to

$K_j$ converge to a measure $\mu_j$ supported on $K_j$. A further diagonal argument then implies that there exists a subsequence $\{\mu_{n_k}\}$ and a limit measure $\mu$ such that for each $\varphi \in C_b(\mathcal{U})$ and $j \geq 0$,

$$\lim_{k\to\infty} \int_{K_j} \varphi \, \mathrm{d}\mu_{n_k} = \int_{K_j} \varphi \, \mathrm{d}\mu.$$

Using the fact that $\mu_{n_k}(\mathcal{U}\backslash K_j) \leq 1/j$ and by the Portmanteau theorem $\mu(\mathcal{U}\backslash K_j) < 1/j$, we find,

$$\begin{aligned}\left|\int_{\mathcal{U}} \varphi \mathrm{d}\mu_{n_k} - \int_{\mathcal{U}} \varphi \mathrm{d}\mu\right| &\leq \left|\int_{K_j} \varphi \mathrm{d}\mu_{n_k} - \int_{K_j} \varphi \mathrm{d}\mu\right| + \left|\int_{\mathcal{U}\backslash K_j} \varphi \mathrm{d}\mu_{n_k} - \int_{\mathcal{U}\backslash K_j} \varphi \mathrm{d}\mu\right| \\ &\leq \left|\int_{K_j} \varphi \mathrm{d}\mu_{n_k} - \int_{K_j} \varphi \mathrm{d}\mu\right| + 2\|\varphi\|_{C_b}/j.\end{aligned}$$

First sending $k \to \infty$ and then sending $j \to \infty$ completes the proof.

□

**Exercise 2.5.** Show the converse statement in Theorem a relatively compact set in $\mathscr{P}(\mathcal{U})$ is tight. Hint: follow the proof of Proposition 2.20.

**Remark 2.23.** The narrow topology can be metrized by a number of metrics. Indeed, $\mathscr{P}(\mathcal{U})$ inherits most of it's properties from $\mathcal{U}$. Namely if $\mathcal{U}$ is separable and metrizable (Polish) then so it $\mathscr{P}(\mathcal{U})$. One of the most commonly used metrics is the *Wasserstein-1 metric* (or the dual Lipschitz metric) defined by

$$\mathcal{W}(\mu_1, \mu_2) = \sup_{\|\varphi\|_{\mathrm{Lip}} \leq 1} \left|\int_{\mathcal{U}} \varphi \mathrm{d}\mu_1 - \int_{\mathcal{U}} \varphi \mathrm{d}\mu_2\right|,$$

where $\|\varphi\|_{\mathrm{Lip}} = \sup_{x\neq y} |\varphi(x) - \varphi(y)|/d(x,y)$, and $d(x,y)$ is *some* distance-like metric on $\mathcal{U}$. This metric is very useful in study of ergodic theory of Markov processes. Something that we will get to later.

It is natural to ask whether, given a probability measure $\mu \in \mathscr{P}(\mathcal{U})$, can one find a random variable $X$ on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ whose law coincides with $\mu$. In general, this is related to the question of sampling of probability measures. The following theorem, due to Skorohod, links the concept of tightness of probability measures with that of almost sure convergence of random variables on a certain probability space. One can think of this as a generalization of the inverse CDF approach to sampling probability distributions on $\mathbb{R}$.

**Theorem 2.24** (Skorohod)**.** *Let $\{\mu_n\} \subset \mathscr{P}(\mathcal{U})$ converge weakly to $\mu \in \mathscr{P}(\mathcal{U})$. Then there exists a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and a sequence of random variables, $\{X_n\}$ and $X$ with $\mathrm{Law}(X_n) = \mu_n$ and $\mathrm{Law}(X) = \mu$ such that $X_n \to X$ $\mathbf{P}$-almost surely.*

*Proof.* The proof is well-known, but rather complicated and can be found in [Bil99] pg 70. The main idea, similar to the inverse CDF approach is the let $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}([0,1])$, and $\mathbf{P} = \mathrm{Leb}$ and then build $X$ by using separbility to build covering of $\mathcal{U}$ that refines down to small scales centered around the points in the separable set. This covering maps to a lexicographicall ordered covering of $[0,1]$ under $\mu$ and $\mu_n$. The random variables are then built to take values in the separable set associated with with each set covering $[0,1]$. □

**Remark 2.25.** If one starts with a sequence of $\mathcal{U}$-valued random variables $X_n$ on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and if the laws $\mu_n = \mathrm{Law}(X_n)$ are tight on $\mathcal{U}$, then by Skorohod, there exists a *different* probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbf{P}})$ and random variables $\{\tilde{X}_n\}$ and $\tilde{X}$ with the same laws such that $\tilde{X}_n \to \tilde{X}$, $\mathbf{P}$-almost surely.

### 2.4.2 Total variation metric

There is another natural stronger topology on $\mathscr{P}(\mathcal{U})$ called the total variation topology. Recall that for any signed Borel measure $\eta$ on $\mathcal{U}$ admits a Hahn-Jordan decomposition $\eta = \eta^+ - \eta^-$, where $\eta^+$ and $\eta^-$ are two mutually singular non-negative Borel measures on $\mathcal{U}$ . The total variation norm of the measure $\eta$ is then defined by

$$\|\eta\|_{TV} = \frac{1}{2}(\eta^+(\mathcal{U}) + \eta^-(\mathcal{U})).$$

Now suppose that $\mu, \nu \in \mathscr{P}(\mathcal{U})$, the total variation norm gives a natural metric $\|\mu - \nu\|_{TV}$ between $\mu$ and $\nu$. Note that since $\mu$ and $\nu$ have the same mass, we have

$$\|\mu - \nu\|_{TV} = (\mu - \nu)^+(\mathcal{U}) = (\mu - \nu)^-(\mathcal{U}).$$

**Remark 2.26.** The factor of $1/2$ in the definition of the TV norm is customary when defining the total variation distance between two probability measures and is to ensure that the total variation distance always satisfies

$$\|\mu - \nu\|_{TV} \leq 1.$$

This convention is not always taken in the literature as it differs from the usual definition of the total variation of signed measure.

Given two measures $\mu, \nu \in \mathscr{P}(\mathcal{U})$, it is known that if $\lambda \in \mathscr{P}(\mathcal{U})$ is such that $\mu$ and $\lambda$ are both absolutely continuous with respect to $\lambda$ (we can always take $\lambda = \frac{1}{2}(\mu + \nu)$), then the following formula holds

$$\frac{\mathrm{d}(\mu - \nu)^\pm}{\mathrm{d}\lambda} = \left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} - \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right)^\pm. \tag{2.5}$$

**Exercise 2.6.** Prove that equation (2.5) holds.

Along these lines, it is convenient to define the non-negative overlap measure $\mu \wedge \nu$ by

$$\frac{\mathrm{d}(\mu \wedge \nu)}{\mathrm{d}\lambda} = \min\left\{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}, \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right\}.$$

So that we have the decompositions

$$\mu = (\mu - \nu)^+ + \mu \wedge \nu = (\nu - \mu)^- + \mu \wedge \nu$$

and

$$\nu = (\nu - \mu)^+ + \mu \wedge \nu = (\mu - \nu)^- + \mu \wedge \nu.$$

Note that since $\mu$ and $\nu$ are probability measures, this "overlap" measure $\mu \wedge \nu$ always has mass $\leq 1$. It follows that we have the following alternate formula for the total variation norm

$$\|\mu - \nu\|_{TV} = 1 - (\mu \wedge \nu)(\mathcal{U}). \tag{2.6}$$

The formula (2.6) is useful because it shows how the distance between two probability measures relates to how much they "overlap". In particular it gives the following useful characterization.

**Lemma 2.27.** *Let $\mu, \nu \in \mathscr{P}(\mathcal{U})$. Then $\mu = \nu$ if and only if $\|\mu - \nu\|_{TV} = 0$. Additionally $\mu$ and $\nu$ are mutually singular if and only if $\|\mu - \nu\|_{TV} = 1$.*

*Proof.* Clearly if $\mu = \nu$ then $\|\mu - \nu\|_{TV} = 0$. Next suppose that $\|\mu - \nu\|_{TV} = 0$. It follows that $(\mu - \nu)^+ = 0$ and therefore $\mu = \mu \wedge \nu = \nu$.

Next suppose that $\mu$ and $\nu$ are mutually singular. Then by the uniqueness of the Hahn-Jordan decomposition $(\mu - \nu)^+ = \mu$ and therefore $\|\mu - \nu\|_{TV} = \mu(\mathcal{U}) = 1$. Next suppose that $\|\mu - \nu\|_{TV} = 1$, then $\mu \wedge \nu = 0$ and therefore

$$\mu = (\mu - \nu)^+ \quad \text{and} \quad \nu = (\mu - \nu)^-,$$

which are mutually singular by the properties of the Hahn-Jordan decomposition theorem. □

There are several useful equivalent ways to define the total variation metric

**Lemma 2.28.** *Let $\mu, \nu \in \mathscr{P}(\mathcal{U})$. Then the following are all equivalent definitions of the total variation distance,*

$$\begin{aligned}
\|\mu - \nu\|_{TV} &= (\mu - \nu)^+(\mathcal{U}) = (\mu - \nu)^-(\mathcal{U}) = 1 - (\mu \wedge \nu)(\mathcal{U}) && (2.7)\\
&= \frac{1}{2}\left\|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} - \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right\|_{L^1(\lambda)} && (2.8)\\
&= \frac{1}{2}\sup_{\|\varphi\|_\infty \le 1} |\mu(\varphi) - \nu(\varphi)| && (2.9)\\
&= \sup_{A \in \mathscr{B}(\mathcal{U})} |\mu(A) - \nu(A)|, && (2.10)
\end{aligned}$$

*where $\lambda \in \mathscr{P}(\mathcal{U})$ is such that $\mu$ and $\nu$ are absolutely continuous with respect to $\lambda$ and the supremum in (2.9) is over all bounded measurable functions with $\|\varphi\|_\infty \le 1$.*

*Proof.* Of course we have already shown (2.7). Equation (2.8) follows easily from the fact that

$$\left|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} - \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\right| = \frac{\mathrm{d}(\mu - \nu)^+}{\mathrm{d}\lambda} + \frac{\mathrm{d}(\mu - \nu)^-}{\mathrm{d}\lambda}.$$

For (2.9) note that for any bounded measurable function $\varphi$ with $\|\varphi\|_\infty \le 1$ we have

$$\begin{aligned}
|\mu(\varphi) - \nu(\varphi)| &\le |(\mu - \nu)^+(\varphi)| + |(\mu - \nu)^-(\varphi)|\\
&\le (\mu - \nu)^+(\mathcal{U}) + (\mu - \nu)^-(\mathcal{U})\\
&\le 2\|\mu - \nu\|_{TV},
\end{aligned}$$

with equality for $\varphi = \mathbb{1}_{\mathcal{U}^+} - \mathbb{1}_{\mathcal{U}^-}$, with $\mathcal{U}^+, \mathcal{U}^- \subset \mathcal{U}$ be the Hahn decomposition of $\mathcal{U}$ relative to $\mu - \nu$.

The proof of (2.10) is left as an exercise.

□

**Exercise 2.7.** Prove (2.10). (Note that there is no factor of 1/2)

**Exercise 2.8.** Let $\ell \in \mathcal{U}^*$ and $\mu, \nu \in \mathscr{P}(\mathcal{U})$. Prove that

$$\|\ell_* \mu - \ell_* \nu\|_{TV} \le \|\mu - \nu\|_{TV}.$$

# 3 Gaussian measures

## 3.1 Preliminaries

Our first step towards dealing with stochastic equations in infinite dimensions is to study Gaussian measures on Banach spaces. This foundation will give us to tools to study an analyze Wiener processes in infinite dimensions, which are crucial for defining stochastic differential equations in infinite dimensions. Much of what we will construct can also be done on locally convex topological vector spaces, but this can get technical and simply complicates the presentation (and is also not needed for our purposes). A more general discussion of Gaussian measures in this general setting can be found in the book by Bogachev [Bog98].

We now turn to the construction and study of Gaussian measures on a separable Banach space $\mathcal{U}$ with respect to $\mathscr{B}(\mathcal{U})$. Naturally in light of Proposition 2.15 we can define such a measure uniquely in terms of it's projection under linear functionals. Recall, a Gaussian measure on $\mathbb{R}$ is a measure $\mu \in \mathscr{P}(\mathbb{R})$ of the form

$$\mu(\mathrm{d}x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mathrm{d}x.$$

The quantities $m$ and $\sigma^2$ are referred to as the *mean* and *variance* respectively. When $\sigma^2 = 0$, the measure is concentrated at $m$ and is given by $\gamma = \delta_m$. We recall some useful properties of Gaussian measures on $\mathbb{R}$

1. The mean and variance are given by

$$m = \int_{\mathbb{R}} x\mu(\mathrm{d}x), \quad \sigma^2 = \int_{\mathbb{R}} (x-m)^2 \mu(\mathrm{d}x).$$

2. The characteristic function is

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}} e^{i\lambda x} \mu(\mathrm{d}x) = \exp\left(im\lambda - \frac{1}{2}\lambda^2\sigma^2\right).$$

3. The moment generating function is

$$M_\mu(\lambda) := \int_{\mathbb{R}} e^{\lambda x} \mu(\mathrm{d}x) = \exp\left(\lambda m + \frac{1}{2}\sigma^2\lambda^2\right).$$

4. Exponential moments are given by

$$\int_{\mathbb{R}} e^{\lambda x^2} \mu(\mathrm{d}x) = \begin{cases} \frac{1}{\sqrt{1-2\lambda\sigma^2}} & \lambda < \frac{1}{2\sigma^2} \\ +\infty & \lambda \geq \frac{1}{2\sigma^2} \end{cases}.$$

**Definition 3.1.** We say a measure $\mu \in \mathscr{P}(\mathcal{U})$ is a *Gaussian measure* if for each $\ell \in \mathcal{U}$, the measure $\ell_*\mu \in \mathscr{P}(\mathbb{R})$ is a Gaussian measure. We say that a Gaussian measure $\mu$ is *centered* (or mean zero) if $\ell_*\mu$ is a centered Gaussian measure.

For a Gaussian measure $\mu$ on $\mathcal{U}$ it is natural to define the mean $m_\mu : \mathcal{U} \to \mathbb{R}$ and covariance $C_\mu : \mathcal{U}^* \times \mathcal{U}^* \to \mathbb{R}$ as linear maps

$$m_\mu(\ell) := \int_{\mathcal{U}} \ell(u)\mu(\mathrm{d}u)$$

and

$$C_\mu(\ell_1, \ell_2) := \int_{\mathcal{U}} (\ell_1(u) - m_\mu(\ell_1))(\ell_2(u) - m_\mu(\ell_2))\mu(\mathrm{d}u).$$

Note that both of these quantities $m_\mu$ and $C_\mu$ are well-defined by estimates on the Gaussian $\ell_*\mu$, since by Proposition 2.11 we have

$$m_\mu(\ell) = \int_{\mathbb{R}} x\,(\ell_*\mu)(\mathrm{d}x) < \infty,$$

and

$$C_\mu(\ell,\ell) = \left(\int_{\mathbb{R}} (x - m_\mu(\ell))^2\,(\ell_*\mu)(\mathrm{d}x)\right) < \infty,$$

which is sufficient by Cauchy-Schwartz $C_\mu(\ell_1,\ell_2) \le C_\mu(\ell_1,\ell_1)^{1/2} C_\mu(\ell_2,\ell_2)^{1/2}$.

**Exercise 3.1.** Show that a measure $\mu \in \mathscr{P}(\mathcal{U})$ is Gaussian if and only if the characteristic function $\hat{\mu}(\ell)$ is given by

$$\hat{\mu}(\ell) = \exp\left(i m_\mu(\ell) - \frac{1}{2} C_\mu(\ell,\ell)\right),$$

for some linear function $m_\mu$ on $\mathcal{U}^*$ and $C_\mu$ a symmetric bilinear function on $\mathcal{U}^*$ such that $\ell \mapsto C_\mu(\ell,\ell)$ is non-negative.

**Example 3.2.** Let $\mathcal{U} = \mathbb{R}^n$ and let $m \in \mathbb{R}^n$ and $\mathcal{Q}$ be a symmetric non-negative definite matrix on $\mathbb{R}^n$. Let $\mu$ be the Gaussian measure on $\mathbb{R}^n$ with characteristic function

$$\hat{\mu}(\lambda) = \exp\left(i\langle \lambda, m\rangle - \frac{1}{2}\langle Q\lambda, \lambda\rangle\right)$$

Taking derivatives of the characteristic function it is easy to deduce that for all $i, j = 1, \dots n$

$$\int_{\mathbb{R}_n} x_i \mu(\mathrm{d}x) = m_i, \quad \int_{\mathbb{R}_n} (x_i - m_i)(x_j - m_j)\mu(\mathrm{d}x) = Q_{ij}.$$

Notice that since $Q$ is only positive *semi-definite*, the measure $\mu$ may by singular on some subspaces and therefore does not have a well-defined density. However if $Q$ is indeed invertible, then $\mu$ has a density

$$\frac{1}{\sqrt{(2\pi)^n \det Q}} \exp\left(-\frac{1}{2}\langle Q^{-1}(x-m), (x-m)\rangle\right).$$

The quantities $m_\mu$ and $C_\mu$ are also continuous on $\mathcal{U}^*$, which can be seen using a version of the Pettis integral, however this is more machinery than we actually need. The following Lemma is sufficient:

**Lemma 3.3.** *Suppose that $\mu \in \mathscr{P}(\mathcal{U})$ such that there is some $k \in \mathbb{N}$, $k \ge 1$ such that for all $\ell \in \mathcal{U}^*$*

$$\int_{\mathcal{U}} |\ell(u)|^k \mu(\mathrm{d}u) < \infty,$$

*for some $\epsilon > 0$, then for $\ell_1, \dots \ell_k \in \mathcal{U}^*$ we have*

$$\left|\int_{\mathcal{U}} \ell_1(u) \dots \ell_k(u)\,\mu(\mathrm{d}u)\right| \le C_k \|\ell_1\|_* \dots \|\ell_k\|_*.$$

*Proof.* We only show it for the $k = 1$ case. The general $k \in \mathbb{N}$ is similar. We define a linear operator $T : \mathcal{U}^* \to L^1(\mu)$ by

$$(T\ell)(u) = \ell(u).$$

The goal is to show that $T$ is bounded. Indeed, in light of the closed graph theorem and the linearity of $T$, we only need to show that $T$ is closed. This is left as an exercise. □

**Exercise 3.2.** Show that $T$ defined in the proof above is a closed operator. Namely, if $\ell_n \to \ell$ in $\mathcal{U}^*$, and $\ell_n(u) \to \bar{\ell}(u)$ in $L^1(\mu)$ then $T\ell = \bar{\ell}$.

**Exercise 3.3.** Extend the proof of Lemma 3.3 to general $k \in \mathbb{N}$, $k \geq 1$. (Hint: show that there exists a $C_k$ such that

$$\left(\int_{\mathcal{U}} |\ell(u)|^k \mathrm{d}\mu\right)^{1/k} \leq C_k \|\ell\|_*,$$

and use Hölder's inequality.)

We now have an immediate corollary of Lemma 3.3.

**Corollary 3.4.** *$m_\mu$ and $C_\mu$ are continuous on $\mathcal{U}^*$. Namely, there exist constants $\|m_\mu\|$ and $\|C_\mu\|$, which are the smallest constants such that*

$$|m_\mu(\ell)| \leq \|m_\mu\|\|\ell\|_* \quad |C_\mu(\ell_1, \ell_2)| \leq \|C_\mu\|\|\ell_1\|_*\|\ell_2\|_*.$$

**Exercise 3.4.** Give an alternative proof of Corollary 3.4 using the characteristic function $\hat{\mu}(\ell)$, Exercise 3.1, and Lebesgue's dominated convergence theorem. Use this to deduce that $\ell \mapsto m_\mu(\ell)$ and $\ell \mapsto C_\mu(\ell, \ell)$ are continuous in the weak-* topology on $\mathcal{U}^*$.

**Remark 3.5.** This continuity given in Corollary 3.4 allows us to identify $m_\mu$ as an element of $\mathcal{U}^{**}$ as well as to define the operator $\hat{C}_\mu : \mathcal{U}^* \to \mathcal{U}^{**}$ by the relation

$$(\hat{C}_\mu \ell_1)(\ell_2) = C_\mu(\ell_1, \ell_2).$$

**Remark 3.6.** It is natural to attempt to define the mean $\hat{m}_\mu \in \mathcal{U}$ and covariance $\hat{C}_\mu : \mathcal{U}^* \to \mathcal{U}$ to be the Bochner integrals

$$m_\mu = \int_{\mathcal{U}} u\, \mu(\mathrm{d}u), \quad \hat{C}_\mu \ell = \int_{\mathcal{U}} u\, \ell(u) \mu(\mathrm{d}u)$$

so that $m_\mu(\ell) = \ell(\hat{m}_\mu)$ and $C_\mu(\ell_1, \ell_2) = \ell_2(\hat{C}_\mu \ell_1)$, and in general this is correct. However, there is an obstacle to these definitions, namely we don't know a priori that $\|u\|$ and $\|u\|^2$ are integrable with respect to $\mu$ and therefore Bochner integrability eludes us.

These integrability issues can be resolved by the following theorem due to Fernique that gives exponential moments for Gaussian meaures.

**Theorem 3.7** (Fernique [Fer71])**.** *Let $\mu \in \mathscr{P}(\mathcal{U})$ be a centered Gaussian measure with covariance $C_\mu$. Then for $\lambda < 1/(2\|C_\mu\|)$ we have*

$$\int_{\mathcal{U}} \exp\left(\lambda \|u\|^2\right) \mu(\mathrm{d}u) < \infty.$$

**Remark 3.8.** This incredible result is very powerful and in fact applies to a much more general class of measures that have a rotation invariance for their product measure. The proof is rather complicated (see Hairer's notes for more details).

An immediate corollary of this result is a better classification of the covariance $\hat{C}_\mu$.

**Corollary 3.9.** *Let $\mu \in \mathscr{P}(\mathcal{U})$ be a centered Gaussian, then $\hat{C}_\mu$ is a bounded linear operator from $\mathcal{U}^*$ to $\mathcal{U}$ with norm $\|C_\mu\|$.*

*Proof.* This follows by recognizing that one has the identity

$$\hat{C}_\mu \ell = \int_{\mathcal{U}} u\, \ell(u)\, \mu(\mathrm{d}u),$$

which is a well defined Bochner integral by the fact that $\|u\ell(u)\| \leq \|u\|^2\|\ell\|_*$ is integrable. Moreover we have

$$\|\hat{C}_\mu \ell\| = \sup_{\|\ell_1\|_*=1} |\ell_1(\hat{C}_\mu \ell)| = \sup_{\|\ell_1\|_*=1} |C_\mu(\ell_1, \ell)| \leq \|C_\mu\|\|\ell\|_*$$

□

**Remark 3.10.** The proof that $m_\mu$ belongs to $\mathcal{U}$ is more subtle, since we have only defined exponential moments for centered Gaussian measures and a priori it is not clear that every non-centered Gaussian measure can be shifted . However, it is true and can be found in Bogachev [Bog98].

## 3.2 The Cameron-Martin space

For a given centered Gaussian measure $\mu$ it is useful to observe that the covariance structure of $\mu$ gives a natural embedding of $\mathcal{U}^*$ into $L^2(\mu)$ since

$$\|\ell\|^2_{L^2(\mu)} = C_\mu(\ell, \ell) < \infty.$$

Denote the corresponding closure $\mathcal{R}_\mu$ of $\mathcal{U}^*$ in $L^2(\mu)$. The space $\mathcal{R}_\mu$ is a Hilbert space with the natural $L^2$ inner product and is therefore separable since $L^2(\mu)$ is. Elements in $\mathcal{R}_\mu$ can be seen as linear functionals away from some measure-zero linear subspace.

**Proposition 3.11.** *For every $\ell \in \mathcal{R}_\mu$, there exists a Borel linear subspace $V_\ell \subseteq \mathcal{U}$ of full measure $\mu(V_\ell) = 1$ such that $\ell|_{V_\ell}$ is a linear map.*

*Proof.* By definition $\ell \in \mathcal{R}_\mu$ is the $L^2(\mu)$ limit of elements in $\mathcal{U}^*$ and therefore since sequences that converge in $L^2$ have $\mu$ almost sure convergence subsequences, we can find a sequence $\{\ell_n\} \subseteq \mathcal{U}^*$ such that $\ell_n \to \ell$ $\mu$-almost surely. Note that we can always take the full measure set of convergence $V_\ell$ to be linear because $\ell_n$ are linear and therefore if $\ell_n$ on a set of points, it also converges on all linear combinations of those points, therefore $\ell|_{V_\ell}$ is linear being the pointwise limit of linear maps. □

**Exercise 3.5.** Show that the operator $\hat{C}_\mu : \mathcal{U}^* \to \mathcal{U}$ can be uniquely extended to a continuous linear operator from $\mathcal{R}_\mu$ to $\mathcal{U}$ and that for each $\ell \in \mathcal{R}_\mu$, the following formula still holds as a Bochner integral

$$\hat{C}_\mu \ell = \int_{\mathcal{U}} u\, \ell(u)\, \mu(\mathrm{d}u).$$

(Hint: show that sequences in $\mathcal{U}^*$ which are Cauchy in $L^2(\mu)$ are Cauchy in $\mathcal{U}$ under $\hat{C}_\mu$)

**Exercise 3.6.** Show that the extension $\hat{C}_\mu : \mathcal{R}_\mu \to \mathcal{U}$ is compact , injective and has dense range.

Since we can extend $\hat{C}_\mu : \mathcal{U}^* \to \mathcal{U}$ uniquely to a continuous linear mapping from $\mathcal{R}_\mu$ to $\mathcal{U}$ , the image of $\mathcal{R}_\mu$ under this map, $\mathcal{H}_\mu := \hat{C}_\mu(\mathcal{R}_\mu) \subseteq \mathcal{U}$ naturally defines another Hilbert space called the *Cameron Martin space* with inner product $\langle \cdot, \cdot \rangle_\mu$, inherited from $L^2(\mu)$ by

$$\langle \hat{C}_\mu \ell_1, \hat{C}_\mu \ell_2 \rangle_\mu := \langle \ell_1, \ell_2 \rangle_{L^2(\mu)}.$$

Moreover by Exercise $\mathcal{H}_\mu$ is a densely and compactly embedded into $\mathcal{U}$ (meaning bounded subsets of $\mathcal{H}_\mu$ are pre-compact in $\mathcal{U}$). It is important to note that using properties of the Bochner integral (3.5) we still have the following formula for each $\ell \in \mathcal{U}^*$ and $\ell' \in \mathcal{R}_\mu$

$$\ell(\hat{C}_\mu \ell') = \langle \ell, \ell' \rangle_{L^2(\mu)}.$$

**Remark 3.12** (Reproducing Kernel Hilbert Space)**.** The term *reproducing kernel Hilbert space* is commonly used in the theory of Gaussian measures to talk about the spaces $\mathcal{R}_\mu$ and $\mathcal{H}_\mu$, and there is considerable confusion about the terminology's use (or misuse) in the literature. For instance, Bogachev [Bog98] (which we are partially following) defines $\mathcal{R}_\mu$ to be the reproducing kernel Hilbert space associated with $\mu$, while DaPrato/Zabczyck call the Cameron Martin space $\mathcal{H}_\mu$ the reproducing kernel Hilbert space associated to $\mu$

Traditionally, a reproducing kernel Hilbert space is a Hilbert space of functions on a set where pointwise evaluation (the delta functional) is a continuous functional and can therefore be represented by a square integrable kernel. Hence the integration of the kernel against a function "reproduces" it pointwise. A simple example of such a space is the Sobolev space $H_0^1([0,1])$ with the kernel $C(s,t) = \min\{s,t\} - st$.

In general neither $\mathcal{R}_\mu$ nor $\mathcal{H}_\mu$ is a reproducing kernel Hilbert space in the traditional sense, since pointwise evaluations of functions $\ell \in \mathcal{R}_\mu$ are not necessarily continuous at all points and $\mathcal{H}_\mu$ is not even a set of functions.However, the connection can be made more apparent if $\mathcal{U}$ is a space of continuous functions (which it is in most cases of interest), for instance $\mathcal{U} = C([0,1],\mathbb{R})$. In this case delta functions $\{\delta_s\}_{s\in[0,1]}$ belong to $\mathcal{U}^*$ and one can show that the Cameron-Martin space $\mathcal{H}_\mu$ is a reproducing kernel Hilbert space in the traditional sense with reproducing kernel $C(s,t) = C_\mu(\delta_s,\delta_t)$.

The following equivalent definitions of $\|\cdot\|_\mu$ and $C_\mu(\ell,\ell)$ are useful.

**Lemma 3.13.** *For each $h \in \mathcal{H}_\mu$*

$$\|h\|_\mu = \sup\left\{|\ell(h)| \,:\, \ell \in \mathcal{U}^*,\, C_\mu(\ell,\ell) \le 1\right\},$$

*and for each and $\ell \in \mathcal{U}^*$*

$$C_\mu(\ell,\ell) = \sup\left\{|\ell(h)|^2 \,:\, h \in \mathcal{H}_\mu,\, \|h\|_\mu \le 1\right\}. \tag{3.1}$$

*Proof.* Let $h \in \mathcal{H}_\mu$. By the definition of $\mathcal{H}_\mu$ there exists a unique $\ell_h \in \mathcal{R}_\mu$ such that $h = \hat{C}_\mu \ell_h$ and therefore using the fact that $\mathcal{R}_\mu$ is a Hilbert space and $\mathcal{U}^*$ is dense in $\mathcal{R}_\mu$ we have

$$\begin{aligned}\|h\|_\mu = \|\ell_h\|_{L^2(\mu)} &= \sup\left\{|\langle \ell,\ell_h\rangle_{L^2(\mu)}| \,:\, \ell \in \mathcal{R}_\mu,\, \|\ell\|_{L^2(\mu)} \le 1\right\}\\ &= \sup\left\{|\ell(\hat{C}_\mu \ell_h)| \,:\, \ell \in \mathcal{U}^*,\, C_\mu(\ell,\ell) \le 1\right\}\\ &= \sup\left\{|\ell(h)| \,:\, \ell \in \mathcal{U}^*,\, C_\mu(\ell,\ell) \le 1\right\}.\end{aligned}$$

Additionally let $\ell \in \mathcal{U}^*$ the again by the Hilbert space structure of $\mathcal{R}_\mu$

$$\begin{aligned}C_\mu(\ell,\ell) = \|\ell\|^2_{L^2(\mu)} &= \sup\left\{\langle \ell,\ell'\rangle^2_{L^2(\mu)} \,:\, \ell' \in \mathcal{R}_\mu,\, \|\ell'\|_{L^2(\mu)} \le 1\right\}\\ &= \sup\left\{|\ell(\hat{C}_\mu(\ell'))|^2 \,:\, \ell' \in \mathcal{R}_\mu,\, \|\hat{C}_\mu \ell'\|_\mu \le 1\right\}\\ &= \sup\left\{|\ell(u)|^2 \,:\, u \in \mathcal{H}_\mu,\, \|u\|_\mu \le 1\right\}.\end{aligned}$$

□

**Remark 3.14.** It is interesting to see that (3.1) implies that for a given $\ell \in \mathcal{U}^*$, the variance $C_\mu(\ell,\ell)$ can be determined simply by calculating the norm of $\ell$ on $\mathcal{H}_\mu$.

As it turns out, given a Banach space $\mathcal{U}$, a Gaussian measure is uniquely determined by it's Cameron Martin space $\mathcal{H}_\mu$.

**Proposition 3.15.** *Let $\mu$ and $\nu$ be two centered Gaussian meaures on $\mathcal{U}$ and suppose that $\mathcal{H}_\mu = \mathcal{H}_\nu$ and $\|h\|_\mu = \|h\|_\nu$ for all $h \in \mathcal{H}_\mu$, then $\mu = \nu$.*

*Proof.* By Lemma 3.13 we immediately find that for each $\ell \in \mathcal{U}^*$

$$C_\mu(\ell, \ell) = C_\nu(\ell, \ell),$$

which implies that $\hat{\mu}(\ell) = \hat{\nu}(\ell)$. □

**Remark 3.16.** The fact that a Gaussian measure on a Banach space can be completely determined by it's Cameron-Martin space is very powerful and allows one to specify a Gaussian measure $\mu$ on $\mathcal{U}$ simply by specifying a Hilbert space $(\mathcal{H}_\mu, \|\cdot\|_\mu)$ that embeds into $\mathcal{U}$ in a certain way.

This embedding $J : \mathcal{H}_\mu \to \mathcal{U}$ along with $(\mathcal{H}_\mu, \|\cdot\|_\mu)$ is referred to as an *abstract Wiener space* and the measure $\mu$ the canonical measure on $\mathcal{U}$. The theory of abstract Wiener spaces was introduced by Leonard Gross [Gro67] to general the construction of the Wiener measure given by Norbert Wiener.

In light of the defining properties of the Cameron-Martin space, a rather startling fact is that that in infinite dimensions $\mathcal{H}_\mu$ is actually $\mu$ measure zero.

**Proposition 3.17.** *If* $\dim(\mathcal{U}) = \infty$ *then* $\mu(\mathcal{H}_\mu) = 0$.

*Proof.* If $\dim(\mathcal{U}) = \infty$, let $\{\hat{\ell}_k\}_{k\in\mathbb{N}}$ be an orthonormal basis for $\mathcal{R}_\mu$ and note that if $h \in \mathcal{H}_\mu$, then

$$\sum_{k\in\mathbb{N}} |\hat{\ell}_k(h)|^2 = \sum_{k\in\mathbb{N}} \langle \hat{\ell}_k, \ell_h \rangle = \|\ell_h\|^2_{L^2(\mu)} = \|h\|^2_\mu < \infty.$$

Since $\{\hat{\ell}_k\}_{k\in\mathbb{N}}$ form a family of iid standard Gaussian random variables on the probability space $(\mathcal{U}, \mathscr{B}(\mathcal{U}), \mu)$, by the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k\in\mathbb{N}} |\hat{\ell}_k|^2 = 1 \quad \mu\text{-almost surely}$$

In particular this implies that

$$\sum_{k\in\mathbb{N}} |\hat{\ell}_k|^2 = \infty \quad \mu\text{-almost surely},$$

which implies that $\mu(\mathcal{H}_\mu) = 0$. □

As it turns out, in infinite dimensions, it is extremely easy for measures to be mutually singular. Indeed, even simple translations and dilations of Gaussian measures can be mutually singular with respect to each other (in contrast to the finite dimensional case). This is illustrated in the following result.

**Proposition 3.18.** *Let* $\mu$ *be a centered Gaussian measure on* $\mathcal{U}$ *with* $\dim(\mathcal{U}) = \infty$. *For each* $c \in \mathbb{R}$ *define* $D_c u = cu$ *to be the dilation of* $u \in \mathcal{U}$ *by* $c$. *Then if* $c \neq \pm 1$, $(D_c)_*\mu$ *and* $\mu$ *are mutually singular.*

*Proof.* To prove this, let $\{\hat{\ell}_k\}_{k\in\mathbb{N}}$ be a an orthonormal basis for $\mathcal{R}_\mu$. Again, by the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} |\hat{\ell}_k|^2 \to 1 \quad \mu \text{ almost surely.}$$

On the other hand under the measure $(D_c)_*\mu$ the set $\{\hat{\ell}_k\}$ are still iid Gaussian random variables, however by linearity of $\hat{\ell}_k$ they have variance $\|\hat{\ell}_k\|^2_{L^2(D_c^*\mu)} = c^2$. Therefore the strong law of large numbers also gives

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} |\hat{\ell}_k|^2 \to c^2 \quad (D_c)_*\mu \text{ almost surely.}$$

This implies that if $c \neq \pm 1$ then the measures $\mu$ and $(D_c)_*\mu$ are mutually singular since the set on which (3.2) holds must be zero measure for $(D_c)_*\mu$ and vice-versa. □

For shifts of centered Gaussian measures, we have the following complete characterization in terms of the Cameron-Martin space.

**Theorem 3.19** (Cameron-Martin)**.** *Let $\mu$ be a centered Gaussian measure on $\mathcal{U}$. For each $h \in \mathcal{U}$ define $T_h u = u + h$ to be the shift of $u \in \mathcal{U}$ by $h$. Then $\mu_h := (T_h)_*\mu$ is absolutely continuous with respect to $\mu$ if and only if $h \in \mathcal{H}_\mu$ with Radon-Nikodym derivative*

$$\frac{\mathrm{d}\mu_h}{\mathrm{d}\mu}(u) = \exp\left(\ell_h(u) - \frac{1}{2}\|h\|_\mu^2\right), \tag{3.2}$$

*where $\ell_h$ is the unique element of $R_\mu$ such that $h = \hat{C}_\mu \ell_h$.*

*Proof.* Suppose that $h \in \mathcal{H}_\mu$, then by studying the characteristic function, we easily find

$$\hat{\mu}_h(\ell) = e^{i\ell(h)}\hat{\mu}(\ell) = \exp\left(i\ell(h) - \frac{1}{2}C_\mu(\ell, \ell)\right).$$

Using the fact that $\ell(h) = C_\mu(\ell, \ell_h) = C_\mu(\ell_h, \ell)$ and $C_\mu(\ell_h, \ell_h) = \|h\|_\mu^2$, we easily find

$$i\ell(h) - \frac{1}{2}C_\mu(\ell, \ell) = \frac{1}{2}C_\mu(\ell - i\ell_h, \ell - i\ell_h) - \frac{1}{2}\|h\|_\mu^2.$$

Therefore

$$\begin{aligned}\hat{\mu}_h(\ell) &= \exp\left(C_\mu(\ell - i\ell_h, \ell - i\ell_h) - \frac{1}{2}\|h\|_\mu^2\right) = \hat{\mu}(\ell - i\ell_h)\exp\left(-\frac{1}{2}\|h\|_\mu^2\right) \\ &= \int_{\mathcal{U}} e^{i\ell(u)} \exp\left(\ell_h(u) - \frac{1}{2}\|h\|_\mu^2\right)\mu(\mathrm{d}u).\end{aligned}$$

which implies absolute continuity of $\mu_h$ with respect to $\mu$ with the Radon-Nikodym derivative (3.2).

To show the converse. Suppose that $h \notin \mathcal{H}_\mu$. Then by Lemma 3.13 there exists a sequence of $\ell_k \in \mathcal{U}^*$ with $C_\mu(\ell_k, \ell_k) = 1$ such that $|\ell_k(u)| \to \infty$ as $k \to \infty$. It follows that

$$\|\mu_h - \mu\|_{TV} \geq \|(\ell_k)_*\mu_h - (\ell_k)_*\mu\|_{TV} = 1 - ((\ell_k)_*\mu_h \wedge (\ell_k)_*\mu)(\mathbb{R}).$$

Using the fact that

$$\begin{aligned}((\ell_k)_*\mu_h \wedge (\ell_k)_*\mu)(\mathbb{R}) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \min\left\{e^{-\frac{x^2}{2}}, e^{-\frac{(x-\ell_k(h))^2}{2}}\right\}\mathrm{d}x \\ &= \frac{2}{\sqrt{2\pi}}\int_{x > |\ell_k(h)|/2} e^{-\frac{x^2}{2}}\mathrm{d}x\end{aligned}$$

Therefore $((\ell_k)_*\mu_h \wedge (\ell_k)_*\mu)(\mathbb{R}) \to 0$ as $k \to \infty$. Passing the limit as $k \to \infty$ on both sides of (3.2) implies that $\|\mu_h - \mu\|_{TV} = 1$ and therefore $\mu_h$ and $\mu$ are mutually singular. □

A natural consequence of Theorem 3.19 is that shifts in Cameron-Martin directions preserve full-measure and zero-measure sets. This gives the following characterization of the Cameron-Martin space.

**Proposition 3.20.** *Let $\mu$ be a centered Gaussian measure on $\mathcal{U}$. Then $\mathcal{H}_\mu$ is the intersection of all linear subspaces of full measure.*

*Proof.* As mentioned already, if $L$ is a full measure subspace of $\mathcal{U}$, then by Theorem 3.19 $L + h$ is also full measure for each $h \in \mathcal{H}_\mu$ and therefore $h \in L$.

Alternatively, suppose that $h \notin \mathcal{H}_\mu$ and let $\{\ell_k\} \subseteq \mathcal{U}$ be such that $|\ell_k(h)| > k$ and $C_\mu(\ell_k, \ell_k) = 1$. Since

$$\mathbf{E} \sum_{k\in\mathbb{N}} \frac{1}{k^2} |\ell_k(u)|^2 = \sum_{k\in\mathbb{N}} \frac{1}{k^2} < \infty.$$

If follows that the linear space

$$L = \left\{ u \in \mathcal{U} \,:\, \sum_{k\in\mathbb{N}} \frac{1}{k^2} |\ell_k(u)|^2 < \infty \right\}$$

has full measure and since by construction $\sum_{k\in\mathbb{N}} \frac{1}{k^2} |\ell_k(h)|^2 > \infty$ we have $h \notin L$. □

## 3.3 Hilbert space case

In the case when $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space (and it's dual is identified with $\mathcal{H}$ by Riesz representation), the behavior of the mean $m_\mu(u)$ and the covariance $C_\mu(u,u)$ can be shown to be much nicer and the Cameron Martin-space identifies more explicitly. Indeed, Corollary 3.4 implies by Riesz representation that there exists a vector $m \in \mathcal{H}$ and bounded symmetric positive semi-definite operator $Q : \mathcal{H} \to \mathcal{H}$ such that

$$m_\mu(u) = \langle m, u \rangle, \quad C_\mu(u,u) = \langle Qu, u \rangle.$$

In this case we will refer to $m$ and $Q$ as the mean and covariance of $\mu$ respectively.

In fact on Hilbert spaces, Gaussian measures can be completely classified in terms of the space of trace-class (or nuclear) symmetric positive semi-definite operators.

**Definition 3.21.** A trace-class (or nuclear) operator $T$ on $\mathcal{H}$ is such that for each orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$, we have

$$\|T\|_1 := \sum_{k\in\mathbb{N}} |\langle Te_k, e_k \rangle| < \infty.$$

This norm is independent of the choice of orthonormal basis and denotes a Banach space of trace-class operators, denoted by $\mathcal{L}_1(\mathcal{H})$. In particular the trace operator

$$\operatorname{Tr} T := \sum_{k\in\mathbb{N}} \langle Te_k, e_k \rangle$$

is a well defined continuous linear functional on $\mathcal{L}_1(\mathcal{H})$. The space of trace-class *symmetric positive semi-definite* operators is denoted by $\mathcal{L}_1^+(\mathcal{H})$.

The following proposition is classical and will allow us to completely characterize Gaussian measures on a Hilbert space.

**Proposition 3.22** (Hilbert-Schmidt)**.** *Let $Q \in \mathcal{L}_1(\mathcal{H})^+$, then there exists an orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ of $\mathcal{H}$ and $\{q_k\}_{k\in\mathbb{N}}$, $q_k \geq 0$ such that for each $k \in \mathbb{N}$*

$$Qe_k = q_k e_k.$$

**Theorem 3.23** (Characterization of Gaussian measures)**.** *A probability measure $\mu$ on a separable Hilbert space $\mathcal{H}$ is Gaussian if and only if there exists a $m \in \mathcal{H}$ and $Q \in \mathcal{L}_1^+(\mathcal{H})$ such that*

$$\hat{\mu}(u) = \exp\left( i\langle m, u \rangle - \frac{1}{2} \langle Qu, u \rangle \right).$$

*Moreover, for every $m \in \mathcal{H}$ and $Q \in \mathcal{L}_1^+(\mathcal{H})$ there exists a Gaussian measure with mean $m$ and covariance $Q$.*

*Proof.* For the first part of the proposition, by exercise 3.1, we simply need to show that $Q$ is trace-class. To see this, we note that since $m \in \mathcal{H}$ we may assume that $\mu$ is mean zero since we can always shift by $m$. It follows that for any orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$

$$\operatorname{Tr} Q = \sum_{k\in\mathbb{N}} \langle Qe_n, e_n\rangle = \sum_{k\in\mathbb{N}} \int_{\mathcal{H}} \langle u, e_k\rangle^2 \mu(\mathrm{d}u) = \int_{\mathcal{H}} \|u\|^2 \mu(\mathrm{d}u) < \infty,$$

by Fernique's Theorem 3.7.

To prove that there exists a Gaussian measure with mean $m$ and covariance $Q$, we let $\{e_k\}_{k\in\mathbb{N}}$ by the eigenfunctions of $Q$ with eigenvalues $\{q_k\}_{k\in\mathbb{N}}$ satisfying $Qe_k = q_k e_k$ (such a spectral decomposition follows from the Hilbert-Schmidt theorem). Then we let $\{\xi_k\}_{k\in\mathbb{N}}$ be a sequence of iid real standard Gaussians (mean zero and unit variance) on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$ (such a family exists by Kolmogorov extension) and define the random variable $\xi \in \mathcal{H}$

$$\xi(\omega) = m + \sum_{k\in\mathbb{N}} \sqrt{q_k}\xi_k(\omega)e_k. \tag{3.3}$$

Note that this series converges almost surely in $\mathcal{H}$, since

$$\sum_{k\in\mathbb{N}} q_k \mathbf{E}|\xi_k|^2 = \operatorname{Tr} Q < \infty$$

and therefore converges by the Kolmogorov series test. Obviously $\xi$ is measurable since it is the limit of measurable functions. We then define $\mu = \operatorname{Law}(\xi)$. Using independence of $\xi_k$, the characteristic function of $\mu$ has the form

$$\begin{aligned}
\hat{\mu}(u) = \mathbf{E}\exp\left(i\langle u, \xi\rangle\right) &= \exp\left(i\langle u, m\rangle\right) \prod_{k\in\mathbb{N}} \mathbf{E}\exp\left(\sqrt{q_k}\langle u, e_k\rangle\xi_k\right) \\
&= \exp\left(i\langle u, m\rangle - \frac{1}{2}\sum_{k\in\mathbb{N}} q_k\langle u, e_k\rangle^2\right) \\
&= \exp\left(i\langle u, m\rangle - \frac{1}{2}\langle Qu, u\rangle\right),
\end{aligned}$$

so that $\mu$ is a Gaussian with mean $m$ and covariance $Q$. □

**Remark 3.24.** This can actually be proved without Fernique's theorem using the characteristic function and more crude estimates on Gaussian tails.

**Exercise 3.7.** Suppose that $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space show and $\mu$ is centered Gaussian measure with covariance $Q$ and let $\{e_k\}_{k\in\mathbb{N}}$ and $\{q_k\}_{k\in\mathbb{N}}$ be the eigen functions and associated eigen values. Show that $\mathcal{H}_\mu = Q^{1/2}(\mathcal{H})$ and that the set $\{\sqrt{q_k}e_k\}$ form an orthonormal basis for $\mathcal{H}_\mu$ and that $Q^{-1/2} : \mathcal{H}_\mu \to \mathcal{H}$ forms Hilbert-Schmidt embedding of $\mathcal{H}_\mu$ into $\mathcal{H}$. In addition, show that

$$\langle h_1, h_2\rangle_\mu = \langle Q^{-1/2}h_1, Q^{-1/2}h_2\rangle = \langle Q^{-1}h_1, h_2\rangle = \sum_k \langle h_1, e_k\rangle\langle h_2, e_k\rangle q_k^{-1},$$

where the last sum is over all $k$ such that $q_k > 0$. (Hint: Show that $\mathcal{R}_\mu$ can be identified with $Q^{-1/2}(\mathcal{H}\backslash\operatorname{Ker}(Q))$).

As a consequence of this theorem, we see that there is no such thing a "standard" Gaussian on an infinite dimensional Hilbert space.

**Corollary 3.25.** *Suppose that* $\dim(\mathcal{H}) = \infty$ *then there is no Gaussian measure on* $\mathcal{H}$ *with covariance* $Q = \mathrm{Id}$.

Additionally, we can obtain a more quantitative version of Fernique's theorem.

**Proposition 3.26.** *Let* $\mu$ *be a centered Gaussian measure with covariance* $Q \in \mathcal{L}_1^+(\mathcal{H})$. *Then for* $\lambda \leq 1/(2\operatorname{Tr} Q)$, *we have*

$$\int_{\mathcal{H}} \exp\left(\lambda \|u\|^2\right) \mu(\mathrm{d}u) = \exp\left(-\frac{1}{2}\operatorname{Tr}\log(1-2\lambda Q)\right) \leq \frac{1}{\sqrt{1-2\lambda \operatorname{Tr} Q}}$$

*Proof.* Let $\xi$ by the random variable defined by (3.3) then since $\lambda q_k \leq 1/2$

$$\begin{aligned}\int_{\mathcal{H}} \exp\left(\lambda \|u\|^2\right) \mu(\mathrm{d}u) &= \prod_{k\in\mathbb{N}} \mathbf{E}\exp(\lambda q_k \|\xi_k\|^2) = \prod_k \frac{1}{\sqrt{1-2\lambda q_k}} \\ &= \exp\left(-\frac{1}{2}\sum_k \log(1-2\lambda q_k)\right).\end{aligned}$$

The final inequality follows from the fact that

$$\sum_{k\in\mathbb{N}} \log(1-2\lambda q_k) \geq \log\left(1 - 2\lambda \sum_{k\in\mathbb{N}} q_k\right),$$

which can be proved by induction. □

Finally we can obtain a more powerful version of the Cameron-Martin theorem.

**Theorem 3.27** (Feldman-Hajek theorem)**.** *Let* $\mu_1$ *and* $\mu_2$ *be two Gaussian measures with mean and covariance* $(m_1, Q_1)$ *and* $(m_2, Q_2)$ *respectively. The measures are equivalent (mutually absolutely continuous) if and only if*

1. $Q_1^{1/2}(\mathcal{H}) = Q_2^{1/2}(\mathcal{H}) = \mathcal{H}_0$
2. $m_1 - m_2 \in \mathcal{H}_0$
3. *The operator* $(Q_1^{-1/2}Q_2^{1/2})(Q_1^{1/2}Q_2^{1/2})^* - I$ *is a Hilbert-Schmidt operator on* $\overline{\mathcal{H}_0}$.

*Otherwise they are mutually singular.*

The proof can be found in [DPZ14] Theorem 2.25.

## 3.4 Spaces of continuous functions: Gaussian processes and regularity

### 3.4.1 Real valued functions

Another common case is when $\mathcal{U} = C([0,1])$ is the space of real valued continuous functions on $[0,1]$. Namely continuous functions $\varphi : [0,1] \to \mathbb{R}$ with norm

$$\|\varphi\| = \sup_{t\in[0,1]} |\varphi(t)|.$$

In this case we know that the dual space is $\mathcal{U}^* = \mathcal{M}([0,1])$ , the set of all finite variation signed Borel measures on $[0,1]$ (see [Yos95], p119). Moreover, let $D \subseteq \mathcal{M}([0,1])$ be the subspace of all finite linear combinations of delta measures $\nu = \sum_k a_k \delta_{t_k}$, then $D$ is a dense set in the weak-* topology.

**Lemma 3.28.** *$D$ is dense in $\mathcal{M}([0,1])$ in the weak-* topology.*

*Proof.* Let $\varphi \in C[0,1]$, and let $P = \{0 = t_0 < \dots t_n = 1\}$ be a partition of $[0,1]$ and $|P|$ be the size of the partition. Let

$$\varphi_P(t) = \sum_{j=1}^{n} \varphi(t_j) \mathbb{1}_{(t_{j-1}, t_j]}(t)$$

be a step function approximation of $\varphi$. Since $\varphi$ is uniform continuous, we know that $\varphi_P \to \varphi$ in $C([0,1])$ as $|P| \to 0$ and therefore $\lim_{|P|\to 0} \nu(\varphi_P) = \nu(\varphi)$, which is equivalent to

$$\lim_{|P|\to 0} \sum_{j=1}^{n} \nu((t_{j-1}, t_j])\delta_{t_j} = \nu \quad \text{weakly-*}.$$

□

As a consequence, we see that a Gaussian random variable in $C([0,1])$ is determined by finite linear combinations of point evaluations.

**Corollary 3.29.** *A probability measure $\mu$ on $C([0,1])$ is Gaussian if and only if $\ell^*\mu$ is Gaussian for each $\ell \in D$. Equivalently a random variable $X \in C([0,1])$ is Gaussian if and only if for each $t_1, t_2, \dots t_n \in [0,1]$, $(X_{t_1}, X_{t_2}, \dots X_{t_n})$ is a Gaussian random vector in $\mathbb{R}^n$.*

*Proof.* Let $\ell \in \mathcal{U}^*$ and $\{\ell_k\}_{k\in\mathbb{N}}$ be a sequence converging to $\ell$ in the weak-* topology, then for each $u \in \mathcal{U}$, $\ell_k(u) \to \ell(u)$. Since $\{\ell_k\}$ is a sequence of Gaussian random variables converging pointwise, the limit must be Gaussian. □

In light of Corollary 3.29 the mean $m_\mu$ and covariance $C_\mu$ can naturally be described in terms of a mean and covariance function $m : [0,1] \to \mathbb{R}$ and $C : [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$m(t) := m_\mu(\delta_s), \quad C(s,t) := C_\mu(\delta_s, \delta_t). \tag{3.4}$$

**Exercise 3.8.** Show that the covariance function $C(s,t)$ defined by (3.4) is symmetric and continuous in each variable and *positive definite* in the sense that for each $\{t_j\}_{j=1}^n \subset [0,1]$ and $\{a_j\}_{j=1}^n \subset \mathbb{R}$,

$$\sum_{i,j=1}^{n} a_i a_j C(t_i, t_j) \geq 0. \tag{3.5}$$

Moreover it completely determines $C_\mu$ by the formula,

$$C_\mu(\nu_1, \nu_2) = \int_0^1 \int_0^1 C(t,s)\nu_1(\mathrm{d}s)\nu_2(\mathrm{d}t),$$

for $\nu_1, \nu_2 \in \mathcal{M}([0,1])$

As it turns out the covariance functions $C(s,t)$ plays the role of a reproducing kernel for the Cameron-Martin space $\mathcal{H}_\mu \subset C([0,1])$.

**Lemma 3.30** (Reproducing kernel property)**.** *Let $h \in \mathcal{H}_\mu$, then $C(t,\cdot) \in \mathcal{H}_\mu$ and satisfies*

$$\langle C(t,\cdot), h\rangle_\mu = h(t).$$

*Proof.* Note that for each $\nu \in M([0,1]) = \mathcal{U}^*$ we have

$$(\hat{C}_\mu \nu)(s) = C_\mu(\delta_t, \nu) = \int_0^1 a(t,s)\nu(\mathrm{d}t).$$

There choosing $\nu = \delta_t$ we have $\hat{C}_\mu \delta_t = a(t,\cdot)$. This implies that

$$\langle C(t,\cdot), h\rangle_\mu = \langle \hat{C}_\mu \delta_t, h\rangle_\mu = \delta_t(h) = h(t)$$

□

The following are well-known examples of Gaussian measures on $C([0,1])$. And will be constructed in generality in the next section.

**Example 3.31** (Wiener process)**.** An example of a Gaussian measure on $C([0,1])$ it the one associated to the Wiener process $W(t)$ on $[0,1]$ called the *Wiener measure* defined to have $m(t) = 0$ and $C(s,t) = \mathbf{E}W(t)W(s) = \min\{s,t\}$.

**Example 3.32** (Brownian bridge)**.** The Brownian bridge $B(t)$ is a Gaussian process with $m(t) = 0$ and $C(s,t) = \mathbf{E}B(s)B(t) = \min\{s,t\} - st$

**Example 3.33** (Ornstein-Uhlenbeck process)**.** The Ornstein-Uhlenbeck process $O(t)$, which has $m(t) = 0$ and $C(s,t) = \mathbf{E}O(t)O(s) = e^{-|t-s|/2}$. Such a process with covariance that depends only on $|t-s|$ is known as a stationary process. Moreover it can be shows that the Ornstein Uhlenbeck-process can be obtained related to the Wiener process by

$$O(t) = e^{-t/2}W(e^t).$$

**Exercise 3.9.** Show that the Cameron-Martin space $\mathcal{H}_\mu$ associated with the Wiener measure $\mu$ defined by $m(t) = 0$, $C(s,t) = \min\{s,t\}$. Coincides with the Sobolev space $H_0^1([0,1])$ of absolutely continuous functions $h \in C([0,1])$ that can be represented as

$$h(t) = \int_0^t f(s)\mathrm{d}s$$

for some $f \in L^2([0,1])$ with inner product

$$\langle h_1, h_2\rangle_\mu = \int_0^1 h_1'(s)h_2'(s)\mathrm{d}s = \int_0^1 f_1(s)f_2(s)\mathrm{d}s.$$

### 3.4.2 Kolmogorov extension and continuity criterion

It is important to understand whether a given mean $m(t)$ and a symmetric positive definite covariance function $C(s,t)$ can be realized as the mean and covariance of a certain Gaussian measure on $C([0,1])$. It is easy to see that for each finite collection of times $T = \{t_i\}_{i=1}^n \subset [0,1]$ there is a Gaussian measure $\mu_T$ on $\mathbb{R}^n$ with mean $(m_T)_i = m(t_i)$ and covariance $(Q_T)_{ij} = C(t_i,t_j)$. Moreover, it is easy to check that the measures

$$\{\mu_T : T \subset [0,1] \text{ finite collection}\}$$

form a consistent set of measures and therefore by the Kolmogorov extension theorem, there exists a probability measure $\mathbf{P}$ on the measurable space $\Omega = \mathbb{R}^{[0,1]}$, endowed with the product sigma algebra $\mathscr{F} = \mathscr{B}(\mathbb{R})^{\otimes[0,1]}$, such that

$$\int_{\mathbb{R}^{[0,1]}} \omega(t)\,\mathbf{P}(\mathrm{d}\omega) = m(t) \quad \text{and} \quad \int_{\mathbb{R}^{[0,1]}} (\omega(t) - m(s))(\omega(s) - m(s))\mathbf{P}(\mathrm{d}\omega) = C(s,t).$$

When $\Omega = \mathbb{R}^{[0,1]}$ is equipped with the measure $\mathbf{P}$, we can associate to point $\omega$ a *Gaussian stochastic process* $X = \{X(t) : t \in [0,1]\}$ with values in $\mathbb{R}$ (known as the canonical process), defined by

$$X(t,\omega) = \omega(t). \tag{3.6}$$

**Remark 3.34.** It is extremely important to recognize that the measure $\mathbf{P}$ is constructed by the axiom of choice, and is only defined on the *product sigma-algebra* $\mathscr{B}(\mathbb{R})^{\otimes[0,1]}$ defined to be the smallest sigma-algebra containing product sets $\bigtimes_{t\in[0,1]} A_t$ for $A_t \in \mathscr{B}([0,1])$. Such a sigma-algebra is extremely weak and is *much* coarser than the Borel sigma-algebra associated with the product sigma-algebra.

We can extend the idea of a stochastic process $X = \{X(t) : t \in [0,1]\}$ to more general Banach space valued valued processes on more general index sets.

**Definition 3.35.** Given a separable Banach space $\mathcal{U}$ and an index set $I \subseteq \mathbb{R}$, a $\mathcal{U}$-valued *stochastic process* $X = \{X(t) : t \in I\}$ on $(\Omega, \mathscr{F}, \mathbf{P})$ is such that $X(t)$ is a $\mathcal{U}$-valued random variable for each $t \in I$. We say the the process $X$ a *measurable* if $X : I \times \Omega \to \mathcal{U}$ is measurable from $\mathscr{B}(I) \otimes \mathscr{F} \to \mathscr{B}(\mathcal{U})$.

We would like to understand when a process $X$ belongs to a space of continuous functions $C(I;\mathcal{U})$. One natural form of continuity that is easily obtained is known as mean square continuity. We define it below for general

**Definition 3.36.** A $\mathcal{U}$-valued process $X = \{X(t) : t \in I\}$ is *mean square continuous* if

$$\lim_{\substack{t\in I\\ t\to t_0}} \mathbf{E}\|X(t) - X(t_0)\|^2 = 0, \quad \text{for each } t_0 \in I.$$

**Proposition 3.37.** *Assume that $t \mapsto m(t)$ and $(t,s) \mapsto C(t,s)$ are continuous. Then $X(t)$ defined by (3.6) is mean square continuous.*

*Proof.* Note that it suffices to assume that $m(t) = 0$ since we mean square continuity of $X(t) - m(t)$ implies mean square continuity of $X(t)$. We note that by expanding the square,

$$\mathbf{E}|X(t) - X(t_0)|^2 = C(t,t) - 2C(t,t_0) + C(t_0,t_0)$$

and therefore mean-square continuity follows from continuity of $C(t,s)$. □

In general it was realized by Kolmogorov that mean square continuity of a process $X$ was not enough to prove sample path continuity of $t \mapsto X(t)$. Indeed, the process $X$ defined by (3.6) is not even a measurable process. Specifically, events that involves pathwise statements on $X$ may are not measurable. For instance, for each $A > 0$ the event

$$\left\{\omega \in \mathbb{R}^{[0,1]} \;:\; \sup_{t\in[0,1]} |\omega(t)| \leq A\right\} = \bigcap_{t\in[0,1]} \left\{\omega \in \mathbb{R}^{[0,1]} \;:\; |\omega(t)| < A\right\}$$

*is not* in the product sigma algebra $\mathscr{F}$ on $\mathbb{R}^{[0,1]}$ since it involves countably many intersections of measurable events. This means that the even the statement "$X(t)$ is continuous" is not a measurable event. In order to get around this, one needs the idea of a modification.

**Definition 3.38.** Given a $\mathcal{U}$-valued process $X = \{X(t) : t \in I\}$ on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$, a $\mathcal{U}$-valued process $\bar{X} = \{\bar{X}(t) : t \in I\}$ is a modification of $X$ if

$$\mathbf{P}\left(\bar{X}(t) = X(t)\right) = 1 \quad \text{for all } t \in I.$$

If the modification is almost surely continuous, then it is called a *continuous modification*.

**Remark 3.39.** On $\mathbb{R}^{[0,1]}$ one may view a modification as a family of measurable functions $\{I_t\}_{t\in[0,1]}$, $I_t : \mathbb{R}^{[0,1]} \to \mathbb{R}$ such that $I_t(\omega) = \omega(t)$ $\mathbf{P}$-almost surely. Namely $I_t$ is the evaluation function away from measure zero set. Moreover, a continuous modification is maps almost every $\omega \in \mathbb{R}^{[0,1]}$ to a continuous function $t \mapsto I_t(\omega)$.

**Remark 3.40.** It is important to note that we may assume a continuous modification is 0 on any probability zero set and that the finite dimensional distributions match those of $X$. Consequently the law of $\{X(t) : t \in I\}$ is a measure on $C(I;\mathcal{U})$ and shares the same covariance function as $X$.

The following theorem is due to Kolmogorov and gives a sufficient criterion for existence of a continuous modification

**Theorem 3.41** (Kolmogorov continuity)**.** *Let $X = \{X(t) : t \in [0,1]\}$ be a $\mathcal{U}$-valued process that satisfies*

$$\mathbf{E}\|X(t) - X(s)\|^{\beta} \leq C|t-s|^{1+\alpha} \tag{3.7}$$

*for some positive constants $\alpha, \beta, C$, then $X$ has a continuous modification.*

*Proof.* To get around the measurability problem, we restrict ourselves to dyadic times $\mathbb{D} \subset [0,1]$. Denote $\mathbb{D}_n$ to be times of the form $t = j2^{-n}$ for $0 \leq j \leq 2^n$ then $\mathbb{D}$ is defined by $\mathbb{D} = \bigcup_{n\in\mathbb{N}} \mathbb{D}_n$, and is clearly a countable dense subset of $[0,1]$. To continue, we will denote the modulus of continuity of $X$ on $\mathbb{D}$ by

$$\delta_n(X) := \sup_{1\leq j\leq 2^n} \|X(j2^{-n}) - X((j-1)2^{-n})\|$$

Note that $\delta_n(X)$ is measurable since it only involves finitely many time evaluations of $X$. If we can show that there exists $\gamma > 0$ and $\delta > 0$ such that

$$\mathbf{P}\left(\delta_n(X) > 2^{-n\gamma}\right) \leq C2^{-n\delta}, \tag{3.8}$$

then by Borel-Cantelli, $\lim_{n\to\infty} \delta_n(X) = 0$ $\mathbf{P}$-almost surely and therefore $t \mapsto X(t)$ is $\mathbf{P}$-almost surely uniformly continuous on $\mathbb{D}$. Since $\mathbb{D}$ is dense, it follows that we can define a continuous modification on $[0,1]$ by,

$$\bar{X}(t,\omega) = \lim_{\substack{s\in\mathbb{D}\\ s\to t}} X(s,\omega)\mathbb{1}_{\Omega_0}(\omega).$$

It remains to prove (3.8). This can be estimated as follows

$$\begin{aligned}
\mathbf{P}\left(\delta_n(X) > 2^{-n\gamma}\right) &\leq 2^n \sup_{1\leq j\leq 2^n} \mathbf{P}\left(\|X(j2^{-n}) - X((j-1)2^{-n})\| > 2^{-n\gamma}\right) \\
&\leq 2^{n(1+\beta\gamma)} \sup_{1\leq j\leq 2^n} \mathbf{E}\|X(j2^{-n}) - X((j-1)2^{-n})\|^{\beta} \\
&\leq C2^{n(1+\beta\gamma)}2^{-n(1+\alpha)} \\
&= C2^{-n(\alpha-\beta\gamma)} \leq C2^{-n\delta}
\end{aligned}$$

If one chooses $\gamma$ small enough so that $0 \leq \delta = \alpha - \beta\gamma$. □

As a natural corollary we have the following existence theorem for Gaussian measures on $C([0,1])$.

**Corollary 3.42.** *Let $m(t)$ and $C(s,t)$ satisfy*

$$|m(s) - m(t)| \leq C_1|t-s|^{\alpha}, \quad |C(s,s) - 2C(t,s) - C(t,t)| \leq C_2|t-s|^{2\alpha}, \tag{3.9}$$

*for positive constants $C_1, C_2$ and $\alpha \in (0,1)$ and let $C(s,t)$ be positive definite in the sense of (3.5). Then there exists a Gaussian measure $\mu$ on $C([0,1])$ with mean $m(t)$ and covariance $C(s,t)$.*

*Proof.* Following our construction above, we can build our canonical Gaussian process $X = \{X(t) : t \in [0,1]\}$ with the right covariance and mean. Furthermore conditions (3.9) imply that

$$\mathbf{E}|X(t) - X(s)|^2 \leq C|t-s|^{2\alpha}.$$

This alone is not enough to apply Theorem 3.41. However since our process is Gaussian we can bound $p$-th moments in terms of second moments.

$$\mathbf{E}|X(t) - X(s)|^p \leq C_p \left(\mathbf{E}|X(t) - X(s)|^2\right)^{p/2} \leq C_p|t-s|^{\alpha p}$$

Therefore, choosing $p$ large enough so that $\alpha p > 1$, we see that Kolmogorov's criterion is satisfied. □

### 3.4.3 Modulus of continuity

We can deduce a more precise modulus of continuity on a continuous process $X$. The following inequality is due to Garsia, Rodemich, Rumsey and Rosenblatt and is essentially a more quantitative Sobolev embedding theorem. It allows us to transfer an integral bound to a pointwise estimate on the modulus of continuity,

**Theorem 3.43** (Garcia-Rodemich-Rumsey-Rosenblatt [GRRR70])**.** *Let $\Psi$ and $p$ be continuous strictly increasing functions on $\mathbb{R}_+$ with $p(0) = 0$ and $\Psi$ convex. Let $\mathcal{U}$ be a separable Banach space and $f \in C([0,1];\mathcal{U})$ satisfy*

$$B = \int_0^1 \int_0^1 \Psi\left(\frac{\|f(t) - f(s)\|}{p(|t-s|)}\right) \mathrm{d}s\mathrm{d}t < \infty. \tag{3.10}$$

*Then for all $s, t \in [0,1]$*

$$\|f(t) - f(s)\| \leq 4\int_0^{|t-s|} \Psi^{-1}\left(\frac{B}{u^2}\right) \mathrm{d}p(u).$$

*Proof.* Let $I \subseteq [0,1]$ be a closed interval and $|I|$ denote it's length. Then by monotonicity of $\Psi$ and $p$ we have

$$\int_I \int_I \Psi\left(\frac{\|f(t) - f(s)\|}{p(|I|)}\right) \mathrm{d}s\mathrm{d}t \leq B.$$

Let $\{I_j\}_{j\in\mathbb{N}}$ be a nested sequence of intervals $I_0 \supset I_1 \supset \ldots$ satisfying

$$p(|I_j|) = \tfrac{1}{2}p(|I_{j-1}|),$$

and denote the Bochner integral average of $f$ over $I_j$ by

$$(f)_j := \frac{1}{|I_j|}\int_{I_j} f(t)\mathrm{d}t.$$

Note that by the monotonicity of $p$, $|I_j| \to 0$ as $j \to \infty$ monotonically. It follows by the Bochner triangle inequality that

$$\|(f)_j - (f)_{j-1}\| \leq \frac{1}{|I_j||I_{j-1}|}\int_{I_j}\int_{I_{j-1}} \|f(t) - f(s)\|\mathrm{d}t\mathrm{d}s,$$

and therefore by Jensen's inequality (since $\Psi$ is convex) and monotonicity of $\{|I_j|\}$,

$$\begin{aligned}
\Psi\left(\frac{\|(f)_j - (f)_{j-1}\|}{p(|I_{j-1}|)}\right) &\leq \frac{1}{|I_j||I_{j-1}|}\int_{I_j}\int_{I_{j-1}} \Psi\left(\frac{\|f(t) - f(s)\|}{p(|I_{j-1}|)}\right)\mathrm{d}s\mathrm{d}t \\
&\leq \frac{1}{|I_j|^2}\int_{I_{j-1}}\int_{I_{j-1}} \Psi\left(\frac{\|f(t) - f(s)\|}{p(|I_{j-1}|)}\right)\mathrm{d}s\mathrm{d}t \\
&\leq \frac{B}{|I_j|^2}.
\end{aligned}$$

Since $\Psi$ is invertible and since $p(|I_{j-1}|) = 2(p(|I_{j-1}|) - p(|I_j|))$ we find

$$\|(f)_j - (f)_{j-1}\| \leq 2(p(|I_{j-1}|) - p(|I_j|))\Psi^{-1}\left(\frac{B}{|I_j|^2}\right).$$

Summing over $j \in \mathbb{N}$, we recognize the left-hand side as a telescoping series and the right-hand side as a Riemann-Stieltjes integral, thereby obtaining

$$\limsup_{j\to\infty} \|(f)_j - (f)_0\| \leq \sum_{j\in\mathbb{N}} \|(f)_j - (f)_{j-1}\| \leq 2\int_0^{|I_0|} \Psi^{-1}\left(\frac{B}{u^2}\right) \mathrm{d}p(u).$$

To complete the proof, we fix $s, t \in [0,1]$ with $s < t$ and let $I_0 = [s,t]$ and choose a sequence of intervals $\{I_j\}$ to decrease to the point $\{s\}$ so that by continuity of $f$ in $\mathcal{U}$, we have

$$(f_j) \to f(s) \quad \text{in} \quad \mathcal{U},$$

and therefore

$$\|f(s) - (f)_0\| \leq 2\int_0^{|t-s|} \Psi^{-1}\left(\frac{B}{u^2}\right) \mathrm{d}p(u).$$

Making the same argument with $s$ replaced by $t$ concludes the proof since

$$\|f(t) - f(s)\| \leq \|f(t) - f_0\| + \|f(s) - f_0\|.$$

□

**Remark 3.44.** Note that the both the integrals in (3.10) are singular and there is some competition between the choice of $\Psi$ and $p$ which allow both integrals to converge.

This theorem can be applied to give almost sure quantitative regularity improvements to Kolmogorov's theorem, specifically Hölder regularity. Define the space of Hölder continuous functions $C^\gamma([0,1];\mathcal{U})$ to be the set of $f \in C([0,1];\mathcal{U})$ such that

$$\|f\|_{C^\gamma} := \sup_{t\neq s} \frac{\|f(t) - f(s)\|}{|t-s|^\gamma} < \infty.$$

**Exercise 3.10.** Show that Theorem 3.43 can be used to prove Morrey's inequality. That is for each $s \in (0,1)$ and $p \geq 1$, if $\gamma = s - \frac{1}{p} > 0$, and $f \in C^\infty([0,1])$, then

$$\|f\|_{C^\gamma} \lesssim \left(\int_0^1 \int_0^1 \frac{|f(t) - f(s)|^p}{|t-s|^{sp+1}} \mathrm{d}s\mathrm{d}t\right)^{\frac{1}{p}}.$$

This is an example of a Sobolev embedding. Specifically that the Sobolev space $W^{s,p}([0,1])$ continuously embeds into a space of Hölder continuous functions $C^\gamma([0,1])$ if $\gamma = s - 1/p > 0$.

**Theorem 3.45.** *Let $X = \{X(t) : t \in [0,T]\}$ be as in Theorem 3.41 with condition (3.7) satisfied and suppose that $X(t)$ is already almost surely continuous. Then for any $\gamma < \frac{\alpha}{\beta}$, we have*

$$\mathbf{P}\left(\|X\|_{C^\gamma} < \infty\right) = 1.$$

*Proof.* We apply Theorem 3.43 with $\Psi(x) = x^\beta$ and $p(u) = u^{\frac{2}{\beta}+\gamma}$. By Fubini's theorem and (3.7) we have

$$\begin{aligned}\mathbf{E}B &= \int_0^1 \int_0^1 \mathbf{E}\frac{|X(t)-X(s)|^\beta}{|t-s|^{2+\gamma\beta}}\mathrm{d}s\mathrm{d}t \\ &\leq C\int_0^1\int_0^1 |t-s|^{\alpha-\gamma\beta-1}\mathrm{d}s\mathrm{d}t,\end{aligned}$$

which converges if $\alpha > \gamma\beta$ and therefore $B$ is almost surely finite. On the other hand

$$\begin{aligned}|f(t+r)-f(r)| &\leq \int_0^r \Psi^{-1}\left(\frac{B}{u^2}\right)\mathrm{d}p(u) \\ &= \frac{2+\gamma}{\beta}B^{1/\beta}\int_0^r u^{\gamma-1}\mathrm{d}u \\ &= \frac{2+\gamma}{\gamma\beta}B^{1/\beta}r^\gamma\end{aligned}$$

and therefore we have an estimate of the form

$$\|X\|_{C^\gamma([0,1])} \lesssim_{\gamma,\beta} B^{1/\beta} < \infty, \quad \mathbf{P}\text{-almost surely}$$

for $\gamma < \frac{\alpha}{\beta}$. □

**Exercise 3.11.** Show that the Gaussian measure $\mu$ constructed in Corollary 3.42 satisfies

$$\mu(C^\gamma([0,1])) = 1, \quad \text{for all } \gamma < \alpha.$$

In particular, show that this means the Wiener process $W = \{W(t) : t \in [0,1]\}$ with covariance function $C(s,t) = \min\{s,t\}$ almost surely belongs to $C^\gamma([0,1])$ for any $\gamma < 1/2$.

**Exercise 3.12.** Let $W = \{W(t) : t \in [0,1]\}$ be the Wiener process with covariance $C(s,t) = \min\{s,t\}$. Show that the choice of $\Psi = \exp(x^2/4)$ and $p(x) = \sqrt{x}$ in Theorem 3.43 gives the more precise modulus of continuity for $r \in (0,1)$ small enough

$$|W(s+r) - W(s)| \leq \sqrt{2r\log\left(\sqrt{B}/r\right)}.$$

Note that this is a strict improvement over the Hölder modulus of continuity of the Brownian motion. (Hint: Use exponential moments of $|W(t)-W(s)|/\sqrt{|t-s|}$ and the change of variables $y = \log(\sqrt{B}/u)$ in the integral for the modulus of continuity)

### 3.4.4 Continuous functions with values in a Hilbert space, $Q$-Wiener process

Much of what has been mentioned so far can be extended to Gaussian measures on $C([0,1];\mathcal{U})$, namely continuity functions on $[0,1]$ with values in $\mathcal{U}$ a separable Banach space. In this case, the dual of $C([0,1];\mathcal{U})$ can be identified with $M([0,1];\mathcal{U}^*)$ the space of all finite variation $\mathcal{U}^*$-valued measures (see [BL74]). Consequently for each $s,t \in [0,1]$ we can define a covariance function $C(s,t) \in \mathcal{L}(\mathcal{U}^*,\mathcal{U})$ for each $\ell_1,\ell_2 \in \mathcal{U}^*$ by

$$\ell_1(C(s,t)\ell_2) = C_\mu(\ell_1\delta_s,\ell_2\delta_t).$$

and satisfies for each $\nu_1,\nu_2 \in M([0,1];\mathcal{U}^*)$

$$C_\mu(\nu_1,\nu_2) = \int_{[0,1]}\int_{[0,1]}\nu_1(\mathrm{d}s)(C(s,t)\nu_2(\mathrm{d}t)).$$

At this stage, things are very similar to the case of a general Gaussian measure on $\mathcal{U}$ and there is not much hope of classifying such covariances $C(s,t) \in \mathcal{L}(\mathcal{U}^*, \mathcal{U})$ without imposing an additional structure on $\mathcal{U}$. However, if $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space then things are much nicer. Indeed it can be seen that in this case, for each $s,t \in [0,1]$, $C(s,t) \in \mathcal{L}_1(\mathcal{H})$ (trace-class) since for any orthonormal basis $\{e_k\}$ of $\mathcal{H}$ we have by monotone convergence, Cauchy-Schwartz and Fernique's theorem that

$$\begin{aligned}
\operatorname{Tr}|C(s,t)| &= \sum_k |\langle C(s,t)e_k, e_k\rangle_{\mathcal{H}}| = \sum_k |C_\mu(e_k\delta_s, e_k\delta_s)| \\
&= \int_{C([0,1];\mathcal{H})} \sum_k |\langle \varphi(s), e_k\rangle_{\mathcal{H}} \langle \varphi(t), e_k\rangle_{\mathcal{H}}| \, \mu(\mathrm{d}\varphi) \\
&\leq \int_{C([0,1];\mathcal{H})} \|\varphi(s)\|_{\mathcal{H}} \|\varphi(t)\|_{\mathcal{H}} \, \mu(\mathrm{d}\varphi) \\
&\leq \int_{C([0,1];\mathcal{H})} \|\varphi\|^2 \, \mu(\mathrm{d}\varphi) < \infty.
\end{aligned}$$

Moreover $C(s,t)$ is symmetric in the sense that for each $h_1, h_2 \in C([0,1];\mathcal{H})$

$$\langle C(s,t)h_1, h_1\rangle_{\mathcal{H}} = \langle h_1, C(t,s)h_2\rangle_{\mathcal{H}} \tag{3.11}$$

and positive definite in the sense that for each $\{a_i\}_{i=1}^n \subset \mathbb{R}$ , $\{t_i\}_{i=1}^n \subset [0,1]$ and $h \in \mathcal{H}$

$$\sum_{ij} a_i a_j \langle C(t_i, t_j)h, h\rangle_{\mathcal{H}} \geq 0. \tag{3.12}$$

In later sections it will be important to construct a Gaussian measure $\mu$ on $C([0,1];\mathcal{H})$ with a given symmetric, positive definite, Hölder continuous covariance $C : [0,1] \times [0,1] \to \mathcal{L}_1(\mathcal{H})$. Indeed, we have the following analogue of Corollary 3.4.

**Theorem 3.46.** *Let $m : [0,1] \to \mathcal{H}$ and $C(s,t) : [0,1] \times [0,1] \to \mathcal{L}_1(\mathcal{H})$, be symmetric and positive definite in the sense of (3.11) and (3.12) above and suppose that there exists $\alpha \in (0,1)$ and constants $C_1$ and $C_2$ such that*

$$\|m(t) - m(s)\|_{\mathcal{H}} \leq C_1 |t-s|^\alpha \quad \|C(t,t) - 2C(s,t) + C(s,s)\|_{\mathcal{L}_1(\mathcal{H})} \leq C_2 |t-s|^{2\alpha}.$$

*Then there exists a Gaussian measure $\mu$ on $C([0,1];\mathcal{U})$ with mean and covariance $m$ and $C$ given above and for each $\gamma < \alpha$ we have*

$$\mu(C^\gamma([0,1];\mathcal{H})) = 1.$$

*Proof.* We will use a similar Kolmogorov extension construction to the real valued case. For each finite set of times $T = \{t_i\}_{i=1^n} \subset [0,1]$ we can build a Gaussian measure $\mu_T$ on the product Hilbert space $\mathcal{H}^n$ with mean $(m_T)_i = m(t_i) \in \mathcal{H}$ and covariance $(Q_T)_{ij} = C(t_i, t_j) \in \mathcal{L}_1(\mathcal{H})$. The collection of measures $\{\mu_T : T \subset [0,1] \text{ finite}\}$ form a consistent collection of measures and therefore by the Kolmogorov extension Theorem there exists measure $\mathbf{P}$ on $\Omega = \mathcal{H}^{[0,1]}$ equipped with product sigma-algebra $\mathscr{F} = \mathscr{B}(\mathcal{H})^{\otimes[0,1]}$ such that for each $s,t \in [0,1]$ and each $h_1, h_2 \in \mathcal{H}$ and $X_t(\omega) = \omega(t)$ we have

$$m(t) = \mathbf{E}X(t), \quad \langle C(s,t)h_1, h_2\rangle_{\mathcal{H}} = \mathbf{E}\langle X(s) - m(s), h_1\rangle_{\mathcal{H}} \langle X(t) - m(t), h_2\rangle_{\mathcal{H}}.$$

As in the proof of Corollary 3.4 we may assume $m(t) = 0$ by shifting and it follows that for $\{e_k\}$ an orthonormal basis for $\mathcal{H}$

$$\begin{aligned}\mathbf{E}\|X(t) - X(s)\|_{\mathcal{H}}^2 &= \sum_k \mathbf{E}|\langle X(t) - X(s), e_k\rangle_{\mathcal{H}}|^2 \\ &= \sum_k \left(\langle C(t,t)e_k, e_k\rangle_{\mathcal{H}} - 2\langle C(s,t)e_k, e_k\rangle_{\mathcal{H}} + \langle C(s,s)e_k, e_k\rangle_{\mathcal{H}}\right) \\ &= \|C(t,t) - 2C(s,t) + C(s,s)\|_{\mathcal{L}_1(\mathcal{H})} \\ &\le C_2|t-s|^{2\alpha}.\end{aligned}$$

And therefore by Gaussianity we have

$$\mathbf{E}\|X(t) - X(s)\|_{\mathcal{H}}^p \le C_p|t-s|^{\alpha p}.$$

It is now a simple matter to apply Theorem 3.41 and Theorem 3.43 to show that $X = \{X(t) : t \in [0,1]\}$ has a continuous modification $\bar{X}$ whose law is a Gaussian measure $\mu$ on $C([0,1];\mathcal{H})$ such that $\mu(C^\gamma([0,1];\mathcal{H})) = 1$ for all $\gamma < \alpha$. □

It is not hard to see that all of the examples of real-valued continuous Gaussian processes on $[0,1]$ can be extended to $\mathcal{H}$-valued processes.

**Example 3.47.** Let $Q \in \mathcal{L}_1^+(\mathcal{H})$ and suppose that $C(s,t) = \min\{s,t\}Q$, then the associated cannonical process $W = \{W(t) : t \in [0,1]\}$ is called a $Q$-Wiener process on $[0,1]$. Analogous examples hold for the bridge and Ornstein-Uhlenbeck processes.

## 3.5 White noise expansion

Formally speaking, on a separable Hilbert space $\mathcal{H}$ with orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$, *white noise* can be thought of as a random variable of the form

$$\xi = \sum_{k\in\mathbb{N}} \xi_k e_k$$

for $\{\xi_k\}_{k\in\mathbb{N}}$ a collection of iid standard Gaussian random variables. Such a "Gaussian" random variable has the property that each of it's "modes" are decorrelated since formally speaking

$$\mathbf{E}\langle \xi, e_i\rangle\langle \xi, e_j\rangle = \mathbf{E}\xi_i\xi_j = \delta_{ij}. \tag{3.13}$$

However, there is some trouble in interpreting such a random variable since the series (3.5) does not converge almost surely in the Hilbert space $\mathcal{H}$, since by the law of large numbers

$$\sum_{k\in\mathbb{N}} |\xi_k|^2 = \infty, \quad \mathbf{P}\text{-almost surely.}$$

Indeed we showed in Corollary 3.25 that the covariance of a Gaussian measure on an infinite dimensional Hilbert space cannot be given by (3.13). The following theorem which says that a canonical Gaussian random variable on a Banach space $\mathcal{U}$ always looks like a white noise on it's Cameron-Martin space.

**Theorem 3.48.** *Let $\mu$ be a Gaussian measure on $\mathcal{U}$ and $\{e_k\}_{k\in\mathbb{N}}$ be an orthonormal basis for it's Cameron-Martin space $\mathcal{H}_\mu$ and $\{\hat{\ell}_k\}_{k\in\mathbb{N}}$ the associated basis in $\mathcal{R}_\mu$, with $e_k = \hat{C}_\mu\hat{\ell}_k$, then*

$$u = \sum_{k\in\mathbb{N}} \hat{\ell}_k(u)e_k, \quad \mu\text{- almost surely.}$$

*Proof.* Denote

$$\xi_{\le n}(u) := \sum_{k\le n} \hat{\ell}_k(u) e_k, \quad \text{and} \quad \xi(u) = u,$$

Then our goal is to show that $\xi_{\le n} \to \xi$, $\mu$- almost surely. Note that $\{\xi_{\le n}\}_{n\in\mathbb{N}}$ is a $\mathcal{U}$-valued martingale with respect to the filtration $\mathscr{F}_n = \sigma(\hat{\ell}_1, \dots, \hat{\ell}_n)$ since for $m \le n$

$$\mathbf{E}_\mu[\xi_{\le n} - \xi_{\le m} | \mathscr{F}_m] = 0.$$

Furthermore for each $\ell \in \mathcal{U}^*$, we have

$$\ell(\xi_{\le n}(u)) = \sum_{k\le n} \hat{\ell}_k(u) \ell(e_k) = \sum_{k\le n} \hat{\ell}_k(u) \langle \ell, \hat{\ell}_k \rangle_{L^2(\mu)},$$

and since $\{\hat{\ell}_k\}$ is an orthonormal basis for $\mathcal{R}_\mu$ it is easy to see that

$$\lim_{n\to\infty} \ell(\xi_{\le n}) = \ell(\xi) \quad \mu\text{-almost surely.}$$

By independence of $\{\hat{\ell}_k\}$ this implies that $\mu$-almost surely we have

$$\ell(\xi_{\le n}) = \mathbf{E}_\mu[\, \ell(\xi) \,|\mathscr{F}_n].$$

A simple application of Fernique's theorem implies that $\xi$ is Bochner integrable with respect to $\mathbf{E}_\mu[\,\cdot\,|\mathscr{F}_n]$ and therefore satisfies $\mathbf{E}_\mu[\, \ell(\xi) \,|\mathscr{F}_n] = \ell\left(\mathbf{E}_\mu[\, \xi \,|\mathscr{F}_n]\right)$, which implies

$$\ell\left(\xi_{\le n} - \mathbf{E}_\mu[\xi|\mathscr{F}_n]\right) = 0, \quad \mu\text{-almost surely.}$$

Using the separability of $\mathcal{U}$ and choosing $\{\ell_k\} \subseteq \mathcal{U}^*$ as in Exercise (2.1) so that $\|u\| = \sup_{k\in\mathbb{N}} \ell_k(u)$ we find that (note that countability is crucial here since there is potentially a different measure 0 set for each $\ell_k$)

$$\xi_{\le n} = \mathbf{E}_\mu[\, \xi \,|\mathscr{F}_n] \quad \mu\text{-almost surely.}$$

Another application of Fernique's theorem implies that $\|\xi\| \in L^2(\mu)$ and so we can apply the Martingale convergence theorem ([HvNVW16] Theorem 3.3.2) for Banach space valued martingales generated by $\xi$ to conclude

$$\xi_{\le n} \to \xi \quad \mu\text{-almost surely.}$$

□

An immediate corollary of this theorem is the following.

**Corollary 3.49.** *If $\{\xi_k\}_{k\in\mathbb{N}}$ is a sequence of iid standard Gaussian random variables on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and $\{e_k\}_{k\in\mathbb{N}}$ an orthonormal basis for $\mathcal{H}_\mu$ associated to some Gaussian measure $\mu$ on a Banach space $\mathcal{U}$ then*

$$\xi(\omega) = \sum_{k\in\mathbb{N}} \xi_k(\omega) e_k$$

*converges $\mathbf{P}$-almost surely in $\mathcal{U}$ with* $\mathrm{Law}(\xi) = \mu$.

**Remark 3.50.** If $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space, then Theorem 3.48 is much simpler and follows immediately from the Hilbert-Schmidt Theorem. Indeed, let $Q \in \mathcal{L}_1^+$ be the covariance of $\mu$ and by the

Hilbert-Schmidt Theorem let $\{e_k\}_{k\in\mathbb{N}}$ the orthonormal basis of eigenfunctions and $\{q_k\}_{k\in\mathbb{N}}$ be the associated eigenvalues satisfying $Qe_k = q_k e_k$. Then each $h \in \mathcal{H}$ can be written as

$$h = \sum_{k\in\mathbb{N}} \hat{\ell}_k(h)\sqrt{q_k}e_k,$$

where $\hat{\ell}_k(h) = \frac{1}{\sqrt{q_k}}\langle e_k, h\rangle$. Since $\mathcal{H}_\mu = Q^{1/2}(\mathcal{H})$, we see that $\{\sqrt{q_k}e_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}_\mu$ and

$$\hat{C}_\mu \hat{\ell}_k = \frac{1}{\sqrt{q_k}} Q^{1/2}(e_k) = e_k.$$

so that $\{\hat{\ell}_k\}$ are iid standard Gaussian random variables.

**Example 3.51.** Let $W = \{W(t) : t \in [0,1]\}$ be a $Q$-Wiener process on $[0,1]$, and let $\{e_k\}_{k\in\mathbb{N}}$ be the orthonormal basis of eigenvectors and $\{q_k\}_{k\in\mathbb{N}}$ be corresponding eigenvalues satisfying $Qe_k = q_k e_k$, then for each $t \geq 0$ we can write

$$W(t) = \sum_{k\in\mathbb{N}} \sqrt{q_k} e_k W_k(t)$$

where $\{W_k(t)\}$, $W_k(t) := \frac{1}{\sqrt{q_k}}\langle W(t), e_k\rangle$ are a standard iid real valued Wiener processes on $[0,1]$.

**Remark 3.52.** Suppose that $\xi(t)$ is a square integrable Gaussian process with covariance $\mathbf{E}\xi(t)\xi(t) = C(s,t)$, then 3.49 is a special case of the *Karhunen-Loève expansion*

$$\xi(t) = \sum_k \xi_k \lambda_k \varphi_k(t)$$

where $\{\xi\}$ are iid standard Gaussian processes and $\{\varphi_k\}$ are a orthonormal basis for $L^2([0,1])$ given by eigenfunctions of the integral operator

$$C\varphi(t) = \int_0^1 C(t,s)\varphi(s)\mathrm{d}s.$$

and $\{\lambda_k\}$ are the associated eigen-values and the series converges in $L^2([0,1])$.

**Exercise 3.13.** Show that the Wiener process $W(t)$ on $[0,1]$ can be represented in terms of the random Fourier series

$$W(t) = \xi_0 t + \sqrt{2}\sum_{k=1}^{n} \xi_k \frac{\sin(\pi k t)}{\pi k}$$

where $\{\xi_k\}_{k\in\mathbb{N}}$ are cannonical iid Gaussian random variables and the convergence holds in $C([0,1])$. (Hint: use Exercise 3.9 and find a basis for the Cameron-Martin space using the appropriate Fourier basis in $L^2([0,1])$).

## 3.6 Hilbert-Schmidt embeddings and cylindrical Wiener process

Now given a separable Hilbert space $\mathcal{H}$, we would like to find a way to interpret the white-noise

$$\xi = \sum_{k\in\mathbb{N}} \xi_k e_k.$$

In light of Theorem 3.48 if $\mathcal{H}$ of is the Cameron-Martin space of some Banach space $\mathcal{U}$, then we are done. Such an approach is taken in the theory of abstract Wiener spaces. However, in practice, due to our characterization of Gaussian measures on Hilbert spaces, it is much easier to find a Hilbert space $\mathcal{H}_1$ that contains $\mathcal{H}$ such that $\mathcal{H}$ is the cameron-Martin space for some Gaussian measure on $\mathcal{H}_1$.

Indeed, suppose that there is a separable Hilbert space $\mathcal{H}_1$ such $\mathcal{H}$ is densely embedded in $\mathcal{H}_1$ and the inclusion map

$$J : \mathcal{H} \to \mathcal{H}_1.$$

is a Hilbert-Schmidt operator, namely $JJ^* : \mathcal{H}_1 \to \mathcal{H}_1$ is trace-class on $\mathcal{H}_1$.

**Exercise 3.14.** One can always find a Hilbert space $\mathcal{H}_1$ with the above properties. Indeed define for a given orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ for $\mathcal{H}$ define the norm

$$\|h\|^2_{\mathcal{H}_1} := \sum_{k\in\mathbb{N}} (1+k^2)^{-1} |\langle h, e_k\rangle_{\mathcal{H}}|^2$$

and let $\mathcal{H}_1$ be the closure of $\mathcal{H}$ with respect to the $\|\cdot\|_{\mathcal{H}_1}$. This can be viewed as an example of a negative Sobolev space. Show that the inclusion $J : \mathcal{H} \to \mathcal{H}_1$ is Hilbert-Schmidt.

**Exercise 3.15.** For $\mathcal{H}$ and $\mathcal{H}_1$ as above, show that the centered Gaussian measure $\mu$ with covariance $Q = JJ^*$ has $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ as it's Cameron-Martin space.

This way of embedding a Hilbert space $\mathcal{H}$ into a large Hilbert space $\mathcal{H}_1$ in a Hilbert-Schmidt way can also be used to define the notion of a process which is both white-in-time and white in the Hilbert space $\mathcal{H}$ called a *cylindrical Wiener process*

**Definition 3.53.** With $\mathcal{H}$ and $\mathcal{H}_1$ defined as above with Hilbert-Schmidt embedding $J$. Then a $JJ^*$- Wiener process in $\mathcal{H}_1$ , $W = \{W(t) : t \in [0,1]\}$ is called a *cylindrical Wiener process* on $\mathcal{H}$ and satisfies for $h_1, h_2$ in $\mathcal{H}_1$ and $s, t \in [0,1]$

$$\mathbf{E}\langle W(s), h_1\rangle_{\mathcal{H}_1} \langle W(t), h_2\rangle_{\mathcal{H}_1} = \min\{s,t\}\langle J^*h_1, J^*h_2\rangle_{\mathcal{H}}.$$

It is important to remark that a cylindrical Wiener process $W$ on $\mathcal{H}$ *does not* actually take values in $\mathcal{H}$. This is the reason for embedding $\mathcal{H}$ into a larger Hilbert space $\mathcal{H}_1$. However in many ways the statistics of $W(t)$ don't depend on the space $\mathcal{H}_1$. Indeed by Theorem 3.48 we can write

$$W(t) = \sum_{k\in\mathbb{N}} e_k W_k(t) \tag{3.14}$$

where $\{W_k(t)\}_{k\in\mathbb{N}}$ are a sequence of iid Wiener processes on $\mathbb{R}$ and $\{e_k\}_{k\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$ and the convergence happens in $\mathcal{H}_1$. However, the formula (3.14) doesn't explicitly depend on $\mathcal{H}_1$ and the convergence clearly holds in any other Hilbert space $\mathcal{H}''$ such that $\mathcal{H}$ embeds into $\mathcal{H}''$ via a Hilbert-Schmidt inclusion.

### 3.6.1 Hilbert-Schmidt mapping of a cylindrical Wiener process

In practice given a cylindrical Wiener process $W(t)$ on a Hilbert space $\mathcal{W} \subset \mathcal{W}_1$, and $A$ a Hilbert-Schmidt operator from $\mathcal{W}$ to another Hilbert space $\mathcal{H}$ (denote such operators by $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$) we will often make reference to the process $AW(t)$ even though $W(t)$ doesn't take values in $\mathcal{W}$. However, in light of Theorem 3.48 and equation 3.14, we can naturally extend $A$ to $\mathcal{W}_1$ in such a way that

$$AW(t) = \sum_{k\in\mathbb{N}} Ae_k W_k(t).$$

It is not surprising that the above series converges in $\mathcal{H}$ since

$$\sum_{k\in\mathbb{N}} \|Ae_k\|^2_{\mathcal{H}} = \mathrm{Tr}(A^*A) = \mathrm{Tr}(AA^*) < \infty,$$

and that $AW(t)$ is an $AA^*$-Wiener process. This can be made more precise in the following theorem,

**Theorem 3.54.** *Let $\mu$ be a centered Gaussian measure on a separable Banach space $\mathcal{U}$ and suppose that $A \in \mathcal{L}_2(\mathcal{H}_\mu, \mathcal{H})$ for some separable Hilbert space $\mathcal{H}$, then there exists a measurable extension $\hat{A} : \mathcal{U} \to \mathcal{H}$ of $A$ such that $\hat{A}$ is $\mu$ almost surely a continuous linear operator and $\hat{A}_*\mu$ is a Gaussian measure on $\mathcal{H}$ with covariance $AA^*$.*

*Proof.* Let $\{e_k\}$ be an orthonormal basis in $\mathcal{H}_\mu$ and $\hat{\ell}_k = \hat{C}_\mu^{-1} e_k$ be the orthonormal basis in $\mathcal{R}_\mu$ corresponding to $\{e_k\}_{k\in\mathbb{N}}$. In light of Theorem 3.48 we would like to define the extension $\hat{A}$ to be

$$\hat{A}u := \sum_{k\in\mathbb{N}} \hat{\ell}_k(u) A e_k.$$

To make sense of this, we note that $S_n(u) = \sum_{k\le n} \hat{\ell}_k(u) A e_k$ is a $\mathcal{H}$-valued Martingale with respect to the filtration $\mathscr{F}_n = \{\hat{\ell}_1, \ldots, \hat{\ell}_n\}$ and that

$$\sup_n \mathbf{E}_\mu \|S_n(u)\|_{\mathcal{H}}^2 \le \sum_{k\in\mathbb{N}} \|Ae_k\|_{\mathcal{H}}^2 = \mathrm{Tr}(AA^*) < \infty.$$

Therefore, by Doob's martingale convergence theorem we find that $\lim_{n\to\infty} S_n(u)$ converges converges $\mu$ almost surely in $\mathcal{H}$ define this limit to be $\hat{A}u$ and $0$ otherwise. Moreover in light of Proposition 3.11 we may take the full measure set to be linear since up to a zero-measure set the limit is a limit of linear functions on $\mathcal{U}$. It follows that $\hat{A}$ is $\mu$ almost surely linear. To see that $\hat{A}_*\mu$ is Gaussian with covariance $AA^*$, we note that for each $h \in \mathcal{H}$ the sum

$$\langle \hat{A}u, h\rangle_{\mathcal{H}} = \sum_{k\in\mathbb{N}} \hat{\ell}_k(u) \langle Ae_k, h\rangle_{\mathcal{H}}$$

converges in $\mathcal{R}_\mu$ since $\{\hat{\ell}_k(u)\}$ are a complete orthonormal basis and therefore $\langle \hat{A}u, h\rangle$ is a Gaussian random variable. Moreover for each $h_1, h_2 \in \mathcal{H}$ we have

$$\begin{aligned}
\int_{\mathcal{H}} \langle u, h_1\rangle_{\mathcal{H}} \langle u, h_2\rangle_{\mathcal{H}} \hat{A}_*\mu(\mathrm{d}u) &= \int_{\mathcal{U}} \langle \hat{A}u, h_1\rangle_{\mathcal{H}} \langle \hat{A}u, h_2\rangle_{\mathcal{H}} \mu(\mathrm{d}u) \\
&= \sum_{k\in\mathbb{N}} \langle Ae_k, h_1\rangle_{\mathcal{H}} \langle Ae_k, h_2\rangle_{\mathcal{H}} \\
&= \sum_{k\in\mathbb{N}} \langle e_k, A^*h_1\rangle_\mu \langle e_k, A^*h_2\rangle_\mu \\
&= \langle A^*h_1, A^*h_2\rangle_\mu = \langle AA^*h_1, h_2\rangle_{\mathcal{H}}.
\end{aligned}$$

and therefore $\hat{A}_*\mu$ is a Gaussian measure with covariance $AA^*$.

□

This gives the following corollary for the cylindrical Wiener process.

**Corollary 3.55.** *Let $W(t)$ be a cylindrical Wiender process on $\mathcal{W}$ with Hilbert-Schmidt embedding $J : \mathcal{W} \to \mathcal{W}_1$, and let $A : \mathcal{W} \to \mathcal{H}$ be Hilbert-Schmidt for some other Hilbert space $\mathcal{H}$. Then there exists a measurable linear extention $\hat{A} : \mathcal{W}_1 \to \mathcal{H}$ such that*

$$AW(t) := \hat{A}W(t) = \sum_{k\in\mathbb{N}} Ae_k W_k(t).$$

*defines an $AA^*$-Wiener process on $\mathcal{H}$.*

# 4 The stochastic integral

In this section we consider two separable Hilbert spaces $\mathcal{H}$ and $\mathcal{W}$ and a cylindrical Wiener process $W(t)$ on $\mathcal{W}$ and a certain process $\Phi(t)$ taking values in Hilbert-Schmidt operators from $\mathcal{W}$ to $\mathcal{H}$ that doesn't "see" into the future. Our goal is to define the infinite dimensional Itô stochastic integral

$$\int_0^\infty \Phi(s)\mathrm{d}W(s),$$

and give it's basic properties. This will prove useful in establishing a version of Itô formula in infinite dimensions.

## 4.1 Cylindrical Wiener process on $\mathbb{R}_+$

As we have see, if we define the covariance operator

$$C(s,t) = \min\{s,t\}Q$$

where $Q \in \mathcal{L}_1^+(\mathcal{H})$ (symmetric, positive definite, trace-class). Then the canonical stochastic process $W = \{W(t) : t \in [0,1]\}$ associated with this Gaussian measure is known as a $Q$-*Wiener process* on $[0,1]$.

In many applications we would like to construct a process $W = \{W(t) : t \in \mathbb{R}_+\}$ which lives on the whole time interval $\mathbb{R}_+$ and not just on $[0,1]$. A common way to do this is through the following definition.

**Definition 4.1.** A $\mathcal{H}$-valued stochastic process $W = \{W(t) : t \in \mathbb{R}_+\}$ on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ is called is called a $Q$-*Wiener process* over $\mathbb{R}_+$ for some $Q \in \mathcal{L}_1^+(\mathcal{H})$ if it satisfies the following criteria:

1. $W(0) = 0$,
2. $t \mapsto W(t)$ is almost surely continuous,
3. $W$ has independent increments, that is for each $0 \le s_1 < t_1 \le \ldots \le s_n < t_n < \infty$ the random variables
$$W(t_1) - W(s_1), \ldots W(t_n) - W(s_n)$$
are independent random variables,
4. for each $0 \le s \le t \le 1$, the law of $W(t) - W(s)$ is a Gaussian random variable with mean $0$ and covariance $(t-s)Q$.

**Exercise 4.1.** Show that a continuous $\mathcal{H}$ valued process $W = \{W(t) : t \in \mathbb{R}_+\}$ is a $Q$-Wiener process in the sense of Definition 4.1 if and only if

1. for each $\{t_i\}_{i=1}^n \subseteq \mathbb{R}_+$, $(W(t_1), \ldots, W(t_n))$ is a Gaussian random vector in $\mathcal{H}^n$,
2. for each $t, s \in \mathbb{R}_+$ and $h_1, h_2 \in \mathcal{H}$ we have
$$\mathbf{E}W(t) = 0 \quad \text{and} \quad \mathbf{E}\langle W(s), h_1\rangle_{\mathcal{H}} \langle W(t), h_2\rangle_{\mathcal{H}} = \min\{s,t\}\langle Qh_1, h_2\rangle_{\mathcal{H}}.$$

It remains to construct a $Q$-Wiener process over $\mathbb{R}_+$. Of course we could use Kolmogorov extension theorem and then continuity criterion tools from the previous section. However, as it turns out, the Wiener process has a lot of self-similarity and essentially contains a version of the process on $\mathbb{R}_+$ inside of $[0,1]$. In essence, by constructing a Q-Wiener process over $[0,1]$ we have already constructed one over $\mathbb{R}_+$. The following proposition makes this precise.

**Proposition 4.2.** *Let $W = \{W(t) : t \in [0,1]\}$ be the canonical Q-Wiener process over $[0,1]$. Then the process $\widehat{W} = \{\widehat{W}(t) : t \in \mathbb{R}_+\}$ defined for each $t \in \mathbb{R}_+$ by*

$$\widehat{W}(t) = (t+1)W\left(\frac{1}{1+t}\right) - W(1)$$

*is a Q-Wiener process on $\mathbb{R}_+$.*

*Proof.* First we remark that obviously $\widehat{W}(0) = 0$ and $t \mapsto \widehat{W}(t)$ is clearly continuous on $\mathbb{R}_+$ since $t \mapsto W(t)$ is continuous on $[0,1]$. Moreover, for each $t \in \mathbb{R}_+$, $\widehat{W}(t)$ is still a mean-zero Gaussian and therefore for each $\{t_i\}_{i=1}^n \subseteq \mathbb{R}_+$, $(\widehat{W}(t_1), \dots, \widehat{W}(t_n))$ is a Gaussian random vector in $\mathcal{H}^n$. Finally we note that

$$\begin{aligned}\mathbf{E}\langle \widehat{W}(s), h_1\rangle_{\mathcal{H}} \langle \widehat{W}(t), h_2\rangle_{\mathcal{H}} &= \Bigg( (1+t)(1+s)\min\left\{\frac{1}{1+s}, \frac{1}{1+t}\right\} - (1+s)\min\left\{\frac{1}{(1+s)}, 1\right\} \\ &\qquad - (1+t)\min\left\{\frac{1}{(1+s)}, 1\right\} + 1\Bigg)\langle Qh_1, h_2\rangle_{\mathcal{H}} \\ &= \left(\min\{1+s, 1+t\} - 1\right)\langle Qh_1, h_2\rangle_{\mathcal{H}} \\ &= \min\{s,t\}\langle Qh_1, h_2\rangle_{\mathcal{H}}.\end{aligned}$$

□

**Remark 4.3.** It is also possible to construct the $Q$-Wiener process over $\mathbb{R}_+$ using the theory developed in the previous section using a weighted Banach space

$$C_W(\mathbb{R}_+; \mathcal{H}) := \left\{\varphi \in C(\mathbb{R}_+; \mathcal{H}) : \lim_{t\to\infty} W(t)/t = 0\right\}.$$

It is not hard to see that such a space is a Banach space with the norm

$$\|\varphi\|_{C_W} = \sup_{t\in\mathbb{R}_+} \frac{\|\varphi(t)\|}{1+t}.$$

Additionally in view of the law of large numbers any Wiener process can be shown to satisfy $\lim_{t\to\infty} \frac{1}{t}W(t) = 0$ almost surely so that it's paths almost surely belong to $C_W(\mathbb{R}_+; \mathcal{H})$.

Of course now that we have defined a $Q$-Wiener process over $\mathbb{R}_+$, we can define a cylindrical Wiener process in $\mathcal{H}$ over $\mathbb{R}_+$ in the natural way.

**Definition 4.4.** Given a separable Hilbert space $\mathcal{H} \subseteq \mathcal{H}_1$ and a Hilbert-Schmidt inclusion $J : \mathcal{H} \to \mathcal{H}_1$ a *cylindrical Wiener process* $W = \{W(t) : t \in \mathbb{R}_+\}$ on $\mathcal{H}$ over $\mathbb{R}_+$ is a $JJ^*$-Wiener process on $\mathcal{H}_1$ over $\mathbb{R}_+$.

## 4.2 Admissible filtrations

Before discussing the stochastic integral, we need to introduce the concept of a filtration. In what follows, we will assume that $\mathcal{U}$ is separable Banach space unless stated otherwise.

**Definition 4.5** (Filtration)**.** A collection of sub sigma-algebras $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$, $\mathscr{F}_t \subset \mathscr{F}$ is known as a *filtration* if it forms an increasing collection of sigma-algebra's

$$\mathscr{F}_s \subseteq \mathscr{F}_t \quad \text{for} \quad s \leq t.$$

**Definition 4.6.** We say that a $\mathcal{U}$-valued stochastic process $X = \{X(t) : t \in \mathbb{R}_+\}$ is *adapted* to the filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ if for each $t \in \mathbb{R}_+$, $X(t)$ is measurable from $\mathscr{F}_t$ to $\mathscr{B}(\mathcal{U})$.

**Remark 4.7.** One has to be very careful when considering modifications of stochastic processes, which involves changing a process $X$ at each time on $\mathbf{P}$-null sets. Indeed if $X$ adapted to a filtration $(\mathscr{F}_t)_{t\in I}$, it is not necessarily true that any modification $\bar{X}$ is still adapted to $(\mathscr{F}_t)_{t\in I}$ since $\mathscr{F}_t$ may not contain all $\mathbf{P}$ null sets. In this case is it often necessary to augment a filtration to include the $\mathbf{P}$-null sets. The current presentation specifically avoids doings this.

For a given $\mathcal{U}$-valued stochastic process $X = \{X(t) : t \in \mathbb{R}_+\}$ define the *natural filtration*

$$\mathscr{F}_t^X := \sigma(X_s : s \leq t)$$

to be the smallest sub-sigma algebra of $\mathscr{F}$ such that $X(s)$ is measurable for all $s \in [0,t]$. Then $(\mathscr{F}_t^X)_{t\in\mathbb{R}_+}$ is a filtration and $X$ is adapted to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. We interpret an event $A \in \mathscr{F}_t^X$ to be such that by time $t$ an observer of $X$ would be able to tell whether $A$ has occurred or not.

**Exercise 4.2.** Let $W = \{W(t) : t \in \mathbb{R}_+\}$ be a $Q$-Wiener process in a separable Hilbert space $\mathcal{H}$ and let

$$M(t) = \sup_{s\in[0,t]} \|W(t)\|_{\mathcal{H}}.$$

Show that for each $a \in \mathbb{R}_+$, the event $\{M(t) > a\}$ belongs to $\mathscr{F}_t^W$. (Hint: use pathwise continuity to write $\{M(t) > a\}$ as a countable union of events in $\mathscr{F}_t^W$).

In general, the natural filtration $(\mathscr{F}_t^W)_{t\in\mathbb{R}_+}$ can be used to state a very important property of the Wiener process, called the *Markov property*.

**Proposition 4.8** (Markov property)**.** *Let $W = \{W(t) : t \in \mathbb{R}_+\}$ be a $Q$-Wiener process in a separable Hilbert space $\mathcal{H}$ and let $(\mathscr{F}_t^W)_{t\in\mathbb{R}_+}$ be the natural filtration. Then $W(t) - W(s)$ is independent of $\mathscr{F}_s^W$ for all $s < t$.*

*Proof.* The proof is a simple consequence of independence of increments for $W(t)$ and is left as an exercise. □

**Exercise 4.3.** Prove Proposition 4.8.

In general, the natural filtration is not the only filtration of interest. It is often the case that we need to augment the filtration by another stochastic process or by $\mathbf{P}$-null sets. This is for instance the case when studying weak solutions to SDE's or in coupling arguments. When dealing with Wiener processes, we will often consider the more general notion of an *admissible filtration* which is compatible with it's independent increments.

**Definition 4.9** (Admissible filtration for Wiener process)**.** Let $W = \{W(t) : t \in \mathbb{R}_+\}$ be a $Q$-Wiener process on a separable Hilbert space $\mathcal{H}$. A filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ is *admissible* for $W$ if

1. $W$ is adapted to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$
2. For every $s < t$, $W(t) - W(s)$ is independent of $\mathscr{F}_s$.

## 4.3 Progressively measurable processes

In general to define the stochastic integral, we will also need the stronger notion of *progressive measurability*.

**Definition 4.10** (Progressive measurability)**.** Given a filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$, define the progressive sigma-algebra $\mathscr{P} \subseteq \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F}$ by

$$\mathscr{P} = \bigcap_{t\in\mathbb{R}_+} \big\{A \in \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F} \,:\, A \cap ([0,t] \times \Omega) \in \mathscr{B}([0,t]) \otimes \mathscr{F}_t\big\}.$$

The we say that a process $X$ is *progressively measurable* if $X : \mathbb{R}_+ \times \Omega \to \mathcal{U}$ is measurable from $\mathscr{P}$ to $\mathscr{B}(\mathcal{U})$. Alternatively for each $t \geq 0$, $X|_{[0,t]} : [0,t] \times \Omega \to \mathcal{U}$ is measurable from $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$ to $\mathscr{B}(\mathcal{U})$.

It is a simple consequence of the definitions above that any progressively measurable process $X$ is *measurable* in the sense that the process $X : \mathbb{R}_+ \times \Omega \to \mathcal{U}$ is measurable from $\mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F} \to \mathscr{B}(\mathcal{U})$, as well as adapted to $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. As far as the converse goes, we have the following, somewhat surprising, result due to Ondreját and Seidler [OS13].

**Theorem 4.11.** *Suppose that $\mathcal{U}$ is a Polish space (separable, metrizable space), and $X = \{X(t) \,:\, t \in \mathbb{R}_+\}$ is a measurable, $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ adapted process with values in $\mathcal{U}$. Then there exists an $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ progressively measurable process $\bar{X}$ with values in $\mathcal{U}$ such that $\bar{X}$ is a modification of $X$.*

**Exercise 4.4.** In practice we rarely need to apply Theorem 4.11 since nearly all of the processes we will be dealing with are continuous. Show that a process $X = \{X(t) \,:\, t \in \mathbb{R}_+\}$ which is continuous and $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ adapted is also $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ progressively measurable.

Note that viewing the process as a measurable function $X : \mathbb{R}_+ \times \Omega \to \mathcal{U}$ allows us to define equivalence classes of such $\mathcal{U}$-valued functions and therefore can define Lebesgue/Bochner spaces with respect to the measure $\lambda = \mathrm{d}t \times \mathbf{P}$. Specifically, denote

$$L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$$

to be the equivalence class of all functions $X : \mathbb{R}_+ \times \Omega \to \mathcal{U}$ measurable from $\mathscr{B}(\mathbb{R}_+) \times \mathscr{F}$ to $\mathscr{B}(\mathcal{U})$ that are Bochner integrable with norm

$$\|X\|_{L^2(\mathbb{R}_+\times\Omega\,;\mathcal{U})} := \left(\int_{\mathbb{R}_+\times\Omega} \|X\|\,\mathrm{d}\lambda\right)^{1/2} < \infty.$$

It well-known that the space $L^2(\mathbb{R}_+ \times \Omega;\mathcal{U})$ is complete and defines Hilbert space in the usual way just as in the real-valued case. Additionally, given a filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$, we define the subspace

$$L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega\,;\mathcal{U}) \subseteq L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U}).$$

of progressively measurable processes. More precisely, we view an equivalence class of $L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$ as an element of $L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$ if it contains a member which is measurable from $\mathscr{P}$ to $\mathscr{B}(\mathcal{U})$.

**Proposition 4.12.** *Let $\mathcal{U}$ be a separable Banach space, then the space $L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$ is a closed subspace of $L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$.*

*Proof.* Suppose that $\{X_n\}_{n\geq 0}$ is a sequence of processes in $L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$, converging to $X \in L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$. Since the limit $X$ belongs to $L^2(\mathbb{R}_+ \times \Omega\,;\mathcal{U})$, it is measurable from $\mathscr{B}([0,T]) \otimes \mathscr{F}$ to $\mathscr{B}(\mathcal{U})$ but might not be something which is measurable from $\mathscr{P}$ to $\mathscr{B}(\mathcal{U})$, in fact, it probably isn't. However, we only need to show that $X$ is equivalent to a progressively measurable process to prove the proposition.

It is a simple extension of the usual property of Lebesgue spaces to show that, up to a subsequence, $X_n(t,\omega) \to X(t,\omega)$ in $\mathcal{U}$ for $\mu$ almost every $(t,\omega)$. Indeed, if $([0,t] \times \Omega, \mathscr{B}([0,t]) \otimes \mathscr{F}_t, \mathrm{d}t \times \mathbf{P})$ were a complete measure space, then we would be able to argue that $X$ was progressively measurable since pointwise almost everywhere convergence from a complete measure space to a measurable space implies measurable. However, since $([0,t]\times\Omega, \mathscr{B}([0,t])\otimes\mathscr{F}_t, \mu)$ is *not* complete, we must argue with more care. Note that the set

$$A = \{(t,\omega)\in[0,T]\times\Omega\,;\ \lim_{n\to\infty} X_n(t,\omega) \text{ exists}\}.$$

is progressively measurable in the sense that $\mathbb{1}_A(t,\omega)$ is progressively measurable. It follows that the process

$$\bar{X}(t,\omega) = \begin{cases} \lim_{n\to\infty} X(t,\omega) & \text{if } x\in A \\ 0 & \text{if } x\notin A. \end{cases}$$

is the pointwise *everywhere* limit of progressively measurable processes, and is therefore progressively measurable. Since $X$ is $\mathscr{B}([0,T])\otimes\mathscr{F}$ measurable, then the set

$$B = \{(t,\omega)\in[0,T]\times\Omega\,;\ \lim_{n\to\infty} X_n(t,\omega) = X(t,\omega)\}\subset A$$

is $\mathscr{B}([0,T])\otimes\mathscr{F}$ measurable and $\mathrm{d}t\times\mathbf{P}(B^c)=0$. Therefore $\bar{X}$ may be modified on a $\mu$ null set to equal $X$. □

**Remark 4.13.** It is natural to also define the space $L^2_{\text{ad}}(\mathbb{R}_+\times\Omega\,;\mathcal{U})$ of equivalence classes which contain one adapted processes. Indeed, Theorem 4.11 implies that adapted and progressively measurable processes are equivalent and therefore $L^2_{\text{ad}}(\mathbb{R}_+\times\Omega\,;\mathcal{U}) = L^2_{\text{pr}}(\mathbb{R}_+\times\Omega\,;\mathcal{U})$, even though an adapted process (unless right-continuous) is certainly *not* progressively measurable.

## 4.4 Continuous martingales

Before continuing it will be convenient to review some basics about continuous time Martingales in Banach spaces. In what follows let $(\Omega,\mathscr{F},\mathbf{P})$ be a probability space and $\mathcal{U}$ be a separable Banach space and $I\subseteq\mathbb{R}_+$ and index set. We will mostly be interested in the case when $I$ is an uncountable subset as the assumption is that the reader is already familiar with discrete time Martingales.

**Definition 4.14.** A stochastic process $M=\{M(t)\,:\,t\in I\}$ with values in $\mathcal{U}$ is a *martingale* with respect to a filtration $(\mathscr{F}_t)_{t\in I}$ if

1. $M$ is adapted to $(\mathscr{F}_t)_{t\in I}$,
2. $\mathbf{E}\|M(t)\|<\infty$ for all $t\in I$,
3. $\mathbf{E}[M(t)|\mathscr{F}_s] = M(s)$ for all $s\le t$, $s,t\in I$.

**Exercise 4.5.** Let $W=\{W(t)\,:\,t\in\mathbb{R}_+\}$ be a cylindrical Wiener process on a Hilbert space $\mathcal{W}$ and let $\mathscr{F}_t=\sigma(W|_{[0,t]})$. Show that $W=\{W(t)\,:\,t\in\mathbb{R}_+\}$ has the property that for $t,s\in\mathbb{R}_+$ the increment $W(t+s)-W(t)$ is independent of $\mathscr{F}_t$ and therefore $W$ is a continuous $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ martingale with values in $\mathcal{W}_1$

Many of the theorems for $\mathbb{R}$ valued martingales can be proven for certain functionals of Banach space valued Martingales.

**Definition 4.15.** A *real-valued* process $M=\{M(t)\,:\,t\in I\}$ is said to be a *sub-martingale* if both 1. and 2. hold, but instead 3. is replaced by

$$\mathbf{E}[M(t)|\mathscr{F}_s]\ge M(s).$$

Alternatively a $M$ is a *super-martingale* if $-M$ is a sub-martingale.

**Exercise 4.6.** Suppose that $M = \{M(t) : t \in I\}$ is a $\mathcal{U}$-valued martingale with respect to $(\mathscr{F}_t)_{t\in I}$. Show that $\{\|M(t)\| : t \in I\}$ is sub-martingale. In addition if $g : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing convex function such that $\mathbf{E}g(\|M(t)\|) < \infty$ show that $\{g(\|M(t)\|) : t \in I\}$ also a sub-martingale.

The following theorem follows from a theorem by Doob for real-valued sub-martingales. The proof is classical and can be found in [KS91] Theorem 3.8.

**Theorem 4.16.** *Suppose that $M = \{M(t) : t \in I\}$ is a $\mathcal{U}$ valued martingale with continuous paths $t \mapsto M(t)$ (if $I$ is discrete then this is always true) then*

$$\mathbf{P}\left(\sup_{t\in I}\|M(t)\| \geq \lambda\right) \leq \frac{1}{\lambda}\sup_{t\in I}\mathbf{E}\|M(t)\|.$$

*Moreover, if for each $p > 1$, $\mathbf{E}\|M(t)\|^p < \infty$ then*

$$\mathbf{E}\left(\sup_{t\in I}\|M(t)\|^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{t\in I}\mathbf{E}\|M(t)\|^p.$$

In it convenient to study spaces of Martingales. Specifically, fix a $T > 0$ and denote by $\mathcal{M}^2_T(\mathcal{U})$ the space of all continuous $\mathcal{U}$-valued square integrable Martingales with $M(0) = 0$. By square integrable, we mean $\mathbf{E}\|M(t)\|^2 < \infty$ for every $t \in [0, T]$. We can define a norm for $\mathcal{M}^2_T(\mathcal{U})$ by

$$\|M\|_{\mathcal{M}^2_T(\mathcal{U})} := \left(\mathbf{E}\sup_{t\in[0,T]}\|M(t)\|^2\right)^{1/2}.$$

Note that since $\|M(t)\|$ is sub-martingale, then we have for all $t < T$

$$\mathbf{E}\|M(t)\| \leq \mathbf{E}\|M(T)\|.$$

Therefore by Doobs inequality

$$\|M\|^2_{\mathcal{M}^2_T(\mathcal{U})} \leq 4\mathbf{E}\|M(T)\|^2,$$

and therefore $\|\cdot\|_{\mathcal{M}^2_T(\mathcal{U})}$ is a well defined norm. A important fact about Martingales is that they are stable under convergence with respect to $\|\cdot\|_{\mathcal{M}^2_T(\mathcal{U})}$.

**Proposition 4.17.** *The space $\mathcal{M}^2_T(\mathcal{U})$ with norm $\|\cdot\|_{\mathcal{M}^2_T(\mathcal{U})}$ is a Banach space.*

*Proof.* Let $\{M_n\}$ be a Cauchy sequence in $\mathcal{M}^2_T(\mathcal{U})$. Note that this implies that there exists a subsequence $\{M_{n_k}\}$ such that

$$\mathbf{P}\left(\sup_{t\in[0,T]}\|M_{n_k} - M_{n_{k-1}}\| > 2^{-k}\right) < 2^{-k}.$$

It follows by Borel-Cantelli and completeness of $C([0, T]; \mathcal{U})$, that there is a continuous limit $M(t)$ such that $M_{n_k}$ converges to $M$ in $C([0, T]; \mathcal{U})$. Using the completeness of $L^2(\Omega; \mathcal{U})$ we deduce that for each $t$ $M_{n_k}(t) \to M(t)$ in $L^2(\Omega; \mathcal{U})$. Therefore since $M_{n_k}$ is a martingale we can pass the limit

$$\mathbf{E}\left(M(t)|\mathscr{F}_s\right) = M(s)$$

$\mathbf{P}$-almost surely to conclude that $M(s)$ is a continuous Martingale. It follows that $M(t) - M_{n_k}(t)$ is a continuous Martingale and by Doobs inequality

$$\|M - M_{n_k}\|^2_{\mathcal{M}^2_T(\mathcal{U})} \leq 4\mathbf{E}\|M(T) - M_{n_k}(T)\|^2 \to 0.$$

Therefore since $\{M_n\}$ was Cauchy, we have that $M_n \to M$ in $\mathcal{M}^2_T(\mathcal{U})$. □

## 4.5 Constructing the stochastic integral

In what follows, let $\mathcal{W}$ be a separable Hilbert space and let $W = \{W(t) \,:\, t \in \mathbb{R}_+\}$ be a cylindrical Wiener process on $\mathcal{W}$ with respect to some probability space $(\Omega, \mathscr{F}, \mathbf{P})$. In light of Theorem 3.54, we will often refer to a cylindrical Wiener process on a separable Hilbert space $\mathcal{W}$ without referencing the Hilbert space $\mathcal{W}_1$ or the Hilbert-Schmidt inclusion $J$ and will implicitly assume it takes values in some larger Hilbert $\mathcal{W}_1$.

Let $\mathcal{H}$ be another separable Hilbert space and denote $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ the space of Hilbert-Schmidt operators from $\mathcal{W}$ to $\mathcal{H}$, namely the space of all operators $A : \mathcal{W} \to \mathcal{H}$ such that $AA^* : \mathcal{H} \to \mathcal{H}$ is trace class.

**Proposition 4.18.** *The space $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ is a separable Hilbert space with the inner product*

$$\langle A_1, A_2 \rangle_{\mathcal{L}_2} = \mathrm{Tr}(A_1 A_2^*).$$

*Proof.* The Hilbert space property is obvious from the definition and that fact that the trace class operators form Banach space. To see separability, let $\{e_k\}_{k\in\mathbb{N}}$ be a complete orthonormal system in $\mathcal{W}$ and $\{h_k\}_{k\in\mathbb{N}}$ be a complete orthonormal system in $\mathcal{H}$. Then the mapping

$$A \mapsto (\langle Ae_k, h_j \rangle_{\mathcal{H}})_{k,j\in\mathbb{N}}$$

defines a linear isomorphism between $\mathcal{L}_2(\mathcal{W}, \mathcal{U})$ and $\ell^2(\mathbb{N} \times \mathbb{N})$. Separability follows since $\ell^2(\mathbb{N} \times \mathbb{N})$ is separable. ☐

In order to build the stochastic integral, just as with any integral, we will start by defining it on simpler class of functions and then extend it to the large class. Many of the properties of the stochastic integral will be established on this smaller and persist for the extension.

**Definition 4.19.** A process $\Phi = \{\Phi(t) \,:\, t \in \mathbb{R}_+\}$ with values in $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ is called an *elementary process* if there exists a sequence of times $\{t_i\}_{i=1}^n \subset \mathbb{R}_+$ with $t_0 < \ldots < t_n$, and a collection of $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ -valued random variables $\{\Phi_k\}_{k=0}^{n-1}$ with $\Phi_k$ being $\mathscr{F}_{t_k}$ measurable and $\mathbf{E}\|\Phi_k\|_{\mathcal{L}_2}^2 < \infty$ such that $\Phi$ can be written as

$$\Phi(t) = \sum_{k=1}^{n} \Phi_k \mathbb{1}_{(t_{k-1}, t_k]}(t).$$

It is easy to see that an elementary process is adapted to the filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. and that it belongs to the Lebesgue/Bochner space $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$. Moreover an elementary process is also a simple $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ valued function on $\mathbb{R}_+ \times \Omega$. Denote the subspace of elementary processes by

$$L^2_{\mathrm{el}}(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H})).$$

We have the following density result:

**Proposition 4.20.** *The elementary processes are dense in $L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$.*

*Proof.* Let $\Phi \in L^2_{\mathrm{pr}}(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ be a progressively measurable representative. We may assume with out loss of generality that $(t, \omega) \mapsto \|\Phi(t, \omega)\|_{\mathcal{U}}$ is bounded, since by dominated convergence

$$\lim_{R\to\infty} \int \|\Phi\|_{\mathcal{L}_2}^2 \mathbb{1}_{\{\|\Phi\|_{\mathcal{L}_2} > R\}} \,\mathrm{d}\lambda = 0.$$

Moreover in view of the time-integrability of $\Phi(t)$, we can also assume that $t \mapsto \Phi(t)$ is supported on $[0, T]$ since

$$\lim_{T\to\infty} \int \|\Phi\|_{\mathcal{L}_2}^2 \mathbb{1}_{[0,T]} \mathrm{d}\lambda = \int \|\Phi\|_{\mathcal{L}_2}^2 \mathrm{d}\lambda$$

We begin by regularizing the process and defining the continuous process $\hat{\Phi}^\epsilon(t)$ by

$$\hat{\Phi}^\epsilon(t,\omega) = \epsilon^{-1}\int_{(t-\epsilon)\vee 0}^{t} \Phi(s,\omega)\mathrm{d}s.$$

We claim that $\hat{\Phi}^\epsilon$ is progressively measurable. Indeed since $\Phi$ is progressively measurable, the process $\Phi(s)\mathbb{1}_{[(t-\epsilon)\vee 0,t]}(s)$ is measurable with respect to $\mathscr{B}([0,t])\otimes\mathscr{F}_t$. Furthermore, the integrability of $\Phi$ with respect to $\lambda$ means that we may apply Fubini's theorem to $\Phi(s)\mathbb{1}_{[(t-\epsilon)\vee 0,t]}(s)$ to conclude that

$$\omega \mapsto \int_0^\infty \Phi\mathbb{1}_{[(t-\epsilon)\vee 0,t]}(s,\omega)\mathrm{d}s$$

is $\mathscr{F}_t$ measurable, hence $\hat{\Phi}^\epsilon(t)$ is adapted. Since $\hat{\Phi}^\epsilon(t)$ is now continuous and adapted, we conclude that $\hat{\Phi}^\epsilon(t)$ is also progressively measurable. Now, the fact that $\hat{\Phi}^\epsilon$ is continuous, means that the elementary process

$$\Phi^{n,\epsilon}(t) = \sum_{k=0}^{2^n-1}\hat{\Phi}^\epsilon(kT2^{-n},\omega)\mathbb{1}_{(kT2^{-n},(k+1)T2^{-n}]}(t)$$

converges pointwise in $(t,\omega)$ to $\hat{\Phi}^\epsilon$ in $\mathcal{L}_2(\mathcal{W},\mathcal{H})$. Since both $\hat{\Phi}^\epsilon$ and $\Phi^{n\epsilon}$ are uniformly bounded we may apply the bounded convergence theorem to conclude

$$\lim_{n\to\infty}\int \|\hat{\Phi}^\epsilon - \Phi^{n,\epsilon}\|^2_{\mathcal{L}_2}\mathrm{d}\lambda = 0.$$

Finally, since for almost every $t\in[0,T]$ and every $\omega\in\Omega$, $\hat{\Phi}^\epsilon_t(\omega)\to\Phi_t(\omega)$ in $\mathcal{U}$ as $\epsilon\to 0$, we may again use the bounded convergence theorem to show

$$\lim_{\epsilon\to 0}\int \|\Phi - \hat{\Phi}^\epsilon\|^2_{\mathcal{L}_2}\mathrm{d}\lambda = 0.$$

□

**Remark 4.21.** Naturally, in view of Remark 4.13, since $\mathcal{L}_2(\mathcal{W},\mathcal{U})$ is separable, this implies that simple processes are also dense in $L^2_{\mathrm{ad}}(\mathbb{R}_+\times\Omega;\mathcal{U})$.

If $\Phi$ is an elementary process in $\mathcal{L}_2(\mathcal{W},\mathcal{H})$ then we define it's *stochastic integral* with respect to $W$ as the $\mathcal{H}$-valued random variable defined by

$$\int_0^\infty \Phi(s)\mathrm{d}W(s) := \sum_{k=1}^n \Phi_{k-1}\Delta W_k, \quad \Delta W_k := W(t_k) - W(t_{k-1}).$$

Note that $W(t_k)$ and $W(t_{k-1})$ are cylindrical Wiener processes on $\mathcal{W}$ and so the product $\Phi_{k-1}\Delta W_k$ has to be interpreted with care since $W(t)$ doesn't actually take take values in $\mathcal{W}$. Naturally we would like to apply Theorem 3.54 and define it in terms of a measurable linear extension to some large Hilbert space $\mathcal{W}_1$. Indeed, for a given fixed $\omega_1\in\Omega$ let $\hat{\Phi}_{k-1}(\omega_1)$ be the measurable extension to $\mathcal{W}_1$ given by Corollary 3.55 then we define

$$[\Phi_{k-1}\Delta W_k](\omega) := \hat{\Phi}_{k-1}(\omega)\Delta W_k(\omega).$$

Of course since $\hat{\Phi}_{k-1}$ is random, the random variable $\Phi_{k-1}\Delta W_k$ is *not* a Gaussian random variable. However since $\Phi_{k-1}$ and $\Delta W_k$ are independent it behaves like a Gaussian with respect to the conditional expectation $\mathbf{E}[\cdot|\mathscr{F}_{t_{k-1}}]$.

**Lemma 4.22.** *Let $\Phi_k$ and $\Delta W_k$ be as above. Then the following hold for each $k \in \mathbb{N}$*

$$\mathbf{E}[\Phi_{k-1}\Delta W_k | \mathscr{F}_{t_{k-1}}] = 0$$

*and*

$$\mathbf{E}\left[\|\Phi_{k-1}\Delta W_k\|_{\mathcal{H}}^2 | \mathscr{F}_{t_{k-1}}\right] = \|\Phi_{k-1}\|_{\mathcal{L}_2}^2 (t_k - t_{k-1}).$$

*Proof.* Note that since $\Phi_{k-1}$ and $\Delta W_k$ are independent, we have for each $\omega_1 \in \Omega$

$$\mathbf{E}[\Phi_{k-1}\Delta W_k | \mathscr{F}_{t_{k-1}}](\omega_1) = \mathbf{E}\Phi_{k-1}(\omega_1)\Delta W_k = 0$$

and

$$\begin{aligned}
\mathbf{E}\left[\|\Phi_{k-1}\Delta W_k\|_{\mathcal{H}}^2 | \mathscr{F}_{t_{k-1}}\right](\omega_1) &= \mathbf{E}\|\Phi_{k-1}(\omega_1)\Delta W_k\|_{\mathcal{H}}^2 \\
&= \mathrm{Tr}(\Phi_{k-1}(\omega_1)\Phi_{k-1}(\omega_1)^*)(t_k - t_{k-1}) \\
&= \|\Phi_{k-1}(\omega_1)\|_{\mathcal{L}_2}^2 (t_k - t_{k-1}),
\end{aligned}$$

since for each $\omega_1 \in \Omega$, $\Phi_{k-1}(\omega_1)\Delta W_k$ is a centered Gaussian random variable in $\mathcal{H}$ with covariance

$$\Phi_{k-1}(\omega_1)\Phi_{k-1}^*(\omega_1)(t_k - t_{k-1}).$$

□

The following result is fundamental to the definition of the Itô integral, known as the *Itô isometry*.

**Proposition 4.23** (Itô Isometry)**.** *Let $\Phi$ be an elementary process, then the following identity holds*

$$\mathbf{E}\left\|\int_0^\infty \Phi(s)\mathrm{d}W(s)\right\|_{\mathcal{H}}^2 = \mathbf{E}\int_0^\infty \|\Phi(s)\|_{\mathcal{L}_2}^2 \mathrm{d}s.$$

*Proof.* We have

$$\mathbf{E}\left\|\int_0^\infty \Phi(s)\mathrm{d}W(s)\right\|_{\mathcal{H}}^2 = \sum_{j,k=1}^n \mathbf{E}\langle \Phi_{k-1}\Delta W_k, \Phi_{j-1}\Delta W_j\rangle_{\mathcal{H}}$$

First we argue that if $k \neq j$ then

$$\mathbf{E}\langle \Phi_{k-1}\Delta W_k, \Phi_{j-1}\Delta W_j\rangle_{\mathcal{H}} = 0.$$

Indeed, without loss of generality, we assume that $k < j$ and therefore

$$\begin{aligned}
\mathbf{E}\langle \Phi_{k-1}\Delta W_k, \Phi_{j-1}\Delta W_j\rangle_{\mathcal{H}} &= \int_\Omega \mathbf{E}[\langle \Phi_{k-1}\Delta W_k, \Phi_{j-1}\Delta W_j\rangle_{\mathcal{H}} | \mathscr{F}_{t_{j-1}}](\omega)\mathrm{d}\mathbf{P}(\omega) \\
&= \int_\Omega \mathbf{E}\langle [\Phi_{k-1}\Delta W_k](\omega), \hat{\Phi}_{j-1}(\omega)\Delta W_j\rangle_{\mathcal{H}}\mathrm{d}\mathbf{P}(\omega) \\
&= \int_\Omega \left\langle \hat{\Phi}_{k-1}(\omega)\Delta W_k(\omega), \mathbf{E}\left(\hat{\Phi}_{j-1}(\omega)\Delta W_j\right)\right\rangle_{\mathcal{H}} \mathrm{d}\mathbf{P}(\omega) \\
&= 0,
\end{aligned}$$

where we used the fact that $\Phi_{k-1}\Delta W_k$ and $\hat{\Phi}_k$ are $\mathscr{F}_{t_{j-1}}$ measurable and $\Delta W_j$ is independent of $\mathscr{F}_{t_{j-1}}$ as well as Lemma 4.22. It follows that

$$\begin{aligned}
\mathbf{E}\left\|\int_0^\infty \Phi(s)\mathrm{d}W(s)\right\|_{\mathcal{H}}^2 &= \sum_{k=1}^n \mathbf{E}\|\Phi_{k-1}\Delta W_k\|_{\mathcal{H}}^2 \\
&= \sum_{k=1}^n \int_\Omega \mathbf{E}[\|\Phi_{k-1}\Delta W_k\|_{\mathcal{H}}^2|\mathscr{F}_{t_{k-1}}](\omega)\mathrm{d}\mathbf{P}(\omega) \\
&= \sum_{k=1}^n \int_\Omega \mathbf{E}\|\hat{\Phi}_{k-1}(\omega)\Delta W_k\|_{\mathcal{H}}^2\mathrm{d}\mathbf{P}(\omega) \\
&= \sum_{k=1}^n \mathbf{E}\|\Phi_{k-1}\|_{\mathcal{L}_2}^2(t_k - t_{k-1}) \\
&= \mathbf{E}\int_0^\infty \|\Phi(s)\|_{\mathcal{L}_2}^2\mathrm{d}s.
\end{aligned}$$

□

The Itô isometry recieves it's name because it implies that the mapping $I : L^2_{\text{el}}(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H})) \to L^2(\Omega; \mathcal{H})$, given by

$$I(\Phi) := \int_0^\infty \Phi(s)\mathrm{d}W(s), \tag{4.1}$$

is an isometry. Combining this with the density of elementary processes in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ given by Proposition 4.20 and the completeness of Lebesgue/Bochner spaces, we find:

**Corollary 4.24.** *The mapping $I$ defined in (4.1) can be uniquely extended to an isometry from $L^2_{\text{el}}(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H})) \to L^2(\Omega; \mathcal{H})$. Specifically, for each progressively measurable $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$- valued process $\Phi = \{\Phi(t) : t \in \mathbb{R}_+\}$, we denote this extension by*

$$\int_0^\infty \Phi(s)\mathrm{d}W(s),$$

*and call it the Itô stochastic integral of $\Phi$ with respect to $W$.*

For each $t \in \mathbb{R}_+$ we define the running time integral by

$$\int_0^t \Phi(s)\mathrm{d}W(s) := \int_0^\infty \mathbb{1}_{[0,t]}(s)\Phi(s)\mathrm{d}W(s) = \sum_{k=1}^n \Phi_{k-1}(W(t_k \wedge t) - W(t_{k-1} \wedge t)),$$

where we have used the notation $a \wedge b = \min\{a, b\}$. In this case the process

$$t \mapsto \int_0^t \Phi(s)\mathrm{d}W(s)$$

is a continuous $\mathcal{H}$-valued stochastic process. Moreover it is a martingale.

**Lemma 4.25.** *Let $\Phi = \{\Phi(t) : t \in \mathbb{R}_+\}$ be an elementary $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$-valued process. Then*

$$t \mapsto \int_0^t \Phi(s)\mathrm{d}W(s)$$

*is a continuous $\mathcal{H}$-valued square integrable martingale and satisfies*

$$\mathbf{E}\sup_{t\in[0,T]}\left\|\int_0^t \Phi(s)\mathrm{d}W(s)\right\|_{\mathcal{H}}^2 \le 4\mathbf{E}\int_0^T \|\Phi(s)\|_{\mathcal{L}_2}^2\mathrm{d}s$$

*Proof.* The fact that the process is continuous follows from the fact that $t \mapsto W(t_k \wedge t)$ is continuous in $\mathcal{W}_1$. Square integrability also follows from the Itô isometry. Hence we really just need to show the Martingale property

$$\mathbf{E}\left[\int_0^t \Phi(r)\mathrm{d}W(r) \,\middle|\, \mathscr{F}_s\right] = \int_0^s \Phi(r)\mathrm{d}W(r).$$

This is relatively straight-forward and is left as an exercise. □

**Exercise 4.7.** Prove that $t \mapsto \int_0^t \Phi(s)\mathrm{d}W(s)$ in Lemma 4.25 is a Martingale. The final bound then follows from Doob's inequality and Itô' isometry.

Naturally, through corollary 4.24 the martingale property is preserved by passing limits in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{H})$. Indeed, for each $T > 0$ define a mapping $I_T$ from elementary processes $L^2_{\mathrm{el}}([0,T] \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ to the space of $\mathcal{H}$-valued continuous square integrable Martingales $\mathcal{M}^2_T(\mathcal{H})$, by

$$\hat{I}_T(\Phi) := \int_0^\cdot \Phi(s)\mathrm{d}W(s).$$

Since $\hat{I}_T$ is continuous and $\mathcal{M}^2_T$ is a complete space, we deduce the following easy consequence.

**Proposition 4.26.** *Let $\Phi$ be a progressively measurable process in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$, then the process*

$$t \mapsto \int_0^t \Phi(s)\mathrm{d}W(s) = \int_0^\infty \mathbb{1}_{[0,t]}\Phi(s)\mathrm{d}W(s),$$

*is a continuous square-integrable $\mathcal{H}$-valued martingale.*

**Remark 4.27.** It is possible to extend the class of integrands $\Phi$ to progressively measurable $\mathcal{L}_2$ beyond $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ to progressively measurable processes $\Phi$ taking values in $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$ such that

$$\mathbf{P}\left(\int_0^\infty \|\Phi(s)\|^2_{\mathcal{L}_2}\mathrm{d}s < \infty\right) = 1,$$

called *stochastically integrable* processes. In this case the associated process $t \mapsto \int_0^t \Phi(s)\mathrm{d}W(s)$ is a *local martingale*. We choose note to cover this general case here.

## 4.6 Properties of the stochastic integral

The stochastic integral has several properties that we will find useful. The first allows us to pass linear operators through the stochastic integral.

**Lemma 4.28.** *Let $\Phi$ be a progressively measurable process in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ and let $W = \{W(t) : t \in \mathbb{R}_+\}$ be a cylindrical Wiener process on $\mathcal{W}$, then for any other separable Hilbert space $\hat{\mathcal{H}}$ and a bounded operator $A \in \mathcal{L}(\mathcal{H}, \hat{\mathcal{H}})$, we have $\mathbf{P}$-almost surely*

$$A\left(\int_0^\infty \Phi(s)\mathrm{d}W(s)\right) = \int_0^\infty A\Phi(s)\mathrm{d}W(s). \tag{4.2}$$

*Proof.* Note that $A\Phi$ is an $\mathcal{L}_2(\mathcal{W}, \hat{\mathcal{H}})$-valued progressively measurable process. Moreover if $\Phi_n$ is an elementary $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$-valued process, then $A\Phi_n$ is an elementary $\mathcal{L}_2(\mathcal{W}, \hat{\mathcal{H}})$-valued process. Moreover, it is easy to see that (4.2) holds for elementary processes. Now suppose that $\Phi_n \to \Phi$ in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ then $A\Phi_n \to A\Phi$ in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \hat{\mathcal{H}}))$ and so by the definition of the stochastic integral we can pass the limit almost surely in both sides of

$$A\left(\int_0^\infty \Phi_n(s)\mathrm{d}W(s)\right) = \int_0^\infty A\Phi_n(s)\mathrm{d}W(s)$$

and use the continuity of $A$ to conclude. □

**Proposition 4.29** (Stochastic Fubini). *Fill me in*

Another useful result is the following representation of the stochastic integral in terms of standard one-dimensional stochastic integrals.

**Proposition 4.30.** *Let $\Phi$ be an $L^2([0,T] \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$ progressively measurable process, then for each each orthonormal basis $\{e_k\}_{k\in\mathbb{N}}$ of $\mathcal{W}$, there exists a sequence of standard 1-d Wiener processes $\{W_k(t)\}_{k\in\mathbb{N}}$ such that for each $t \in [0,T]$, we have*

$$\int_0^t \Phi(s)\mathrm{d}W(s) = \sum_{k\in\mathbb{N}} \int_0^t \Phi(s)(e_k)\,\mathrm{d}W_k(t)$$

$\mathbf{P}$*-almost surely and the convergence happens in $L^2(\Omega; C([0,T]; \mathcal{H}))$.*

*Proof.* Let $\Pi_{\leq n}$ denote the projection on $\mathcal{W}$ to the first $n$ elements of the basis $\{e_k\}$

$$\Pi_{\leq n} h = \sum_{k\leq n} h_k e_k,$$

and let $\Pi_{\geq n} = \mathrm{Id} - \Pi_{\leq n}$. It is not hard to see that for elementary processes $\Phi$ and the fact that we can write the measurable linear extension of $\Phi(s)\Pi_n \in \mathcal{L}_2(\mathcal{W}, \mathcal{H})$ acting on $W(t)$ as

$$\widehat{\Phi(s)\Pi_n} W(t) = \sum_{k\leq n} \Phi(s)(e_k) W_k(t).$$

and therefore

$$\int_0^t \Phi(s)\Pi_n \mathrm{d}W(s) = \sum_{k\leq n} \int_0^t \Phi(s)(e_k)\mathrm{d}W_k(s).$$

This can easily be extended to progressively measurable processes by realizing that if $\Phi_n$ is an elementary function approximation of some progressively measurable processes $\Phi$, then $\Phi_n(e_k)$ is an elementary function approximation of the $\mathcal{H}$-valued progressively measurable process $\Phi(e_k)$. Next we note that

$$t \mapsto \int_0^t \Phi(s)\Pi_{\geq n}\mathrm{d}W(s) = \int_0^t \Phi(s)\mathrm{d}W(s) - \int_0^t \Phi(s)\Pi_{\leq n}\mathrm{d}W(s)$$

is a continuous $\mathcal{H}$-valued square integrable Martingale, so that by Doobs inequality

$$\begin{aligned}
\mathbf{E} \sup_{t\in[0,T]} \left\| \int_0^t \Phi(s)\Pi_{\geq n}\mathrm{d}W(s) \right\|_{\mathcal{H}}^2 &\leq 4\mathbf{E} \left\| \int_0^T \Phi(s)\Pi_{\geq n}\mathrm{d}W(s) \right\|_{\mathcal{H}}^2 \\
&= 4\mathbf{E} \int_0^T \|\Phi(s)\Pi_{\geq n}\|_{\mathcal{L}_2}^2 \mathrm{d}s \\
&= 4\mathbf{E} \int_0^T \sum_{k\geq n} \|\Phi(s)(e_k)\|_{\mathcal{H}}^2 \mathrm{d}s \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

by dominated convergence since $\|\Phi\|_{\mathcal{L}_2}$ belongs to $L^2([0,T] \times \Omega)$. □

## 4.7 Burkhold-Davis-Gundy inequality

To be filled out

## 4.8 Itô's formula

One of the most powerful features of the stochastic integral is that certain functions of it can be written as another stochastic integral, effectively allows one to use a chain rule. For instance, we will see that the following formula holds

$$\left\| \int_0^t \Phi(s)\mathrm{d}W(s) \right\|_{\mathcal{H}}^2 = \int_0^t \|\Phi(s)\|_{\mathcal{L}_2}^2 \mathrm{d}s + \int_0^t \left\langle \int_0^s \Phi(r)\mathrm{d}W(r), \Phi(s)\mathrm{d}W(s) \right\rangle_{\mathcal{H}},$$

which implies that $\int_0^t \|\Phi(s)\|_{\mathcal{L}_2}^2 \mathrm{d}s$ is the quadratic variation associated with the martingale $t \to \int_0^t \Phi(s)\mathrm{d}W(s)$ and easily implies Itô's isometry.

In what follows, we consider an $\mathcal{H}$-valued process $X = \{X(t)\,;\, t \in \mathbb{R}_+\}$ given by

$$X(t) = X(0) + \int_0^t \varphi(s)\mathrm{d}s + \int_0^t \Phi(s)\mathrm{d}W(s), \tag{4.3}$$

where $\varphi$ is a progressively measurable $\mathcal{H}$-valued process in $L^1(\mathbb{R}_+ \times \Omega; \mathcal{H})$ and $\Phi$ is a progressively measurable $\mathcal{L}_2(\mathcal{W}, \mathcal{H})$-valued process in $L^2(\mathbb{R}_+ \times \Omega; \mathcal{L}_2(\mathcal{W}, \mathcal{H}))$.

**Remark 4.31.** Such a process $X(t)$ is often referred to as an *Itô process* or *Itô diffusion* and is often written using stochastic differentials as a short-hand

$$\mathrm{d}X(t) = \varphi(t)\mathrm{d}t + \Phi(t)\mathrm{d}W(t),$$

to be interpreted in the time integrated sense (4.3) above.

Additionally let $F : \mathbb{R}_+ \times \mathcal{H} \to \mathbb{R}$ be a twice Fréchet differentiable function whose derivatives are given by

$$\begin{aligned} &\partial_t F : \mathbb{R}_+ \times \mathcal{H} \to \mathbb{R} \\ &DF : \mathbb{R}_+ \times \mathcal{H} \to \mathcal{H}^* \cong \mathcal{H} \\ &D^2 F : \mathbb{R}_+ \times \mathcal{H} \to \mathcal{L}(\mathcal{H}), \end{aligned}$$

and are uniformly continuous on bounded subsets of $\mathbb{R}_+ \times \mathcal{H}$. For the process $X(t)$ given above and the Fréchet differentiable function $F$, we have the following result.

**Theorem 4.32.** *The following formula holds* $\mathbf{P}$*-almost surely for each* $t \in \mathbb{R}_+$.

$$\begin{aligned} F(t, X(t)) = F(0, X(0)) &+ \int_0^t \langle DF(s, X(s), \Phi(s)\mathrm{d}W(s)\rangle \\ &+ \int_0^t \partial_t F(s, X(s)) + \langle DF(s, X(s)), \varphi(s)\rangle \mathrm{d}s \\ &+ \int_0^t \mathrm{Tr}\left(D^2 F(s, X(s))\Phi(s)\Phi(s)^*\right) \mathrm{d}s. \end{aligned}$$

*Proof.* By density of elementary functions it suffices to consider $\varphi$ and $\Phi$ to be elementary functions. Furthermore, by restricting to intervals of constant $\varphi$ and $\Phi$, we may reduce to the case where $X(t)$ is given by

$$X(t) - X(T_0) = \varphi_0(t - T_0) + \Phi_0(W(t) - W(T_0))$$

for all $t \in [T_0, T_1] \subset \mathbb{R}_+$ and $\varphi_0$ and $\Phi_0$ are constant random variables measurable with respect to $\mathscr{F}_{T_0}$.

Next, let $T_0 = t_0 < t_1 < \ldots < t_n = t$ be a partition of $[T_0, t] \subset [T_0, T_1]$. For each $k \in \{1, \ldots n\}$, we denote by

$$\Delta t_k = t_k - t_{k-1}, \quad \Delta_k X = X(t_k) - X(t_{k-1}), \quad \Delta W_k = W(t_k) - W(t_{k-1}),$$

and note that we can write

$$\begin{aligned} F(t, X(t)) - F(T_0, X(T_0)) &= \sum_{k=1}^{n} F(t_k, X(t_k)) - F(t_{k-1}, X(t_k)) \\ &\quad + \sum_{k=1}^{n} F(t_{k-1}, X(t_k)) - F(t_{k-1}, X(t_{k-1})) \end{aligned}$$

Using Taylor expansion, we can write this as

$$\begin{aligned} F(t, X(t)) - F(T_0, X(T_0)) &= \sum_{k=1}^{n} \partial_t F(t_k, X(t_k))\Delta t_k + \sum_{k=1}^{n} \langle DF(t_{k-1}, X(t_{k-1})), \Delta X_k\rangle \\ &\quad + \frac{1}{2}\sum_{k=1}^{n} \langle D^2 F(t_{k-1}, X(t_{k-1}))\Delta X_k, \Delta X_k\rangle + R_n \end{aligned}$$

where

$$\begin{aligned} R_n = &\sum_{k=1}^{n} \int_0^1 [\partial_t F(t_{k-1} + \theta\Delta t_k, X(t_k)) - \partial_t F(t_k, X(t_k))]\Delta t_k \mathrm{d}\theta \\ &+ \frac{1}{2}\sum_{k=1}^{n} \int_0^1 \langle [D^2 F(t_{k-1}, X(t_{k-1}) + \theta\Delta X_k) - D^2 F(t_{k-1}, X(t_{k-1}))]\Delta X_k, \Delta X_k\rangle \mathrm{d}\theta \end{aligned}$$

Clearly sending the the partition size $\sup_k \Delta t_k \to 0$, we see that $\mathbf{P}$-almost surely

$$\sum_{k=0}^{n-1} \partial_t F(t_{k+1}, X(t_{k+1}))\Delta t_k \to \int_0^t \partial_t F(s, X(s))\mathrm{d}s$$

and using the fact that $\Delta X_k = \varphi_0 \Delta t_k + \Phi_0 \Delta W_k$ we also have $\mathbf{P}$-almost surely

$$\sum_{k=1}^{n} \langle DF(t_{k-1}, X(t_{k-1})), \Delta X_{k-1}\rangle \to \int_0^t \langle DF(s, X(s)), \varphi_0\rangle \mathrm{d}s + \int_0^t \langle DF(s, X(s)), \Phi_0 \mathrm{d}W(s)\rangle.$$

It is not hard to show that

$$\begin{aligned} |R_n| \leq &\sup_{s \leq \sup_k \Delta t_k} \|\partial_t F(\cdot + s, \cdot) - \partial_t F(\cdot, \cdot)\|_\infty (t - T_0) \\ &+ \sup_{|y| \leq \sup_k |\Delta X_k|} \|D^2 F(\cdot, \cdot + y) - D^2 F(\cdot, \cdot)\|_\infty \sum_{k=1}^{n} \|\Delta X_k\|^2 \end{aligned}$$

and so using the fact that $\partial_t F$ and $D^2 F$ are uniformly continuous that $X(t)$ is continuous and $\sum_k |\Delta X_k|^2$ is almost surely bounded as $\Delta t_k \to 0$ we deduce that $\mathbf{P}$-almost surely

$$|R_n| \to 0$$

as the partition size $\sup_k \Delta t_k \to 0$.

It remains to deal with the term

$$\begin{aligned}\frac{1}{2}\sum_{k=1}^{n}\langle D^2F(t_{k-1},X(t_{k-1}))\Delta X_k,\Delta X_k\rangle &= \frac{1}{2}\sum_{k=1}^{n}\langle \Phi_0^* D^2F(t_{k-1},X(t_{k-1}))\Phi_0\Delta W_k,\Delta W_k\rangle \\ &+\sum_{k=1}^{n}\langle D^2F(t_{k-1},X(t_{k-1}))\Phi_0\Delta W_k,\varphi_0\rangle\Delta t_k \\ &+\frac{1}{2}\sum_{k=1}^{n}\langle D^2F(t_{k-1},X(t_{k-1}))\varphi_0,\varphi_0\rangle(\Delta t_k)^2 \\ &= I_1^n+I_2^n+I_3^n\end{aligned}$$

It is easy to see using independence

$$\mathbf{E}[\langle D^2F(t_{k-1},X(t_{k-1}))\Phi_0\Delta W_k,\varphi_0\rangle\Delta t_k]^2=\mathbf{E}\langle D^2F(t_{k-1},X(t_{k-1}))\Phi_0\Phi_0^*\varphi_0,\varphi_0\rangle(\Delta t_k)^3$$

and therefore it is not hard to deduce that as $\sup_k \Delta t_k \to 0$

$$I_2^n \to 0$$

in $L^2(\Omega)$ (and therefore a subsequence goes $\mathbf{P}$-almost surely) and similarly due to boundedness $D^2F$ we have

$$I_3^n \to 0$$

$\mathbf{P}$-almost surely. For the remaining term $I_1^n$, we denote $G_k := \Phi_0^* D^2F(t_{k-1},X(t_{k-1}))\Phi_0$ and note that if $k \neq j$ we have

$$\mathbf{E}\left(\langle G_{k-1}\Delta W_k,\Delta W_k\rangle - \mathrm{Tr}\left(G_k\right)\Delta t_k\right)\left(\langle G_{j-1}\Delta W_j,\Delta W_j\rangle - \mathrm{Tr}\left(G_j\right)\Delta t_j\right)=0$$

and therefore

$$\begin{aligned}\mathbf{E}\left(\sum_{k=1}^{n}\langle G_{k-1}\Delta W_k,\Delta W_k\rangle - \mathrm{Tr}\left(G_k\right)\Delta t_k\right)^2 &= \mathbf{E}\sum_{k=1}^{n}\left(\langle G_{k-1}\Delta W_k,\Delta W_k\rangle - \mathrm{Tr}\left(G_{k-1}\right)\Delta t_k\right)^2 \\ &\lesssim \sum_{k=1}^{n}(\Delta t_k)^2 \to 0.\end{aligned}$$

as $\sup_k \Delta t_k \to 0$. Therefore up to a subsequence we have shown that

$$I_1^n \to \int_0^t \mathrm{Tr}\left(D^2F(s,X(s))\Phi(s)\Phi(s)^*\right)\mathrm{d}s$$

$\mathbf{P}$-almost surely as $\sup_k \Delta t_k \to 0$.

□

# 5 Strongly continuous one-parameter semigroups

## 5.1 $C_0$ semigroups

Before we study stochastic linear evolution equations we will need to cover some preliminaries for deterministic linear evolution equations in Banach spaces, known as the theory of $C_0$ semigroups, which will serve as the theoretical underpinning for much of our discussion of linear and semilinear SPDE. Much of the presentation here follows the book by Engel and Nagel [EN99] as well as the notes by Hairer.

$\mathcal{U}$ of the form

$$\partial_t u(t) = Lu(t), \quad u(0) = u \in \mathcal{U} \tag{5.1}$$

where $L$ is a possibly unbounded operator on $\mathcal{U}$ with dense domain $\mathrm{Dom}(L)$. The time derivative above is interpreted as the following strong limit in $\mathcal{U}$

$$\partial_t u(t) = \lim_{h \to 0} \frac{u(t+h) - u(t)}{h}.$$

This problem (5.1) is refered to as the abstract Cauchy problem for $L$.

We will consider the following notion of well-posedness for the above equation.

**Definition 5.1.** We say that the Cauchy problem (5.1) is strongly *well-posed* if for each $u \in \mathrm{Dom}(L)$ there exists a unique solution $u(t)$ to (5.1) starting from $u$ such that $u(t)$ is strongly differentiable in time and if $\{u_n\} \subseteq \mathrm{Dom}(L)$ is such that $u_n \to 0$ in $\mathcal{U}$ and $u_n(t)$ are the associated solutions then

$$u_n(t) \to 0$$

in $\mathcal{U}$ uniformly on compact sets of $t$.

If (5.1) is strongly well-posed, then the solution $u(t)$ starting from initial data $u \in \mathrm{Dom}(L)$ should be given by

$$u(t) = S(t)u$$

where $\{S(t) : t \in \mathbb{R}_+\}$ is a family of linear operators on $\mathrm{Dom}(L)$ which map the initial data $u$ to the solution $u(t)$ (it is linear because the evolution equation is linear).

**Exercise 5.1.** Show that if (5.1) is well-posed, then $S(t)$ can be uniquely extended to a bounded linear operator on $\mathcal{U}$.

Well-posedness also means that the operators extension of the operators $S(t)$ to $\mathcal{U}$ should satisfy $S(0) = \mathrm{Id}$, while uniqueness implies that they have a *semigroup property* $S(t+s) = S(t)S(s)$, meaning that evolving $u$ forward by $t+s$ should be the same as evolving $u(t)$ forward by $s$ or $u(s)$ forward by $t$. This motivates the following definition.

**Definition 5.2.** A family of linear operators $\{S(t) : t \in \mathbb{R}_+\}$ is called a $C_0$ semigroup if

1. $S(0) = \mathrm{Id}$ and for all $t, s \in \mathbb{R}_+$,
$$S(t+s) = S(t)S(s),$$
2. $S(t)$ is *strongly continuous*, namely for each $u \in \mathcal{U}$ the following limit holds in $\mathcal{U}$
$$\lim_{t \to 0} S(t)u = u.$$

**Remark 5.3.** Note that strong continuity of $S(t)$ is equivalent to convergence of $S(t) \to \mathrm{Id}$ in the *strong operator topology* and can generally be relaxed to showing that $S(t)u \to \mathrm{Id}$ for all $u$ in a dense subset of $\mathcal{U}$. However, in general it is not necessarily that case that $S(t)$ converges to Id in the operator norm. (We will see an example of this later).

**Example 5.4** (Heat equation)**.** An easy example of a linear evolution equation on a Banach space is the heat equation

$$\partial_t u - \partial_x^2 u = 0,$$

on $[0,1]$ with Dirichlet conditions $u(0) = u(1) = 0$. In this case one can take $\mathcal{U} = C([0,1])$ (or any $L^p$ space) and the semigroup is given by convolution the Green's function $K(t,x)$

$$S(t)f := K(t,\cdot) \star f.$$

**Example 5.5** (Transport equation)**.** suppose that $b(x)$ is a smooth vector field on the torus $\mathbb{T}^d$, then the transport equation

$$\partial_t u = b \cdot \nabla u.$$

is a linear evolution equation and the semigroup is given by the method of characteristics

$$S(t)u(x) = u(\phi_t(x)),$$

where $\phi_t(x)$ is the flow associated with the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_t(x) = b(\phi_t(x)), \quad \phi_0(x) = x \in \mathbb{T}^d.$$

**Remark 5.6.** Note that the two examples studied about are fundamentally different since one canfact that we are studying semigroups (as opposed to groups) is important here since the evolution of the equation (3.37) may lose information over time. A typical examply of this is the heat equation

$$\partial_t u - \partial_x^2 u = 0.$$

However the semigroup is not a group since although it is possible to define a left inverse for $S(t)$ the associated inverse is not defined for all continuous functions (i.e. the backwards in time heat equation is not well-defined).

**Proposition 5.7.** *Let $S(\cdot)$ be a $C_0$ semigroup on $\mathcal{U}$*

1. *There exists constants $C > 0, \omega \in \mathbb{R}$ such that*

$$\|S(t)\| \leq Ce^{\omega t}. \tag{5.2}$$

2. *For every $u \in \mathcal{U}$, $t \mapsto S(t)u$ is a strongly continuous function.*

*Proof.* First we claim that $\|S(t)\|$ is uniformly bounded in a neighborhood of $t = 0$. Indeed if not then there would exist a sequence of times $t_n \to 0$ such that $\|S(t_n)\| \to \infty$. Then by the uniform boundedness principle (or it's converse) there must exist a $u \in \mathcal{U}$ such that $\|S(t_n)u\| \to \infty$ as $n \to \infty$, contradicting strong continuity. Therefore there exists an $r > 0$ and $C > 0$ such that

$$\sup_{t\in[0,r]} \|S(t)\| < \infty.$$

Using the semigroup property we can see that for each $t \in \mathbb{R}_+$, there exists an integer $n \leq 1/r$ such that $t = nr + s$ where $s \in [0, r)$ and therefore $S(t) = S(r)^n S(s)$. This implies that for each $u \in \mathcal{U}$

$$\|S(t)\| \leq \|S(r)\|^n \sup_{s\in[0,r]} \|S(s)\| \leq C^{n+1} \leq Ce^{\omega t},$$

where $\omega = \frac{1}{r}\log C$

To show continuity we note that for any $s < t \in \mathbb{R}_+$ and $u \in \mathcal{U}$ by the semigroup property we have

$$S(t)u - S(s)u = S(s)\left[S(t-s)u - u\right]$$

and so by (5.2)

$$\|S(t)u - S(s)u\| \leq Ce^{\omega s}\|S(t-s)u - u\|$$

sending $t \to s$ or $s \to t$ completes the proof. □

The bound (5.3) says that the semigroup can't grow (or decay) faster than an exponential, however more can be said about asymptotic exponential growth (or decay) rates of $S(t)$ as $t \to \infty$. Indeed, by the semigroup property we have for $t, s \in \mathbb{R}_+$

$$\log(\|S(t+s)\|) \leq \log(\|S(t)\|) + \log(\|S(s)\|)$$

and therefore the function $t \mapsto \log(\|S(t)\|)$ is sub-additive. It follows by Fekete's subadditive limit theorem that $\lim_{t\to\infty} \frac{1}{t} \log(\|S(t)\|)$ exists and is given by

$$\omega_0 := \inf_t \frac{\log(\|S(t)\|)}{t}. \tag{5.3}$$

The value $\omega_0$ is often called the *growth type* or asymptotic growth of $S(t)$. Clearly we have that $\omega_0 \leq \omega$ (where $\omega$ is from (5.2)) and that

$$\|S(t)\| = \mathcal{O}(e^{\omega_0 t}), \quad \text{as} \quad t \to \infty.$$

**Exercise 5.2.** Show that for each $\epsilon > 0$ there exists an $M_\epsilon \geq 1$ such that we have upper and lower bounds

$$e^{\omega_0 t} \leq \|S(t)\| \leq M_\epsilon e^{(\omega_0+\epsilon)t}$$

Given a $C_0$ semigroup $S(\cdot)$ we can always associate a linear operator, called it's infinitesimal generator (or just generator) which is essentially it's derivative at $t = 0$.

**Definition 5.8.** The *infinitesimal generator* $L$ of $S(t)$ is a linear operator $L$ defined by

$$Lu := \lim_{t\to 0} \frac{S(t)u - u}{t}$$

on the set $\mathrm{Dom}(L)$ of elements $u \in \mathcal{U}$ where the above limit exists.

Recall that a linear operator $A : \mathrm{Dom}(A) \subseteq \mathcal{U} \to \mathcal{U}$ is *closed* if for every $\{x_n\} \subset \mathrm{Dom}(A)$ such that $x_n \to x$ in $\mathcal{U}$ and such that $Ax_n \to y$, then $x \in \mathrm{Dom}(A)$ and $Ax = y$, or equivalently if it's graph

$$\mathcal{G}_A = \{(x, y) \in \mathcal{U} \times \mathcal{U} \;:\; x \in \mathrm{Dom}(A), y = Ax\}$$

is closed in $\mathcal{U} \times \mathcal{U}$. $L$ is said *closable* if the closure of $\mathcal{G}_A$ is still the graph of a linear operator, refered to as the closure of $L$.

The following theorem shows that the generator $L$ is always closed and that $u(t) = S(t)u$ solves (5.1) in a strong or time integrated sense, depending on whether $u \in \mathrm{Dom}(L)$ or $\mathcal{U}$.

**Theorem 5.9.** *Given a $C_0$ semigroup $S(\cdot)$ on $\mathcal{U}$ the infinitesimal generator $L$ is a densely defined closed linear operator on $\mathcal{U}$. Moreover, the $\mathrm{Dom}(L)$ is invariant under $S(t)$ and for every $u \in \mathrm{Dom}(L)$, we have*

$$\partial_t S(t)u = LS(t)u = S(t)Lu. \tag{5.4}$$

*Additionally, for every $u \in \mathcal{U}$, and $t > 0$, $\int_0^t S(s)u \,\mathrm{d}s \in \mathrm{Dom}(L)$ and we have*

$$S(t)u - u = L \int_0^t S(s)u \,\mathrm{d}s \tag{5.5}$$

*Proof.* First, we note that we can write the difference quotient in two ways

$$\frac{S(t+h)u - S(t)u}{h} = S(t)\frac{S(h)u - u}{h} = \frac{S(h) - \mathrm{Id}}{h} S(t)u$$

when $u \in \mathrm{Dom}(L)$, the middle equality above converges to $S(t)Lu$ and therefore both the left and the right equalities also converge implying the invariance of $\mathrm{Dom}(L)$ under $S(t)$ as well as formula (5.4). Moreover, given a $u \in \mathcal{U}$ we have

$$\begin{aligned} h^{-1}(S(h) - \mathrm{Id}) \int_0^t S(s)u \,\mathrm{d}s &= \frac{1}{h}\int_0^t S(s+h)u - S(s)u\mathrm{d}s \\ &= \frac{1}{h}\int_t^{t+h} S(s)u\mathrm{d}s - \frac{1}{h}\int_0^h S(s)u\mathrm{d}s \end{aligned}$$

Since $t \mapsto S(t)u$ is strongly continuous, sending $h \to 0$ we see that the right-hand side converges strongly and therefore $\int_0^t S(s)u\mathrm{d}s \in \mathrm{Dom}(L)$ with

$$L\int_0^t S(s)u \,\mathrm{d}s = S(t)u - u$$

To see density of $\mathrm{Dom}(L)$, for a given $u \in \mathcal{U}$ and $t > 0$ define a "regularized" $u_n$ by

$$u_n = n\int_0^{1/n} S(s)u \,\mathrm{d}s.$$

and note that for each $n$ $u_n \in \mathrm{Dom}(L)$ and by strong continuity $u_n \to u$ in $\mathcal{U}$ implying density of $\mathrm{Dom}(L)$.

To see closedness, take $\{u_n\} \subseteq \mathrm{Dom}(L)$ converging to $u \in \mathcal{U}$ such that $Lu_n \to y \in \mathcal{U}$ then by time-integrating (5.4) we find

$$\frac{S(h)u_n - u_n}{h} = \frac{1}{h}\int_0^h S(s)Lu_n \,\mathrm{d}s.$$

Sending $n \to \infty$ on both sides above we find

$$\frac{S(h)u - u}{h} = \frac{1}{h}\int_0^h S(s)y \,\mathrm{d}s.$$

By strong continuity, the right-hand side converges as $h \to 0$ and therefore $u \in \mathrm{Dom}(L)$ and we deduce $Lu = y$ as required. □

**Exercise 5.3.** Show that for any $n \in \mathbb{N}$ the $\mathrm{Dom}(L^n)$ is dense in $\mathcal{U}$. (Hint: Mollify $S(t)u$ in time by integrating against a smooth approximation $\phi_\epsilon(t)$ of $\delta_0$).

**Exercise 5.4.** Show that a $C_0$ semigroup is uniquely determined by it's generator.

## 5.2 Characterization of the generator: Hille-Yosida theorem

In order to build a $C_0$ semigroup from a closed generator $L$, we will need some elementary spectral theory.

**Definition 5.10.** Let $A$ be a closed operator. We say that $\lambda \in \mathbb{C}$ belongs to the *resolvent set* $\rho(A)$ if $\lambda\,\mathrm{Id} - A$ maps $\mathrm{Dom}(A)$ one-to-one and onto $\mathcal{U}$. The complement of $\rho(A)$ in $\mathbb{C}$ is called the *spectrum* of $A$ and is denoted by $\sigma(A)$.

For a given closed operator, if $\lambda \in \rho(A)$, then $(\lambda \operatorname{Id} - L)$ is invertible. It's inverse

$$R(\lambda) = (\lambda \operatorname{Id} - A)^{-1}$$

is called the *resolvent* of $A$ and maps $\mathcal{U}$ onto $\operatorname{Dom}(A)$. Since $A$ is closed it is easy to see that $R(\lambda)$ is also closed, and therefore by the Closed Graph Theorem $R(\lambda)$ is a bounded linear operator on $\mathcal{U}$.

**Exercise 5.5.** Suppose that $\rho(A)$ is non-empty and $\lambda \in \rho(A)$. Show that $\gamma \in \sigma(A)$ if and only if $(\lambda - \gamma)^{-1}$ belongs to $\sigma(R(\lambda))$. Deduce from this the $\sigma(A)$ is a closed subset of $\mathbb{C}$.

**Exercise 5.6.** Show that for each $\lambda, \mu \in \rho(A)$ we have the resolvent identity

$$R(\lambda) - R(\mu) = (\mu - \lambda)R(\mu)R(\lambda).$$

and therefore $R(\mu)$ and $R(\lambda)$ commute.

**Exercise 5.7.** Use Neumann series expansion to show that that if $\lambda, \lambda_0 \in \rho(A)$ are such that $|\lambda - \lambda_0| \leq \|R(\lambda_0)\|^{-1}$, then

$$R(\lambda) = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n R(\lambda_0)^{n+1},$$

and the series converges uniformly. Deduce that $R(\lambda)$ is analytic on $\rho(A)$.

When $L$ is the generator of a $C_0$ semigroup, the resolvent $R(\lambda)$ of $L$ can always be directly computed from the semigroup via Laplace transform.

**Proposition 5.11.** *Let $S(t)$ be a $C_0$ semigroup with generator $L$. If $\lambda \in \mathbb{C}$ satisfies $\operatorname{Re}\lambda > \omega_0$, then $\lambda \in \rho(L)$ and the following formula holds*

$$R(\lambda) = \int_0^\infty e^{-\lambda t} S(t)\, \mathrm{d}t,$$

*where the right-hand side is understood as an improper Riemann integral.*

*Proof.* Let $\epsilon = \frac{1}{2}(\operatorname{Re}\lambda - \omega_0)$, then for $T_\epsilon$ chosen suitably large we can see that for $t \geq T_\epsilon$

$$\|e^{-\lambda t} S(t)\| \leq e^{-\epsilon t}$$

and therefore, combined with (5.2) impies that the Laplace transform

$$Z(\lambda)u = \int_0^\infty e^{-\lambda t} S(t) u\, \mathrm{d}t$$

is well-defined. Next we claim that $Z(\lambda) = R(\lambda)$. To show this, it suffices to realized that that $e^{-\lambda t}S(t)$ is in fact another $C_0$ semigroup on $\mathcal{U}$ with generator $L - \lambda I$ and domain $\operatorname{Dom}(L)$. Therefore for each $u \in \operatorname{Dom}(L)$

$$e^{-\lambda t} S(t) u - u = \int_0^t e^{-\lambda s} S(s)(L - \lambda \operatorname{Id}) u\, \mathrm{d}s$$

sending $t \to \infty$ and using the fact that $\operatorname{Re}\lambda > \omega_0$ we conclude $u = Z(\lambda)(\lambda \operatorname{Id} - L)u$. This implies that $Z(\lambda)$ is a left-inverse of $(\lambda \operatorname{Id} - L)$. To see it is a right inverse take $u \in \mathcal{U}$ and conclude that

$$e^{-\lambda t} S(t) u - u = (L - \lambda \operatorname{Id}) \int_0^t e^{-\lambda s} S(s) u\, \mathrm{d}s.$$

Sending $t \to \infty$ again gives $u = (\lambda \operatorname{Id} - L)Z(\lambda)u$, whereby we deduce that $\lambda \in \rho(L)$ and that $Z(\lambda) = R(\lambda)$.

□

We are now ready to prove the following characterization theorem for generators of $C_0$ semigroups. The version originally was due to

**Theorem 5.12** (Hille Yosida). *Let $L$ be a closed, densely defined operator on $\mathcal{U}$ with*

$$s(L) = \sup\{\text{Re}\lambda \,:\, \lambda \in \rho(L)\}.$$

*The following are equivalent.*

1. *$L$ is the generator of a $C_0$ semigroup $S(\cdot)$ with growth type $\omega_0 \leq s(L)$.*
2. *For each $\omega > s(L)$, there exists an $M$ such that for each $\lambda > \omega$ and $n \geq 1$*

$$\|R(\lambda)^n\| \leq M(\text{Re}\,\lambda - \omega)^{-n}.$$

*Proof.* To see that 1 implies 2, we set $\epsilon = \omega - \omega_0$ and note that there exists an $M$ such that

$$\|S(t)\| \leq Me^{(\omega_0+\epsilon)t},$$

the resolvent bound then follows from the fact that for $\text{Re}\,\lambda > \omega$

$$R(\lambda)^n = \int_0^\infty \cdots \int_0^\infty e^{-\lambda(t_1+\ldots+t_n)} S(t_1 + \ldots + t_n)\text{d}t_1 \ldots \text{d}t_n.$$

To show that 2 implies 1, we will need to construct a semigroup from $L$ to do this, we define the *Yosida approximation* $L_n := nLR(n)$. Note that the identity

$$L_n = n^2 R(n) - n\,\text{Id}$$

shows that $L_n$ is a bounded operator. Next we note that for each $u \in \text{Dom}(L)$

$$nR(n)u - u = R(n)Lu$$

and therefore

$$\|nR(n)u - u\| \leq \|R(n)Lu\| \leq M(n-\omega)^{-1}\|Lx\| \to 0,$$

as $n \to \infty$. It follows by the fact that $nR(n)$ is uniformly bounded in $n$ and density of $\text{Dom}(L)$ that for each $u \in \mathcal{U}$

$$\lim_{n\to\infty} nR(n)u = u.$$

It follows that for each $u \in \text{Dom}(L)$ that

$$\|L_n u - Lu\| = \|(nR(n) - \text{Id})Lu\| \to 0,$$

so that $L_n$ is indeed an approximation of $L$.

To build the semigroup, since $L_n$ is bounded we define the approximation

$$S_n(t) = e^{tL_n} = \sum_{k=0}^\infty \frac{t^k}{k!} L_n^k,$$

which clearly converges for all $t \in \mathbb{R}_+$. It follows from the formula $L_n = n^2R(n) - n\,\text{Id}$ that we have the uniform in $n$

$$\|S_n(t)\| \leq e^{-nt} \sum_{k=0}^\infty \frac{t^k n^{2k}}{k!} \|R^k(n)\| \leq M \exp\left(-nt + \frac{n^2}{n-\omega} t\right) = M \exp\left(\frac{n\omega t}{n-\omega}\right)$$

It follows that $\limsup_{n\to\infty} \|S_n(t)\| \le Me^{\omega t}$. We now want to show that $S_n(t)$ converges. To estimate the difference of $S_n(t)$ and $S_m(t)$ we use the fact that $L_n$ and $L_m$ commute with $S_n(t)$ and $S_m(t)$ to conclude for $u \in \mathrm{Dom}(L)$

$$\partial_s S_n(t-s)S_m(s)u = S_n(t-s)S_m(s)[L_m - L_n]u.$$

Integrating this in $s$ from 0 to $t$ and choosing $n, m$ large enough (relative to $\omega$) so that $\|S_n(t)\|, \|S_m(t)\| \le e^{2\omega t}$ we find

$$\|S_n(t)u - S_m(t)u\| \le e^{2\omega t}\|L_m u - L_n u\|.$$

Using the fact that $L_n u \to Lu$ we deduce that $S_n(t)u$ has a limit as $n \to \infty$ and therefore we define $S(t)u$ by

$$S(t)u = \lim_{n\to\infty} S_n(t)u.$$

Moreover the limit exists uniformly on bounded sets in $t$ and therefore strong continuity of $S_n(t)u$ implies strong continuity of $S(t)u$. It follows from the uniform boundedness of $S_n(t)$ that $\|S(t)\| \le Me^{\omega t}$ and clearly $S(t)$ inherits the semigroup property from $S_n(t)$.

It remains to show the generator of $S(t)$ is $L$. To see this we write

$$S_n(t)u - u = \int_0^t S_n(s)L_n u \mathrm{d}s.$$

If $u \in \mathrm{Dom}(L)$ we can take the limit as $n \to \infty$ to deduce

$$S(t)u - u = \int_0^t S(s)Lu \mathrm{d}s.$$

Let $\hat{L}$ be the generator of $S(t)$. If we divide the above expression by $t$ and take $t \to 0$ we conclude that $\mathrm{Dom}(L) \subseteq \mathrm{Dom}(\hat{L})$ and that $\hat{L}$ is an extension of $L$. However since $\lambda > \omega$ belongs to $\rho(\hat{L})$ and $\rho(L)$, $\hat{L}$ cannot be a proper extension of $L$.

□

### 5.3 Adjoint semigroups

It will be useful later to use duality methods for semigroups to study weak solutions. This involves understanding the evolution of the adjoint $S^*(t) : \mathcal{U}^* \to \mathcal{U}^*$ of a $C_0$ semigroup. Recall the definition of an adjoint on a Banach space.

**Definition 5.13.** Let $A : \mathrm{Dom}(A) \subset \mathcal{U} \to \mathcal{U}$ be a densely defined linear operator. Then the adjoint $A^* : \mathrm{Dom}(A^*) : \mathcal{U}^* \to \mathcal{U}^*$ is is defined by

$$\langle A^*\ell, u\rangle = \langle \ell, Au\rangle$$

for all $u \in \mathrm{Dom}(A)$ and $\ell \in \mathrm{Dom}(A^*)$ define by

$$\mathrm{Dom}(A^*) = \{\ell \in \mathcal{U}^* \,:\, u \mapsto \langle \ell, Au\rangle \text{ is continuous on } \mathrm{Dom}(A)\}\,.$$

It is an easy exercise to show that the adjoint of any densely defined linear operator $A$ is well-defined and closed. However it may not be densely defined.

**Exercise 5.8.** Show that $A^*$ is closed even if $A$ is not closed.

As a consequence of this, it is possible that the adjoint $S^*(t)$ of a $C_0$ semigroup may fail to be continuous on $\mathcal{U}$. Indeed the adjoint $L^*$ of the generator $L$ of $S(t)$ may not be densely defined and *cannot* be the generator of a $C_0$ semigroup on $S(t)$. The problem is that the space $\mathcal{U}^*$ may be too big. This can be resolved by closing $\mathrm{Dom}(A^*)$ in $\mathcal{U}^*$ (which is potentially smaller than $\mathcal{U}^*$) and restricting the semigroup $S^*(t)$ to this space.

**Proposition 5.14.** *Let $S(\cdot)$ be a $C_0$ semigroup on $\mathcal{U}$ and L be it's generator. Define $\mathcal{U}^\dagger$ to be the closure of $\mathrm{Dom}(L^*)$ in $\mathcal{U}$. Then $S^*(t)$ maps $\mathcal{U}^\dagger$ into itself and is a $C_0$ semigroup on $\mathcal{U}^\dagger$ with generator $L^\dagger$ with domain*

$$\mathrm{Dom}(L^\dagger) = \left\{ \ell \in \mathrm{Dom}(L^*) \ : \ L^*\ell \in \mathcal{U}^\dagger \right\}.$$

*Proof.* Recall that for each $u \in \mathrm{Dom}(L)$, we have $S(t)Lu = LS(t)u$. Therefore for each $\ell \in \mathrm{Dom}(L^*)$ and $u \in \mathrm{Dom}(L)$ we have

$$|\langle S^*(t)\ell, Lu\rangle| = |\langle L^*\ell, S(t)u\rangle| \leq \|L^*\ell\|\|S(t)\|\|u\|$$

and so $S(t)\ell$ belongs to $\mathrm{Dom}(L^*)$. Since $S^*(t)$ is also bounded on $\mathcal{U}$ we easily see that $S^*(t)$ is a well-defined bounded operator on $\mathcal{U}^\dagger$. Note that formula (6.2) implies that for each $\ell \in \mathrm{Dom}(L^*)$ we have

$$S^*(t)\ell - \ell = \int_0^t S^*(s)L^*\ell \mathrm{d}s$$

It follows that $S^*(t)\ell \to \ell$ in $\mathcal{U}^*$ as $t \to 0$ for all $\ell \in \mathrm{Dom}(L^*)$ since $S^*(t)$ is bounded near zero on $\mathcal{U}^\dagger$, by density of $\mathrm{Dom}(L^*)$ in $\mathcal{U}^\dagger$, we get that $S(t)$ is a $C_0$ semigroup on $\mathcal{U}^\dagger$. Finally, note that the resolvent $R^\dagger(\lambda)$ on $\mathcal{U}^\dagger$ is given by

$$R^\dagger(\lambda) = \int_0^\infty e^{-\lambda t} S^*(s)\mathrm{d}s$$

and therefore $R^\dagger(\lambda)$ is simply the restriction of $R^*(\lambda)$ to $\mathcal{U}^\dagger$. It follows that

$$\mathrm{Dom}(L^\dagger) = \mathrm{Ran}(R^\dagger(\lambda)) = \mathrm{Ran}(R^*(\lambda)|_{\mathcal{U}^\dagger}) = \left\{ \ell \in \mathrm{Dom}(L^*) \ : \ (\lambda - L^*)\ell \in \mathcal{U}^\dagger \right\},$$

which corresponds to the stated domain of $L^\dagger$. ☐

Although $\mathcal{U}^\dagger$ may not be dense in $\mathcal{U}^*$ in the strong topology. It is dense in the weak-* topology and therefore can be used to separate points in $\mathcal{U}$. The following results shows this.

**Lemma 5.15.** *$\mathcal{U}^\dagger$ is weak-* dense in $\mathcal{U}^*$, namely for each $\ell \in \mathcal{U}^*$, there exists $\{\ell_n\}$ such that for every $u \in \mathcal{U}$, $\langle \ell_n, u\rangle \to \langle \ell, u\rangle$.*

*Proof.* For each $\ell \in \mathcal{U}^*$ define $\ell_n = nR^*(n)\ell$. It is clear that $\ell_n \in \mathrm{Dom}(L^*) \subseteq \mathcal{U}^\dagger$. However since

$$|\langle \ell_n - \ell, u\rangle| = |\langle \ell, nR(n)u - u\rangle| \leq \|\ell\|_*\|nR(n)u - u\| \to 0$$

since we know from the proof of the Hille-Yoside theorem that $nR(n)u \to u$ in $\mathcal{U}$ as $n \to \infty$. ☐

### 5.4 Analytic semigroups

In many cases, we are interested in the case when $S(t)$ has some smoothing properties. This is for instance the case for parabolic PDE like the heat equation. At the level of generality of $C_0$ semigroups, this smoothing behavior is often associated with an important class of semigroups called *analytic semigroups*. In addition to smoothing properties, this type of semigroup has a better perturbation theory with regards to infinitesimal generator of the semigroup as well as a better spectral theory.

Our approach will define analytic semigroups in terms of spectral conditions on the generator. Specifically, let $\omega \in \mathbb{R}$ and $\theta \in (0, \pi/2)$, and define the *sector* $S_{\omega,\theta}$ in $\mathbb{C}$ with angle $\theta$ by

$$S_{\omega,\theta} = \{\lambda \in \mathbb{C} \,:\, |\arg(\omega - \lambda)| < \theta\}.$$

An illustration of $S_{\omega,\theta}$ is included below shaded blue.

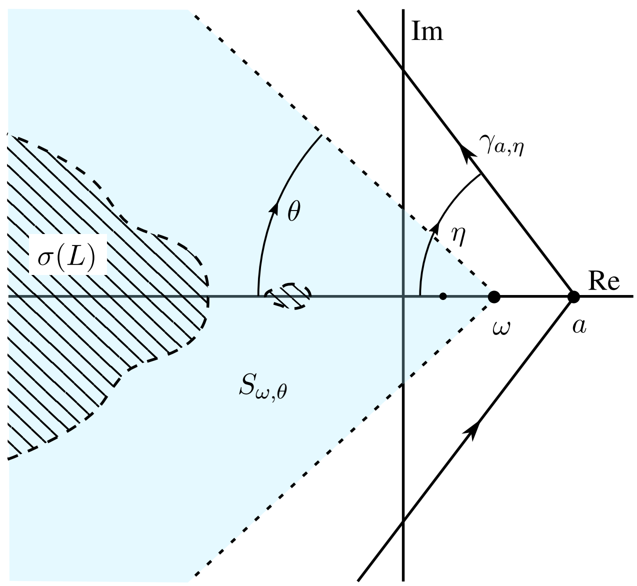


Figure 1: Diagram of $S_{\omega,\theta}$ and $\gamma_{a,\eta}$

**Definition 5.16** (Sectorial operator)**.** We say that a densely defined closed linear operator $L : \mathrm{Dom}(L) \subseteq \mathcal{U} \to \mathcal{U}$ is *sectorial* with angle $\theta$ if

1. there is an $\omega \in \mathbb{R}$ such that $\sigma(L) \subseteq S_{\theta,\omega}$
2. there exists an $M \geq 1$ such that the resolvent $R(\lambda) = (\lambda \operatorname{Id} - L)^{-1}$ satisfies

$$\|R(\lambda)\| \leq M(|\lambda - \omega|)^{-1}$$

for all $\lambda \notin S_{\omega,\theta}$.

**Definition 5.17** (Analytic semigroup)**.** Given a sectorial operator $L$ we can define a semigroup $S(t)$ on $\mathcal{U}$ known as an *analytic semigroup* by the inverse Laplace transform

$$S(t) := \frac{1}{2\pi i} \int_{\gamma_{a,\eta}} e^{\lambda t} R(\lambda) \,\mathrm{d}\lambda \tag{5.6}$$

where for $a > \omega$ and $\eta > \theta$, $\gamma_{a,\eta}$ is the curve given by

$$\gamma_{a,\eta} = \{\lambda \in \mathbb{C} \,:\, |\arg(a - \lambda)| = \eta\},$$

oriented counter clockwise. See Figure 1 for a precise illustration of $\gamma$.

Since $\gamma_{a,\eta} \notin S_{\omega,\theta}$, the resolvent bound implies that for $\lambda \in \gamma_{a,\eta}$ we have that $R(\lambda)$ is uniformly bounded and therefore the integral is well-defined. Moreover the fact that $R(\lambda)$ is analytic and Cauchy integral formula implies that the definition of $S(t)$ does not depend on the choice of $a$ and $\eta$.

**Example 5.18.** Let $L$ be a self-adjoint operator and suppose that for all $u \in \mathrm{Dom}(L)$, $u \neq 0$, we have $\langle Lu, u\rangle < 0$ then $L$ is sectorial.

**Proposition 5.19.** *Let $L$ be a sectorial operator with sector $S_{\omega,\theta}$ and let $S(t)$ be defined by (5.6). Then $S(t)$ defines a $C_0$ semigroup on $\mathcal{U}$ with generator $L$.*

*Proof.* The semigroup property follows from the resolvent identity $R(\lambda)R(\mu) = (R(\lambda) - R(\mu))/(\mu - \lambda)$ and the fact that for $t, s, > 0$ we have for $\omega < a < a'$

$$\begin{aligned} S(t)S(s) &= \frac{1}{(2\pi i)^2}\int_{\gamma_{a',\eta}}\int_{\gamma_{a,\eta}} e^{\lambda t}e^{\mu s}\frac{R(\lambda)-R(\mu)}{\lambda-\mu}\mathrm{d}\lambda\mathrm{d}\mu \\ &= \frac{1}{(2\pi i)^2}\int_{\gamma_{a,\eta}}\left(\int_{\gamma_{a',\eta}}\frac{e^{\mu s}}{\mu-\lambda}\mathrm{d}\mu\right)e^{\lambda t}R(\lambda)\mathrm{d}\lambda + \frac{1}{(2\pi i)^2}\int_{\gamma_{a',\eta}}\left(\int_{\gamma_{a,\eta}}\frac{e^{\lambda t}}{\lambda-\mu}\mathrm{d}\lambda\right)e^{\mu s}R(\mu)\mathrm{d}\mu \end{aligned}$$

Note that we needed to take $a < a'$ so that the inner integrals above are well defined. Indeed with these choices of contours, we can close the contours $\gamma_{a,\eta}$ and $\gamma_{a',\eta}$ on the left by circles with increasing diameter so that by Cauchy's integral formula we have

$$\frac{1}{2\pi i}\int_{\gamma_{a',\eta}}\frac{e^{\mu s}}{\mu-\lambda}\mathrm{d}\mu = e^{\lambda s} \quad \text{and} \quad \frac{1}{2\pi i}\int_{\gamma_{a,\eta}}\frac{e^{\lambda t}}{\lambda-\mu}\mathrm{d}\lambda = 0.$$

Therefore

$$S(t)S(s) = \frac{1}{2\pi i}\int_{\gamma_{a,\eta}} e^{\lambda(t+s)}R(\lambda)\,\mathrm{d}\lambda = S(t+s)$$

To show strong continuity, we note that it suffices to show that for all $u \in \mathrm{Dom}(L)$

$$\lim_{t\to 0} S(t)u = u.$$

It follows from the Cauchy integral formula that

$$\frac{1}{2\pi i}\int_{\gamma_{a,\eta}}\frac{e^{\lambda t}}{\lambda-\omega}\mathrm{d}\lambda = 1$$

and therefore

$$\begin{aligned} S(t)u - u &= \frac{1}{2\pi i}\int_{\gamma_{a,\eta}} e^{\lambda t}\left(R(\lambda) - \frac{1}{\lambda-\omega}\right)u\,\mathrm{d}\lambda \\ &= \frac{1}{2\pi i}\int_{\gamma_{a,\eta}}\frac{e^{\lambda t}}{\lambda-\omega}R(\lambda)(L-\omega\,\mathrm{Id})u\,\mathrm{d}\lambda \end{aligned}$$

It follows from the resolvent bound that

$$\|e^{\lambda t}(\lambda-\omega)^{-1}R(\lambda)(L-\omega)u\| \leq Me^{\mathrm{Re}\lambda t}|\lambda-\omega|^{-2}\|(L-\omega)u\|$$

and therefore by dominated convergence we have

$$\lim_{t\to 0} S(t)u - u = \frac{1}{2\pi i}\int_{\gamma_{a,\eta}}\frac{R(\lambda)(L-\omega\,\mathrm{Id})u}{\lambda-\omega}\mathrm{d}\lambda = 0,$$

where the last equality follows by closing the contour $\gamma_{a,\eta}$ on the right using cirles of increasing diameter and using the resolvent bound. Therefore $S(\cdot)$ defines a $C_0$ semigroup.

Let $\hat{L}$ denotes it's generator and $\hat{R}(\lambda)$ its resovent. We now want to show that $\hat{L} = L$. To do this, it suffices to show that

$$R(\lambda) = \hat{R}(\lambda)$$

for $\lambda$ sufficiently larger than $\omega$. Indeed, by Proposition (5.11) we know that

$$\hat{R}(\lambda) = \int_0^\infty e^{-\lambda t} S(t) \mathrm{d}t$$

and therefore using Fubini's theorem and taking $\lambda > a$, we have

$$\begin{aligned} \hat{R}(\lambda) &= \frac{1}{2\pi i} \int_{\gamma_{a,\eta}} \int_0^\infty e^{(\mu-\lambda)t} R(\mu) \mathrm{d}t \mathrm{d}\mu \\ &= \int_{\gamma_{a,\eta}} \frac{1}{\lambda - \mu} R(\mu) \mathrm{d}\mu = R(\lambda) \end{aligned}$$

where we used Cauchy's integral formula and closed the contour on the right. □

**Remark 5.20.** Let $L$ be a sectorial operator and define

$$s_0 = \sup\{\mathrm{Re}\lambda \, : \, \lambda \in \sigma(L)\}.$$

Then it is not hard to see that by deforming the contour in definition 5.6 implies that for each $\epsilon > 0$ there exists $M_\epsilon$ such that

$$\|S(t)\| \le M_\epsilon e^{(s_0+\epsilon)t}$$

and therefore by density of $\mathrm{Dom}(L)$ in $\mathcal{U}$, $s_0$ coincides with the type $\omega_0$ of $S(\cdot)$.

As it turns out, semigroups generated by sectorial operators have smoothing properties. Indeed such semigroups have smooth dynamics in time. The consequence of this is that for $t > 0$ $S(t)u$ instantly takes values in $\mathrm{Dom}(L^n)$ for each $n \in \mathbb{N}$, which can be interpreted as a smoothing property.

**Proposition 5.21.** *Let $L$ be a sectorial operator with sector $S_{\omega,\theta}$ then $t \mapsto S(t)$ is analytic and for each $u \in \mathcal{U}$, $t > 0$ and $n \in \mathbb{N}$, $S(t)u \in \mathrm{Dom}(L^n)$ and*

$$\partial_t^n S(t)u = L^n S(t)u.$$

*Moreover for each there exists a constant $M$ and $\omega > 0$ such that*

$$\|LS(t)\| \le \left(\frac{1}{t} + \omega_0\right) M e^{\omega_0 t} \tag{5.7}$$

*Proof.* Analyticity follows immediately from formula (5.6), the analyticity of $t \mapsto e^{\lambda t}$, the fact that $e^{\lambda t}$ decays exponentially along $\gamma_{a,\eta}$ and the fact that the contour integral converges uniformly in $\mathcal{L}(\mathcal{U})$. This means that for each $u \in \mathcal{U}$, $S(t)u$ is infinitely differentiable. In particular the limit

$$\lim_{h\to 0} \frac{S(t+h)u - S(t)u}{h} = \lim_{h \to 0} \frac{S(h) - \mathrm{Id}}{h} S(t)u$$

exists for each $u \in \mathcal{U}$ and therefore $S(t)u \in \mathrm{Dom}(L)$ and

$$\partial_t S(t)u = LS(t)u.$$

Moreover, using the fact that for $u \in \mathrm{Dom}(L)$, $LS(t)u = S(t)Lu$ we have that for $u \in \mathcal{U}$

$$\lim_{h\to 0} \frac{\partial_t S(t+h) - \partial_t S(t)}{h} = \lim_{h\to 0} \frac{S(h) - \mathrm{Id}}{h} LS(t)$$

exists and therefore $LS(t)u \in \mathrm{Dom}(L)$ which implies $S(t)u \in \mathrm{Dom}(L^2)$ and

$$\partial_t^2 S(t)u = L^2 S(t)u.$$

Repeating this argument by induction proves that $\partial_t^n S(t)u = L^n S(t)u$.

To prove (5.7) we can assume without-loss of generality that $\omega_0 = 0$ since we can always study the rescaled semigroup $e^{-\omega_0 t}S(t)$ with generator $L - \omega_0 \,\mathrm{Id}$. Since for each $t > 0$, the operator $LS(t)$ is closed with domain $\mathrm{Dom}(LS(t)) = \mathcal{U}$ and therefore by the closed graph theorem it is bounded. To estimate it's norm, we note that by definition of $S(t)$ and the fact that $LR(\lambda) = \lambda R(\lambda) - \mathrm{Id}$ we have for each $a > 0$

$$\begin{aligned} LS(t) &= \frac{1}{2\pi i}\int_{\gamma_{a,\eta}} e^{\lambda t} LR(\lambda)\mathrm{d}\lambda \\ &= \frac{1}{2\pi i}\int_{\gamma_{a,\eta}} e^{\lambda t} (\lambda R(\lambda) - \mathrm{Id})\mathrm{d}\lambda \\ &= \frac{1}{2\pi i}\int_{\gamma_{a,\eta}} e^{\lambda t} \lambda R(\lambda)\mathrm{d}\lambda \end{aligned}$$

where in the last equality we closed the contour on the right and used Cauchy's integral formula. It follows that after parameterizing $\gamma_{a,\eta}$ by $z(r) = a + e^{\pm i\eta} r$ and using the resolvent bound $\lambda\|R(\lambda)\| \le M$, that

$$\|LS(t)\| \lesssim \int_0^\infty e^{(a + r\sin\eta)t}\mathrm{d}r = \frac{e^{at}}{t\sin\eta}.$$

The estimate follows upon taking $a \to 0$ and using the fact that $\sin\eta > 0$.

□

**Exercise 5.9.** Show that (5.7) implies that for all $t > 0$ and $n \in \mathbb{N}$ we have

$$\|L^n S(t)\| \le \left(\frac{1}{t} + \omega_0\right)^n M^n e^{n\omega_0 t}.$$

The analyticity of $t \mapsto S(t)$ is an important property for semigroups generated by sectorial operators. This is the reason why semigroups generated by sectorial operators are known as *analytic semigroups*.

**Exercise 5.10.** Show that in addition to being real analytic in time. $S(t)$ as defined by 5.6 can be extended to a map $z \mapsto S(z)$ which is analytic in the sector

$$\Sigma_\theta = \{0\} \cup \{\lambda \in \mathbb{C} \,:\, |\arg(\lambda)| < \theta\},$$

and that for all $|\eta| < \theta$, $S_\eta(t) = S(e^{i\eta}t)$ is a strongly continuous semigroup.

Indeed it is possible to prove the following characterization of analytic semigroups analagous to the Hille-Yosida theorem.

**Theorem 5.22.** *Let $S(t)$ be a $C_0$ semigroup on $\mathcal{U}$ with generator $L$ and growth type $\omega_0$. The following are equivalent:*

1. *There exists a $\theta \in (0, \pi/2)$ such that $t \mapsto S(t)$ can be extended to an analytic mapping on $\Sigma_\theta$ and for all $|\eta| < \theta$, $S_\eta(t) = S(e^{i\eta}t)$ is a strongly continuous semigroup with growth type $\omega_0$*
2. *There exists a $\theta \in (0, \pi/2)$ such that $L$ is sectorial with sector $S_{\omega,\theta}$ for each $\omega > \omega_0$.*

*Proof.* We have already proved that 2. implies 1. The proof that 1. implies 2. is simple an uses the fact that the generator of $S_\eta(t)$ is $L_\eta = e^{i\eta}L$, where $L$ is the generator of $S(t)$. □

## 5.5 Variational generators and rigged Hilbert spaces

It is useful to consider a class of generators commonly found in PDE which generate analytic semigroups. In what follows, we assume that $\mathcal{H}$ is a Hilbert space and that there exist another Hilbert space $\mathcal{V} \subseteq \mathcal{H}$ which is continuously and densely embedded in $\mathcal{H}$. Furthermore, by Riesz representation, we identify $\mathcal{H}$ with it's dual $\mathcal{H}^*$. Since we assumed $\mathcal{V}$ is dense in $\mathcal{H}$, it is possible to embed $H$ in $V^*$ in through the duality pairing given by the $H$ inner product

$$\langle v, h\rangle_{\mathcal{H}}, \quad v \in \mathcal{V},\, h \in \mathcal{H},$$

so that we get the embedding sandwich

$$V \subseteq H \subseteq V^*$$

each densely and continuously embedding into the next. Such Hilbert space pair $(\mathcal{H}, \mathcal{V})$ is often called a *rigged Hilbert space* or *Gelfand triple*. In general, one can take $V$ to be a topological vector space instead of Hilbert.

**Example 5.23.** A simple example commonly used in PDE is given by Sobolev spaces

$$H^s \subseteq L^2 \subseteq H^{-s},$$

for $s > 0$.

**Definition 5.24** (Variational operator)**.** Given a rigged Hilbert space $\mathcal{V} \subseteq \mathcal{H}$ we say that a densely defined closed operator $L$ on $\mathrm{Dom}(L) \subseteq \mathcal{H}$ is *variational* if

1. There exists a continuous Bilinear form $B(\cdot, \cdot)$ on $\mathcal{V}$ satisfying the coercivity bound

$$B(v, v) + \lambda\|v\|_{\mathcal{H}}^2 \geq \alpha\|v\|_{\mathcal{V}}^2$$

   for some constants $\lambda \geq 0$ and $\alpha > 0$.

2. We have

$$B(u, v) = -\langle Lu, v\rangle_{\mathcal{H}}, \quad \text{for all } u \in \mathrm{Dom}(L) \text{ and } v \in \mathcal{V}$$

   and

$$\mathrm{Dom}(L) = \{v \in \mathcal{V} \,:\, h \mapsto B(v, h) \text{ is continuous on } \mathcal{H}\}.$$

**Exercise 5.11.** Suppose that $a = a_{ij}{}_{ij=1}^{d}$ and $b = \{b_j\}_{j=1}^{d}$ be smooth bounded functions on $\mathbb{T}^d$ and suppose that $a$ is uniformly elliptic in the sense that there exists a constant $C \geq 1$ such that for each $\xi = \{\xi_i\} \in \mathbb{R}^d$ and $x \in \mathbb{T}^d$

$$C^{-1}|\xi|^2 \leq \sum_{ij} \xi_i \xi_j a_{ij}(x) \leq C|\xi|^2.$$

Let $L$ be the elliptic partial differential operator given by

$$L = \sum_{ij} \partial_i(a_{ij}\partial_j) + b \cdot \nabla.$$

Show that $L$ is a variational operator with rigged Hilbert space $\mathcal{H} = L^2$ and $\mathcal{V} = H^1$. (Hint: use Poincaré's inequality).

As it turns out variations generators are a convenient way to check that an operator generates an analytic semigroup without needing to check any spectral conditions.

**Proposition 5.25.** *Let $L$ be a variational operator on $\mathcal{H}$, then $L$ is the generator of an analytic semigroup $S(t)$ with the bound*

$$\|S(t)\| \leq e^{\alpha t}.$$

*Moreover if $B$ is symmetric, then $L$ is self-adjoint.*

*Proof.* See ([Tan79] Theorem 3.6.1) □

## 5.6 Interpolation spaces, fractional powers and regularity

To study the precise regularity properties of analytic semigroups it is convenient to introduce a continuum scale of subspaces of $\mathcal{U}$ corresponding to various levels of regularity. For simplicity of the notation we will assume that $S(t)$ is an analytic semigroup of *negative type* $\omega_0 < 0$, so that we can always take the sector to be $S_{0,\theta}$ and let $L$ be it's sectorial generator. As we've seen several times already, we can always extend to the general case by rescaling the semigroup by $e^{-(\omega_0+\epsilon)t}S(t)$ which corresponds to shifting the spectrum of the generator to the left by $\omega_0 + \epsilon$.

In this case since $\|S(t)\|$ decays exponentially, it is possible to define the inverse of $-L$ through the resolvent formula

$$(-L)^{-1} = \int_0^\infty S(t)\mathrm{d}t.$$

This motivates the following definition for negative fractional powers of $-L$ by

$$(-L)^{-\alpha} := \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1} S(t)\mathrm{d}t, \tag{5.8}$$

where $\alpha > 0$ and $\Gamma(\alpha)$ is the gamma function defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-\alpha t} t^{\alpha-1}\mathrm{d}t.$$

. This definition makes sense for any $\alpha > 0$ because of the exponential decay of $S(t)$.

**Exercise 5.12.** Show that (5.8) is equivalent to the following functional calculus definition

$$(-L)^{-\alpha} = \frac{1}{2\pi i} \int_{\gamma_{a,\eta}} (-\lambda)^{-\alpha} R(\lambda)\mathrm{d}\lambda,$$

where $\omega_0 < a < 0$ and $(-\lambda)^{-\alpha} = e^{-\alpha \log(-\lambda)}$ is defined by it's principle branch.

**Exercise 5.13.** Show that $((-L)^{-\alpha})_{\alpha>0}$ forms a semigroup. That is, for each $\alpha, \beta > 0$ we have the identity

$$(-L)^{-\alpha}(-L)^{-\beta} = (-L)^{-\alpha-\beta}$$

and that for each $\alpha > 0$, $(-L)^{-\alpha}$ is injective. (Hint: use change of variables and the fact that

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t,$$

holds for any $\alpha, \beta > 0$).

Since $(-L)^{-\alpha}$ is injective, we denote $(-L)^{\alpha}$ its inverse with domain $\mathrm{Dom}((-L)^{\alpha})$. It is not difficult to see that $(-L)^{\alpha}$ is closed with dense domain and that if $0 < \alpha < \beta$ then

$$\mathrm{Dom}((-L)^{\beta}) \subseteq \mathrm{Dom}((-L)^{\alpha})$$

and that the semigroup property still holds for $((-L)^{\alpha})_{\alpha \geq 0}$ on a suitable domain. Naturally this gives rise to the following family of subspaces associated to $L$.

**Definition 5.26.** Let $L$ be a sectorial operator with $\omega < 0$. For each $\alpha > 0$ define the *interpolation space* $\mathcal{U}^{\alpha} := \mathrm{Dom}((-L)^{\alpha})$ with norm

$$\|u\|_{\alpha} := \|(-L)^{\alpha}u\|,$$

and $\mathcal{U}^{-\alpha}$ the completion of $\mathcal{U}$ with respect to the norm

$$\|u\|_{-\alpha} := \|(-L)^{-\alpha}u\|.$$

**Example 5.27.** Let $\mathcal{U}$ be the Hilbert space $L^2(\mathbb{T}^d)$ and $L = \Delta$ (the Laplacian) so that $S(t)$ is the heat semi-group on $L^2$. Then the spaces $\mathcal{U}^\alpha$ correspond to the Sobolev spaces $H^\alpha(\mathbb{T}^d)$.

**Exercise 5.14.** Let $\alpha \in (0,1)$. Show that if $u \in \mathrm{Dom}(L)$ then

$$\|u\|_\alpha \lesssim_\alpha \|Lu\|^\alpha \|u\|^{1-\alpha}. \tag{5.9}$$

(Hint: show that

$$(-L)^\alpha u = \frac{1}{\Gamma(1-\alpha)} \int_0^\infty t^{-\alpha} LS(t)u\mathrm{d}t,$$

which is well defined by (5.7), and split the integral $\int_0^\infty = \int_0^T + \int_T^\infty$ and optimize in $T$.)

**Remark 5.28.** The inequality (5.9) is known as an *interpolation inequality* and shows that in some sense $\mathcal{U}^\alpha$ is and interpolation of $\mathrm{Dom}(L)$ and $\mathcal{U}$. Moreover, a straight forward generalization of the previous exercise shows that for $\alpha < \beta \in [0,1]$ we have for $u \in \mathcal{U}^\beta$,

$$\|u\|_\alpha \lesssim_{\alpha,\beta} \|u\|_\beta^{\alpha/\beta} \|u\|^{1-\alpha/\beta}.$$

It is also possible to show a generalization of (5.7) for $S(t)$ in $\mathcal{U}^\alpha$ showing that $S(t)$ instantaneously gains regularity for all $t > 0$.

**Proposition 5.29.** *Assume $S(t)$ is an analytic semigroup of negative type $\omega_0 < 0$, then for each $\alpha \geq 0$ there exists a constant $M_\alpha$ such that for $t > 0$ and $u \in \mathcal{U}$ we have $S(t)u \in \mathcal{U}^\alpha$ and*

$$\|S(t)u\|_\alpha \leq \frac{M_\alpha}{t^\alpha} \|u\|. \tag{5.10}$$

*Proof.* It suffices to prove this for $\alpha \in [0,1]$. Since we can use the semi-group property to iterate the estimate to obtain any $\alpha \geq 0$. In this case we write

$$(-L)^\alpha = (-L)^{\alpha-1}(-L)$$

and therefore using the fact that $S(t)$ commutes with any power of the generator, we find

$$(-L)^\alpha S(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^\infty s^{-\alpha} LS(t+s)\mathrm{d}s.$$

Using (5.7), we find for $\omega_0 < \omega < 0$,

$$\|(-L)^\alpha S(t)\| \lesssim \int_0^\infty s^{-\alpha} \frac{e^{-\omega(t+s)}}{t+s} \mathrm{d}s = t^{-\alpha} \int_0^\infty \frac{s^{-\alpha}}{1+s} \mathrm{d}s,$$

where in the last equality we made the substitution $s \to ts$, and we are done since $\alpha < 1$. $\square$

It will also be useful to understand the time regularity of $S(t)u$ when $u \in \mathcal{U}^\alpha$. Indeed we know that when $u \in \mathrm{Dom}(L)$, then $S(t)u$ is in $C^1(\mathcal{U})$. In general, we will see that there is close connection between $u \in \mathcal{U}^\alpha$, estimate and Hölder regularity of $t \mapsto S(t)u$.

**Proposition 5.30** (Hölder time-regularity)**.** *Let $S(t)$ be an analytic semi-group of negative type with generator $L$, then for each $\alpha \in (0,1)$ and $s,t \in [0,1]$ we have*

$$\|S(t)u - S(s)u\| \lesssim_\alpha |t-s|^\alpha \|u\|_\alpha.$$

*and therefore $t \mapsto S(t)u$ is Hölder continuous in time, namely*

$$S(\cdot)u \in C^\alpha([0,1];\mathcal{U}).$$

*Proof.* It suffices to take $s = 0$ since if $t > s$ we have $S(t) - S(s) = S(t)(S(t-s) - \mathrm{Id})$ and $\|S(t)\| \lesssim 1$ on $[0,1]$. By density of $\mathrm{Dom}(L) \subseteq \mathcal{U}_\alpha$ it suffices to show this for each $u \in \mathrm{Dom}(L)$. Note that for such $u$ and $t > 0$, we have by (5.10) that

$$\|S(t)Lu\| = \|(-L)^{1-\alpha}S(t)(-L)^{\alpha}u\| \le \|(-L)^{1-\alpha}S(t)\|\|u\|_\alpha \lesssim t^{\alpha-1}\|u\|_\alpha,$$

It follows that

$$\|S(t)u - u\| = \left\| \int_0^t S(s)Lu\mathrm{d}s \right\| \lesssim \left( \int_0^t s^{1-\alpha}\mathrm{d}s \right) \|u\|_\alpha \lesssim_\alpha t^\alpha \|u\|_\alpha.$$

□

**Remark 5.31.** In general it is not true that $u \in \mathcal{U}_\alpha$ are the values for which $t \mapsto S(t)u$ is Hölder continuous. In the theory of semi-groups, it is natural to define the *Favard space* $F_\alpha$ of initial $u$ leading to Hölder continuous paths with norm

$$\|u\|_{F_\alpha} = \sup_{t>0} t^{-\alpha}\|S(t)u - u\| < \infty,$$

so that Proposition (5.30) is equivalent to the bound

$$\|u\|_{F_\alpha} \lesssim \|u\|_\alpha.$$

# 6 Linear evolution equations with additive noise

We are now equipped with the tools necessary to start studying stochasic PDE. In this section we will study the simplest case, namely linear equations with additive noise of the form

$$\mathrm{d}u = Lu\,\mathrm{d}t + B\mathrm{d}W(t), \quad u(0) = u \in \mathcal{U} \tag{6.1}$$

where $W(t)$ is a cylindrial Wiener process in some Hilbert space $\mathcal{W}$, $L$ is the generator of a $C_0$ semigroup on $\mathcal{U}$, and $B : \mathcal{W} \to \mathcal{U}$ is a bounded operator.

Note that immediatly, there are issues with interpreting this equation. Indeed the differential 'd' notation in stochastic analysis is short for the time-integrated equation

$$u(t) = u + \int_0^t Lu(s)\mathrm{d}s + BW(t).$$

However, just as we encountered with deterministic evolutions equations, $u(s)$ may not take values in $\mathrm{Dom}(L)$ unless $u \in \mathrm{Dom}(L)$ so that $Lu(s)$ may not be defined (unless of course $L$ generates and analytic semi-group). Additionally, since we have only assumed $B$ to be bounded and $W(t)$ is a cylindrical Wiener process, we don't expect $BW(t)$ to take values in $\mathcal{U}$.

In order to get around some of these difficulties, it is convenient to weaken our notion of solution. We can do this be pairing both sides of the equation with a suitable "test function" $\ell \in \mathrm{Dom}(L^*)$ and removing some of the regularity requirements on $u(t)$. Specifically we introduce the following notion of weak solution.

**Definition 6.1** (Weak solution)**.** A $\mathcal{U}$-valued process $\{u(t) : t \in \mathbb{R}_+\}$ is said to be a *weak solution* to (6.1) if $t \mapsto u(t)$ is almost surely locally Bochner integrable and for every $\ell \in \mathrm{Dom}(L^*)$, the following identity holds $\mathbf{P}$ almost surely for each $t \in \mathbb{R}_+$

$$\langle \ell, u(t) \rangle = \langle \ell, u \rangle + \int_0^t \langle L^*\ell, u(s) \rangle \mathrm{d}s + \langle B^*\ell, W(s) \rangle_{\mathcal{W}}. \tag{6.2}$$

In the above identity (6.2) $\langle \cdot, \cdot \rangle$ (without any subscript) denotes the dual pairing between $\mathcal{U}'$ and $\mathcal{U}$ and the local time-integrability is required to make sense of the time integral in (6.2). Additionally, the mapping

$$w \mapsto \langle B^* \ell, w \rangle_{\mathcal{W}}$$

is Hilbert-Schmidt from $\mathcal{W} \to \mathbb{R}$, since for any orthonormal basis $\{e_k\}$ of $\mathcal{W}$

$$\sum_k |\langle B^* \ell, e_k \rangle_{\mathcal{W}}|^2 = \|B^* \ell\|_{\mathcal{W}}^2 < \infty.$$

Therefore by Corollary 3.55, we can view $\langle B^* \ell, W(s) \rangle_{\mathcal{W}}$ in terms of a measurable linear extension so that

$$\langle B^* \ell, W(s) \rangle_{\mathcal{W}} = \sum_{k \in \mathbb{N}} \langle B^* \ell, e_k \rangle_{\mathcal{W}} W^k(t)$$

for a sequence of iid standard Wiener processes $\{W_k(t)\}$.

**Remark 6.2.** The terminology *weak solution* means weak in the analytic (or PDE) sense. This should not be confused with the notion of a probablistically weak solution in the theory of stochastic differential equations. This terminology collision is an unfortunate consequence of the meeting of PDE and SDE. As a result probabilistically weak solutions for stochastic PDE are often refered to as *martingale solutions* to avoid confusion.

In this linear setting, we can view $BW(t)$ as an inhomogeneity for the linear evolution equation. For instance, if $f(t)$ is a function that takes values in $\mathrm{Dom}(L)$, then the deterministic linear equation

$$\partial_t u(t) = Lu(t) + f(t) \quad u(0) = u \tag{6.3}$$

can be solved explicitly when $u \in \mathrm{Dom}(L)$ via *Duhamel's formula*

$$u(t) = S(t)u + \int_0^t S(t-s) f(s) \mathrm{d}s.$$

Indeed since $u, f(s) \in \mathrm{Dom}(L)$, we can take the time derivative and find that

$$\partial_t u(t) = LS(t)u + \int_0^t LS(t-s) f(s) \mathrm{d}s + f(s) = Lu(t) + f(s).$$

Of course the formula for $u(t)$ doesn't care if $u$ or $f(s)$ don't take values in $\mathrm{Dom}(L)$ and therefore gives an appropriate generalized notion of solution to (6.3) for any initial data $u$ and $f(t) \in \mathcal{U}$, called a *mild solution*.

This strategy can also be applied to the stochastic setting (6.1) and motivates the following explicit formula

$$u(t) = S(t)u + \int_0^t S(t-s) B \mathrm{d}W(s). \tag{6.4}$$

A solution to (6.1) of the form (6.4) is sometimes called a *mild solution*. The process

$$W_L(t) := \int_0^t S(t-s) B \mathrm{d}W(s)$$

on the right-hand side of 6.4 is called the *stochastic convolution* and plays a very important role in both linear and non-linear theory.

**Remark 6.3.** From our theory of the stochastic integral, we can see that $W_L(t)$ is well-defined if we can find a Hilbert space $\mathcal{H}$ that $\mathcal{U} \subseteq \mathcal{H}$ continuously embeds into with embedding $J : \mathcal{U} \to \mathcal{H}$ and for each $t > 0$ $JS(t)B$ is a Hilbert-Schmidt mapping from $\mathcal{W} \to \mathcal{H}$ and satisfies for each $t > 0$

$$\int_0^t \|JS(s)B\|^2_{\mathcal{L}^2(\mathcal{W},\mathcal{U})} \mathrm{d}s < \infty.$$

However, at the level of generality we are dealing with it is not clear what conditions are sufficient for $\int_0^T \|u(t)\| \mathrm{d}t < \infty$ almost surely since we do not have an Itô isometry in Banach spaces. None-the-less this question can be answered when $\mathcal{U}$ is a Banach space of continuous functions (see [DPZ14] 5.5).

We can now show in what sense (6.4) solves equation (6.1).

**Proposition 6.4.** *Let $u(t)$ be given by (6.4), such that $\int_0^T \|u(s)\| \mathrm{d}s < \infty$ almost surely, then $u(t)$ is a weak solution to (6.1).*

*Proof.* The proof uses a duality method and the dual semi-group $S^*(t)$. Note that using the fact that $S(t)u$ solves

$$\langle \ell, S(t)u \rangle = \int_0^t \langle L^*\ell, S(s)u \rangle \mathrm{d}s,$$

it suffices to show that $W_L(t)$ is a weak solution with zero initial data. To do this we choose $\ell \in \mathrm{Dom}(L^\dagger)$, and note that by stochastic Fubini, we have

$$\int_0^t \langle L^*\ell, W_L(s) \rangle \mathrm{d}s = \int_0^t \int_0^s \langle L^*\ell, S(s-r)B\mathrm{d}W(r) \rangle \mathrm{d}s = \int_0^t \left\langle \int_r^t S^*(s-r)L^*\ell \mathrm{d}s, B\mathrm{d}W(r) \right\rangle$$

Since $S^*(t)$ is a strongly continuous semi-group on $\mathcal{U}^\dagger$ and $\ell \in \mathrm{Dom}(L^\dagger)$ we find

$$\int_r^t S^*(t-r)L^*\ell \mathrm{d}s = \int_0^{t-r} S^*(s)L^*\ell \mathrm{d}s = S^*(t-r)\ell - \ell$$

and therefore

$$\begin{aligned} \int_0^t \langle L^*\ell, W_L(s) \rangle \mathrm{d}s &= \int_0^t \langle S^*(t-r)\ell, B\mathrm{d}W(r) \rangle - \int_0^t \langle \ell, B\mathrm{d}W(r) \rangle \\ &= \langle \ell, W_L(t) \rangle - \langle \ell, BW(r) \rangle. \end{aligned}$$

Thus $W_L(t)$ is a weak solution for each $\ell \in \mathrm{Dom}(L^\dagger)$. To extend this to all $\ell \in \mathrm{Dom}(L^*)$ we follow as similar proof to that of Lemma 5.15. Let $\ell \in \mathrm{Dom}(L^*)$ and define $\ell_n = nR^*(n)\ell$. It follows that $L^*\ell_n = nR^*(n)L^*\ell \in \mathrm{Dom}(L^*) \subseteq \mathcal{U}^\dagger$ and therefore $\ell_n \in \mathrm{Dom}(L^\dagger)$ (by the definition of $\mathrm{Dom}(L^\dagger)$). As in Lemma5.15 we have $\ell_n \to \ell$ weakly-*. Moreover we find that for each $u \in \mathcal{U}$

$$|\langle L^*\ell_n - L^*\ell, u \rangle| = |\langle L^*\ell, nR(n)u - u \rangle| \le \|L^*\ell\|_* \|nR(u) - u\| \to 0,$$

since $\mathrm{Dom}(L)$ is dense in $\mathcal{U}$. Therefore $L^*\ell_n \to L^*\ell$ weakly-*. Choosing $\ell_n$ as an approximating sequence and passing to the limit gives the result. □

Moreover, we can show that every weak solution must be given by (6.4). To do this, we first need the following Lemma.

**Lemma 6.5.** *Let $u(t)$ be weak solution to (6.1), then for each $\psi(\cdot) \in C^1([0,T]; \mathrm{Dom}(L^*))$ we have*

$$\langle \psi(t), u(t) \rangle = \langle \psi(0), u \rangle + \int_0^t \langle \partial_s \psi(s) + L^*\psi(s), u(s) \rangle \mathrm{d}s + \int_0^t \langle \psi(s), B\mathrm{d}W(s) \rangle.$$

*Proof.* It suffices to prove this for functions $\psi(t) = \varphi(t)\ell$, where $\ell \in \mathrm{Dom}(L^*)$ and $\varphi(t)$ is $C^1$ since such functions are linearly dense in $C^1([0,T];\mathrm{Dom}(L^*))$. Applying Itô's formula to $\varphi(t)\langle \ell, u(t)\rangle$ (note that this is real-valued process), we obtain

$$\varphi(t)\langle \ell, u(t)\rangle = \varphi(0)\langle \ell, u\rangle + \int_0^t \langle \partial_s\varphi(s)\ell + \varphi(s)L^*\ell, u(s)\rangle \mathrm{d}s + \int_0^t \varphi(s)\langle \ell, B\mathrm{d}W(s)\rangle,$$

which is the desired formula. □

We are now ready to prove the uniqueness property for (6.1).

**Proposition 6.6.** *Every weak solution takes the form* (6.4).

*Proof.* We again use a duality approach and for a fixed $\ell \in \mathrm{Dom}(L^\dagger)$, let $\psi(s) = S^*(t-s)\ell$ be a backwards in time solution to the adjoint problem. Since $S^*(t)$ is $C_0$ semi-group on $\mathcal{U}^\dagger$ and $\mathrm{Dom}(L^\dagger)$ is the domain of it's generator, we see that $\psi(\cdot) \in C^1([0,t];\mathrm{Dom}(L^\dagger))$ and satisfies the backwards evolution equation

$$\partial_s\psi(s) + L^*\psi(s) = 0, \quad \psi(0) = S^*(t)\ell, \quad \psi(t) = \ell$$

Therefore applying Lemma (6.5), we find

$$\begin{aligned}\langle \ell, u(t)\rangle &= \langle S^*(t)\ell, u\rangle + \int_0^t \langle \partial_s\psi(s) + L^*\psi(s), u(s)\rangle \mathrm{d}s + \int_0^t \langle S^*(t-s)\ell, B\mathrm{d}W(s)\rangle \\ &= \langle \ell, S(t)u\rangle + \left\langle \ell, \int_0^t S(t-s)\mathrm{d}W(s)\right\rangle.\end{aligned}$$

Using the fact that $\mathrm{Dom}(L^\dagger)$ is dense in $\mathcal{U}^\dagger$ and that $\mathcal{U}^\dagger$ is weak-* dense in $\mathcal{U}^*$, we conclude the proof. □

Note that we are just shy of an existence and uniqueness theorem in $\mathcal{U}$ since we have not identified a condition under which $W_L(t)$ is locally Bochner integrable in time and therefore we don't know that every term in (6.4) actually makes sense as a weak solution. Indeed, this is non-trivial at the level general abstract Banach spaces. However when $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space by Itô's Isometry and Fubini, we have

$$\mathbf{E}\int_0^T \|JW_L(t)\|^2 \mathrm{d}t = \int_0^T \int_0^t \|S(s)B\|^2_{\mathcal{L}_2(\mathcal{W},\mathcal{U})}\mathrm{d}t < \infty,$$

which implies that if we have for each $t > 0$

$$\int_0^t \|S(s)B\|^2_{\mathcal{L}_2(\mathcal{W},\mathcal{H})}\mathrm{d}s < \infty$$

then any mild solution $u(t)$ is well-defined and locally Bochner integrable in time This gives the following existence and uniqueness corollary in the Hilbert space case.

**Corollary 6.7.** *Let $\mathcal{U} = \mathcal{H}$ be a separable Hilbert space, then* (6.1) *has a unique weak solution given by* (6.4).

## 6.1 Regularity of the stochastic convolution in Hilbert spaces

In what follows, we will restrict our attention to the case when $\mathcal{U} = \mathcal{H}$ is a separable Hilbert space. Since much of this relies on Itô's formula, we In $\mathcal{H}$, $W_A(t)$ inherits many properties from the stochastic integral.

**Theorem 6.8.** *Suppose that*

$$\int_0^t \|S(s)B\|^2_{\mathcal{L}^2(\mathcal{W},\mathcal{H})} \mathrm{d}s < \infty.$$

*Then $W_L(t)$ is mean-square continuous and adapted to $\mathscr{F}_t$ and the trajectories $t \mapsto JW_L(t)$ are almost surely square integrable with $\mathrm{Law}(W_L(\cdot))$ a centered Gaussian measure on $L^2([0,T];\mathcal{H})$ with covariance*

$$C(t,s) = \int_0^{\min\{s,t\}} S(t-r)BB^*S(t-r)\mathrm{d}r.$$

*Proof.* The proof is an easy exercise using the properties of the stochastic integral. See [DPZ14] Theorem 5.2. □

## References


[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[BL74] J K Brooks and P W Lewis. Linear operators and vector measures. *Trans. Amer. Math. Soc.*, 192:139–139, 1974.

[Bog98] Vladimir Igorevich Bogachev. *Gaussian Measures*. American Mathematical Soc., 1998.

[DPZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, April 2014.

[EN99] Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer Science & Business Media, October 1999.

[Fer71] Xavier Fernique. Régularité de processus gaussiens. *Invent. Math.*, 12(4):304–320, December 1971.

[Fol99] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[Gro67] Leonard Gross. Abstract wiener spaces. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*. The Regents of the University of California, 1967.

[GRRR70] A M Garsia, E Rodemich, H Rumsey, and M Rosenblatt. A real variable lemma and the continuity of paths of some gaussian processes. *Indiana Univ. Math. J.*, 20(6):565–578, 1970.

[HvNVW16] Tuomas Hytönen, Jan van Neerven, Mark Veraar, and Lutz Weis. *Analysis in Banach Spaces : Volume I: Martingales and Littlewood-Paley Theory*. Springer, Cham, 2016.

[Jak97] A. Jakubowski. The almost sure Skorokhod representation for subsequences in nonmetric spaces. *Teor. Veroyatnost. i Primenen.*, 42(1):209–216, 1997.

[KS91] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media, 1991.

[OS13] Martin Ondreját and Jan Seidler. On existence of progressively measurable modifications. *Electron. Commun. Probab.*, 18:no. 20, 6, 2013.

[Roy88] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[Tan79] Hiroki Tanabe. *Equations of evolution*, volume 6 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1979. Translated from the Japanese by N. Mugibayashi and H. Haneda.

[Yos95] Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.