An Introduction to Stochastic PDEs

Martin Hairer

Contents

1

Foreword

These notes are based on a postgraduate course given at the University of Warwick in spring 2008 and 2011, at the Courant Institute in spring 2009, and at Imperial College London in autumn 2018. It is an attempt to give a reasonably self-contained presentation of the basic theory of stochastic partial differential equations, taking for granted elementary measure theory, functional analysis and probability theory, but nothing else. Since the aim was to present most of the material covered in these notes during a 30-hours series of postgraduate lectures, such an attempt is doomed to failure unless drastic choices are made. This is why many important facets of the theory of stochastic PDEs are missing from these notes. In particular, we treat equations with multiplicative noise only superficially, we do *not* treat equations driven by Lévy noise, we do *not* consider equations with "rough" (that is not locally Lipschitz, even in a suitable space) nonlinearities, we do *not* treat measure-valued processes, we do *not* consider hyperbolic or elliptic problems, we do *not* cover Malliavin calculus and densities of solutions, etc. The reader who is interested in a more detailed exposition of some of these more advanced aspects of the theory is advised to read the excellent works [DPZ92b, DPZ96, PZ07, PR07, SS05] or to dive into the current literature. These notes also do not cover at all the various recently developed approaches to treating *singular SPDEs*, namely SPDEs that require "renormalisation" in order to make sense. The reader interested in this topic is referred to the articles [Hai14, GIP15, Hai18].

Instead, the approach taken in these notes is to focus on semilinear *parabolic* problems driven by *additive* noise. These can be treated as stochastic evolution equations in some infinite-dimensional Banach or Hilbert space that usually have nice regularising properties and they already form (in my humble opinion) a very rich class of problems with many interesting properties. Furthermore, this class of problems has the advantage of allowing to completely pass under silence many subtle technical problems (both functional analytical and measure theoretical) arising from stochastic integration in infinite-dimensional spaces and from the general theory of stochastic processes.

1.1 Acknowledgements

These notes would never have been completed, were it not for the enthusiasm of the attendants of the course. Many dozens of typos and mistakes were spotted and corrected. I am particularly indebted to David Epstein and Jochen Voß who carefully worked their way through a very preliminary version of these notes when they were still in a state of wilderness. Special thanks are also due to Pavel Bubak and Hendrik Weber who organised the tutorials for the courses given at the University of Warwick.

2

Some Motivating Examples

In this chapter, we will discuss two situations in which stochastic PDEs of the type discussed in these notes arise naturally. We will then take the simplest of these examples, the stochastic heat equation, and perform a few explicit calculations to start to develop a "gut feeling" for the behaviour of the solutions of these objects.

2.1 A model for a random string (polymer)

Our first motivating example is a toy model for the evolution of a polymer in a fluid. For simplicity, we model the polymer by a chain of particles with harmonic nearest-neighbour interactions. In other words, we consider $N+1$ particles with positions u_n and mass m immersed in a fluid and assume that nearest-neighbours are connected by harmonic springs. If the particles are furthermore subject to an external force field F, their equations of motion are given by

$$\begin{aligned}
m\frac{d^2u_0}{dt^2} &= k(u_1 - u_0) - \frac{du_0}{dt} + F(u_0)\;,\\
m\frac{d^2u_n}{dt^2} &= k(u_{n+1} + u_{n-1} - 2u_n) - \frac{du_n}{dt} + F(u_n)\;, \quad n = 1, \dots, N-1\;,\\
m\frac{d^2u_N}{dt^2} &= k(u_{N-1} - u_N) - \frac{du_N}{dt} + F(u_N)\;.
\end{aligned}$$

Here, the terms $-\frac{du_n}{dt}$ are used to model the friction caused by the interaction between the particles and the fluid.

At this stage, we simplify the model even further by assuming that we are in the overdamped regime where the forces acting on the particle are more important than inertia. Formally, we obtain this by setting the masses of the particles to zero, so that our equations of motion become

$$\begin{aligned}
\frac{du_0}{dt} &= k(u_1 - u_0) + F(u_0)\;,\\
\frac{du_n}{dt} &= k(u_{n+1} + u_{n-1} - 2u_n) + F(u_n)\;, \quad n = 1, \dots, N-1\;,\\
\frac{du_N}{dt} &= k(u_{N-1} - u_N) + F(u_N)\;.
\end{aligned}$$

This model does however not take into account the effect of the molecules of water that would randomly "kick" the particles that make up our polymer. Assuming that these kicks occur randomly and independently at high rate, this effect can be modelled in first instance by independent white noises acting on all degrees of freedom of our model. We thus obtain a system of coupled stochastic differential equations:

$$\begin{aligned}
du_0 &= k(u_1 - u_0)\,dt + F(u_0)\,dt + \sigma\,dw_0(t)\;,\\
du_n &= k(u_{n+1} + u_{n-1} - 2u_n)\,dt + F(u_n)\,dt + \sigma\,dw_n(t)\;, \quad n = 1, \dots, N-1\;,\\
du_N &= k(u_{N-1} - u_N)\,dt + F(u_N)\,dt + \sigma\,dw_N(t)\;,
\end{aligned}$$

where the w_n are independent Wiener processes. At a formal level at least, this means that dw_n/dt are independent centred Gaussian "processes" (actually distributions) such that

$$\mathbf{E}\frac{dw_n(t)}{dt}\frac{dw_m(s)}{ds} = \delta_{m,n}\delta(t-s)\;, \tag{2.1}$$

where δ denotes the Dirac distribution.

We would like to take a continuum limit for this process. Setting $n \approx Nx$ with $x \in [0,1]$, we see that the interaction between nearest neighbours converges to a second derivative with respect to x, provided that we choose k such that $k \approx \nu N^2$. It is less obvious to guess what is the correct scaling for σ to obtain a non-trivial limit. A hint can be taken from (2.1): it seems plausible to look for a scaling such that, in the continuous limit, our equation is driven by a "process" W such that $\mathbf{E}\frac{dW(x,t)}{dt}\frac{dW(y,s)}{ds} = \delta(x-y)\delta(t-s)$. Since, with our scaling, an integral over x is approximated by N^{-1} times a sum over n, we conclude from (2.1) that the correct scaling to consider is $\sigma \approx \sqrt{N}$.

This heuristic discussion suggests that, if N is very large and we choose $k \approx \nu N^2$ and $\sigma \approx \sqrt{N}$, our toy model is well described by the solution to a stochastic *partial differential equation*

$$du(x,t) = \nu\partial_x^2 u(x,t)\,dt + F(u(x,t))\,dt + dW(x,t)\;, \tag{2.2}$$

endowed with the boundary conditions $\partial_x u(0,t) = \partial_x u(1,t) = 0$ and such that, at least on a formal level, $\mathbf{E}\frac{dW(x,t)}{dt}\frac{dW(y,s)}{ds} = \delta(x-y)\delta(t-s)$. The precise meaning of this formula will be discussed in Chapter 5.

2.2 The stochastic Navier-Stokes equations

The Navier-Stokes equations describing the evolution of the velocity field $u(x,t)$ of an incompressible viscous fluid are given by

$$\frac{du}{dt} = \nu\Delta u - (u\cdot\nabla)u - \nabla p + f\;, \tag{2.3}$$

complemented with the (algebraic) incompressibility condition $\operatorname{div} u = 0$. Here, f denotes some external force acting on the fluid, whereas the pressure p is given implicitly by the requirement that $\operatorname{div} u = 0$ at all times.

While it is not too difficult in general to show that solutions to (2.3) exist in some weak sense, in the case where $x \in \mathbf{R}^d$ with $d \geq 3$, their *uniqueness* (globally in time) is an extremely hard open problem that has been endowed with a $1,000,000 prize by the Clay Mathematics Institute. We will of course not attempt to solve this long-standing problem here, so we are going to restrict ourselves to the case $d = 2$. (The case $d = 1$ makes no sense since there the condition $\operatorname{div} u = 0$ would imply that u is constant. However, one could also consider Burgers' equation which has similar features to the Navier-Stokes equations.)

For simplicity, we consider solutions that are periodic in space, so that we view u as a function from $\mathbf{T}^2 \times \mathbf{R}_+$ to $\mathbf{R}^2$. In the absence of external forcing f, one can use the incompressibility assumption to see that

$$\frac{d}{dt}\int_{\mathbf{T}^2}|u(x,t)|^2\,dx = -2\nu\int_{\mathbf{T}^2}\operatorname{tr} Du(x,t)^* Du(x,t)\,dx \leq -2\nu\int_{\mathbf{T}^2}|u(x,t)|^2\,dx\;,$$

where we used the Poincaré inequality to obtain the last bound (assuming that $\int_{\mathbf{T}^2} u(x,t)\,dx = 0$). Therefore, by Gronwall's inequality, the solutions decay to 0 exponentially fast. This shows that energy needs to be pumped into the system continuously if one wishes to maintain an interesting regime.

One way to achieve this from a mathematical point of view is to add a force f that is randomly fluctuating. We are going to show in Chapter 6 that if one takes a random force that is Gaussian and such that

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

for some correlation function C then, provided that C is sufficiently regular, one can show that (2.3) has solutions for all times. Furthermore, these solutions do not blow up in the sense that one can find a constant K such that, for any solution to (2.3), one has

$$\limsup_{t\to\infty} \mathbf{E}\|u(\,\cdot\,,t)\|^2 \le K\ ,$$

for some suitable norm $\|\,\cdot\,\|$. This allows to provide a construction of a model for homogeneous turbulence which is amenable to mathematical analysis.

2.3 The stochastic heat equation

In this section, we focus on the particular example of the stochastic heat equation, namely (2.2) with $F = 0$. We will perform a number of calculations that give us a feeling for what the solutions to this equation look like. These calculations will not be completely rigorous but could be made so with minimal additional effort. Most tools required to make them rigorous will be introduced later on.

2.3.1 Setup

Recall that the *heat equation* is the partial differential equation:

$$\partial_t u = \Delta u\ , \quad u\colon \mathbf{R}^n \times \mathbf{R}_+ \to \mathbf{R}\ . \tag{2.4}$$

Given any bounded continuous initial condition $u_0\colon \mathbf{R}^n \to \mathbf{R}$, there exists a unique solution u to (2.4) which is continuous on $\mathbf{R}^n \times \mathbf{R}_+$ and such that $u(x,0) = u_0(x)$ for every $x \in \mathbf{R}^n$.

It is a tedious but straightforward calculation to check that this solution is given by the formula

$$u(x,t) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbf{R}^n} e^{-\frac{|x-y|^2}{4t}} u_0(y)\,dy\ .$$

We will denote this by the shorthand $u(\,\cdot\,,t) = e^{\Delta t}u_0$ by analogy with the solution to an $\mathbf{R}^d$-valued linear equation of the type $\partial_t u = Au$.

Let us now go one level up in difficulty by considering (2.4) with an additional "forcing term" f:

$$\partial_t u = \Delta u + f\ , \quad u\colon \mathbf{R}^n \times \mathbf{R}_+ \to \mathbf{R}\ . \tag{2.5}$$

From the variations of constants formula, we obtain that the solution to (2.5) is given by

$$u(\,\cdot\,,t) = e^{\Delta t}u_0 + \int_0^t e^{\Delta(t-s)} f(\,\cdot\,,s)\,ds\ . \tag{2.6}$$

Since the kernel defining $e^{\Delta t}$ is very smooth, this expression actually makes sense for a large class of distributions f, in particular it still makes sense if f is "space-time white noise". We do not define this rigorously for the moment, but characterise it as a (distribution-valued) centred Gaussian process ξ such that $\mathbf{E}\xi(x,s)\xi(y,t) = \delta(t-s)\delta(x-y)$.

The stochastic heat equation is then the stochastic partial differential equation

$$\partial_t u = \Delta u + \xi\ , \quad u\colon \mathbf{R}_+ \times \mathbf{R}^n \to \mathbf{R}\ . \tag{2.7}$$

Consider the simplest case $u_0 = 0$, so that its solution is given from (2.6) by

$$u(t,x) = \int_0^t \frac{1}{(4\pi|t-s|)^{n/2}} \int_{\mathbf{R}^n} e^{-\frac{|x-y|^2}{4(t-s)}} \xi(y,s)\,dy\,ds\ . \tag{2.8}$$

This is again a centred Gaussian process, but its covariance function is more complicated. The aim of this section is to get some idea about the space-time regularity properties of (2.8). While the solutions to ordinary stochastic differential equations are in general α-Hölder continuous (in time) for every $\alpha < 1/2$ but not for $\alpha = 1/2$, we will see that in dimension $n = 1$, u as given by (2.8) is only "almost"

$1/4$-Hölder continuous in time and "almost" $1/2$-Hölder continuous in space. In higher dimensions, it is not even function-valued... The reason for this lower time-regularity is that the driving space-time white noise is not only very singular as a function of time, but also as a function of space. Therefore, some of the regularising effect of the heat equation is required to turn it into a continuous function in space.

Heuristically, the appearance of the Hölder exponents $1/2$ for space and $1/4$ for time in dimension $n = 1$ can be understood by the following argument. If we were to remove the term $\partial_t u$ in (2.7), then u would have the same time-regularity as ξ, but two more derivatives of space regularity. If on the other hand we were to remove the term Δu, then u would have the same space regularity as ξ, but one more derivative of time regularity. The consequence of keeping both terms is that we can "trade" space-regularity against time-regularity at a cost of one time derivative for two space derivatives. Now we know that white noise (that is the centred Gaussian process η with $\mathbf{E}\eta(t)\eta(s) = \delta(t-s)$) is the time derivative of Brownian motion, which itself is "almost" $1/2$-Hölder continuous. Therefore, the regularity of η requires "a bit more than half a derivative" of improvement if we wish to obtain a continuous function.

Turning back to ξ, we see that it is expected to behave like η both in the space direction and in the time direction. So, in order to turn it into a continuous function of time, roughly half of a time derivative is required. This leaves a bit less than half of a time derivative, which we trade against one spatial derivative, thus concluding that for fixed time, u is expected to be almost $1/2$-Hölder continuous in space. Concerning the time regularity, we note that half of a space derivative is required to turn ξ into a continuous function of space, thus leaving one and a half space derivative. These can be traded against $3/4$ of a time derivative, thus explaining the $1/4$-Hölder continuity in time.

In Section 5.1, we are going to see more precisely how the space-regularity and the time-regularity interplay in the solutions to linear SPDEs, thus allowing us to justify rigorously this type of heuristic arguments. For example, one can interpret Theorem 5.18 as stating that a fraction δ of a time derivative can be traded against an index δ in the scale of interpolation spaces corresponding to the linear operator. Since in our case this linear operator is the Laplacian, an index of one in the corresponding interpolation spaces corresponds precisely to two spatial derivatives. For the moment however, let us justify these statements on the regularity of the solution by an explicit calculation in the particular case of the stochastic heat equation.

2.3.2 A formal calculation

Define the covariance for the solution to the stochastic heat equation by

$$C(s,t,x,y) = \mathbf{E}u(x,s)u(y,t)\;, \tag{2.9}$$

where u is given by (2.8).

By (statistical) translation invariance, it is clear that $C(s,t,x,y) = C(s,t,0,x-y)$. Using (2.8) and the expression for the covariance of ξ, one has

$$\begin{aligned}
&C(s,t,0,x)\\
&\quad= \frac{1}{(4\pi)^n}\mathbf{E}\int_0^t\int_0^s\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \frac{1}{|s-r'|^{n/2}|t-r|^{n/2}} e^{-\frac{|x-y|^2}{4(t-r)}-\frac{|y'|^2}{4(s-r')}}\xi(y,r)\xi(y',r')\,dy\,dy'\,dr'\,dr\\
&\quad= \frac{1}{(4\pi)^n}\int_0^{s\wedge t}\int_{\mathbf{R}^n} \frac{1}{|s-r|^{n/2}|t-r|^{n/2}} e^{-\frac{|x-y|^2}{4(t-r)}-\frac{|y|^2}{4(s-r)}}\,dy\,dr\\
&\quad= \frac{1}{(4\pi)^n}\int_0^{s\wedge t}\int_{\mathbf{R}^n} \frac{1}{|s-r|^{n/2}|t-r|^{n/2}}\\
&\qquad\times \exp\Big(-\frac{|x|^2}{4(t-r)}-\frac{\langle x,y\rangle}{2(t-r)}-\frac{|y|^2}{4(s-r)}-\frac{|y|^2}{4(t-r)}\Big)\,dy\,dr\;.
\end{aligned}$$

The integral over y can be performed explicitly by "completing the square" and one obtains

$$C(s,t,0,x) = \frac{1}{(4\pi)^{n/2}} \int_0^{s\wedge t} (s+t-2r)^{-n/2} \exp\Big(-\frac{|x|^2}{4(s+t-2r)}\Big)\, dr$$
$$= \frac{1}{2(4\pi)^{n/2}} \int_{|s-t|}^{s+t} \ell^{-n/2} \exp\Big(-\frac{|x|^2}{4\ell}\Big)\, d\ell\ . \tag{2.10}$$

We notice that the singularity at $\ell = 0$ is integrable if and only if $n < 2$, so that $C(t,t,0,0)$ is finite only in the one-dimensional case. We therefore limit ourselves to this case in the sequel.

Remark 2.1 Even though the random variable u defined by (2.8) is not function-valued in dimension 2, it is "almost" the case since the singularity in (2.10) diverges only logarithmically. The stationary solution to (2.7) is called the *Gaussian free field* and has been the object of intense studies over the last few years, especially in dimension 2. Its interest stems from the fact that many of its features are conformally invariant (as a consequence of the conformal invariance of the Laplacian), thus linking probability theory to quantum field theory on one hand and to complex geometry on the other hand. The Gaussian free field also relates directly to the Schramm-Loewner evolutions (SLEs) for the study of which Werner was awarded the Fields medal in 2006, see [Law04, SS06]. For more information on the Gaussian free field, see for example the review article by Sheffield [She07].

The regularity of u is determined by the behaviour of C near the "diagonal" $s = t$, $x = y$. We first consider the time regularity. We therefore set $x = 0$ and compute

$$C(s,t,0,0) = \frac{1}{4\sqrt{\pi}} \int_{|s-t|}^{s+t} \ell^{-1/2}\, d\ell = \frac{1}{2\sqrt{\pi}} (|s+t|^{\frac{1}{2}} - |s-t|^{\frac{1}{2}})\ .$$

This shows that, in the case $n = 1$ and for $s \approx t$, one has the asymptotic behaviour

$$\mathbf{E}|u(0,s) - u(0,t)|^2 \approx |t-s|^{\frac{1}{2}}\ .$$

Comparing this with the standard Brownian motion for which $\mathbf{E}|B(s) - B(t)|^2 = |t-s|$, we conclude that the process $t \mapsto u(x,t)$ is, for fixed x, almost surely α-Hölder continuous for any exponent $\alpha < 1/4$ but *not* for $\alpha = 1/4$. This argument is a special case of Kolmogorov's celebrated continuity test, of which we will see a version adapted to Gaussian measures in Section 3.3.

If, on the other hand, we fix $s = t$, we obtain (still in the case $n = 1$) via the change of variables $z = |x|^2/4\ell$, the expression

$$C(t,t,0,x) = \frac{|x|}{8\sqrt{\pi}} \int_{\frac{|x|^2}{8t}}^{\infty} z^{-\frac{3}{2}} e^{-z}\, dz\ .$$

Integrating by parts, we get

$$C(t,t,0,x) = \frac{\sqrt{t}}{\sqrt{2\pi}} e^{-\frac{|x|^2}{8t}} - \frac{|x|}{4\sqrt{\pi}} \int_{\frac{|x|^2}{8t}}^{\infty} z^{-\frac{1}{2}} e^{-z}\, dz\ ,$$

so that to leading order we have for small values of x:

$$C(t,t,0,x) \approx \sqrt{\frac{t}{2\pi}} - \frac{|x|}{4\sqrt{\pi}} \int_0^{\infty} z^{-\frac{1}{2}} e^{-z}\, dz = \sqrt{\frac{t}{2\pi}} - \frac{|x|}{4} + \mathcal{O}(|x|^2/\sqrt{t})\ .$$

This shows that, at any fixed instant t, the solution to (2.7) looks like a Brownian motion in space over lengthscales of order $t^{1/2}$ (this is the scale at which $|x|^2/\sqrt{t}$ becomes comparable to $|x|$). Note that over such a lengthscale the Brownian motion fluctuates by about $t^{1/4}$, which is precisely the order of magnitude of $\mathbf{E}|u(x,t)|$.

2.3.3 What have we learned?

Here are some conclusions that we can draw from these calculations:

1. At a "hand-waving" level, we have forced our equation with a term that has a temporal evolution resembling white noise, so that one would naively expect its solutions to have a temporal regularity resembling Brownian motion. However, for any fixed location in space, the solution to the stochastic heat equation has a time-regularity which is only almost Hölder-$\frac{1}{4}$, as opposed to the almost Hölder-$\frac{1}{2}$ time-regularity of Brownian motion. This is due to the fact that the forcing term is also very "rough" in space, and not just in time.
2. Unlike the solutions to an ordinary parabolic PDE, the solutions to a stochastic PDE tend to be spatially "rough". It is therefore not obvious *a priori* how the formal expression that we obtained is to be related to the original equation (2.7), since even for positive times, the map $x \mapsto u(t,x)$ is certainly not twice differentiable.
3. Even though the deterministic heat equation has the property that $e^{\Delta t}u \to 0$ as $t \to \infty$ for every $u \in L^2$, the solution to the stochastic heat equation has the property that $\mathbf{E}|u(x,t)|^2 \to \infty$ for fixed x as $t \to \infty$. This shows that in this particular case, the stochastic forcing term pumps energy into the system faster than the deterministic evolution can dissipate it.

Exercise 2.2 Perform the same calculation, but for the equation

$$\partial_t u = \Delta u - au + \xi\ , \quad u\colon \mathbf{R}_+ \times \mathbf{R} \to \mathbf{R}\ .$$

Show that as $t \to \infty$, the law of its solution converges to the law of an Ornstein-Uhlenbeck process (if the space variable is viewed as "time"):

$$\lim_{t\to\infty} \mathbf{E}u(x,t)u(y,t) = Ce^{-c|x-y|}\ .$$

Compute the constants C and c as functions of the parameter a.

3

Elements of Gaussian Measure Theory

While most of this chapter is devoted to Gaussian measure theory, we start with an introductory section on the general theory of probability measures on Polish (that is complete, separable, metric) spaces. This theory is of course much too vast to be done any sort of justice in these few short pages. We therefore refer the interested reader to the excellent and quite extensive treatise by Bogachev [Bog07] and to the much shorter but maybe more readily accessible book by Billingsley [Bil68] which still covers a large part of the material required for the basic study of stochastic PDEs.

3.1 Convergence of probability measures

The focus of this section will be mainly on the question of convergence of probability measures on Polish spaces. We will introduce a number of different topologies on the space of probability measures on an arbitrary Polish space and we will discuss the relations between these topologies and the metrics that generate them.

Recall that the Borel σ-algebra $\mathscr{B}(\mathcal{X})$ of a Polish space $\mathcal{X}$ is the smallest collection of sets containing all the open sets of $\mathcal{X}$ that is furthermore closed under countable unions and taking complements. A probability measure μ on $\mathcal{X}$ is then a map $\mu\colon \mathscr{B}(\mathcal{X}) \to [0,1]$ such that $\mu(\mathcal{X}) = 1$ and $\mu(\bigcup A_i) = \sum \mu(A_i)$ for any countable collection of mutually disjoint sets $A_i \in \mathscr{B}(\mathcal{X})$. We denote by $\mathscr{P}(\mathcal{X})$ the set of all (Borel) probability measures on $\mathcal{X}$. In order of increasing strength, here are the two main notions of convergence that will be used in this book:

- **Weak convergence:** A sequence of probability measures μ_n converges weakly to a limiting probability measure μ if $\lim_{n\to\infty} \int \varphi(x)\,\mu_n(dx) = \int \varphi(x)\,\mu(dx)$ for every bounded continuous function $\varphi\colon \mathcal{X} \to \mathbf{R}$.
- **Total variation convergence:** This notion of convergence is defined by the total variation metric given by

$$\|\mu - \nu\|_{\text{TV}} = \sup_{\|\varphi\|_\infty \le 1} \left| \int \varphi(x)\,\mu(dx) - \int \varphi(x)\,\nu(dx) \right| . \tag{3.1}$$

  Here, $\|\varphi\|_\infty$ denotes the supremum norm of φ.

Remark 3.1 Another widespread notion of convergence is that of *strong convergence*: a sequence μ_n converges strongly to a limiting measure μ if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for every $A \in \mathscr{B}(\mathcal{X})$. However, this notion of convergence does have somewhat more pathological properties. For example, it is possible to find several *non-equivalent* topologies on the space of signed measures on $\mathcal{X}$ giving rise to this notion of convergence for countable sequences, see [Bog07, Section 4.7(v)].

Furthermore, these topologies are not "nice". For example, if we consider the set $\mathscr{M}(\mathcal{X})$ of finite signed measures on $\mathcal{X}$ endowed with the total variation norm, then this is a Banach space (call it $\mathscr{M}_{\text{TV}}$). It is then a very unfortunate fact that one of the topologies on $\mathscr{M}(\mathcal{X})$ giving rise to the notion of strong convergence for sequences is actually the *weak* (in the usual sense of functional analysis) topology on $\mathscr{M}_{\text{TV}}$. Since, just like $\mathscr{M}_{\text{TV}}$ itself, the dual of $\mathscr{M}_{\text{TV}}$ is not separable (unless $\mathcal{X}$ is finite of course), it

follows from general principles (see for example [FHH+01, Ziz03]) that this topology is not metrisable (not even on bounded sets), which greatly limits its use in practice.

Remark 3.2 If $\mathcal{X}$ is a compact metric space, then the Riesz-Markov theorem [Rud91] tells us that the dual of $\mathcal{C}_b(\mathcal{X})$, the space of bounded continuous functions, is precisely given by the space $\mathscr{M}(\mathcal{X})$ of finite signed measures on $\mathcal{X}$. Furthermore, in this language, the topology of weak convergence is nothing but the weak-* topology on $\mathscr{M}(\mathcal{X})$, viewed as the dual of $\mathcal{C}_b(\mathcal{X})$. Since the dual of an infinite-dimensional Banach space is never metrisable for the weak-* topology [FHH+01, Ziz03], one may think then that the notion of weak convergence for probability measure suffers from the same problems as those pointed out in the previous remark for the strong convergence. Fortunately, it turns out that *bounded subsets* of the dual of a separable Banach space are weak-* metrisable, which is sufficient for our purpose since we are mostly interested in probability measures.

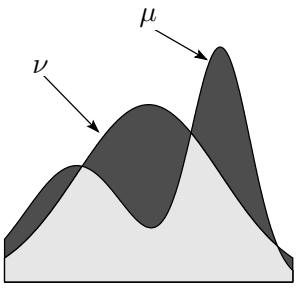
We will sometimes use slightly different notions of convergence, but they will be only minor variations on the general themes given here. The above notions of convergence give rise to two non-equivalent metrisable topologies on $\mathscr{P}(\mathcal{X})$. This is obvious for the notion of total variation convergence, but much less so for the notion of weak convergence. Before we turn to the construction of metrics for weak convergence, let us give a few classical examples illustrating the differences between them.

Example 3.3 Let $\mathcal{X} = \mathbf{R}$ and let μ_n be the Dirac measure located at $1/n$. Then it converges weakly, but neither strongly nor in total variation to the Dirac measure located at the origin.

Example 3.4 Let μ_n be the measure on $[0, \pi]$ given by $\mu_n(dx) = \frac{2}{\pi} \sin^2(nx)\, dx$. Then, μ_n converges to the normalised Lebegsue measure both in the weak and the strong sense. However, the total variation distance between μ_n and its limit $\frac{dx}{\pi}$ is equal to $\frac{2}{\pi}$ for every n.

3.1.1 Total variation convergence

The total variation distance between two probability measures μ and ν is relatively straightforward to comprehend: it consists of the total amount of mass that doesn't overlap between μ and ν. Since a picture is worth a thousand words, we illustrate this by the figure shown on the right: the total variation distance between μ and ν is given by the dark gray area.



If μ and ν have densities $\mathcal{D}_\mu$ and $\mathcal{D}_\nu$ with respect to some common positive reference measure π (by the Radon-Nikodym theorem, it is always possible to take $\pi = \frac{1}{2}(\mu + \nu)$ for example), then one has the identity

$$\|\mu - \nu\|_{\text{TV}} = \int_{\mathcal{X}} |\mathcal{D}_\mu(x) - \mathcal{D}_\nu(x)|\, \pi(dx)\,, \tag{3.2}$$

which is also sometimes taken as the definition of the total variation distance.

Exercise 3.5 Show that the characterisation (3.2) of the total variation distance does not depend on the particular choice of a reference measure π and that it does indeed agree with the definition previously given in (3.1). **Hint:** Consider the test function $\varphi(x)$ such that $\varphi(x) = 1$ if $\mathcal{D}_\mu(x) > \mathcal{D}_\nu(x)$ and $\varphi(x) = -1$ otherwise.

As a consequence of the characterisation (3.2), we have the following very important fact:

Corollary 3.6 *For μ and ν two probability measures, one has $\|\mu - \nu\|_{\text{TV}} = 2$ if and only if μ and ν are mutually singular and $\|\mu - \nu\|_{\text{TV}} < 2$ otherwise.*

Proof. Let μ and ν be mutually singular, denoted by $\mu \perp \nu$. Then there exists a set A such that $\mu(A) = 1$ and $\nu(A) = 0$. Setting $\varphi(x) = 2\,\mathbf{1}_A(x) - 1$, it follows from (3.1) that $\|\mu - \nu\|_{\text{TV}} \geq 2$. On the other hand, one has $\|\mu - \nu\|_{\text{TV}} \leq 2$ as a consequence of the definition, so that the first implication follows.

To show the converse, assume that μ and ν are not mutually singular and denote by $\mathcal{D}_\mu$ and $\mathcal{D}_\nu$ their densities with respect to a common reference measure π. Let $A_\mu = \{x : \mathcal{D}_\mu(x) > 0\}$ and similarly for A_ν, and set $A = A_\mu \cap A_\nu$. With this notation, one must have $\pi(A) > 0$, for otherwise $\mu \perp \nu$. Since for two positive numbers a and b we have the identity $|a - b| = a + b - 2(a \wedge b)$, it follows from (3.2) that

$$\|\mu - \nu\|_{\text{TV}} = 2 - 2\int_A (\mathcal{D}_\mu(x) \wedge \mathcal{D}_\nu(x))\,\pi(dx)\ .$$

Since $\pi(A) > 0$ and $\mathcal{D}_\mu(x) \wedge \mathcal{D}_\nu(x) > 0$ for $x \in A$ by definition, the claim follows. □

There is a third very useful (and more probabilistically appealing) interpretation of the total variation distance between two probability measures. Indeed, the total variation distance between two probability measures μ and ν on a Polish space $\mathcal{X}$ is given by

$$\|\mu - \nu\|_{\text{TV}} = 2 \inf_{\pi \in \mathscr{C}(\mu,\nu)} \pi(\{x \neq y\})\ , \tag{3.3}$$

where the infimum runs over the set $\mathscr{C}(\mu, \nu)$ of all probability measures π on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν. (This set is also called the set of all *couplings* of μ and ν.) In other words, if the total variation distance between μ and ν is smaller than 2ε, then it is possible to construct $\mathcal{X}$-valued random variables X and Y with respective laws μ and ν such that $X = Y$ with probability $1 - \varepsilon$. This gives a straightforward probabilistic interpretation of the total variation distance as twice the probability that a random sample drawn from μ can be distinguished from a sample drawn at random from ν.

Exercise 3.7 Show that the identity (3.3) holds and that the infimum is attained. **Hint:** The optimal coupling can be constructed explicitly by considering a combination of the measure $(\mathcal{D}_\mu(x) \wedge \mathcal{D}_\nu(x))\,\pi(dx)$ on the diagonal (x, x) and the measure $(\mathcal{D}_\mu(x) - \mathcal{D}_\nu(x))_+(\mathcal{D}_\nu(y) - \mathcal{D}_\mu(y))_+\,\pi(dx)\,\pi(dy)$ off the diagonal.

Remark 3.8 Some authors define the total variation distance between measures as the expression (3.3), but without the factor 2. Being aware of this helps to navigate a literature that could otherwise cause some confusion.

Exercise 3.9 Given a function $V\colon \mathcal{X} \to \mathbf{R}_+$, we can also define a weighted total variation distance by

$$\|\mu - \nu\|_{\text{TV},V} = \int_{\mathcal{X}} (1 + V(x))|\mathcal{D}_\mu(x) - \mathcal{D}_\nu(x)|\,\pi(dx)\ .$$

By (3.2), we recover the usual total variation distance as a special case when $V = 0$. This distance is only defined on the subspace of finite signed measures that integrate V and turns this subspace into a Banach space. Show that one does have, similarly to (3.3), the characterisation

$$\|\mu - \nu\|_{\text{TV},V} = \inf_{\pi \in \mathscr{C}(\mu,\nu)} \int_{x \neq y} (2 + V(x) + V(y))\,\pi(dx, dy)\ . \tag{3.4}$$

3.1.2 Weak convergence

It is obvious that if $\mathcal{X}$ is uncountable, then $\mathscr{P}(\mathcal{X})$ endowed with the total variation metric is not a separable space. Indeed, the collection $\{\delta_x\}_{x \in \mathcal{X}}$ yields an uncountable set of elements that are all at distance 2 of each other. On the other hand, if we endow $\mathscr{P}(\mathcal{X})$ with the topology of weak convergence, then it turns out to be separable. Even better, it is actually itself a Polish space and one can construct a

number of natural distance functions that generate its topology. In this section, we collect without proofs a few important results about the properties of the weak convergence topology. For a more detailed account, including complete proofs, we refer for example to [Bil68, Bog07, Vil03].

If one had to choose one, the single most import result in the theory of weak convergence of probability measures would probably be Prohorov's characterisation of those subsets of the set of probability measures that are precompact for the weak convergence topology. Before we state this theorem, we introduce the concept of a *tight* family of probability measures, which is a fundamental concept in the theory of weak convergence:

Definition 3.10 Given a collection $M \subset \mathscr{P}(\mathcal{X})$ of probability measures on a Polish space $\mathcal{X}$, we say that M is *tight* if, for every $\varepsilon > 0$, there exists a compact set $K \subset \mathcal{X}$ such that $\mu(K) \geq 1 - \varepsilon$ for every $\mu \in M$.

In other words, M is tight if its elements are uniformly concentrated on compact sets. It turns out that sets comprising of a single measure (or finitely many measures) are always tight. While this is obvious if the space $\mathcal{X}$ can be covered by a countable collection of compact sets (like it is the case for $\mathbf{R}^n$ for example), it is not so obvious if $\mathcal{X}$ is an infinite-dimensional space. Using the Heine-Borel theorem, it is however not to difficult to prove it, and this is the content of the next lemma:

Lemma 3.11 *Let $\mu \in \mathscr{P}(\mathcal{X})$ for a Polish space $\mathcal{X}$. Then the singleton $\{\mu\}$ is tight.*

Proof. Fix $\varepsilon > 0$. Since $\mathcal{X}$ is separable it can be covered by countably many balls of fixed, but arbitrary, radius. Therefore, for every $n > 0$, one can find a set K_n consisting of finitely many balls of radius $1/n$ and such that $\mu(K_n) > 1 - 2^{-n}\varepsilon$. Setting $K = \bigcap_{n>0} K_n$, it follows that $\mu(K) \geq 1 - \varepsilon$, which concludes the proof since K is totally bounded[1] and therefore compact by the Heine-Borel theorem which characterises compact sets in Polish spaces as closed totally bounded sets. □

Another interesting fact is that tightness follows from the following property which may appear weaker at first sight:

Lemma 3.12 *Let $M \subset \mathscr{P}(\mathcal{X})$ for $\mathcal{X}$ a Polish space width metric d. Assume that, for every $\varepsilon > 0$, there exists $K \subset \mathcal{X}$ compact such that $\mu(K^\varepsilon) \geq 1 - \varepsilon$, where K^ε is the ε-fattening of K: $K^\varepsilon = \{x \in \mathcal{X} : d(x, K) \leq \varepsilon\}$. Then M is tight.*

Proof. The proof works as before since, for every n, we can find K_n such that $\mu(K_n^{\varepsilon 2^{-n}}) \geq 1 - \varepsilon 2^{-n}$ by assumption, and then set $K = \bigcap_{n>0} K_n$ as before. □

So why is tightness so important? The following theorem due to Prohorov [Pro56] shows that tight families of probability measure coincide with precompact subsets of $\mathscr{P}(\mathcal{X})$ in the topology of weak convergence:

Theorem 3.13 (Prohorov) *A subset $M \subset \mathscr{P}(\mathcal{X})$ is precompact for the topology of weak convergence if and only if it is tight.*

Before we give the proof of this theorem, let us prove the following particular case:

Lemma 3.14 *If $\mathcal{X}$ is compact, then both $\mathscr{P}(\mathcal{X})$ and the unit ball in $\mathscr{M}(\mathcal{X})$ are compact for the topology of weak convergence.*

[1] A subset of a metric space is totally bounded if, for every $\varepsilon > 0$, it can be covered by finitely many balls of radius ε.

Proof. Take any sequence μ_n of uniformly bounded measures on $\mathcal{X}$. Since $\mathcal{X}$ is compact, the space $\mathcal{C}(\mathcal{X})$ is separable, so we can find a dense countable subset φ_n of the unit ball in $\mathcal{C}(\mathcal{X})$. A simple diagonal extraction argument, combined with the fact that there exists C such that $\mu_n(\varphi_m) \leq C$ by assumption, allows to extract a subsequence μ_{n_k} such that $\mu_{n_k}(\varphi_n) \to c_n$ for every n. By density of the φ_n and boundedness of the μ_n, it follows that there exists a continuous linear functional μ on $\mathcal{C}(\mathcal{X})$ such that $\mu_{n_k}(\varphi) \to \mu(\varphi)$ for every $\varphi \in \mathcal{C}(\mathcal{X})$. The conclusion then follows from the Riesz-Markov theorem that identifies the dual of $\mathcal{C}(\mathcal{X})$ with $\mathscr{M}(\mathcal{X})$.

Note that in particular the set of probability measures is compact since, by testing against the constant function 1 and positive functions, we conclude that the limit for any converging subsequence is again a probability measure. □

Remark 3.15 The above proof is nothing but a special case of the Banach-Alaoglu theorem [Rud91, Thm 3.15], but its proof is sufficiently short and elementary so that we reproduced it here. It works for the unit ball of the dual space of any separable Banach space, endowed with the weak-* topology.

We are now ready to give the proof of Prohorov's theorem, which follows rather closely the exposition given in [Bog07, Theorem 8.6.2]. The original proof can be found in [Pro56], but see also [Bil68] for a clean proof in the special case $\mathcal{X} = \mathbf{R}$.

Proof of Prohorov's theorem. We first show that tightness is sufficient by extracting a weakly convergent subsequence from M under the assumption that M is tight. By assumption, we can find an increasing sequence of compact sets $K_n \subset \mathcal{X}$ for that $\mu(\mathcal{X} \setminus K_n) \geq 1 - 2^{-n}$ for every $\mu \in M$. Using Lemma 3.14 and a diagonal extraction argument, we can find a sequence $\{\mu_n\}$ in M such that, for every $m > 0$, the restricted sequence $\{\mu_n|_{K_m}\}$ converges to some element $\hat{\mu}_m$, which is a positive measure on K_m with $|\hat{\mu}_m(K_m) - 1| \leq 2^{-m}$. Since $\hat{\mu}_m|_{K_n} = \hat{\mu}_n$ for $n \leq m$, this is an increasing and bounded sequence, which therefore has a limit μ with $\mu|_{K_m} = \hat{\mu}_m$. We conclude that, for every continuous function φ bounded by 1 and every $m \geq 1$, we have

$$|\mu_n(\varphi) - \mu(\varphi)| \leq |(\mu_n|_{K_m})(\varphi) - (\mu|_{K_m})(\varphi)| + |(\mu_n|_{K_m})(\varphi) - \mu_n(\varphi)| + |(\mu|_{K_m})(\varphi) - \mu(\varphi)| \,.$$

The first term converges to 0 and the other two terms are each bounded by 2^{-m}. Since m was arbitrary, this shows that $\{\mu_n\} \to \mu$ as required.

We now show the converse statement, namely that if M is not tight, then it cannot be precompact. Assuming that M is not tight, we use the contrapositive of Lemma 3.12 to conclude that there exists a fixed $\varepsilon > 0$ such that, for every compact $K \subset \mathcal{X}$, there is an element $\mu_K \in M$ such that $\mu_K(\mathcal{X} \setminus K^\varepsilon) \geq \varepsilon$. We now fix a value $\delta > 0$ (think of δ as being much smaller than the ε that we just found), and we construct a sequence of measures $\mu_n \in M$ and two sequences of compact sets (A_n, K_n) recursively in the following way:

- Choose for μ_0 any element of M (it has to contain infinitely many elements since it is not tight), choose K_0 compact such that $\mu_0(K_0) \geq 1 - \delta$, and set $A_0 = K_0$.
- Given data up to the nth index, choose $\mu_{n+1} \in M$ such that $\mu_{n+1}(\mathcal{X} \setminus A_n^\varepsilon) \geq \varepsilon$, which is possible by the lack of tightness of M. Then, choose a compact set $K \subset \mathcal{X}$ such that $\mu_{n+1}(K) \geq 1 - \delta$ and set $A_{n+1} = A_n \cup K$ and $K_{n+1} = K \setminus A_n^\varepsilon$.

Actually, the only properties of this construction that we are going to use are that $\mu_n(K_n) \geq \varepsilon - \delta$, $\mu_n(\bigcup_{k>n} K_k) \leq \delta$, and $K_n^{\varepsilon/2} \cap K_m^{\varepsilon/2} = \emptyset$ for every $n \neq m$.

Our aim is to show that the sequence μ_n constructed in this way contains no convergent subsequence. If $\bar{\mu}_n = \mu_{k_n}$ is an arbitrary subsequence then, by setting $\bar{K}_n = K_{k_n}$ and $\bar{A}_n = A_{k_n}$, the sequence $(\bar{\mu}_n, \bar{A}_n, \bar{K}_n)$ also has the properties mentioned in the last paragraph, so that it suffices to show that any sequence $\{\mu_n\}$ with these properties cannot be convergent.

We do this by exhibiting a continuous test function φ such that $\mu_n(\varphi)$ does not converge. Define first continuous functions φ_n by $\varphi_n(x) = 1$ if $x \in K_n$, $\varphi_n(x) = 0$ if $x \notin K_n^{\varepsilon/2}$, and $\varphi_n(x) =$

$1 - 2\varepsilon^{-1} d(x, K_n)$ otherwise. Note that since these functions all have disjoint supports and are all Lipschitz continuous with the same Lipschitz constant, the function

$$\varphi_\lambda(x) \stackrel{\text{def}}{=} \sum_{n>0} \lambda_n \varphi_n(x) \,,$$

is continuous and bounded for every bounded sequence λ. Given the sequence μ_n, we now construct in a recursive way a sequence λ with $|\lambda_n| \leq 1$ for every n and such that $|\mu_n(\varphi_\lambda) - \mu_{n+1}(\varphi_\lambda)| \geq c$ for some fixed $c > 0$ and for every n.

Choose first $\lambda_0 = 0$, say. For arbitrary $n \geq 0$, once $\lambda_0, \dots, \lambda_n$ are given, it follows from the property $\mu_n(\bigcup_{k>n} K_k) \leq \delta$ and the fact that $|\lambda_i| \leq 1$ that $\mu_n(\varphi_\lambda)$ is determined to within an error of at most δ by $\lambda_0, \dots, \lambda_n$. On the other hand, we have $\mu_{n+1}(K_{n+1}) \geq \varepsilon - \delta$ so that, by adjusting $\lambda_{n+1} \in [-1, 1]$, we can cover a range of values of width at least $2(\varepsilon - \delta)$ for $\mu_{n+1}(\varphi_\lambda)$. This guarantees that we can find λ_{n+1} in such a way that $|\mu_{n+1}(\varphi_{\bar\lambda}) - \mu_n(\varphi_{\bar\lambda})| \geq \varepsilon - 3\delta$ for *every* sequence $\bar\lambda$ such that $\bar\lambda_k = \lambda_k$ for $k \leq n+1$ and $|\bar\lambda_i| \leq 1$. Since ε was fixed but δ was arbitrary in this construction, the claim follows by choosing δ sufficiently small. □

Now that we have some understanding how compact sets look like in $\mathscr{P}(\mathcal{X})$, we turn to the construction of a family of metrics that generate this topology. Given any bounded lower semicontinuous metric d on $\mathcal{X}$ (note that d does not necessarily need to generate the topology of $\mathcal{X}$!), we can "lift" it to the space of probability measures on $\mathcal{X}$ in a natural way by setting:

$$d(\mu, \nu) = \inf_{\pi \in \mathscr{C}(\mu,\nu)} \int_{\mathcal{X}} \int_{\mathcal{X}} d(x, y)\, \pi(dx, dy) \,. \tag{3.5}$$

This distance is called the 1-*Wasserstein distance*[2] for d on $\mathscr{P}(\mathcal{X})$. (The p-Wasserstein distances can be defined similarly for every $p \geq 1$ by setting their pth power equal to the right hand side of (3.5) with d replaced by d^p.) The reason why we assumed that d is lower semicontinuous is the following:

Exercise 3.16 Show that the infimum in (3.5) is achieved. **Hint:** Use the fact that single measures are tight to conclude that the set $\mathscr{C}(\mu, \nu)$ is compact for any two probability measures μ and ν. Then use the lower semicontinuity of d to show that any accumulation point of approximate minimisers must be a minimiser.

Theorem 3.17 *If the metric d is bounded and generates the topology of $\mathcal{X}$, then its p-Wasserstein lift to $\mathscr{P}(\mathcal{X})$ generates the topology of weak convergence.*

Finally, a very useful feature of the Wasserstein-1 distances is that they can also be viewed as the dual norm to the Lipschitz norm on functions. This is the content of the celebrated Monge-Kantorovich-Rubinstein duality theorem (see for example [Vil09]) which we state here without proof.

Theorem 3.18 *For d any lower semicontinuous metric on $\mathcal{X}$, the identity*

$$d(\mu, \nu) = \sup_{\mathrm{Lip}_d(\varphi) \leq 1} \Big(\int_{\mathcal{X}} \varphi(x)\, \mu(dx) - \int_{\mathcal{X}} \varphi(y)\, \mu(dy) \Big) \tag{3.6}$$

holds for all pairs (μ, ν) of probability measures. Here, $\mathrm{Lip}_d(\varphi)$ denotes the best Lipschitz constant for φ with respect to the metric d.

Remark 3.19 The metric d does not need to be bounded in general, so there might be pairs of probability measures for which $d(\mu, \nu)$ is infinite.

[2] This is really a misnomer since these distances were introduced by Kantorovich and the special case $p = 1$ was already studied by Monge. However, the name "Wasserstein distance" is now being used in most of the literature on the subject so we'll stick with it.

Remark 3.20 There exists a generalisation of the duality (3.6) that holds also if d is not a distance function (and therefore also for the p-Wasserstein distances for $p > 1$), but it is slightly more complicated to state. See [Vil03, Vil09] for a very nice treatment of many questions related to Wasserstein distances.

Remark 3.21 It may appear surprising at first sight that an explicit bound on the rate of convergence of integrals of a sequence $\{\mu_n\}$ against Lipschitz continuous functions should yield convergence of the same integrals against any continuous function. However, recall that Prohorov's theorem tells us that any converging sequence of probability measures is essentially concentrated on compact sets. Since on a compact set, any continuous function can be approximated uniformly by Lipschitz continuous functions, this should make it much more plausible that (3.6) does indeed define the topology of weak convergence.

Remark 3.22 Although $\mathscr{P}(\mathcal{X})$ is complete under d, the space of signed measures with finite mass is not complete, if we endow it with the norm defined in (3.6). To see this, take for example $\mathcal{X} = [0, 1]$ and let d be the usual distance function. Then, the sequence

$$f_n = \sum_{k=1}^{n} k^2 (\delta_{3^{-k}} - \delta_{2\cdot 3^{-k}}) \,,$$

is Cauchy, but it obviously does not converge to a measure with finite mass.

One special case of the Monge-Kantorovich-Rubinstein duality is of particular interest. Setting d_{TV} to be the trivial distance function which is equal to 2 for all pairs (x, y) with $x \neq y$, we see that the 1-Wasserstein lift of d_{TV} to $\mathscr{P}(\mathcal{X})$ as in (3.5) is nothing but the total variation distance as characterised in (3.3).

On the other hand, the set of d_{TV}-Lipschitz continuous functions φ with best Lipschitz constant 1 is, up to translations by constants, equal to the set of bounded functions with $\sup_x |\varphi(x)| \leq 1$, so that the dual representation of the 1-Wasserstein lift of d_{TV} as in (3.6) is nothing but the original definition of the total variation distance given in (3.1).

Exercise 3.23 Convince yourself that the identity (3.4) is also a special case of the Monge-Kantorovich-Rubinstein duality. What is the corresponding distance function?

Exercise 3.24 Show that every probability measure μ on $\mathcal{X}$ can be approximated by a finite convex combination of Dirac measures in the topology of weak convergence. **Hint:** Take a dense sequence $\{x_n\}$ in $\mathcal{X}$ and, for given $\varepsilon > 0$, put a Dirac measure at x_n with mass

$$\mu\Big(B(x_n, \varepsilon) \setminus \bigcup_{k<n} B(x_k, \varepsilon)\Big) \,,$$

for sufficiently many values of n so that the total mass adds up to at least $1 - \varepsilon$.

3.2 Basic properties of Gaussian measures

This section is devoted to the study of Gaussian measures on general Banach spaces. Throughout this section and throughout most of the remainder of these notes, we will denote by $\mathcal{B}$ an arbitrary separable Banach space. Recall that a space is separable if it contains a countable dense subset, see for example the monograph [Yos95]. This separability assumption turns out to be crucial for measures on $\mathcal{B}$ to behave in a non-pathological way. It can be circumvented by trickery in most natural situations where non-separable spaces arise, but we choose not to complicate our lives by considering overly general cases in these notes. Another convention that will be used throughout these notes is that all of the measures that we consider are Borel probability measures, meaning that we consider every open set

to be measurable and that we assume that they are normalised in such a way that the whole space has measure 1.

Remark 3.25 Another natural σ-algebra on $\mathcal{B}$ is the one generated by all linear functionals, i.e. the smallest σ-algebra containing $\ell^{-1}(I)$ for every open interval $I \subset \mathbf{R}$ and every $\ell \in \mathcal{B}^*$. We will see in Proposition 3.31 below that this in fact *equals* the Borel σ-algebra of $\mathcal{B}$ thanks to the separability of $\mathcal{B}$.

One additional assumption that would appear to be natural in the context of Gaussian measure theory is that $\mathcal{B}$ be reflexive (that is $\mathcal{B}^{**} = \mathcal{B}$). This is because, for example, the mean of a measure μ appears at first sight to be an element of $\mathcal{B}^{**}$ rather than of $\mathcal{B}$, since the natural[3] way of defining the mean m of μ is to set $m(\ell) = \int_{\mathcal{B}} \ell(x)\,\mu(dx)$ for any $\ell \in \mathcal{B}^*$. This turns out not to be a problem, since the mean of a Gaussian measure on a *separable* Banach space $\mathcal{B}$ is always an element of $\mathcal{B}$ itself, see the monograph [Bog98]. However this result is not straightforward to prove, so we will take here the more pragmatic approach that whenever we consider Gaussian measures with non-zero mean, we simply take the mean $m \in \mathcal{B}$ as given.

Example 3.26 Before we proceed, let us just mention a few examples of Banach spaces. The spaces $L^p(\mathcal{M}, \nu)$ (with $(\mathcal{M}, \nu)$ any countably generated measure space, like for example any Polish space equipped with a positive Radon measure ν) for $p \in (1, \infty)$ are both reflexive and separable. However, reflexivity fails in general for L^1 spaces and both properties fail to hold in general for L^∞ spaces [Yos95]. The space of bounded continuous functions on a compact space is separable, but not reflexive. The space of bounded continuous functions from $\mathbf{R}^n$ to $\mathbf{R}$ is neither separable nor reflexive, but the space of continuous functions from $\mathbf{R}^n$ to $\mathbf{R}$ vanishing at infinity is separable. (The last two statements are still true if we replace $\mathbf{R}^n$ by any Polish space that is locally compact but not compact.) Hilbert spaces are obviously reflexive since $\mathcal{H}^* = \mathcal{H}$ for every Hilbert space $\mathcal{H}$ by the Riesz representation theorem [Yos95]. There exist non-separable Hilbert spaces, but they have rather pathological properties and do not appear very often in practice. An example of a non-separable Hilbert space would be the set of functions $f\colon [0,1] \to \mathbf{R}$ that vanish at all except countably many points, and such that $\|f\|^2 \stackrel{\text{def}}{=} \sum_{t\in[0,1]} |f(t)|^2 < \infty$.

We start with the definition of a Gaussian measure on a Banach space. Since there is no equivalent to Lebesgue measure in infinite dimensions (one could never expect it to be σ-additive), we cannot define it by prescribing the form of its density. However, it turns out that Gaussian measures on $\mathbf{R}^n$ can be characterised by prescribing that the projections of the measure onto any one-dimensional subspace of $\mathbf{R}^n$ are all Gaussian. This is a property that can readily be generalised to infinite-dimensional spaces:

Definition 3.27 A *Gaussian probability measure* μ on a Banach space $\mathcal{B}$ is a Borel measure such that $\ell^\sharp\mu$ is a Gaussian probability measure on $\mathbf{R}$ for every continuous linear functional $\ell\colon \mathcal{B} \to \mathbf{R}$. (Here, Dirac measures are considered to be Gaussian measures with zero variance.) We call it *centred* if $\ell^\sharp\mu$ is centred for every ℓ.

Remark 3.28 We used here the notation $f^\sharp\mu$ for the push-forward of a measure μ under a map f, which is defined by the relation $(f^\sharp\mu)(A) = \mu(f^{-1}(A))$. Some authors use the notation f^* instead, but we reserve this notation for adjoints in a probably futile attempt to minimise confusion.

Remark 3.29 We could also have defined Gaussian measures by imposing that $T^\sharp\mu$ is Gaussian for every bounded linear map $T\colon \mathcal{B} \to \mathbf{R}^n$ and every n. These two definitions are equivalent since probability measures on $\mathbf{R}^n$ are characterised by their Fourier transform and these are constructed from one-dimensional marginals, see Proposition 3.34 below.

[3] Without further assumption, we do not know a priori whether $x \mapsto \|x\|$ is integrable, so that the even more natural definition $m = \int_{\mathcal{B}} x\,\mu(dx)$ is prohibited at this stage. See however Theorem 3.36 below.

Exercise 3.30 Let $\{\xi_n\}$ be a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables and let $\{a_n\}$ be a sequence of real numbers. Show that the law of $(a_0\xi_0, a_1\xi_1, \ldots)$ determines a Gaussian measure on ℓ^2 if and only if $\sum_{n\geq 0} a_n^2 < \infty$.

One first question that one may ask is whether this is indeed a reasonable definition. After all, it only makes a statement about the one-dimensional projections of the measure μ, which itself lives on a huge infinite-dimensional space. However, this turns out to be reasonable since, provided that $\mathcal{B}$ is separable, the one-dimensional projections of any probability measure carry sufficient information to characterise it. This statement can be formalised as follows:

Proposition 3.31 *Let $\mathcal{B}$ be a separable Banach space and let μ and ν be two probability measures on $\mathcal{B}$. If $\ell^\sharp\mu = \ell^\sharp\nu$ for every $\ell \in \mathcal{B}^*$, then $\mu = \nu$.*

Proof. Denote by $\mathrm{Cyl}(\mathcal{B})$ the algebra of cylindrical sets on $\mathcal{B}$, that is $A \in \mathrm{Cyl}(\mathcal{B})$ if and only if there exists $n > 0$, a continuous linear map $T\colon \mathcal{B} \to \mathbf{R}^n$, and a Borel set $\tilde{A} \subset \mathbf{R}^n$ such that $A = T^{-1}\tilde{A}$. It follows from the fact that measures on $\mathbf{R}^n$ are determined by their one-dimensional projections that $\mu(A) = \nu(A)$ for every $A \in \mathrm{Cyl}(\mathcal{B})$ and therefore, by a basic uniqueness result in measure theory (see Lemma II.4.6 in [RW94] or Theorem 1.5.6 in [Bog07] for example), for every A in the σ-algebra $\mathcal{E}(\mathcal{B})$ generated by $\mathrm{Cyl}(\mathcal{B})$. It thus remains to show that $\mathcal{E}(\mathcal{B})$ coincides with the Borel σ-algebra of $\mathcal{B}$. Actually, since every cylindrical set is a Borel set, it suffices to show that all open (and therefore all Borel) sets are contained in $\mathcal{E}(\mathcal{B})$.

Since $\mathcal{B}$ is separable, every open set U can be written as a countable union of closed balls. (Fix any dense countable subset $\{x_n\}$ of $\mathcal{B}$ and check that one has for example $U = \bigcup_{x_n \in U} \bar{B}(x_n, r_n)$, where $r_n = \frac{1}{2}\sup\{r > 0 : \bar{B}(x_n, r) \subset U\}$ and $\bar{B}(x, r)$ denotes the closed ball of radius r centred at x.) Since $\mathcal{E}(\mathcal{B})$ is invariant under translations and dilations, it remains to check that $\bar{B}(0,1) \in \mathcal{E}(\mathcal{B})$. Let $\{x_n\}$ be a countable dense subset of $\{x \in \mathcal{B} : \|x\| = 1\}$ and let ℓ_n be any sequence in $\mathcal{B}^*$ such that $\|\ell_n\| = 1$ and $\ell_n(x_n) = 1$ (such elements exist by the Hahn-Banach extension theorem [Yos95]). Define now $K = \bigcap_{n\geq 0}\{x \in \mathcal{B} : |\ell_n(x)| \leq 1\}$. It is clear that $K \in \mathcal{E}(\mathcal{B})$, so that the proof is complete if we can show that $K = \bar{B}(0,1)$.

Since obviously $\bar{B}(0,1) \subset K$, it suffices to show that the reverse inclusion holds. Let $y \in \mathcal{B}$ with $\|y\| > 1$ be arbitrary and set $\hat{y} = y/\|y\|$. By the density of the x_n's, there exists a subsequence x_{k_n} such that $\|x_{k_n} - \hat{y}\| \leq \frac{1}{n}$, say, so that $\ell_{k_n}(\hat{y}) \geq 1 - \frac{1}{n}$. By linearity, this implies that $\ell_{k_n}(y) \geq \|y\|(1 - \frac{1}{n})$, so that there exists a sufficiently large n so that $\ell_{k_n}(y) > 1$. This shows that $y \notin K$ and we conclude that $K \subset \bar{B}(0,1)$ as required. □

From now on, we will mostly consider centred Gaussian measures, since one can always reduce oneself to the general case by a simple translation. Given a centred Gaussian measure μ, we define a map $C_\mu\colon \mathcal{B}^* \times \mathcal{B}^* \to \mathbf{R}$ by

$$C_\mu(\ell, \ell') = \int_{\mathcal{B}} \ell(x)\ell'(x)\,\mu(dx)\;. \tag{3.7}$$

Remark 3.32 In the case $\mathcal{B} = \mathbf{R}^n$, this is just the covariance matrix, provided that we perform the usual identification of $\mathbf{R}^n$ with its dual.

Remark 3.33 One can identify in a canonical way C_μ with an operator $\hat{C}_\mu\colon \mathcal{B}^* \to \mathcal{B}^{**}$ via the identity $\hat{C}_\mu(\ell)(\ell') = C_\mu(\ell, \ell')$.

The map C_μ will be called the *Covariance operator* of μ. It follows immediately from the definitions that the operator C_μ is bilinear and positive definite, although there might in general exist some ℓ such

that $C_\mu(\ell,\ell) = 0$. Furthermore, the *Fourier transform* of μ is given by

$$\hat{\mu}(\ell) \stackrel{\text{def}}{=} \int_{\mathcal{B}} e^{i\ell(x)}\,\mu(dx) = \exp(-\tfrac{1}{2}C_\mu(\ell,\ell))\;, \tag{3.8}$$

where $\ell \in \mathcal{B}^*$. This can be checked by using the explicit form of the one-dimensional Gaussian measure. Conversely, (3.8) characterises Gaussian measures in the sense that if μ is a measure such that there exists C_μ satisfying (3.8) for every $\ell \in \mathcal{B}^*$, then μ must be centred Gaussian. The reason why this is so is that two distinct probability measures necessarily have distinct Fourier transforms:

Proposition 3.34 *Let μ and ν be any two probability measures on a separable Banach space $\mathcal{B}$. If $\hat{\mu}(\ell) = \hat{\nu}(\ell)$ for every $\ell \in \mathcal{B}^*$, then $\mu = \nu$.*

Proof. In the particular case $\mathcal{B} = \mathbf{R}^n$, if φ is a smooth function with compact support, it follows from Fubini's theorem and the invertibility of the Fourier transform that one has the identity

$$\int_{\mathbf{R}^n} \varphi(x)\,\mu(dx) = \frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \hat{\varphi}(k)e^{-ikx}\,dk\,\mu(dx) = \frac{1}{(2\pi)^n}\int_{\mathbf{R}^n} \hat{\varphi}(k)\,\hat{\mu}(-k)\,dk\;,$$

so that, since bounded continuous functions can be approximated by smooth functions, μ is indeed determined by $\hat{\mu}$. The general case then follows immediately from Proposition 3.31. □

As a simple consequence, we have the following trivial but useful property:

Proposition 3.35 *Let μ be a Gaussian measure on $\mathcal{B}$ and, for every $\varphi \in \mathbf{R}$, define the "rotation" $R_\varphi\colon \mathcal{B}^2 \to \mathcal{B}^2$ by*

$$R_\varphi(x,y) = (x\sin\varphi + y\cos\varphi, x\cos\varphi - y\sin\varphi)\;.$$

Then, one has $R_\varphi^\sharp(\mu\otimes\mu) = \mu\otimes\mu$.

Proof. Since we just showed in Proposition 3.34 that a measure is characterised by its Fourier transform, it suffices to check that $\widehat{\mu\otimes\mu}\circ R_\varphi = \widehat{\mu\otimes\mu}$, which is an easy exercise. □

3.3 A-priori bounds on Gaussian measures

We are going to show now that the operator C_μ has to be bounded, as a straightforward consequence of the fact that $x \mapsto \|x\|^2$ is integrable for any Gaussian measure. Actually, we are going to show much more, namely that there always exists a constant $\alpha > 0$ such that $\exp(\alpha\|x\|^2)$ is integrable! In other words, the norm of any Banach-space valued Gaussian random variable has Gaussian tails, just like in the finite-dimensional case. Amazingly, this result uses the Gaussianity of the measure only indirectly through the rotation invariance shown in Proposition 3.35, and even this property is only used for rotations by the angle $\varphi = \pi/4$. This is the content of the following fundamental result in the theory of Gaussian measures, which can be found in the original article [Fer70]:

Theorem 3.36 (Fernique) *Let μ be a Gaussian probability measure on a separable Banach space $\mathcal{B}$. Then, there exists $\alpha > 0$ such that $\int_{\mathcal{B}} \exp(\alpha\|x\|^2)\,\mu(dx) < \infty$.*

Proof. Amazingly, the only property of Gaussian measures that will be used in the proof is the conclusion of Proposition 3.35 with $\varphi = \pi/4$. Indeed, for any two positive numbers t and τ with $t > \tau$, it implies the bound

$$\mu(\|x\| \le \tau)\,\mu(\|x\| > t) = \int_{\|x\|\le\tau}\int_{\|y\|>t} \mu(dx)\,\mu(dy) = \int_{\|\frac{x-y}{\sqrt{2}}\|\le\tau}\int_{\|\frac{x+y}{\sqrt{2}}\|>t} \mu(dx)\,\mu(dy)$$

$$\leq \int_{\|x\|>\frac{t-\tau}{\sqrt{2}}} \int_{\|y\|>\frac{t-\tau}{\sqrt{2}}} \mu(dx)\,\mu(dy) = \mu\Big(\|x\| > \tfrac{t-\tau}{\sqrt{2}}\Big)^2 . \tag{3.9}$$

In order to go from the first to the second line, we have used the fact that, by the triangle inequality,

$$\min\{\|x\|, \|y\|\} \geq \tfrac{1}{2}(\|x+y\| - \|x-y\|) ,$$

so that $\|x+y\| > \sqrt{2}t$ and $\|x-y\| \leq \sqrt{2}\tau$ do indeed imply that both $\|x\|$ and $\|y\|$ are greater than $\frac{t-\tau}{\sqrt{2}}$. Since $\|x\|$ is μ-almost surely finite, there exists some $\tau > 0$ such that $\mu(\|x\| \leq \tau) \geq \frac{3}{4}$. Set now $t_0 = \tau$ and define t_n for $n > 0$ recursively by the relation $t_n = \frac{t_{n+1}-\tau}{\sqrt{2}}$. It follows from (3.9) that

$$\mu(\|x\| > t_{n+1}) \leq \mu\Big(\|x\| > \tfrac{t_{n+1}-\tau}{\sqrt{2}}\Big)^2 / \mu(\|x\| \leq \tau) \leq \tfrac{4}{3}\mu(\|x\| > t_n)^2 .$$

Setting $y_n = \frac{4}{3}\mu(\|x\| > t_{n+1})$, this yields the recursion $y_{n+1} \leq y_n^2$ with $y_0 \leq 1/3$. Applying this inequality repeatedly, we obtain

$$\mu(\|x\| > t_n) = \frac{3}{4} y_n \leq \frac{3}{4} y_0^{2^n} \leq \frac{1}{4} 3^{-1-2^n} \leq 3^{-2^n} .$$

On the other hand, one can check explicitly that $t_n = \frac{\sqrt{2}^{n+1}-1}{\sqrt{2}-1}\tau \leq 2^{n/2} \cdot (2+\sqrt{2})\tau$, so that in particular $t_{n+1} \leq 2^{n/2} \cdot 5\tau$. This shows that one has the bound

$$\mu(\|x\| > t_n) \leq 3^{-\frac{t_{n+1}^2}{25\tau^2}} ,$$

implying that there exists a *universal* constant $\alpha > 0$ such that the bound $\mu(\|x\| > t) \leq \exp(-2\alpha t^2/\tau^2)$ holds for every $t \geq \tau$. Integrating by parts, we finally obtain

$$\begin{aligned}\int_{\mathcal{B}} \exp\Big(\frac{\alpha\|x\|^2}{\tau^2}\Big)\,\mu(dx) &\leq e^\alpha + \frac{2\alpha}{\tau^2}\int_\tau^\infty t e^{\alpha\frac{t^2}{\tau^2}}\mu(\|x\| > t)\,dt \\ &\leq e^\alpha + 2\alpha\int_1^\infty t e^{-\alpha t^2}\,dt < \infty ,\end{aligned} \tag{3.10}$$

which is the desired result. □

As an immediate corollary of Fernique's theorem, we have

Corollary 3.37 *There exists a constant $\|C_\mu\| < \infty$ such that $C_\mu(\ell,\ell') \leq \|C_\mu\|\|\ell\|\|\ell'\|$ for any $\ell, \ell' \in \mathcal{B}^*$. Furthermore, the operator $\hat{C}_\mu$ defined in Remark 3.33 is a continuous operator from $\mathcal{B}^*$ to $\mathcal{B}$.*

Proof. The boundedness of C_μ implies that $\hat{C}_\mu$ is continuous from $\mathcal{B}^*$ to $\mathcal{B}^{**}$. However, $\mathcal{B}^{**}$ might be strictly larger than $\mathcal{B}$ in general. The fact that the range of $\hat{C}_\mu$ actually belongs to $\mathcal{B}$ follows from the fact that one has the identity

$$\hat{C}_\mu \ell = \int_{\mathcal{B}} x\,\ell(x)\,\mu(dx) . \tag{3.11}$$

Here, the right-hand side is well-defined as a Bochner integral [Boc33, Hil53] because $\mathcal{B}$ is assumed to be separable and we know from Fernique's theorem that $\|x\|^2$ is integrable with respect to μ. □

Remark 3.38 In Theorem 3.36, one can actually take for α any value smaller than $1/(2\|C_\mu\|)$. Furthermore, this value happens to be sharp, see [Led96, Thm 4.1].

Another consequence of the proof of Fernique's theorem is an even stronger result, namely all moments (including exponential moments!) of the norm of a Banach-space valued Gaussian random variable can be estimated in a universal way in terms of its first moment. More precisely, we have

Proposition 3.39 *There exist universal constants $\alpha, K > 0$ with the following properties. Let μ be a centred Gaussian measure on a separable Banach space $\mathcal{B}$ and let $f: \mathbf{R}_+ \to \mathbf{R}_+$ be any measurable function such that $f(x) \leq C_f \exp(\alpha x^2)$ for every $x \geq 0$. Define furthermore the first moment of μ by $M = \int_{\mathcal{B}} \|x\| \, \mu(dx)$. Then, one has the bound $\int_{\mathcal{B}} f(\|x\|/M) \, \mu(dx) \leq K C_f$.*

In particular, the higher moments of μ are bounded by $\int_{\mathcal{B}} \|x\|^{2n} \, \mu(dx) \leq n! K \alpha^{-n} M^{2n}$.

Proof. It suffices to note that the bound (3.10) is independent of τ and that by Chebychev's inequality, one can choose for example $\tau = 4M$. The last claim then follows from the fact that $e^{\alpha x^2} \geq \frac{\alpha^n x^{2n}}{n!}$. □

Actually, the covariance operator C_μ is more than just bounded. If $\mathcal{B}$ happens to be a Hilbert space, one has indeed the following result, which allows us to characterise in a very precise way the set of all centred Gaussian measures on a Hilbert space:

Proposition 3.40 *If $\mathcal{B} = \mathcal{H}$ is a Hilbert space, then the operator $\hat{C}_\mu: \mathcal{H} \to \mathcal{H}$ defined as before by the identity $\langle \hat{C}_\mu h, k \rangle = C_\mu(h, k)$ is trace class and one has the identity*

$$\int_{\mathcal{H}} \|x\|^2 \, \mu(dx) = \operatorname{tr} \hat{C}_\mu \, . \tag{3.12}$$

(Here, we used Riesz's representation theorem to identify $\mathcal{H}$ with its dual.)

Conversely, for every positive trace class symmetric operator $K: \mathcal{H} \to \mathcal{H}$, there exists a Gaussian measure μ on $\mathcal{H}$ such that $\hat{C}_\mu = K$.

Proof. Fix an arbitrary orthonormal basis $\{e_n\}$ of $\mathcal{H}$. We know from Theorem 3.36 that the second moment of the norm is finite: $\int_{\mathcal{H}} \|x\|^2 \, \mu(dx) < \infty$. On the other hand, one has

$$\int_{\mathcal{H}} \|x\|^2 \, \mu(dx) = \sum_{n=1}^{\infty} \int_{\mathcal{H}} \langle x, e_n \rangle^2 \, \mu(dx) = \sum_{n=1}^{\infty} \langle e_n, \hat{C}_\mu e_n \rangle = \operatorname{tr} \hat{C}_\mu \, ,$$

which is precisely (3.12). Here, to pull the sum out of the integral in the first equality, we used Lebesgue's dominated convergence theorem.

In order to prove the converse statement, since K is compact, we can find an orthonormal basis $\{e_n\}$ of $\mathcal{H}$ such that $Ke_n = \lambda_n e_n$ and $\lambda_n \geq 0$, $\sum_n \lambda_n < \infty$. Let furthermore $\{\xi_n\}$ be a collection of i.i.d. $\mathcal{N}(0,1)$ Gaussian random variables (such a family exists by Kolmogorov's extension theorem). Then, since $\sum_n \lambda_n \mathbf{E} \xi_n^2 = \operatorname{tr} K < \infty$, the series $\sum_n \sqrt{\lambda_n} \xi_n e_n$ converges in mean square, so that it has a subsequence converging almost surely in $\mathcal{H}$. One can easily check that the law of the limiting random variable is Gaussian and has the requested covariance. □

No such precise characterisation of the covariance operators of Gaussian measures exists in Banach spaces. One can however show that $\hat{C}_\mu$ is at least a little bit better than bounded, namely that it is always a compact operator. We leave this statement as an exercise for the interested reader, since we will not make any use of it in these notes:

Exercise 3.41 Show that in the case of a Gaussian measure μ on a general separable Banach space $\mathcal{B}$, the covariance operator $\hat{C}_\mu: \mathcal{B}^* \to \mathcal{B}$ is compact in the sense that it maps the unit ball on $\mathcal{B}^*$ into a compact subset of $\mathcal{B}$. **Hint:** Proceed by contradiction by first showing that if $\hat{C}_\mu$ wasn't compact, then it would be possible to find a constant $c > 0$ and a sequence of elements $\{\ell_k\}_{k \geq 0}$ such that $\|\ell_k\| = 1$, $C_\mu(\ell_k, \ell_j) = 0$ for $k \neq j$, and $C_\mu(\ell_k, \ell_k) \geq c$ for every k. Conclude that if this was the case, then the law of large numbers applied to the sequence of random variables $\ell_n(x)$ would imply that $\sup_{k \geq 0} \ell_k(x) = \infty$ μ-almost surely, thus obtaining a contradiction with the fact that $\sup_{k \geq 0} \ell_k(x) \leq \|x\| < \infty$ almost surely.

In many situations, it is furthermore helpful to find out whether a given covariance structure can be realised as a Gaussian measure on some space of Hölder continuous functions. This can be achieved

through the following version of Kolmogorov's continuity criterion, which can be found for example in [RY94, p. 26]:

Theorem 3.42 (Kolmogorov) *For $d > 0$, let $C\colon [0,1]^d \times [0,1]^d \to \mathbf{R}$ be a symmetric function such that, for every finite collection $\{x_i\}_{i=1}^m$ of points in $[0,1]^d$, the matrix $C_{ij} = C(x_i, x_j)$ is positive definite. If furthermore there exists $\alpha > 0$ and a constant $K > 0$ such that $C(x,x) + C(y,y) - 2C(x,y) \leq K|x-y|^{2\alpha}$ for any two points $x, y \in [0,1]^d$ then there exists a unique centred Gaussian measure μ on $\mathcal{C}([0,1]^d, \mathbf{R})$ such that*

$$\int_{\mathcal{C}([0,1]^d,\mathbf{R})} f(x) f(y)\, \mu(df) = C(x,y)\,, \tag{3.13}$$

for any two points $x, y \in [0,1]^d$. Furthermore, for every $\beta < \alpha$, one has $\mu(\mathcal{C}^\beta([0,1]^d, \mathbf{R})) = 1$, where $\mathcal{C}^\beta([0,1]^d, \mathbf{R})$ is the space of β-Hölder continuous functions.

Proof. Set $\mathcal{B} = \mathcal{C}([0,1]^d, \mathbf{R})$ and $\mathcal{B}^*$ its dual, which consists of the set of Borel measures with finite total variation [Yos95, p. 119]. Since convex combinations of Dirac measures are dense (in the topology of weak convergence) in the set of probability measures, it follows that the set of linear combinations of point evaluations is weakly dense in $\mathcal{B}^*$. Therefore, the claim follows if we are able to construct a measure μ on $\mathcal{B}$ such that (3.13) holds and such that, if f is distributed according to μ, then for any finite collection of points $\{x_i\} \subset [0,1]^d$, the joint law of the $f(x_i)$ is Gaussian.

By Kolmogorov's extension theorem, we can construct a measure μ_0 on $\mathcal{X} = \mathbf{R}^{[0,1]^d}$ endowed with the product σ-algebra such that the laws of all finite-dimensional marginals are Gaussian and satisfy (3.13). We denote by X an $\mathcal{X}$-valued random variable with law μ_0. At this stage, one would think that the proof is complete if we can show that X almost surely has finite β-Hölder norm. The problem with this statement is that the β-Hölder norm is not a measurable function on $\mathcal{X}$! The reason for this is that it requires point evaluations of X at uncountably many locations, while functions that are measurable with respect to the product σ-algebra on $\mathcal{X}$ are allowed to depend on at most countably many function evaluations.

This problem can be circumvented very elegantly in the following way. Denote by $\mathcal{D} \subset [0,1]^d$ the subset of dyadic numbers (actually any countable dense subset would do for now, but the dyadic numbers will be convenient later on) and define the event Ω_β by

$$\Omega_\beta = \Big\{ X \,:\, \hat{X}(x) \stackrel{\text{def}}{=} \lim_{\substack{y \to x \\ y \in \mathcal{D}}} X(y) \text{ exists for every } x \in [0,1]^d \text{ and } \hat{X} \text{ belongs to } \mathcal{C}^\beta([0,1]^d, \mathbf{R}) \Big\}\,.$$

Since the event Ω_β can be constructed from evaluating X at only countably many points, it is a measurable set. For the same reason, the map $\iota\colon \mathcal{X} \to \mathcal{C}^\beta([0,1]^d, \mathbf{R})$ given by

$$\iota(X) = \begin{cases} \hat{X} & \text{if } X \in \Omega_\beta, \\ 0 & \text{otherwise} \end{cases}$$

is measurable with respect to the product σ-algebra on $\mathcal{X}$ (and the Borel σ-algebra on $\mathcal{C}^\beta$), so that the claim follows if we can show that $\mu_0(\Omega_\beta) = 1$ for every $\beta < \alpha$. (Take $\mu = \iota^\sharp \mu_0$.) Denoting the β-Hölder norm of X restricted to the dyadic numbers by $M_\beta(X) = \sup_{x \neq y\,:\, x,y \in \mathcal{D}} \{|X(x) - X(y)|/|x-y|^\beta\}$, we see that Ω_β can alternatively be characterised as $\Omega_\beta = \{X \,:\, M_\beta(X) < \infty\}$, so that the claim follows if we can show for example that $\mathbf{E} M_\beta(X) < \infty$.

Denote by $\mathcal{D}_m \subset \mathcal{D}$ the set of those numbers whose coordinates are integer multiples of 2^{-m} and denote by Δ_m the set of pairs $x, y \in \mathcal{D}_m$ such that $|x-y| = 2^{-m}$. In particular, note that Δ_m contains at most $2^{(m+2)d}$ such pairs.

We are now going to make use of our simplifying assumption that we are dealing with Gaussian random variables, so that pth moments can be bounded in terms of second moments. More precisely, for every $p \geq 1$ there exists a constant C_p such that if X is a Gaussian random variable, then one has

the bound $\mathbf{E}|X|^p \le C_p(\mathbf{E}|X|^2)^{p/2}$. Setting $K_m(X) = \sup_{x,y\in\Delta_m} |X(x) - X(y)|$ and fixing some arbitrary $\beta' \in (\beta, \alpha)$, we see that for $p \ge 1$ large enough, there exists a constant K_p such that

$$\begin{aligned}\mathbf{E}K_m^p(X) &\le \sum_{x,y\in\Delta_m} \mathbf{E}|X(x) - X(y)|^p \le C_p \sum_{x,y\in\Delta_m} (\mathbf{E}|X(x) - X(y)|^2)^{p/2} \\ &= C_p \sum_{x,y\in\Delta_m} (C(x,x) + C(y,y) - 2C(x,y))^{p/2} \le \hat{C}_p 2^{(m+2)d-\alpha mp} \\ &\le \hat{C}_p 2^{-\beta' mp} ,\end{aligned}$$

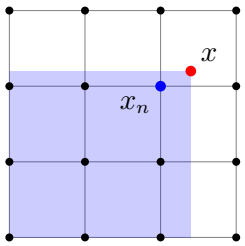
for some constants $\hat{C}_p$. (In order to obtain the last inequality, we had to assume that $p \ge \frac{d}{\alpha-\beta'}\frac{m+2}{m}$ which can always be achieved by some value of p independent of m since we assumed that $\beta' < \alpha$.) Using Jensen's inequality, this shows that there exists a constant K such that the bound

$$\mathbf{E}K_m(X) \le K 2^{-\beta' m} \tag{3.14}$$

holds uniformly in m. Fix now any two points $x, y \in \mathcal{D}$ with $x \ne y$ and denote by m_0 the largest m such that $|x - y| < 2^{-m}$. One can then find sequences $\{x_n\}_{n\ge m_0}$ and $\{y_n\}_{n\ge m_0}$ with the following properties:

1. One has $\lim_{n\to\infty} x_n = x$ and $\lim_{n\to\infty} y_n = y$.
2. Either $x_{m_0} = y_{m_0}$ or both points belong to the vertices of the same "dyadic hypercube" in $\mathcal{D}_{m_0}$, so that they can be connected by at most d "bonds" in Δ_{m_0}.
3. For every $n \ge m_0$, x_n and x_{n+1} belong to the vertices of the same "dyadic hypercube" in $\mathcal{D}_{n+1}$, so that they can be connected by at most d "bonds" in Δ_{n+1} and similarly for (y_n, y_{n+1}).

One way of constructing this sequence is to order elements in $\mathcal{D}_m$ by lexicographic order and to choose $x_n = \max\{\bar{x} \in \mathcal{D}_n : \bar{x}_j \le x_j \ \forall j\}$, as illustrated in the picture to the right. This shows that one has the bound



$$\begin{aligned}|X(x) - X(y)| &\le |X(x_{m_0}) - Y(x_{m_0})| + \sum_{n=m_0}^{\infty} |X(x_{n+1}) - X(x_n)| \\ &\quad + \sum_{n=m_0}^{\infty} |X(y_{n+1}) - X(y_n)| \\ &\le dK_{m_0}(X) + 2d \sum_{n=m_0}^{\infty} K_{n+1}(X) \le 2d \sum_{n=m_0}^{\infty} K_n(X) .\end{aligned}$$

Note at this stage that since we have $x, y \in \mathcal{D}$, the above sums are actually always finite since, if for example $x \in \mathcal{D}_N$, then $x_n = x$ for every $n > N$.

Since m_0 was chosen in such a way that $|x - y| \ge 2^{-m_0-1}$, one has the bound

$$M_\beta(X) \le 2d \sup_{m\ge 0} 2^{\beta(m+1)} \sum_{n=m}^{\infty} K_n(X) \le 2^{\beta+1} d \sum_{n=0}^{\infty} 2^{\beta n} K_n(X) .$$

It follows from this and from the bound (3.14) that

$$\mathbf{E}|M_\beta(X)| \le 2^{\beta+1} d \sum_{n=0}^{\infty} 2^{\beta n} \mathbf{E}K_n(X) \le 2^{\beta+1} dK \sum_{n=0}^{\infty} 2^{(\beta-\beta')n} < \infty ,$$

since β' was chosen strictly larger than β. □

Combining Kolmogorov's continuity criterion with Fernique's theorem, we note that we can apply it not only to real-valued processes, but to any Gaussian Banach-space valued process:

Proposition 3.43 *Let $\mathcal{B}$ be a separable Banach space and let $\{X(x)\}_{x\in[0,1]^d}$ be a collection of $\mathcal{B}$-valued Gaussian random variables such that*

$$\mathbf{E}\|X(x) - X(y)\| \le C|x-y|^\alpha \,,$$

for some $C > 0$ and some $\alpha \in (0,1]$. Then, there exists a unique Gaussian measure μ on $\mathcal{C}([0,1]^d, \mathcal{B})$ such that, if Y is a random variable with law μ, then $Y(x)$ is equal in law to $X(x)$ for every x. Furthermore, $\mu(\mathcal{C}^\beta([0,1]^d, \mathcal{B})) = 1$ for every $\beta < \alpha$.

Proof. The proof is identical to that of Theorem 3.42, noting that the bound $\mathbf{E}\|X(x) - X(y)\|^p \le C_p|x-y|^{\alpha p}$ follows from the assumption and Proposition 3.39. □

Remark 3.44 The space $\mathcal{C}^\beta([0,1]^d, \mathbf{R})$ is not separable. However, the space $\mathcal{C}_0^\beta([0,1]^d, \mathbf{R})$ of Hölder continuous functions that furthermore satisfy $\lim_{y\to x} \frac{|f(x)-f(y)|}{|x-y|^\beta} = 0$ uniformly in x is separable (polynomials with rational coefficients are dense in it). This is in complete analogy with the fact that the space of bounded measurable functions is not separable, while the space of continuous functions is.

It is furthermore possible to check that $\mathcal{C}^{\beta'} \subset \mathcal{C}_0^\beta$ for every $\beta' > \beta$, so that Exercise 3.64 below shows that μ can actually be realised as a Gaussian measure on $\mathcal{C}_0^\beta([0,1]^d, \mathbf{R})$.

Exercise 3.45 Try to find conditions on $G \subset \mathbf{R}^d$ that are as weak as possible and such that Kolmogorov's continuity theorem still holds if the cube $[0,1]^d$ is replaced by G. **Hint:** One possible strategy is to embed G into a cube and then to try to extend $C(x,y)$ to that cube.

Exercise 3.46 Show that if G is as in the previous exercise, $\mathcal{H}$ is a Hilbert space, and $C\colon G \times G \to \mathcal{L}(\mathcal{H}, \mathcal{H})$ is such that $C(x,y)$ positive definite, symmetric, and trace class for any two $x, y \in G$, then Kolmogorov's continuity theorem still holds if its condition is replaced by $\operatorname{tr} C(x,x) + \operatorname{tr} C(y,y) - 2\operatorname{tr} C(x,y) \le K|x-y|^\alpha$. More precisely, one can construct a measure μ on the space $\mathcal{C}^\beta([0,1]^d, \mathcal{H})$ such that

$$\int_{\mathcal{C}^\beta([0,1]^d,\mathbf{R})} \langle h, f(x)\rangle\langle f(y), k\rangle \, \mu(df) = \langle h, C(x,y)k\rangle \,,$$

for any $x, y \in G$ and $h, k \in \mathcal{H}$.

A very useful consequence of Kolmogorov's continuity criterion is the following result:

Corollary 3.47 *Let $\{\eta_k\}_{k\ge 0}$ be countably many i.i.d. standard Gaussian random variables (real or complex). Moreover let $\{f_k\}_{k\ge 0} \subset \operatorname{Lip}(G, \mathbf{C})$ where the domain $G \subset \mathbf{R}^d$ is sufficiently regular for Kolomgorov's continuity theorem to hold. Suppose there is some $\delta \in (0,2)$ such that*

$$S_1^2 = \sum_{k\in I} \|f_k\|_{L^\infty}^2 < \infty \quad \text{and} \quad S_2^2 = \sum_{k\in I} \|f_k\|_{L^\infty}^{2-\delta} \operatorname{Lip}(f_k)^\delta < \infty \,, \tag{3.15}$$

and define $f = \sum_{k\in I} \eta_k f_k$. Then f is almost surely bounded and Hölder continuous for every Hölder exponent smaller than $\delta/2$.

Proof. From the assumptions we immediately derive that $f(x)$ and $f(x) - f(y)$ are a centred Gaussian for any $x, y \in G$. Moreover, the corresponding series converge absolutely. Using that the η_k are i.i.d., we obtain

$$\begin{aligned}\mathbf{E}|f(x) - f(y)|^2 &= \sum_{k\in I} |f_k(x) - f_k(y)|^2 \le \sum_{k\in I} \min\{2\|f_k\|_{L^\infty}^2, \operatorname{Lip}(f_k)^2|x-y|^2\} \\ &\le 2\sum_{k\in I} \|f_k\|_{L^\infty}^{2-\delta} \operatorname{Lip}(f_k)^\delta |x-y|^\delta = 2S_2^2|x-y|^\delta \,,\end{aligned}$$

where we used that $\min\{a, bx^2\} \leq a^{1-\delta/2} b^{\delta/2} |x|^\delta$ for any $a, b \geq 0$. The claim now follows from Kolmogorov's continuity theorem. □

Remark 3.48 One should really think of the f_k's in Corollary 3.47 as being an orthonormal basis of the Cameron-Martin space of some Gaussian measure. (See Section 3.4 below for the definition of the Cameron-Martin space associate to a Gaussian measure.) The criterion (3.15) then provides an effective way of deciding whether the measure in question can be realised on a space of Hölder continuous functions.

3.4 The Cameron-Martin space

Given a Gaussian measure μ on a separable Banach space $\mathcal{B}$, it is possible to associate to it in a canonical way a Hilbert space $\mathcal{H}_\mu \subset \mathcal{B}$, called the Cameron-Martin space of μ. One of the main features of the Cameron-Martin space is that it characterises precisely those directions in $\mathcal{B}$ in which translations leave the measure μ "quasi-invariant" in the sense that the translated measure has the same null sets as the original measure. In general, the space $\mathcal{H}_\mu$ will turn out to be strictly smaller than $\mathcal{B}$. Actually, this is always the case as soon as $\dim \mathcal{H}_\mu = \infty$ and, even worse, we will see that in this case one necessarily has $\mu(\mathcal{H}_\mu) = 0$! Contrast this to the case of finite-dimensional Lebesgue measure which is invariant under translations in any direction! This is a striking illustration of the fact that measures in infinite-dimensional spaces have a strong tendency of being mutually singular.

Another remarkable feature of $\mathcal{H}_\mu$ is that even though it is in general of measure 0, it is sufficient to specify the action of a linear map on $\mathcal{H}_\mu$ to guarantee that it can uniquely be extended to a measurable linear map on a subset of full measure! This will be the content of Section 3.5.1 below.

There are several equivalent definitions of the Cameron-Martin space, and we will see some of its equivalent characterisations in this section. We settle on the following definition as our starting point, as it naturally yields the "correct" norm on $\mathcal{H}_\mu$. Here, we postpone to Remark 3.51 and Proposition 3.54 the verification that $\|h\|_\mu$ is well-defined and that $\|h\|_\mu > 0$ for $h \neq 0$:

Definition 3.49 The *Cameron-Martin space* $\mathcal{H}_\mu$ of μ is the completion of the linear subspace $\mathring{\mathcal{H}}_\mu \subset \mathcal{B}$ defined by

$$\mathring{\mathcal{H}}_\mu = \{h \in \mathcal{B} : \exists\, h^* \in \mathcal{B}^* \text{ with } C_\mu(h^*, \ell) = \ell(h)\ \forall \ell \in \mathcal{B}^*\}\,,$$

under the norm $\|h\|_\mu^2 = \langle h, h\rangle_\mu = C_\mu(h^*, h^*)$. It is a Hilbert space when endowed with the scalar product $\langle h, k\rangle_\mu = C_\mu(h^*, k^*) = h^*(k) = k^*(h)$.

Exercise 3.50 Convince yourself that the space $\mathring{\mathcal{H}}_\mu$ is nothing but the range of the operator $\hat{C}_\mu$ defined in Remark 3.33. This shows that in general, it is not guaranteed that $\mathcal{H}_\mu$ is dense in $\mathcal{B}$. However, we will see in Proposition 3.68 below that this is always the case if μ has full support.

Remark 3.51 Even though the map $h \mapsto h^*$ may not be one to one, the norm $\|h\|_\mu$ is well-defined. To see this, assume that for a given $h \in \mathring{\mathcal{H}}_\mu$, there are two corresponding elements h_1^* and h_2^* in $\mathcal{B}^*$. Then, defining $k = h_1^* + h_2^*$, one has

$$C_\mu(h_1^*, h_1^*) - C_\mu(h_2^*, h_2^*) = C_\mu(h_1^*, k) - C_\mu(h_2^*, k) = k(h) - k(h) = 0\,,$$

showing that $\|h\|_\mu$ does indeed not depend on the choice of h^*.

Exercise 3.52 Show that in the case $\mathcal{B} = \mathbf{R}^n$, the Cameron-Martin space is given by the range of the covariance matrix. Write an expression for $\|h\|_\mu$ in this case.

Exercise 3.53 Show that the Cameron-Martin space of a Gaussian measure determines it. More precisely, if μ and ν are two centred Gaussian measures on $\mathcal{B}$ such that $\mathcal{H}_\mu = \mathcal{H}_\nu$ and such that $\|h\|_\mu = \|h\|_\nu$ for every $h \in \mathcal{H}_\mu$, then they are identical.

For this reason, a Gaussian measure on $\mathcal{B}$ is sometimes given by specifying the Hilbert space structure $(\mathcal{H}_\mu, \|\cdot\|_\mu)$. Such a specification is then usually called an *abstract Wiener space*.

Let us discuss a few properties of the Cameron-Martin space. First of all, we show that it is a subspace of $\mathcal{B}$ despite the completion procedure and that all non-zero elements of $\mathcal{H}_\mu$ have strictly positive norm:

Proposition 3.54 *One has $\mathcal{H}_\mu \subset \mathcal{B}$. Furthermore, one has the bound*

$$\langle h, h\rangle_\mu \geq \|C_\mu\|^{-1}\|h\|^2\,, \tag{3.16}$$

where the norms on the right hand side are understood to be taken in $\mathcal{B}$.

Proof. One has the chain of inequalities

$$\|h\|^2 = \sup_{\ell\in\mathcal{B}^*\setminus\{0\}} \frac{\ell(h)^2}{\|\ell\|^2} = \sup_{\ell\in\mathcal{B}^*\setminus\{0\}} \frac{C_\mu(h^*,\ell)^2}{\|\ell\|^2} \leq \sup_{\ell\in\mathcal{B}^*\setminus\{0\}} \frac{C_\mu(h^*,h^*)\,C_\mu(\ell,\ell)}{\|\ell\|^2} \leq \|C_\mu\|\langle h,h\rangle_\mu\,,$$

which yields the bound on the norms. The fact that $\mathcal{H}_\mu$ is a subset of $\mathcal{B}$ (or rather that it can be interpreted as such) then follows from the fact that $\mathcal{B}$ is complete and that Cauchy sequences in $\mathring{\mathcal{H}}_\mu$ are also Cauchy sequences in $\mathcal{B}$ by (3.16). □

A simple example showing that the correspondence $h \mapsto h^*$ in the definition of $\mathring{\mathcal{H}}_\mu$ is not necessarily unique is the case $\mu = \delta_0$, so that $C_\mu = 0$. If one chooses $h = 0$, then any $h^* \in \mathcal{B}$ has the required property that $C_\mu(h^*, \ell) = \ell(h)$, so that this is an extreme case of non-uniqueness. However, if we view $\mathcal{B}^*$ as a subset of $L^2(\mathcal{B}, \mu)$ (by identifying linear functionals that agree μ-almost surely), then the correspondence $h \mapsto h^*$ is always an isomorphism. One has indeed $\int_\mathcal{B} h^*(x)^2\,\mu(dx) = C_\mu(h^*, h^*) = \|h\|_\mu^2$. In particular, if h_1^* and h_2^* are two distinct elements of $\mathcal{B}^*$ associated to the same element $h \in \mathcal{B}$, then $h_1^* - h_2^*$ is associated to the element 0 and therefore $\int_\mathcal{B}(h_1^* - h_2^*)^2(x)\,\mu(dx) = 0$, showing that $h_1^* = h_2^*$ as elements of $L^2(\mathcal{B}, \mu)$. We have:

Proposition 3.55 *There is a canonical isomorphism $\iota\colon h \mapsto h^*$ between $\mathcal{H}_\mu$ and the closure $\mathcal{R}_\mu$ of $\mathcal{B}^*$ in $L^2(\mathcal{B}, \mu)$. In particular, $\mathcal{H}_\mu$ is separable.*

Proof. We have already shown that $\iota\colon \mathcal{H}_\mu \to L^2(\mathcal{B}, \mu)$ is an isomorphism onto its image, so it remains to show that all of $\mathcal{B}^*$ belongs to the image of ι. For $h \in \mathcal{B}^*$, define $h_* \in \mathcal{B}$ as in (3.11) by

$$h_* = \int_\mathcal{B} x\,h(x)\,\mu(dx)\,. \tag{3.17}$$

This integral converges since $\|x\|^2$ is integrable by Fernique's theorem. Since one has the identity $\ell(h_*) = C_\mu(\ell, h)$, it follows that $h_* \in \mathring{\mathcal{H}}_\mu$ and $h = \iota(h_*)$, as required to conclude the proof.

The separability of $\mathcal{H}_\mu$ then follows immediately from the fact that $L^2(\mathcal{B}, \mu)$ is separable whenever $\mathcal{B}$ is separable, since its Borel σ-algebra is countably generated. □

Remark 3.56 The space $\mathcal{R}_\mu$ defined in Proposition 3.55 is called the *reproducing kernel Hilbert space* for μ (or just *reproducing kernel* for short). However, since it is isomorphic to the Cameron-Martin space in a natural way, there is considerable confusion between the two in the literature. We retain in these notes the terminology from [Bog98], but we urge the reader to keep in mind that there are authors who use a slightly different terminology.

Remark 3.57 In general, there do exist Gaussian measures with non-separable Cameron-Martin space, but they are measures on more general vector spaces. One example would be the measure on $\mathbf{R}^{\mathbf{R}}$ (yes, the space of *all* functions from $\mathbf{R}$ to $\mathbf{R}$ endowed with the product σ-algebra) given by the uncountable product of one-dimensional Gaussian measures. The Cameron-Martin space for this somewhat pathological measure is given by those functions f that are non-zero on at most countably points and such that $\sum_{t\in\mathbf{R}} |f(t)|^2 < \infty$. This is a prime example of a non-separable Hilbert space.

Exercise 3.58 The Wiener measure μ is defined on $\mathcal{B} = \mathcal{C}([0,1],\mathbf{R})$ as the centred Gaussian measure with covariance operator given by $C_\mu(\delta_s,\delta_t) = s \wedge t$. Show that the Cameron-Martin space for the Wiener measure on $\mathcal{B} = \mathcal{C}([0,1],\mathbf{R})$ is given by the space $H_0^{1,2}([0,1])$ of all absolutely continuous functions h such that $h(0) = 0$ and $\int_0^1 \dot{h}^2(t)\,dt < \infty$. **Hint:** If we denote by W a process with law μ, one way of proceeding is to first use the particular form of the covariance of μ to show that elements in $\mathcal{R}_\mu$ can be represented as $W \mapsto \int_0^1 g(s)\,dW(s)$ for $g \in L^2([0,1],\mathbf{R})$. One can then use the map (3.17) to show that the corresponding element $h \in \mathcal{H}_\mu$ is given by $h(t) = \int_0^t g(s)\,ds$.

Exercise 3.59 Let μ be a Gaussian measure on a Hilbert space $\mathcal{H}$ with covariance K and consider the spectral decomposition of K: $Ke_n = \lambda_n e_n$ with $\sum_{n\geq 1}\lambda_n < \infty$ and $\{e_n\}$ an orthonormal basis of eigenvectors. Such a decomposition exists since we already know that K must be trace class from Proposition 3.40.

Assume now that $\lambda_n > 0$ for every n. Show that $\mathring{\mathcal{H}}_\mu$ is given by the range of K and that the correspondence $h \mapsto h^*$ is given by $h^* = K^{-1}h$. Show furthermore that the Cameron-Martin space $\mathcal{H}_\mu$ consists of the image of $K^{1/2}$, *i.e.* those elements h of $\mathcal{H}$ such that $\sum_{n\geq 1}\lambda_n^{-1}\langle h, e_n\rangle^2 < \infty$, and that $\langle h,k\rangle_\mu = \langle K^{-1/2}h, K^{-1/2}k\rangle$.

Exercise 3.60 Show that one has the alternative characterisation

$$\|h\|_\mu = \sup\{\ell(h) \,:\, C_\mu(\ell,\ell) \leq 1\}\,, \tag{3.18}$$

and $\mathcal{H}_\mu = \{h \in \mathcal{B} \,:\, \|h\|_\mu < \infty\}$. **Hint:** Use the fact that in any Hilbert space $\mathcal{H}$, one has $\|h\| = \sup\{\langle k,h\rangle \,:\, \|k\| \leq 1\}$, and that this characterisation still holds if we restrict k to a dense subspace. (Prove it!)

Since elements in $\mathcal{R}_\mu$ are built from the space of all bounded linear functionals on $\mathcal{B}$, it should come as little surprise that its elements are "almost" linear functionals on $\mathcal{B}$ in the following sense:

Proposition 3.61 *For every $\ell \in \mathcal{R}_\mu$ there exists a measurable linear subspace V_ℓ of $\mathcal{B}$ such that $\mu(V_\ell) = 1$ and a linear map $\hat{\ell}\colon V_\ell \to \mathbf{R}$ such that $\ell = \hat{\ell}$ μ-almost surely.*

Proof. Fix $\ell \in \mathcal{R}_\mu$. By the definition of $\mathcal{R}_\mu$ and Borel-Cantelli, we can find a sequence $\ell_n \in \mathcal{B}^*$ such that $\lim_{n\to\infty}\ell_n(x) = \ell(x)$ for μ-almost every $x \in \mathcal{B}$. (Take for example ℓ_n such that $\|\ell_n - \ell\|_\mu^2 \leq n^{-4}$.) It then suffices to define

$$V_\ell = \Big\{x \,:\, \lim_{n\to\infty}\ell_n(x) \text{ exists}\Big\}\,,$$

and to set $\hat{\ell}(x) = \lim_{n\to\infty}\ell_n(x)$ on V_ℓ. □

Another very useful fact about the reproducing kernel space is given by:

Proposition 3.62 *The law of any element $h^* = \iota(h) \in \mathcal{R}_\mu$ is a centred Gaussian with variance $\|h\|_\mu^2$. Furthermore, any two elements h^*, k^* have covariance $\langle h,k\rangle_\mu$.*

Proof. We already know from the definition of a Gaussian measure that the law of any element of $\mathcal{B}^*$ is a centred Gaussian. Let now h^* be any element of $\mathcal{R}_\mu$ and let h_n be a sequence in $\mathcal{R}_\mu \cap \mathcal{B}^*$ such

that $h_n \to h^*$ in $\mathcal{R}_\mu$. We can furthermore choose this approximating sequence such that $\|h_n\|_{\mathcal{R}_\mu} = \|h^*\|_{\mathcal{R}_\mu} = \|h\|_\mu$, so that the law of each of the h_n is equal to $\mathcal{N}(0, \|h\|_\mu^2)$.

Since L^2-convergence implies convergence in law, we conclude that the law of h^* is also given by $\mathcal{N}(0, \|h\|_\mu^2)$. The statement about the covariance then follows by polarisation, since

$$\mathbf{E}h^*k^* = \frac{1}{2}(\mathbf{E}(h^*+k^*)^2 - \mathbf{E}(h^*)^2 - \mathbf{E}(k^*)^2) = \frac{1}{2}(\|h+k\|_\mu^2 - \|h\|_\mu^2 - \|k\|_\mu^2) = \langle h, k\rangle_\mu \,,$$

by the previous statement. □

Remark 3.63 Actually, the converse of Proposition 3.61 is also true: if $\ell\colon \mathcal{B} \to \mathbf{R}$ is measurable and linear on a measurable linear subspace V of full measure, then ℓ belongs to $\mathcal{R}_\mu$. This is not an obvious statement. It can be viewed for example as a consequence of the highly non-trivial fact that every Borel measurable linear map between two sufficiently "nice" topological vector spaces is bounded, see for example [Sch66, Kat82]. (The point here is that the map must be linear on the whole space and not just on some "large" subspace as is usually the case with unbounded operators.) This implies by Proposition 3.67 that ℓ is a measurable linear extension of some bounded linear functional on $\mathcal{H}_\mu$. Since such extensions are unique (up to null sets) by Theorem 3.73 below, the claim follows from Proposition 3.55.

Exercise 3.64 Show that if $\tilde{\mathcal{B}} \subset \mathcal{B}$ is a continuously embedded Banach space with $\mu(\tilde{\mathcal{B}}) = 1$, then the embedding $\mathcal{B}^* \hookrightarrow \mathcal{R}_\mu$ extends to an embedding $\tilde{\mathcal{B}}^* \hookrightarrow \mathcal{R}_\mu$. Deduce from this that the restriction of μ to $\tilde{\mathcal{B}}$ is again a Gaussian measure. In particular, Kolmogorov's continuity criterion yields a Gaussian measure on $\mathcal{C}_0^\beta([0,1]^d, \mathbf{R})$.

The properties of the reproducing kernel space of a Gaussian measure allow us to give another illustration of the fact that measures on infinite-dimensional spaces behave in a rather different way from measures on $\mathbf{R}^n$:

Proposition 3.65 *Let μ be a centred Gaussian measure on a separable Banach space $\mathcal{B}$ such that $\dim \mathcal{H}_\mu = \infty$. Denote by D_c the dilatation by a real number c on $\mathcal{B}$, that is $D_c(x) = cx$. Then, μ and $D_c^\sharp\mu$ are mutually singular for every $c \neq \pm 1$.*

Proof. Since the reproducing Kernel space $\mathcal{R}_\mu$ is a separable Hilbert space, we can find an orthonormal basis $\{e_n\}_{n\geq 0}$. Consider the sequence of random variables $X_N(x) = \frac{1}{N}\sum_{n=1}^N |e_n(x)|^2$ over $\mathcal{B}$. If $\mathcal{B}$ is equipped with the measure μ then, since the e_n are independent under μ, we can apply the law of large numbers and deduce that

$$\lim_{N\to\infty} X_N(x) = 1 \,, \tag{3.19}$$

for μ-almost every x. On the other hand, it follows from the linearity of the e_n that when we equip $\mathcal{B}$ with the measure $D_c^\sharp\mu$, the e_n are still independent, but have variance c^2, so that

$$\lim_{N\to\infty} X_N(x) = c^2 \,,$$

for $D_c^\sharp\mu$-almost every x. This shows that if $c \neq \pm 1$, the set on which the convergence (3.19) takes place must be of $D_c^\sharp\mu$-measure 0, which implies that μ and $D_c^\sharp\mu$ are mutually singular. □

Note that this example also shows that just because two Gaussian measure share the same Cameron-Martin space, this does not mean that they are mutually equivalent! However, as already mentioned earlier, the importance of the Cameron-Martin space is that it represents precisely those directions in which one can translate the measure μ without changing its null sets:

Theorem 3.66 (Cameron-Martin) *For $h \in \mathcal{B}$, define the map $T_h\colon \mathcal{B} \to \mathcal{B}$ by $T_h(x) = x + h$. Then, the measure $T_h^\sharp \mu$ is absolutely continuous with respect to μ if and only if $h \in \mathcal{H}_\mu$. Furthermore, in the latter case, its Radon-Nikodym derivative is given by*

$$\frac{dT_h^\sharp \mu}{d\mu}(x) = \exp(h^*(x) - \tfrac{1}{2}\|h\|_\mu^2) \; .$$

Proof. Fix $h \in \mathcal{H}_\mu$ and let $h^* \in L^2(\mathcal{B}, \mu)$ be the corresponding element of the reproducing kernel. Since the law of h^* is Gaussian by Proposition 3.62, the map $x \mapsto \exp(h^*(x))$ is integrable. Since furthermore the variance of h^* is given by $\|h\|_\mu^2$, the function

$$\mathcal{D}_h(x) = \exp(h^*(x) - \tfrac{1}{2}\|h\|_\mu^2) \tag{3.20}$$

is strictly positive, belongs to $L^1(\mathcal{B}, \mu)$, and integrates to 1. It is therefore the Radon-Nikodym derivative of a measure μ_h that is absolutely continuous with respect to μ. To check that one has indeed $\mu_h = T_h^\sharp \mu$, it suffices to show that their Fourier transforms coincide. Assuming that $h^* \in \mathcal{B}^*$, one has

$$\begin{aligned}\hat{\mu}_h(\ell) &= \int_{\mathcal{B}} \exp(i\ell(x) + h^*(x) - \tfrac{1}{2}\|h\|_\mu^2)\, \mu(dx) = \exp(\tfrac{1}{2}C_\mu(i\ell + h^*, i\ell + h^*) - \tfrac{1}{2}\|h\|_\mu^2) \\ &= \exp(-\tfrac{1}{2}C_\mu(\ell,\ell) - iC_\mu(\ell, h^*)) = \exp(-\tfrac{1}{2}C_\mu(\ell,\ell) + i\ell(h)) \; .\end{aligned}$$

Using Proposition 3.62 for the joint law of ℓ and h^*, it is an easy exercise to check that this equality still holds for arbitrary $h \in \mathcal{H}_\mu$.

On the other hand, we have

$$\begin{aligned}\widehat{T_h^\sharp \mu}(\ell) &= \int_{\mathcal{B}} \exp(i\ell(x))\, T_h^\sharp \mu(dx) = \int_{\mathcal{B}} \exp(i\ell(x+h))\, \mu(dx) = e^{i\ell(h)} \int_{\mathcal{B}} \exp(i\ell(x))\, \mu(dx) \\ &= \exp(-\tfrac{1}{2}C_\mu(\ell,\ell) + i\ell(h)) \; ,\end{aligned}$$

showing that $\mu_h = T_h^\sharp \mu$.

To show the converse, note first that one can check by an explicit calculation that $\|\mathcal{N}(0,1) - \mathcal{N}(h,1)\|_{\text{TV}} \geq 2 - 2\exp(-\frac{h^2}{8})$. Fix now some arbitrary $n > 0$. If $h \notin \mathcal{H}_\mu$ then, by Exercise 3.60, there exists $\ell \in \mathcal{B}^*$ with $C_\mu(\ell,\ell) = 1$ such that $\ell(h) \geq n$. Since the image $\ell^\sharp \mu$ of μ under ℓ is $\mathcal{N}(0,1)$ and the image of $T_h^\sharp \mu$ under ℓ is $\mathcal{N}(-\ell(h), 1)$, this shows that

$$\|\mu - T_h^\sharp \mu\|_{\text{TV}} \geq \|\ell^\sharp \mu - \ell^\sharp T_h^\sharp \mu\|_{\text{TV}} = \|\mathcal{N}(0,1) - \mathcal{N}(-\ell(h),1)\|_{\text{TV}} \geq 2 - 2\exp(-\frac{n^2}{8}) \; .$$

Since this is true for every n, we conclude that $\|\mu - T_h^\sharp \mu\|_{\text{TV}} = 2$, thus showing that they are mutually singular. □

As a consequence, we have the following characterisation of the Cameron-Martin space:

Proposition 3.67 *The space $\mathcal{H}_\mu \subset \mathcal{B}$ is the intersection of all (measurable) linear subspaces of full measure. However, if $\mathcal{H}_\mu$ is infinite-dimensional, then one has $\mu(\mathcal{H}_\mu) = 0$.*

Proof. Take an arbitrary linear subspace $V \subset \mathcal{B}$ of full measure and take an arbitrary $h \in \mathcal{H}_\mu$. It follows from Theorem 3.66 that the affine space $V - h$ also has full measure. Since $(V - h) \cap V = \emptyset$ unless $h \in V$, one must have $h \in V$, so that $\mathcal{H}_\mu \subset \bigcap\{V \subset \mathcal{B} \,:\, \mu(V) = 1\}$.

Conversely, take an arbitrary $x \notin \mathcal{H}_\mu$ and let us construct a linear space $V \subset \mathcal{B}$ of full measure, but not containing x. Since $x \notin \mathcal{H}_\mu$, one has $\|x\|_\mu = \infty$ with $\|\cdot\|_\mu$ extended to $\mathcal{B}$ as in (3.18). Therefore, we can find a sequence $\ell_n \in \mathcal{B}^*$ such that $C_\mu(\ell_n, \ell_n) \leq 1$ and $\ell_n(x) \geq n$. Defining the norm $|y|^2 = \sum_n n^{-2}(\ell_n(y))^2$, we see that

$$\int_{\mathcal{B}} |y|^2\, \mu(dy) = \sum_{n=1}^\infty \frac{1}{n^2} \int_{\mathcal{B}} (\ell_n(y))^2\, \mu(dy) \leq \frac{\pi^2}{6} \; ,$$

so that the linear space $V = \{y : |y| < \infty\}$ has full measure. However, $|x| = \infty$ by construction, so that $x \notin V$.

To show that $\mu(\mathcal{H}_\mu) = 0$ if $\dim \mathcal{H}_\mu = \infty$, consider an orthonormal sequence $e_n \in \mathcal{R}_\mu$ so that the random variables $\{e_n(x)\}$ are i.i.d. $\mathcal{N}(0,1)$ distributed. By the second Borel-Cantelli lemma, it follows that $\sup_n |e_n(x)| = \infty$ for μ-almost every x, so that in particular $\|x\|_\mu^2 \geq \sum_n e_n^2(x) = \infty$ almost surely. □

We conclude by showing that while $\mathcal{H}_\mu$ has measure 0, its closure always has full measure, and is actually the smallest *closed* subspace of $\mathcal{B}$ with that property. Recall that the (topological) support $\operatorname{supp}\mu$ of a Borel measure μ on a complete separable metric space consists of those points x such that $\mu(U) > 0$ for *every* neighbourhood U of x. Alternatively, it consists of the intersection of all closed sets of full measure, and it does itself have full measure, see for example [Bog07, Chapter 7]. One then has

Proposition 3.68 *The support of a Gaussian measure μ coincides with the closure $\bar{\mathcal{H}}_\mu$ of $\mathcal{H}_\mu$ in $\mathcal{B}$.*

Proof. Similarly to the proof of Proposition 3.31, we see that one can find a countable number of linear functionals $\ell_n \in \mathcal{B}^*$ such that $\ell_n(\bar{\mathcal{H}}_\mu) = 0$ and such that $x \in \bar{\mathcal{H}}_\mu$ if and only if $\ell_n(x) = 0$ for all n.

On the other hand, combining Definition 3.49 and Proposition 3.54, we see that $\bar{\mathcal{H}}_\mu$ is given by the closure of the range of $\hat{C}_\mu$, so that $C_\mu(\ell_n, \ell_n) = 0$ for every n. Combining this with (3.8) shows that $\ell_n^\sharp \mu = \delta_0$ for every n, so that $\mu(\bigcap_n \ker \ell_n) = \mu(\bar{\mathcal{H}}_\mu) = 1$, which shows that $\operatorname{supp}\mu \subset \bar{\mathcal{H}}_\mu$.

Conversely, $x \in \operatorname{supp}\mu$ implies that $x + h \in \operatorname{supp}\mu$ for every $h \in \mathcal{H}_\mu$ (and therefore also for every $h \in \bar{\mathcal{H}}_\mu$) by Theorem 3.66. Since $\operatorname{supp}\mu$ cannot be empty, we conclude that $\operatorname{supp}\mu = \mathcal{H}_\mu$. □

3.5 Images of Gaussian measures

It follows immediately from the definition of a Gaussian measure and the expression for its Fourier transform that if μ is a Gaussian measure on some Banach space $\mathcal{B}$ and $A\colon \mathcal{B} \to \mathcal{B}_2$ is a bounded linear map for $\mathcal{B}_2$ some other Banach space, then $\nu = A^\sharp \mu$ is a Gaussian measure on $\mathcal{B}_2$ with covariance given by

$$C_\nu(\ell, \ell') = C_\mu(A^*\ell, A^*\ell') \,. \tag{3.21}$$

In this expression, $A^*\colon \mathcal{B}_2^* \to \mathcal{B}^*$ is the adjoint to A, that is the operator such that $(A^*\ell)(x) = \ell(Ax)$ for every $x \in \mathcal{B}$ and every $\ell \in \mathcal{B}_2^*$.

Recall now that $\mathcal{H}_\mu$ is the intersection over all linear subspaces of $\mathcal{B}$ that have full measure under μ. This suggests that in order to determine the image of μ under a linear map, it is sufficient to know how that map acts on elements of $\mathcal{H}_\mu$. This intuition is made precise by the following theorem:

Theorem 3.69 *Let μ be a centred Gaussian probability measure on a separable Banach space $\mathcal{B}$. Let furthermore $\mathcal{H}$ be a separable Hilbert space and let $A\colon \mathcal{H}_\mu \to \mathcal{H}$ be a Hilbert-Schmidt operator. (That is $AA^*\colon \mathcal{H} \to \mathcal{H}$ is trace class.) Then, there exists a measurable map $\hat{A}\colon \mathcal{B} \to \mathcal{H}$ such that $\nu = \hat{A}^\sharp \mu$ is Gaussian with covariance $C_\nu(h,k) = \langle A^*h, A^*k\rangle_\mu$. Furthermore, there exists a measurable linear subspace $V \subset \mathcal{B}$ of full μ-measure such that $\hat{A}$ restricted to V is linear and $\hat{A}$ restricted to $\mathcal{H}_\mu \subset V$ agrees with A.*

Proof. Let $\{e_n\}_{n\geq 1}$ be an orthonormal basis for $\mathcal{H}_\mu$ and denote by e_n^* the corresponding elements in $\mathcal{R}_\mu \subset L^2(\mathcal{B}, \mu)$ and define $S_N(x) = \sum_{n=0}^N e_n^*(x) A e_n$. Recall from Proposition 3.61 that we can find subspaces V_{e_n} of full measure such that e_n^* is linear on V_{e_n}. Define now a linear subspace $V \subset \mathcal{B}$ by

$$V = \Big\{ x \in \bigcap_{n\geq 0} V_{e_n} \,:\, \text{the sequence } \{S_N(x)\} \text{ converges in } \mathcal{H} \Big\} \,,$$

(the fact that V is linear follows from the linearity of each of the e_n^*) and set

$$\hat{A}(x) = \begin{cases} \lim_{N\to\infty} S_N(x) & \text{for } x \in V, \\ 0 & \text{otherwise.} \end{cases}$$

Since the random variables $\{e_n^*\}$ are i.i.d. $\mathcal{N}(0,1)$-distributed under μ, the sequence $\{S_N\}$ forms an $\mathcal{H}$-valued martingale and one has

$$\sup_N \mathbf{E}_\mu \|S_N(x)\|^2 = \sum_{n=0}^{\infty} \|Ae_n\|^2 \le \operatorname{tr} A^*A < \infty \,,$$

where the last inequality is a consequence of A being Hilbert-Schmidt. It follows that $\mu(V) = 1$ by Doob's martingale convergence theorem.

To see that $\nu = \hat{A}^\sharp\mu$ has the stated property, fix an arbitrary $h \in \mathcal{H}$ and note that the series $\sum_{n\ge1} e_n^*\langle Ae_n, h\rangle$ converges in $\mathcal{R}_\mu$ to an element with covariance $\|A^*h\|^2$. The statement then follows from Proposition 3.62 and the fact that $C_\nu(h,h)$ determines C_ν by polarisation. To check that ν is Gaussian, we can compute its Fourier transform in a similar way. □

The proof of Theorem 3.69 can easily be extended to the case where the image space is a Banach space rather than a Hilbert space. However, in this case we cannot give a straightforward characterisation of those maps A that are "admissible", since we have no good complete characterisation of covariance operators for Gaussian measures on Banach spaces. However, we can take the pragmatic approach and simply assume that the new covariance determines a Gaussian measure on the target Banach space. With this approach, we can formulate the following version for Banach spaces:

Proposition 3.70 *Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two separable Banach space and let μ be a centred Gaussian probability measure on $\mathcal{B}_1$. Let $A\colon \mathcal{H}_\mu \to \mathcal{B}_2$ be a bounded linear operator such that there exists a centred Gaussian measure ν on $\mathcal{B}_2$ with covariance $C_\nu(h,k) = \langle A^*h, A^*k\rangle_\mu$. Then, there exists a measurable map $\hat{A}\colon \mathcal{B}_1 \to \mathcal{B}_2$ such that $\nu = \hat{A}^\sharp\mu$ and such that there exists a measurable linear subspace $V \subset \mathcal{B}$ of full μ-measure such that $\hat{A}$ restricted to V is linear and $\hat{A}$ restricted to $\mathcal{H}_\mu \subset V$ agrees with A.*

Proof. As a first step, we construct a Hilbert space $\mathcal{H}_2$ such that $\mathcal{B}_2 \subset \mathcal{H}_2$ as a Borel subset. Denote by $\mathcal{H}_\nu \subset \mathcal{B}_2$ the Cameron-Martin space of ν and let $\{e_n\} \subset \mathcal{H}_\nu$ be an orthonormal basis of elements such that $e_n^* \in \mathcal{B}_2^*$ for every n. (Such an orthonormal basis can always be found by using the Grahm-Schmidt procedure.) We then define a norm on $\mathcal{B}_2$ by

$$\|x\|_2^2 = \sum_{n\ge1} \frac{e_n^*(x)^2}{n^2\|e_n^*\|^2} \,,$$

where $\|e_n^*\|$ is the norm of e_n^* in $\mathcal{B}_2^*$. It is immediate that $\|x\|_2 < \infty$ for every $x \in \mathcal{B}_2$, so that this turns $\mathcal{B}_2$ into a pre-Hilbert space. We finally define $\mathcal{H}_2$ as the completion of $\mathcal{B}_2$ under $\|\cdot\|_2$.

Denote by ν' the image of the measure ν under the inclusion map $\iota\colon \mathcal{B}_2 \hookrightarrow \mathcal{H}_2$. It follows that the map $A' = \iota \circ A$ satisfies the assumptions of Theorem 3.69, so that there exists a map $\hat{A}\colon \mathcal{B}_1 \to \mathcal{H}_2$ which is linear on a subset of full μ-measure and such that $\hat{A}^\sharp\mu = \nu'$. On the other hand, we know by construction that $\nu'(\mathcal{B}_2) = 1$, so that the set $\{x : \hat{A}x \in \mathcal{B}_2\}$ is of full measure. Modifying $\hat{A}$ outside of this set by for example setting it to 0 and using Exercise 3.64 then yields the required statement. □

To conclude this section, we show that the Cameron-Martin space of the image measure is given by the image of the Cameron-Martin space of the original measure. More precisely, one has:

Proposition 3.71 *In the setting of Proposition 3.70, we have $\mathcal{H}_\nu = A\mathcal{H}_\mu$ and $\|h\|_\nu = \inf\{\|k\|_\mu : Ak = h\}$.*

Proof. It follows immediately from the characterisation (3.18) that $\|Ak\|_\nu \le \|k\|_\mu$, so that $A\mathcal{H}_\mu \subset \mathcal{H}_\nu$. On the other hand, since $\nu(V) \ge \mu(\hat{A}V)$ for every measurable linear subspace V, one has $\mathcal{H}_\nu \subset \bigcap\{\hat{A}V : \mu(V) = 1\} = \hat{A}\mathcal{H}_\mu = A\mathcal{H}_\mu$.

It remains to show that for every $h \in \mathcal{H}_\nu$ there exists $k \in \mathcal{H}_\mu$ such that $Ak = h$ and such that $\|h\|_\nu = \|k\|_\mu$. For this, consider the adjoint map $A^*\colon \mathcal{H}_\nu \to \mathcal{H}_\mu$. If we identify (the dual space of) $\mathcal{H}_\nu$ with $\mathcal{R}_\nu$ via $h \leftrightarrow h^*$ as in Proposition 3.55, then we can check that one has the identity $(A^*h)^*(x) = h^*(\hat{A}x)$. This follows from the fact that for $k \in \mathcal{H}_\mu$ one has

$$(A^*h)^*(k) = \langle A^*h, k\rangle_\mu = \langle h, Ak\rangle_\nu = h^*(Ak)\ ,$$

where the middle identity follows from the definiton of the adjoint and the two remaining identities follow from Definition 3.49. On the other hand, since $\nu = \hat{A}^\sharp\mu$, composition with $\hat{A}$ is an isometry from $L^2(\mathcal{B}_2, \nu)$ into $L^2(\mathcal{B}_1, \mu)$, so that AA^* is the identity. We conclude that $k = A^*h$ has the required properties since $Ak = (AA^*)h = h$ and $\|k\|_\mu^2 = \|A^*h\|_\mu^2 = \|h\|_\nu^2$. □

add proof

Exercise 3.72 Show that if μ is a Gaussian measure on a separable Banach space $\mathcal{B}$ with Cameron-Martin space $\mathcal{H}_\mu$, then every bounded map $A\colon \mathcal{H}_\mu \to \mathcal{H}_\mu$ can be extended uniquely to a measurable map $\hat{A}\colon \mathcal{B} \to \mathcal{B}$ which is linear on a subspace of full measure and such that $\hat{A}^\sharp\mu$ is again Gaussian on $\mathcal{B}$.

3.5.1 Uniqueness of measurable extensions and the isoperimetric inequality

This section is devoted to a proof of the converse of Theorem 3.69 and Proposition 3.70, namely

Theorem 3.73 *Let μ be a Gaussian measure on a separable Banach space $\mathcal{B}_1$ with Cameron-Martin space $\mathcal{H}_\mu$ and let $A\colon \mathcal{H}_\mu \to \mathcal{B}_2$ be a linear map satisfying the assumptions of Proposition 3.70. Then the extension $\hat{A}$ of A is unique (up to sets of measure 0) in the class of maps such that there exists a measurable linear subspace V with $\mu(V) = 1$ such that $\hat{A}$ is linear on V, and $\hat{A}x = Ax$ for $x \in \mathcal{H}_\mu \subset V$.*

Remark 3.74 Note that unlike in the statement of Proposition 3.70, we do *not* a priori impose that any two extensions $\hat{A}_1$ and $\hat{A}_2$ are such that $\hat{A}_1^\sharp\mu = \hat{A}_2^\sharp\mu$! As a consequence of this result, the precise Banach spaces $\mathcal{B}_1$ and $\mathcal{B}_2$ are completely irrelevant when one considers the image of a Gaussian measure under a linear transformation. The only thing that matters is the Cameron-Martin space for the starting measure and the way in which the linear transformation acts on this space. This fact will be used repeatedly in the sequel.

This is probably one of the most remarkable results in Gaussian measure theory. At first sight, it appears completely counterintuitive: the Cameron-Martin space $\mathcal{H}_\mu$ has measure 0, so how can the specification of a measurable map on a set of measure 0 be sufficient to determine it on a set of measure 1? Part of the answer lies of course in the requirement that the extension $\hat{A}$ should be linear on a set of full measure. However, even this requirement would not be sufficient by itself to determine $\hat{A}$ since the Hahn-Banach theorem provides a huge number of different extension of A that do not coincide anywhere except on $\mathcal{H}_\mu$. The missing ingredient that solves this mystery is the requirement that $\hat{A}$ is not just any linear map, but a *measurable* linear map. This additional constraint rules out all of the non-constructive extensions of A provided by the Hahn-Banach theorem and leaves only one (constructive) extension of A.

The main ingredient in the proof of Theorem 3.73 is the Borell-Sudakov-Cirel'son inequality [SC74, Bor75], a general form of isoperimetric inequality for Gaussian measures which is very interesting

and useful in its own right. In order to state this result, we first introduce the notation B_ε for the open $\mathcal{H}_\mu$-ball of radius ε centred at the origin. We also denote by $A + B$ the sum of two sets defined by

$$A + B = \{x + y : x \in A\,, \quad y \in B\}\,,$$

and we denote by Φ the distribution function of the normal Gaussian: $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2}\,ds$. With these notations at hand, we have the following:

Theorem 3.75 (Borell-Sudakov-Cirel'son) *Let μ be a Gaussian measure on a separable Banach space $\mathcal{B}$ with Cameron-Martin space $\mathcal{H}_\mu$ and let $A \subset \mathcal{B}$ be a measurable subset with measure $\mu(A) = \Phi(\alpha)$ for some $\alpha \in \mathbf{R}$. Then, for every $\varepsilon > 0$, one has the bound $\mu(A + B_\varepsilon) \geq \Phi(\alpha + \varepsilon)$.*

Remark 3.76 Theorem 3.75 is remarkable since it implies that even though $\mathcal{H}_\mu$ itself has measure 0, whenever A is a set of positive measure, no matter how small, the set $A + \mathcal{H}_\mu$ has full measure!

Remark 3.77 The bound given in Theorem 3.75 is sharp whenever A is a half space, in the sense that $A = \{x \in \mathcal{B} : \ell(x) \geq c\}$ for some $\ell \in \mathcal{R}_\mu$ and $c \in \mathbf{R}$. In the case where ε is small, $(A + B_\varepsilon) \setminus A$ is a fattened boundary for the set A, so that $\mu(A + B_\varepsilon) - \mu(A)$ can be interpreted as a kind of "perimeter" for A. The statement can then be interpreted as stating that in the context of Gaussian measures, half-spaces are the sets of given perimeter that have the largest measure. This justifies the statement that Theorem 3.75 is an isoperimetric inequality.

We are not going to give a proof of Theorem 3.75 in these notes because this would lead us too far astray from our main object of study. The interested reader may want to look into the monograph [LT91] for a more exhaustive treatment of probability theory in Banach spaces in general and isoperimetric inequalities in particular. Let us nevertheless remark shortly on how the argument of the proof goes, as it can be found in the original papers [SC74, Bor75]. In a nutshell, it is a consequence of the two following remarks:

- Let ν_M be the uniform measure on a sphere of radius $\sqrt{M}$ in $\mathbf{R}^M$ and let $\Pi_{M,n}$ be the orthogonal projection from $\mathbf{R}^M$ to $\mathbf{R}^n$. Then, the sequence of measures $\Pi_{M,n}\nu_M$ converges as $M \to \infty$ to the standard Gaussian measure on $\mathbf{R}^n$. This remark is originally due to Poincaré.
- A claim similar to that of Theorem 3.75 holds for the uniform measure on the sphere, in the sense that the volume of a fattened set $A + B_\varepsilon$ on the sphere is bounded from below by the volume of a fattened "cap" of volume identical to that of A. Originally, this fact was discovered by Lévy, and it was then later generalised by Schmidt, see [Sch48] or the review article [Gar02].

These two facts can then be combined in order to show that half-spaces are optimal for finite-dimensional Gaussian measures. Finally, a clever approximation argument is used in order to generalise this statement to infinite-dimensional measures as well.

An immediate corollary is given by the following zero-one law for Gaussian measures:

Corollary 3.78 *Let $V \subset \mathcal{B}$ be a measurable linear subspace. Then, one has either $\mu(V) = 0$ or $\mu(V) = 1$.*

Proof. Let us first consider the case where $\mathcal{H}_\mu \not\subset V$. In this case, just as in the proof of Proposition 3.67, we conclude that $\mu(V) = 0$, for otherwise we could construct an uncountable collection of disjoint sets with positive measure.

If $\mathcal{H}_\mu \subset V$, then we have $V + B_\varepsilon = V$ for every $\varepsilon > 0$, so that if $\mu(V) > 0$, one must have $\mu(V) = 1$ by Theorem 3.75. □

We have now all the necessary ingredients in place to be able to give a proof of Theorem 3.73:

Proof of Theorem 3.73. Assume by contradiction that there exist two measurable extensions $\hat{A}_1$ and $\hat{A}_2$ of A. In other words, we have $\hat{A}_i x = Ax$ for $x \in \mathcal{H}_\mu$ and there exist measurable subspaces V_i with

$\mu(V_i) = 1$ such that the restriction of $\hat{A}_i$ to V_i is linear. Denote $V = V_1 \cap V_2$ and $\Delta = \hat{A}_2 - \hat{A}_1$, so that Δ is linear on V and $\Delta|_{\mathcal{H}_\mu} = 0$.

Let $\ell \in \mathcal{B}_2^*$ be arbitrary and consider the events $V_\ell^c = \{x : \ell(\Delta x) \le c\}$. By the linearity of Δ, each of these events is invariant under translations in $\mathcal{H}_\mu$, so that by Theorem 3.75 we have $\mu(V_\ell^c) \in \{0, 1\}$ for every choice of ℓ and c. Furthermore, for fixed ℓ, the map $c \mapsto \mu(V_\ell^c)$ is increasing and it follows from the σ-additivity of μ that we have $\lim_{c\to-\infty} \mu(V_\ell^c) = 0$ and $\lim_{c\to\infty} \mu(V_\ell^c) = 1$. Therefore, there exists a unique $c_\ell \in \mathbf{R}$ such that $\mu(V_\ell^c)$ jumps from 0 to 1 at $c = c_\ell$. In particular, this implies that $\ell(\Delta x) = c_\ell$ μ-almost surely. However, the measure μ is invariant under the map $x \mapsto -x$, so that we must have $c_\ell = -c_\ell$, implying that $c_\ell = 0$. Since this is true for every $\ell \in \mathcal{B}_2^*$, we conclude from Proposition 3.31 that the law of Δx is given by the Dirac measure at 0, so that $\Delta x = 0$ μ-almost surely, which is precisely what we wanted. □

The remainder of this chapter is devoted to two applications of the extension theorem for Gaussian measures, Theorem 3.73. First, we show that a Gaussian measure on a product space can always be "disintegrated" against one factor into regular conditional probabilities that are Gaussian on the other factor. Then, we will see how we can take advantage of the extension theorem to construct a theory of stochastic integration with respect to a "cylindrical Wiener process", which is the infinite-dimensional analogue of a standard n-dimensional Wiener process.

3.6 Disintegration of Gaussian measures

The setup considered in the section is the following. We are given a Gaussian measure μ on a space $\mathcal{B}$ which can be written as $\mathcal{B} = \mathcal{B}_1 \oplus \mathcal{B}_2$ for two separable Banach spaces $\mathcal{B}_i$. We denote by $\Pi_i \colon \mathcal{B} \to \mathcal{B}_i$ the canonical projections so that $\Pi_2\Pi_1 = \Pi_1\Pi_2 = 0$. We also denote as before by $\mathcal{H}_\mu \subset \mathcal{B}$ the Cameron-Martin space of μ. The aim is to find a (reasonably) explicit expression for a map $\mathcal{P}$ from $\mathcal{B}_1$ into the set of probability measures on $\mathcal{B}_2$ such that the identity

$$\int_{\mathcal{B}} \varphi(x)\, \mu(dx) = \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} \varphi(x_1 + x_2)\, \mathcal{P}(x_1)(dx_2)\, \mu_1(dx_1)\ ,$$

holds for every measurable bounded function $\varphi \colon \mathcal{B} \to \mathbf{R}$, where $\mu_1 = \Pi_1^\sharp \mu$ is the projection of μ onto the first factor. In order to keep notations simple, we make the slight abuse of notation by identifying the spaces $\mathcal{B}_i$ with closed subspaces of $\mathcal{B}$. In the finite-dimensional case, the measures $\mathcal{P}(x_1)$ always consist of translates of a common centred Gaussian measure μ_2^c by an amount Px_1 for some linear map $P \colon \mathcal{B}_1 \to \mathcal{B}_2$. Maybe not surprisingly, this is still the case for Gaussian measures on arbitrary Banach spaces. The aim of this section is to justify this fact rigorously and to give explicit expressions for the conditional measure μ_2^c and the conditional mean P.

Example 3.79 Consider the measure μ on $\mathbf{R}^2$ with covariance given by

$$C_\mu = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \qquad \Rightarrow \qquad C_\mu^{-1} = \frac{1}{5}\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}.$$

The projected measure μ_1 then has variance 2. Furthermore, the density of μ with respect to Lebesgue measure is proportional to $\exp(-\frac{1}{2}\langle x, C_\mu^{-1} x\rangle)$, so that the conditional measure for x_2 given x_1 is proportional to $\exp(-\frac{1}{5}x_2^2 + \frac{1}{5}x_1x_2)$, which in turn is proportional to $\exp(-\frac{1}{5}(x_2 - \frac{1}{2}x_1)^2)$. In this particular case, we thus see that $\mathcal{P}(x_1)$ is Gaussian with variance $\frac{5}{2}$ and mean $-\frac{1}{2}x_1$.

We start by denoting by $\mathcal{H}_i^p$ the Cameron-Martin spaces of the projected measures $\mu_i = \Pi_i^\sharp \mu$, which by Proposition 3.71 are given by $\mathcal{H}_i^p = \Pi_i \mathcal{H}_\mu \subset \mathcal{B}_i$. We also denote by C_i their covariance operators, which we interpret in this section as operators from $\mathcal{B}_i^*$ to $\mathcal{B}_i$. It follows from (3.21) that these covariances are given by $C_i = \Pi_i C_\mu \Pi_i^*$. Furthermore, we note that the Cameron-Martin norm of elements h in the image of C_i is given by

$$\|h\|_{i,p}^2 = \langle h, C_i^{-1} h\rangle\ ,$$

where we denote by $\langle \cdot, \cdot \rangle$ the pairing between $\mathcal{B}_i$ and $\mathcal{B}_i^*$.

It may be worth to note at this stage that $\mathcal{H}_i^p$ is in general *not* a subspace of $\mathcal{H}_\mu$. This can be seen already in the simplest possible case $\mathcal{B} = \mathbf{R}^2$ and $\mathcal{B}_i = \mathbf{R}$ by taking for μ image of the normal Gaussian $\mathcal{N}(0,1)$ under the map $x \mapsto (x,x)$. However, there exists a natural isomorphism between $\mathcal{H}_i^p$ and some closed subspace of $\mathcal{H}_\mu$ in the following way. For x in the image of C_i, define $U_i x = C_\mu \Pi_i^* C_i^{-1} x \in \mathcal{B}$. One then has:

Lemma 3.80 *For every x in the image of C_i, one has $U_i x \in \mathcal{H}_\mu$. Furthermore, the map U_i extends to an isometry between $\mathcal{H}_i^p$ and $U_i \mathcal{H}_i^p \subset \mathcal{H}_\mu$.*

Proof. Since $U_i x$ belongs to the image of C_μ by construction, one has $\|U_i x\|^2 = \langle U_i x, C_\mu^{-1} U_i x \rangle = \langle C_\mu \Pi_i^* C_i^{-1} x, \Pi_i^* C_i^{-1} x \rangle = \langle C_i C_i^{-1} x, C_i^{-1} x \rangle = \langle x, C_i^{-1} x \rangle = \|x\|_{i,p}^2$. The claim follows from the fact that the image of C_i is dense in $\mathcal{H}_i^p$. □

Remark 3.81 This calculation also shows that although $C_i^{-1} x$ is only defined modulo elements in the kernel of C_i, $U_i x$ is well-defined as an element of $\mathcal{H}_i^p$.

We denote by $\hat{\mathcal{H}}_i^p$ the images of $\mathcal{H}_i^p$ under U_i. If we identify $\mathcal{H}_\mu$ with $\mathcal{R}_\mu$, we see that $\hat{\mathcal{H}}_i^p$ is actually nothing but the closure in $\mathcal{R}_\mu$ of the image of $\mathcal{B}_i^*$ under the dual map Π_i^*. Denoting by $\hat{\Pi}_i^p \colon \mathcal{H}_\mu \to \mathcal{H}_\mu$ the orthogonal projection (in $\mathcal{H}_\mu$) onto $\hat{\mathcal{H}}_i^p$, it is a straightforward calculation to check that one has the identity $\hat{\Pi}_i^p x = U_i \Pi_i x$. On the other hand, it follows from the definition of C_i that $\Pi_i U_i x = x$, so that $\Pi_i \colon \hat{\mathcal{H}}_i^p \to \mathcal{H}_i^p$ is the inverse of the isomorphism U_i.

We can also define subspaces $\mathcal{H}_i^c$ of $\mathcal{H}$ by

$$\mathcal{H}_i^c = \overline{\mathcal{H}_\mu \cap \mathcal{B}_i} = \overline{\mathcal{H}_\mu \cap \mathcal{H}_i^p}\,, \tag{3.22}$$

Here, the closures are taken with respect to the topology of $\mathcal{H}_\mu$. The spaces $\mathcal{H}_i^c$ are again Hilbert spaces. (They inherit their structure from $\mathcal{H}_\mu$, not from $\mathcal{H}_i^p$!) Since they are contained in $\mathcal{H}_\mu$, it follows from Exercise 3.72 that the orthogonal projection $\Pi_i^c \colon \mathcal{H}_\mu \to \mathcal{H}_i^c$ extends to a set of full measure, so that we can construct measures $\mu_i^c = (\Pi_i^c)^\sharp \mu$ on $\mathcal{B}_i$. Note that for $x \in \mathcal{H}_i^c \cap \mathcal{H}_i^p$, one has $\|x\|_\mu \geq \|x\|_{i,p}$, so that the inclusion $\mathcal{H}_i^c \subset \mathcal{H}_i^p$ holds. The relation between the spaces $\mathcal{H}_i^c$ and $\hat{\mathcal{H}}_i^p$ is the following:

Lemma 3.82 *One has $\mathcal{H}_1^c = (\hat{\mathcal{H}}_2^p)^\perp$ and vice-versa.*

Proof. Under the identification $h \leftrightarrow h^*$, $\hat{\mathcal{H}}_2^p$ is nothing but the closure of the image of Π_2^* in $\mathcal{R}_\mu$. On the other hand, $\mathcal{H}_1^c$ is a subset of the image of Π_1. Note now that the scalar product in $\mathcal{H}_\mu$ is an extension of the duality pairing between $\mathcal{B}$ and $\mathcal{B}^*$. □

We now define a (continuous) operator $P \colon \mathcal{H}_1^p \to \mathcal{H}_2^p$ by $Px = \Pi_2 U_1 x$. It follows from the previous remarks that P is unitarily equivalent to the orthogonal projection (in $\mathcal{H}$) from $\hat{\mathcal{H}}_1^p$ to $\hat{\mathcal{H}}_2^p$. Furthermore, one has $Px = U_1 x - x$, so that

$$\|Px\|_\mu \leq \|x\|_{1,p} + \|x\|_\mu\,, \tag{3.23}$$

which, combined with (3.22), shows that P can be extended to a bounded operator from $\mathcal{H}_1^c$ to $\mathcal{H}_2^c$. It follows from the Gaussian measurable extension theorem that P can be extended uniquely to a measurable map $\hat{P} \colon \mathcal{B}_1 \to \mathcal{B}_2$ which is linear on a set of full μ_1-measure. With these notations at hand, the main statement of this section is given by:

Proposition 3.83 *The measure μ admits the disintegration*

$$\int \varphi(x)\, \mu(dx) = \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} \varphi(x + \hat{P}x + y)\, \mu_2^c(dy)\, \mu_1(dx)\,. \tag{3.24}$$

Proof. Denote by ν the measure on the right hand side. Since ν is the image of the Gaussian measure $\mu_1 \otimes \mu_2^c$ under the measurable linear map $A\colon (x,y) \mapsto x + \hat{P}x + y$, the claim follows from Proposition 3.70 if we can show that A is an isometry between $\mathcal{H}_1^p \oplus \mathcal{H}_2^c$ and $\mathcal{H}_\mu$. This is equivalent to the fact that the operator $x \mapsto x + Px = x + \Pi_2 U_1 x$ from $\mathcal{H}_1^p$ to $\mathcal{H}$ is an isometry between $\mathcal{H}_1^p$ and $(\mathcal{H}_2^c)^\perp$. On the other hand, we know from Lemma 3.82 that $(\mathcal{H}_2^c)^\perp = \hat{\mathcal{H}}_1^p$ and we know from Lemma 3.80 that U_1 is an isomorphism between $\mathcal{H}_1^p$ and $\hat{\mathcal{H}}_1^p$. Finally, it follows from the definitions that $\Pi_1 U_1 x = C_1 C_1^{-1} x = x$ for every $x \in \mathcal{H}_1^p$, so that one has $x + \Pi_2 U_1 x = (\Pi_1 + \Pi_2) U_1 x = U_1 x$, which completes the proof. □

Exercise 3.84 Suppose that $\mathcal{B} = \mathbf{R}^n \oplus \mathbf{R}^m$ and that the covariance matrix C of the Gaussian measure μ on $\mathcal{B}$ is given in block form as

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{pmatrix} .$$

Show that in this case, the covariance matrix of μ_1 is given by C_{11}, the covariance of μ_2^c is given by $C_{22} - C_{12}^T C_{11}^{-1} C_{12}$, and the linear map P is given by $P = C_{12}^T C_{11}^{-1/2}$.

3.7 Cylindrical Wiener processes and stochastic integration

Central to the theory of stochastic PDEs is the notion of a *cylindrical Wiener process*, which can be thought of as an infinite-dimensional generalisation of a standard n-dimensional Wiener process. Before we proceed to the definition and construction of such a cylindrical Wiener process, let us recall a few basic facts about stochastic process.

In general, a stochastic process X over a probability space $(\Omega, \mathbf{P})$ and taking values in a separable Banach space $\mathcal{B}$ is nothing but a collection $\{X(t)\}$ of $\mathcal{B}$-valued random variables indexed by time $t \in \mathbf{R}$ (or taking values in some subset of $\mathbf{R}$). By Kolmogorov's extension theorem, we can also view this as a map $X\colon \Omega \to \mathcal{B}^{\mathbf{R}}$, where $\mathcal{B}^{\mathbf{R}}$ is endowed with the product sigma-algebra. A notable special case which will be of interest here is the case where the probability space is taken to be for example $\Omega = \mathcal{C}([0,T], \mathcal{B})$ (or some other space of $\mathcal{B}$-valued continuous functions) endowed with some Gaussian measure $\mathbf{P}$ and where the process X is given by

$$X(t)(\omega) = \omega(t) , \qquad \omega \in \Omega .$$

In this case, X is called the *canonical process* on Ω.

The usual (one-dimensional) Wiener process is a real-valued centred Gaussian process $B(t)$ such that $B(0) = 0$ and $\mathbf{E}|B(t) - B(s)|^2 = |t - s|$ for any pair of times s, t. From our point of view, the Wiener process on any finite time interval I can always be realised as the canonical process for the Gaussian measure on $\mathcal{C}(I, \mathbf{R})$ with covariance function $C(s,t) = s \wedge t = \min\{s,t\}$. (Note that such a measure exists by Kolmogorov's continuity criterion.)

Since the space $\mathcal{C}(\mathbf{R}, \mathbf{R})$ is not a Banach space and we have not extended our study of Gaussian measures to Fréchet spaces, we refrain from defining a measure on it. However, one can define Wiener measure on a separable Banach space of the type

$$\mathcal{C}_\varrho(\mathbf{R}_+, \mathbf{R}) = \Big\{ f \in \mathcal{C}(\mathbf{R}_+, \mathbf{R}) \,:\, \lim_{t\to\infty} f(t)/\varrho(t) \text{ exists} \Big\} , \qquad \|f\|_\varrho = \sup_{t \in \mathbf{R}} \frac{|f(t)|}{\varrho(t)} ,$$

for a suitable weight function $\varrho\colon \mathbf{R} \to [1, \infty)$. For example, we will see that $\varrho(t) = 1 + t^2$ is suitable, and we will therefore define $\mathcal{C}_W = \mathcal{C}_\varrho$ for this particular choice.

Proposition 3.85 *There exists a Gaussian measure μ on $\mathcal{C}_W$ with covariance function $C(s,t) = s \wedge t$.*

Proof. We use the fact that $f \in \mathcal{C}([0,\pi],\mathbf{R})$ if and only if the function $T(f)$ given by $T(f)(t) = (1+t^2)f(\arctan t)$ belongs to $\mathcal{C}_W$. Our aim is then to construct a Gaussian measure μ_0 on $\mathcal{C}([0,\pi],\mathbf{R})$ which is such that $T^\sharp\mu_0$ has the required covariance structure.

The covariance C_0 for μ_0 is then given by

$$C_0(x,y) = \frac{\tan x \wedge \tan y}{(1+\tan^2 x)(1+\tan^2 y)} \; .$$

It is now a straightforward exercise to check that this covariance function does indeed satisfy the assumption of Kolmogorov's continuity theorem. □

The standard n-dimensional Wiener process is simply given by n independent copies of a standard one-dimensional Wiener process, so that its covariance is given by

$$\mathbf{E}W_i(s)W_j(t) = (s\wedge t)\delta_{i,j} \; .$$

In other words, if u and v are any two elements in $\mathbf{R}^n$, we have

$$\mathbf{E}\langle u, W(s)\rangle\langle W_j(t), v\rangle = (s\wedge t)\langle u, v\rangle \; .$$

This is the characterisation that we will now extend to an arbitrary separable Hilbert space $\mathcal{H}$. One natural way of constructing such an extension would be to fix an orthonormal basis $\{e_n\}_{n\geq 1}$ of $\mathcal{H}$ and a countable collection $\{W_n\}$ of independent one-dimensional Wiener processes, and to set

$$W(t) \stackrel{\text{def}}{=} \sum_{n\geq 1} W_n(t)\, e_n \; . \tag{3.25}$$

The problem with such a construction is that this sum will almost surely not converge in $\mathcal{H}$! We therefore have to be a little bit more careful, but (3.25) is nevertheless how one should think of a cylindrical Wiener process on $\mathcal{H}$.

From now on, we fix a (separable) Hilbert space $\mathcal{H}$, as well as a larger Hilbert space $\mathcal{H}'$ containing $\mathcal{H}$ as a dense subset and such that the inclusion map $\iota\colon \mathcal{H} \to \mathcal{H}'$ is Hilbert-Schmidt. Given $\mathcal{H}$, it is always possible to construct a space $\mathcal{H}'$ with this property: choose an orthonormal basis $\{e_n\}$ of $\mathcal{H}$ and take $\mathcal{H}'$ to be the completion of $\mathcal{H}$ under the norm

$$\|x\|_{\mathcal{H}'}^2 = \sum_{n=1}^\infty \frac{1}{n^2}\langle x, e_n\rangle^2 \; .$$

One can check that in the framework of this construction, the map $\iota\iota^*$ is then given by $\iota\iota^* e_n = \frac{1}{n^2}e_n$, so that, since $\{ne_n\}$ is an orthonormal basis of $\mathcal{H}'$, it is indeed trace class.

Definition 3.86 Let $\mathcal{H}$ and $\mathcal{H}'$ be as above. We then call a *cylindrical Wiener process on* $\mathcal{H}$ any $\mathcal{H}'$-valued Gaussian process W such that

$$\mathbf{E}\langle h, W(s)\rangle_{\mathcal{H}'}\langle W(t), k\rangle_{\mathcal{H}'} = (s\wedge t)\langle \iota^* h, \iota^* k\rangle = (s\wedge t)\langle \iota\iota^* h, k\rangle_{\mathcal{H}'} \; , \tag{3.26}$$

for any two times s and t and any two elements $h, k \in \mathcal{H}'$. By Kolmogorov's continuity theorem, this can be realised as the canonical process for some Gaussian measure on $\mathcal{C}_W(\mathbf{R},\mathcal{H}')$.

Alternatively, we could have defined the cylindrical Wiener process on $\mathcal{H}$ as the canonical process associated to any Gaussian measure with Cameron-Martin space $H_0^{1,2}([0,T],\mathcal{H})$, see Exercise 3.58.

Proposition 3.87 *In the same setting as above, the Gaussian measure μ on $\mathcal{H}'$ with covariance $\iota\iota^*$ has $\mathcal{H}$ as its Cameron-Martin space. Furthermore, $\|h\|_\mu^2 = \|h\|^2$ for every $h \in \mathcal{H}$.*

Proof. It follows from the definition of $\mathring{\mathcal{H}}_\mu$ that this is precisely the range of $\iota\iota^*$ and that the map

$h \mapsto h^*$ is given by $h^* = (\iota\iota^*)^{-1}h$. In particular, $\mathring{\mathcal{H}}_\mu$ is contained in the range of ι. Therefore, for any $h, k \in \mathring{\mathcal{H}}_\mu$, there exist $\hat{h}, \hat{g} \in \mathcal{H}$ such that $h = \iota\hat{h}$ and $k = \iota\hat{k}$. Using this, we have

$$\langle h, k\rangle_\mu = \langle (\iota\iota^*)h^*, k^*\rangle_{\mathcal{H}'} = \langle h, (\iota\iota^*)^{-1}k\rangle_{\mathcal{H}'} = \langle \iota\hat{h}, (\iota\iota^*)^{-1}\iota\hat{k}\rangle_{\mathcal{H}'} = \langle \hat{h}, \iota^*(\iota\iota^*)^{-1}\iota\hat{k}\rangle = \langle \hat{h}, \hat{k}\rangle \ ,$$

from which the claim follows. □

The name "cylindrical Wiener process on $\mathcal{H}$" may sound confusing at first, since it is actually *not* an $\mathcal{H}$-valued process. (A better terminology may have been "cylindrical Wiener process over $\mathcal{H}$", but we choose to follow the convention that is found in the literature.) Note however that if h is an element in $\mathcal{H}$ that is in the range of ι^* (so that ιh belongs to the range of $\iota\iota^*$ and $\iota^*(\iota\iota^*)^{-1}\iota h = h$), then

$$\langle h, k\rangle = \langle \iota^*(\iota\iota^*)^{-1}\iota h, k\rangle = \langle (\iota\iota^*)^{-1}\iota h, \iota k\rangle_{\mathcal{H}'} \ .$$

In particular, if we just *pretend* for a moment that $W(t)$ belongs to $\mathcal{H}$ for every t (which is of course not true!), then we get

$$\begin{aligned}\mathbf{E}\langle h, W(s)\rangle\langle W(t), k\rangle &= \mathbf{E}\langle (\iota\iota^*)^{-1}\iota h, \iota W(s)\rangle_{\mathcal{H}'}\langle (\iota\iota^*)^{-1}\iota k, \iota W(t)\rangle_{\mathcal{H}'} \\ &= (s\wedge t)\langle \iota\iota^*(\iota\iota^*)^{-1}\iota h, (\iota\iota^*)^{-1}\iota k\rangle_{\mathcal{H}'} \\ &= (s\wedge t)\langle \iota h, (\iota\iota^*)^{-1}\iota k\rangle_{\mathcal{H}'} = (s\wedge t)\langle h, \iota^*(\iota\iota^*)^{-1}\iota k\rangle_{\mathcal{H}'} \\ &= (s\wedge t)\langle h, k\rangle \ .\end{aligned}$$

Here we used (3.26) to go from the first to the second line. This shows that $W(t)$ should be thought of as an $\mathcal{H}$-valued random variable with covariance given by t times the identity operator. This of course does not make sense since the identity is not a trace class operator if $\mathcal{H}$ is infinite-dimensional, so that such an object cannot exist if $\dim \mathcal{H} = \infty$, but it helps to think of it in this way when calculating with it. Combining Proposition 3.87 with Theorem 3.69, we see furthermore that if $\mathcal{K}$ is some Hilbert space and $A\colon \mathcal{H} \to \mathcal{K}$ is a Hilbert-Schmidt operator, then the $\mathcal{K}$-valued random variable $AW(t)$ is perfectly well-defined.

Here we made an abuse of notation and also used the symbol A for the measurable extension of A to $\mathcal{H}'$, but since we know that such a measurable extension is defined uniquely by the values of A on $\mathcal{H}$, this abuse of notation is rather harmless. Furthermore, the distribution of AW does not depend on the choice of the larger space $\mathcal{H}'$, thus further justifying this abuse of notation.

Example 3.88 (White noise) Recall that we informally defined "white noise" as a Gaussian process ξ with covariance $\mathbf{E}\xi(s)\xi(t) = \delta(t-s)$. In particular, if we denote by $\langle\cdot,\cdot\rangle$ the scalar product in $L^2(\mathbf{R})$, this suggests that

$$\mathbf{E}\langle g, \xi\rangle\langle h, \xi\rangle = \mathbf{E}\iint g(s)h(t)\xi(s)\xi(t)\,ds\,dt = \iint g(s)h(t)\delta(t-s)\,ds\,dt = \langle g, h\rangle \ . \tag{3.27}$$

This calculation shows that white noise can be constructed as a Gaussian random variable on any Hilbert space $\mathcal{H}$ of distributions containing $L^2(\mathbf{R})$ and such that the embedding $L^2(\mathbf{R}) \hookrightarrow \mathcal{H}$ is Hilbert-Schmidt. Furthermore, by Theorem 3.69, integrals of the form $\int g(s)\xi(s)\,ds$ are well-defined random variables, provided that $g \in L^2(\mathbf{R})$. Taking for g the indicator function of the interval $[0,t]$, we can check that the process $B(t) = \int_0^t \xi(s)\,ds$ is a Brownian motion, thus justifying the statement that "white noise is the derivative of Brownian motion".

The interesting fact about this construction is that we can use it to define space-time white noise in exactly the same way, simply replacing $L^2(\mathbf{R})$ by $L^2(\mathbf{R}^2)$.

We can now define a Hilbert space-valued stochastic integral against a cylindrical Wiener process in very much the same way as what is usually done in finite dimensions. In the sequel, we fix a cylindrical Wiener process W on some Hilbert space $\mathcal{H} \subset \mathcal{H}'$, which we realise as the canonical coordinate process on $\Omega = \mathcal{C}_W(\mathbf{R}_+, \mathcal{H}')$ equipped with the measure constructed above. We also denote by $\mathscr{F}_s$ the σ-field on Ω generated by $\{W_r : r \le s\}$.

Consider now a finite collection of *disjoint* intervals $(s_n, t_n] \subset \mathbf{R}_+$ with $n = 1, \dots, N$ and a corresponding finite collection of $\mathscr{F}_{s_n}$-measurable random variables Φ_n taking values in the space $\mathcal{L}_2(\mathcal{H}, \mathcal{K})$ of Hilbert-Schmidt operators from $\mathcal{H}$ into some other fixed Hilbert space $\mathcal{K}$. Let furthermore Φ be the $L^2(\mathbf{R}_+ \times \Omega, \mathcal{L}_2(\mathcal{H}, \mathcal{K}))$-valued function defined by

$$\Phi(t, \omega) = \sum_{n=1}^{N} \Phi_n(\omega)\, \mathbf{1}_{(s_n, t_n]}(t) \,,$$

where we denoted by $\mathbf{1}_A$ the indicator function of a set A. We call such a Φ an *elementary process on* $\mathcal{H}$.

Definition 3.89 Given an elementary process Φ and a cylindrical Wiener process W on $\mathcal{H}$, we define the $\mathcal{K}$-valued stochastic integral

$$\int_0^\infty \Phi(t)\, dW(t) \stackrel{\text{def}}{=} \sum_{n=1}^{N} \Phi_n(W)\, (W(t_n) - W(s_n)) \,.$$

Note that since Φ_n is $\mathscr{F}_{s_n}$-measurable, $\Phi_n(W)$ is independent of $W(t_n) - W(s_n)$, therefore each term on the right hand side can be interpreted in the sense of the construction of Theorems 3.69 and 3.73.

Remark 3.90 If Φ is an elementary process on $\mathcal{H}$ and $t > 0$ is some fixed time, then $s \mapsto \Phi(s)\mathbf{1}_{(0,t]}(s)$ is again an elementary process on $\mathcal{H}$. This allows us to define an indefinite stochastic integral by

$$\int_0^t \Phi(s)\, dW(s) \stackrel{\text{def}}{=} \int_0^\infty \Phi(s)\mathbf{1}_{(0,t]}(s)\, dW(s) \,.$$

Remark 3.91 Thanks to Theorem 3.73, this construction is well-posed without requiring to specify the larger Hilbert space $\mathcal{H}'$ on which W can be realised as an $\mathcal{H}'$-valued process. This justifies the terminology of W being "the cylindrical Wiener process on $\mathcal{H}$" without any mentioning of $\mathcal{H}'$, since the value of stochastic integrals against W is independent of the choice of $\mathcal{H}'$.

It follows from Theorem 3.69 and (3.12) that one has the identity

$$\mathbf{E}\Big\| \int_0^\infty \Phi(t)\, dW(t) \Big\|_{\mathcal{K}}^2 = \sum_{n=1}^{N} \mathbf{E}\, \mathrm{tr}(\Phi_n(W)\Phi_n^*(W))(t_n - s_n) = \mathbf{E} \int_0^\infty \mathrm{tr}\, \Phi(t)\Phi^*(t)\, dt \,, \quad (3.28)$$

which is an extension of the usual Itô isometry to the Hilbert space setting. As a consequence, the stochastic integral that we just defined is an isometry from the subset of elementary processes in $L^2(\mathbf{R}_+ \times \Omega, \mathcal{L}_2(\mathcal{H}, \mathcal{K}))$ to $L^2(\Omega, \mathcal{K})$.

Let now $\mathscr{F}_{\mathrm{pr}}$ be the "predictable" σ-field, that is the σ-field over $\mathbf{R}_+ \times \Omega$ generated by all subsets of the form $(s, t] \times A$ with $t > s$ and $A \in \mathscr{F}_s$. This is the smallest σ-algebra with respect to which all elementary processes are $\mathscr{F}_{\mathrm{pr}}$-measurable. One furthermore has:

Proposition 3.92 *The set of elementary processes is dense in the space $L^2_{\mathrm{pr}}(\mathbf{R}_+ \times \Omega, \mathcal{L}_2(\mathcal{H}, \mathcal{K}))$ of all predictable $\mathcal{L}_2(\mathcal{H}, \mathcal{K})$-valued processes.*

Proof. Denote by $\hat{\mathscr{F}}_{\mathrm{pr}}$ the set of all sets of the form $(s, t] \times A$ with $A \in \mathscr{F}_s$. Denote furthermore by $\hat{L}^2_{\mathrm{pr}}$ the closure of the set of elementary processes in L^2. One can check that $\hat{\mathscr{F}}_{\mathrm{pr}}$ is closed under intersections, so that $\mathbf{1}_G \in \hat{L}^2_{\mathrm{pr}}$ for every set G in the algebra generated by $\hat{\mathscr{F}}_{\mathrm{pr}}$. It follows from the monotone class theorem that $\mathbf{1}_G \in \hat{L}^2_{\mathrm{pr}}$ for every set $G \in \mathscr{F}_{\mathrm{pr}}$. The claim then follows from the definition of the Lebesgue integral, just as for the corresponding statement in $\mathbf{R}$. □

By using the Itô isometry (3.28) and the completeness of $L^2(\Omega, \mathcal{K})$, it follows that:

Corollary 3.93 *The stochastic integral* $\int_0^\infty \Phi(t)\,dW(t)$ *can be uniquely defined for every process* $\Phi \in L^2_{\mathrm{pr}}(\mathbf{R}_+ \times \Omega, \mathcal{L}_2(\mathcal{H}, \mathcal{K}))$.

This concludes our presentation of the basic properties of Gaussian measures on infinite-dimensional spaces. The next chapter deals with the other main ingredient to solving stochastic PDEs, which is the behaviour of deterministic linear PDEs.

4

Elements of Semigroup Theory

This chapter is strongly based on Davies's excellent monograph [Dav80] for the first part on strongly continuous semigroups and very loosely follows [Yos95] and [Lun95] for the second part on analytic semigroups. Another good reference on some of the material covered here is the monograph [Paz83]. The aim of this chapter is to give a rigorous meaning to solutions to linear equations of the type

$$\partial_t x = Lx\ , \quad x(0) = x_0 \in \mathcal{B}\ , \tag{4.1}$$

where x takes values in some Banach space $\mathcal{B}$ and L is a possibly unbounded operator on $\mathcal{B}$. From a formal point of view, if such a solution exists, one expects to be able to describe it by a family of linear operators $S(t)$ that map the initial condition x_0 onto the solution $x(t)$ of (4.1) at time t. If such a solution is unique, then the family of operators $S(t)$ should satisfy $S(0) = 1$ and $S(t) \circ S(s) = S(t+s)$. This is called the *semigroup* property.

Furthermore, such a family of solution operators $S(t)$ should have some regularity as $t \to 0$ in order to give a meaning to the notion of an initial condition. (The family given by $S(t) = 0$ for $t > 0$ and $S(0) = 1$ does satisfy the semigroup property but clearly does not define a family of solution operators for any equation of the type (4.1).)

These considerations motivate the following definition:

Definition 4.1 A semigroup $S(t)$ on a Banach space $\mathcal{B}$ is a family of bounded linear operators $\{S(t)\}_{t\geq 0}$ with the properties that $S(t) \circ S(s) = S(t+s)$ for any $s, t \geq 0$ and that $S(0) = \mathrm{Id}$. A semigroup is furthermore called

- *strongly continuous* if, for every $x \in \mathcal{B}$, the map $t \mapsto S(t)x$ is continuous in $\mathcal{B}$.
- *analytic* if there exists $\theta > 0$ such that the operator-valued map $t \mapsto S(t)$ has an analytic extension to $\{\lambda \in \mathbf{C} : |\arg \lambda| < \theta\}$, satisfies the semigroup property there, and is such that $t \mapsto S(e^{i\varphi}t)$ is a strongly continuous semigroup for every angle φ with $|\varphi| < \theta$.

A strongly continuous semigroup is also sometimes called a $\mathcal{C}_0$-semigroup.

Remark 4.2 It follows from the Banach-Steinhaus theorem [BS27] that if S is strongly continuous, then $\sup_{t\leq 1} \|S(t)\| < \infty$. Surprisingly, if we trade strong continuity for weak continuity, namely continuity of $t \mapsto S(t)x$ in the weak topology, then this property fails. See Exercise 5.20 for an example.

Exercise 4.3 Show that the previous remark implies that if S is a $\mathcal{C}_0$-semigroup, then the map $(t, x) \mapsto S(t)x$ is jointly continuous from $\mathbf{R}_+ \times \mathcal{B}$ into $\mathcal{B}$.

Exercise 4.4 Show that being strongly continuous is equivalent to $t \mapsto S(t)x$ being continuous at $t = 0$ for every $x \in D$ with D some dense subset of $\mathcal{B}$, and the operator norm of $S(t)$ being bounded by Me^{at} for some constants M and a.

Remark 4.5 If we only know the strong continuity of $t \mapsto S(t)x$ for x in a dense subset of $\mathcal{B}$, then the second condition in Exercise 4.4 cannot be relaxed in general. In Exercise 5.20 below, we will see

how to construct a semigroup of *bounded* operators such that $\|S(t)\|$ is unbounded near $t = 0$, even though $\lim_{t\to 0} S(t)x = x$ in the strong topology of $\mathcal{B}$ for a dense set of elements $x \in \mathcal{B}$.

Remark 4.6 Some authors, like [Lun95], do not impose strong continuity in the definition of an analytic semigroup. This can result in additional technical complications due to the fact that the generator may then not have dense domain. The approach followed here has the slight drawback that with our definitions the heat semigroup is not analytic on $L^\infty(\mathbf{R})$. (It lacks strong continuity as can be seen by applying it to a step function.)

In most examples relevant for applications, these kind of problems can be circumvented by a suitable choice of ambient space $\mathcal{B}$. For example, in the case of the heat semigroup, it is strongly continuous on $\mathcal{C}_0(\mathbf{R})$, the space of continuous functions vanishing at infinity.

This section is going to assume some familiarity with functional analysis. All the necessary results can be found for example in the classical monograph by Yosida [Yos95]. Recall that an unbounded operator L on a Banach space $\mathcal{B}$ consists of a linear subspace $\mathcal{D}(L) \subset \mathcal{B}$ called the *domain* of L and a linear map $L\colon \mathcal{D}(L) \to \mathcal{B}$. The *graph* of an operator is the subset of $\mathcal{B} \times \mathcal{B}$ consisting of all elements of the form (x, Lx) with $x \in \mathcal{D}(L)$. An operator is *closed* if its graph is a closed subspace of $\mathcal{B} \times \mathcal{B}$ (endowed with the strong topology). It is *closable* if the closure of its graph is again the graph of a linear operator and that operator is called the *closure* of L.

The domain $\mathcal{D}(L^*)$ of the *adjoint* L^* of an unbounded operator $L\colon \mathcal{D}(L) \to \mathcal{B}$ is defined as the set of all elements $\ell \in \mathcal{B}^*$ such that there exists an element $L^*\ell \in \mathcal{B}^*$ with the property that $(L^*\ell)(x) = \ell(Lx)$ for every $x \in \mathcal{D}(L)$. It is clear that in order for the adjoint to be well-defined, we have to require that the domain of L is dense in $\mathcal{B}$. Fortunately, this will be the case for all the operators that will be considered in these notes.

Exercise 4.7 Show that L being closed is equivalent to the fact that if $\{x_n\} \subset \mathcal{D}(L)$ is Cauchy in $\mathcal{B}$ and $\{Lx_n\}$ is also Cauchy, then $x = \lim_{n\to\infty} x_n$ belongs to $\mathcal{D}(L)$ and $Lx = \lim_{n\to\infty} Lx_n$.

Exercise 4.8 Show that the adjoint of an operator with dense domain is always closed.

The *resolvent set* $\varrho(L)$ of an operator L is defined by

$$\varrho(L) = \{\lambda \in \mathbf{C} \,:\, \text{range}(\lambda - L) \text{ is dense in } \mathcal{B} \text{ and } \lambda - L \text{ has a continuous inverse.}\}\ ,$$

and the *resolvent* R_λ is given for $\lambda \in \varrho(L)$ by $R_\lambda = (\lambda - L)^{-1}$. (Here and in the sequel we view $\mathcal{B}$ as a complex Banach space. If an operator is defined on a real Banach space, it can always be extended to its complexification in a canonical way and we will identify the two without further notice in the sequel.) The spectrum of L is the complement of the resolvent set.

The most important results regarding the resolvent of an operator that we are going to use are that any closed operator L with non-empty resolvent set is defined in a unique way by its resolvent. Furthermore, the resolvent set is open and the resolvent is an analytic function from $\varrho(L)$ to the space $\mathcal{L}(\mathcal{B})$ of bounded linear operators on $\mathcal{B}$ endowed with the topology of convergence in operator norm. Finally, the resolvent operators for different values of λ all commute and satisfy the resolvent identity

$$R_\lambda - R_\mu = (\mu - \lambda)R_\mu R_\lambda\ ,$$

for any two $\lambda, \mu \in \varrho(L)$.

The fact that the resolvent is operator-valued should not be a conceptual obstacle to the use of notions from complex analysis. Indeed, for $D \subset \mathbf{C}$ an open domain, a function $f\colon D \to \hat{\mathcal{B}}$ where $\hat{\mathcal{B}}$ is any complex Banach space (typically the complexification of a real Banach space which we identify with the original space without further ado) is said to be analytic in exactly the same way as usual by imposing that its Taylor series at any point $a \in D$ converges to f uniformly in $\hat{\mathcal{B}}$ on a neighbourhood of a. The same definition applies if $D \subset \mathbf{R}$ and analytic continuation then works in exactly the same way as

for complex-valued functions. In particular, Cauchy's residue theorem, which is the main result from complex analysis that we are going to use later on, works for Banach-space valued functions in exactly the same way as for complex-valued functions.

4.1 Strongly continuous semigroups

We start our investigation of semigroup theory with a discussion of the main results that can be obtained for strongly continuous semigroups. Given a $\mathcal{C}_0$-semigroup, one can associate to it a "generator", which is essentially the derivative of $S(t)$ at $t = 0$:

Definition 4.9 The *generator* L of a $\mathcal{C}_0$-semigroup is given by

$$Lx = \lim_{t \to 0} t^{-1}(S(t)x - x) \,, \tag{4.2}$$

on the set $\mathcal{D}(L)$ of all elements $x \in \mathcal{B}$ such that this limit exists (in the sense of strong convergence in $\mathcal{B}$).

The following result shows that if L is the generator of a $\mathcal{C}_0$-semigroup $S(t)$, then $x(t) = S(t)x_0$ is indeed the solution to (4.1) in a weak sense.

Proposition 4.10 *The domain $\mathcal{D}(L)$ of L is dense in $\mathcal{B}$, invariant under S, and the identities $\partial_t S(t)x = LS(t)x = S(t)Lx$ hold for every $x \in \mathcal{D}(L)$ and every $t \geq 0$. Furthermore, for every $\ell \in \mathcal{D}(L^*)$ and every $x \in \mathcal{B}$, the map $t \mapsto \langle \ell, S(t)x \rangle$ is differentiable and one has $\partial_t \langle \ell, S(t)x \rangle = \langle L^*\ell, S(t)x \rangle$.*

Proof. Fix some arbitrary $x \in \mathcal{B}$ and set $x_t = \int_0^t S(s)x\,ds$. One then has

$$\begin{aligned}\lim_{h \to 0} h^{-1}(S(h)x_t - x_t) &= \lim_{h \to 0} h^{-1}\Big(\int_h^{t+h} S(s)x\,ds - \int_0^t S(s)x\,ds\Big) \\ &= \lim_{h \to 0} h^{-1}\Big(\int_t^{t+h} S(s)x\,ds - \int_0^h S(s)x\,ds\Big) = S(t)x - x \,,\end{aligned}$$

where the last equality follows from the strong continuity of S. This shows that $x_t \in \mathcal{D}(L)$. Since $t^{-1}x_t \to x$ as $t \to 0$ and since x was arbitrary, it follows that $\mathcal{D}(L)$ is dense in $\mathcal{B}$. To show that it is invariant under S, note that for $x \in \mathcal{D}(L)$ one has

$$\lim_{h \to 0} h^{-1}(S(h)S(t)x - S(t)x) = S(t) \lim_{h \to 0} h^{-1}(S(h)x - x) = S(t)Lx \,,$$

so that $S(t)x \in \mathcal{D}(L)$ and $LS(t)x = S(t)Lx$. To show that it this is equal to $\partial_t S(t)x$, it suffices to check that the left derivative of this expression exists and is equal to the right derivative. This is left as an exercise.

To show that the second claim holds, it is sufficient (using the strong continuity of S) to check that it holds for $x \in \mathcal{D}(L)$. Since one then has $S(t)x \in \mathcal{D}(L)$ for every t, it follows from the definition (4.2) of $\mathcal{D}(L)$ that $t \mapsto S(t)x$ is differentiable and that its derivative is equal to $LS(t)x$. □

It follows as a corollary that no two semigroups can have the same generator (unless the semigroups coincide of course), which justifies the notation $S(t) = e^{Lt}$ that we are occasionally going to use in the sequel.

Corollary 4.11 *If a function $x\colon [0,1] \to \mathcal{D}(L)$ satisfies $\partial_t x_t = Lx_t$ for every $t \in [0,1]$, then $x_t = S(t)x_0$. In particular, no two distinct $\mathcal{C}_0$-semigroups can have the same generator.*

Proof. Let $T > 0$ be an arbitrary time. It follows from an argument almost identical to that given in the proof of Proposition 4.10 that the map $t \mapsto S(t)x_{T-t}$ is continuous on $[0,T]$ and differentiable on $(0,T)$. Computing its derivative, we obtain $\partial_t S(t)x_{T-t} = LS(t)x_{T-t} - S(t)Lx_{T-t} = 0$, so that,

evaluating the function at $t = 0$ and $t = T$, we obtain $x_T = S(T)x_0$. Since T was arbitrary, this proves the claim. □

Exercise 4.12 Show that the semigroup $S(t)$ on $L^2(\mathbf{R})$ given by

$$(S(t)f)(\xi) = f(\xi + t) ,$$

is strongly continuous and that its generator is given by $L = \partial_\xi$ with $\mathcal{D}(L) = H^1$. Similarly, show that the heat semigroup on $L^2(\mathbf{R})$ given by

$$(S(t)f)(\xi) = \frac{1}{\sqrt{4\pi t}} \int \exp\Big(-\frac{|\xi - \eta|^2}{4t}\Big) f(\eta)\, d\eta ,$$

is strongly continuous and that its generator is given by $L = \partial_\xi^2$ with $\mathcal{D}(L) = H^2$. **Hint:** Use Exercise 4.4 to show strong continuity.

Remark 4.13 We did not make any assumption on the structure of the Banach space $\mathcal{B}$. *However*, it is a general rule of thumb (although this is *not* a theorem) that semigroups on non-separable Banach spaces tend not to be strongly continuous. For example, it is an excellent exercise to convince oneself that neither the heat semigroup nor the translation semigroup from the previous exercise are strongly continuous on $L^\infty(\mathbf{R})$ or even on $\mathcal{C}_b(\mathbf{R})$, the space of all bounded continuous functions on $\mathbf{R}$.

Recall now that the resolvent set for an operator L consists of those $\lambda \in \mathbf{C}$ such that the operator $\lambda - L$ is one to one. For λ in the resolvent set, we denote by $R_\lambda = (\lambda - L)^{-1}$ the resolvent of L. It turns out that the resolvent of the generator of a $\mathcal{C}_0$-semigroup can easily be computed:

Proposition 4.14 *Let $S(t)$ be a $\mathcal{C}_0$-semigroup such that $\|S(t)\| \leq Me^{at}$ for some constants M and a. If $\mathrm{Re}\lambda > a$, then λ belongs to the resolvent set of L and one has the identity $R_\lambda x = \int_0^\infty e^{-\lambda t} S(t)x\, dt$.*

Proof. By the assumption on the bound on S, the expression $Z_\lambda = \int_0^\infty e^{-\lambda t} S(t)x\, dt$ is well-defined for every λ with $\mathrm{Re}\lambda > a$. In order to show that $Z_\lambda = R_\lambda$, we first show that $Z_\lambda x \in \mathcal{D}(L)$ for every $x \in \mathcal{B}$ and that $(\lambda - L)Z_\lambda x = x$. We have

$$\begin{aligned} LZ_\lambda x &= \lim_{h\to 0} h^{-1}(S(h)Z_\lambda x - Z_\lambda x) = \lim_{h\to 0} h^{-1} \int_0^\infty e^{-\lambda t}(S(t+h)x - S(t)x)\, dt \\ &= \lim_{h\to 0}\Big(\frac{e^{\lambda h} - 1}{h} \int_0^\infty e^{-\lambda t} S(t)x\, dt - \frac{e^{\lambda h}}{h} \int_0^h e^{-\lambda t} S(t)x\, dt\Big) \\ &= \lambda Z_\lambda x - x , \end{aligned}$$

which is the required identity. To conclude, it remains to show that $\lambda - L$ is an injection on $\mathcal{D}(L)$. If it was not, we could find $x \in \mathcal{D}(L) \setminus \{0\}$ such that $Lx = \lambda x$. Setting $x_t = e^{\lambda t}x$ and applying Corollary 4.11, this yields $S(t)x = e^{\lambda t}x$, thus contradicting the bound $\|S(t)\| \leq Me^{at}$ if $\mathrm{Re}\lambda > a$. □

We can deduce from this that:

Proposition 4.15 *The generator L of a $\mathcal{C}_0$-semigroup is a closed operator.*

Proof. We are going to use the characterisation of closed operators given in Exercise 4.7. Shifting L by a constant if necessary (which does not affect it being closed or not), we can assume that $a = 0$. Take now a sequence $x_n \in \mathcal{D}(L)$ such that $\{x_n\}$ and $\{Lx_n\}$ are both Cauchy in $\mathcal{B}$ and set $x = \lim_{n\to\infty} x_n$ and $y = \lim_{n\to\infty} Lx_n$. Setting $z_n = (1 - L)x_n$, we have $\lim_{n\to\infty} z_n = x - y$.

On the other hand, we know that 1 belongs to the resolvent set, so that

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} R_1 z_n = R_1(x - y) .$$

By the definition of the resolvent, this implies that $x \in \mathcal{D}(L)$ and that $x - Lx = x - y$, so that $Lx = y$ as required. □

We are now ready to give a full characterisation of the generators of $\mathcal{C}_0$-semigroups. This is the content of the following theorem:

Theorem 4.16 (Hille-Yosida) *A closed densely defined operator L on the Banach space $\mathcal{B}$ is the generator of a $\mathcal{C}_0$-semigroup $S(t)$ with $\|S(t)\| \leq Me^{at}$ if and only if all λ with $\mathrm{Re}\lambda > a$ lie in its resolvent set and the bound $\|R_\lambda^n\| \leq M\,(\mathrm{Re}\lambda - a)^{-n}$ holds there for every $n \geq 1$.*

Proof. The generator L of a $\mathcal{C}_0$-semigroup is closed by Proposition 4.15. The fact that its resolvent satisfies the stated bound follows immediately from the fact that

$$R_\lambda^n x = \int_0^\infty \cdots \int_0^\infty e^{-\lambda(t_1+\ldots+t_n)} S(t_1 + \ldots + t_n)x\, dt_1 \cdots dt_n$$

by Proposition 4.14.

To show that the converse also holds, we are going to construct the semigroup $S(t)$ by using the so-called "Yosida approximations" $L_\lambda = \lambda L R_\lambda$ for L. Note first that $\lim_{\lambda\to\infty} LR_\lambda x = 0$ for every $x \in \mathcal{B}$: it obviously holds for $x \in \mathcal{D}(L)$ since then $\|LR_\lambda x\| = \|R_\lambda Lx\| \leq \|R_\lambda\|\|Lx\| \leq M\,(\mathrm{Re}\lambda - a)^{-1}\|Lx\|$. Furthermore, $\|LR_\lambda x\| = \|\lambda R_\lambda x - x\| \leq (M\lambda(\lambda - a)^{-1} + 1)\|x\| \leq (M+2)\|x\|$ for λ large enough, so that $\lim_{\lambda\to\infty} LR_\lambda x = 0$ for every x by a standard density argument.

Using this fact, we can show that the Yosida approximation of L does indeed approximate L in the sense that $\lim_{\lambda\to\infty} L_\lambda x = Lx$ for every $x \in \mathcal{D}(L)$. Fixing an arbitrary $x \in \mathcal{D}(L)$, we have

$$\lim_{\lambda\to\infty} \|L_\lambda x - Lx\| = \lim_{\lambda\to\infty} \|(\lambda R_\lambda - 1)Lx\| = \lim_{\lambda\to\infty} \|LR_\lambda Lx\| = 0\ . \tag{4.3}$$

Define now a family of bounded operators $S_\lambda(t)$ by $S_\lambda(t) = e^{L_\lambda t} = \sum_{n\geq 0} \frac{t^n L_\lambda^n}{n!}$. This series converges in the operator norm since L_λ is bounded and one can easily check that S_λ is indeed a $\mathcal{C}_0$-semigroup (actually a group) with generator L_λ. Since $L_\lambda = -\lambda + \lambda^2 R_\lambda$, one has for $\lambda > a$ the bound

$$\|S_\lambda(t)\| \leq e^{-\lambda t} \sum_{n\geq 0} \frac{t^n \lambda^{2n} \|R_\lambda^n\|}{n!} \leq M \exp\Big(-\lambda t + \frac{\lambda^2}{\lambda - a} t\Big) \leq M \exp\Big(\frac{\lambda a t}{\lambda - a}\Big)\ , \tag{4.4}$$

so that $\limsup_{\lambda\to\infty} \|S_\lambda(t)\| \leq Me^{at}$. Let us show next that the limit $\lim_{\lambda\to\infty} S_\lambda(t)x$ exists for every $t \geq 0$ and every $x \in \mathcal{B}$. Fixing λ and μ large enough so that $\max\{\|S_\lambda(t)\|, \|S_\mu(t)\|\} \leq Me^{2at}$, and fixing some arbitrary $t > 0$, we have for $s \in [0, t]$

$$\begin{aligned}\|\partial_s S_\lambda(t-s)S_\mu(s)x\| &= \|S_\lambda(t-s)(L_\mu - L_\lambda)S_\mu(s)x\| = \|S_\lambda(t-s)S_\mu(s)(L_\mu - L_\lambda)x\| \\ &\leq M^2 e^{2at}\|(L_\mu - L_\lambda)x\|\ .\end{aligned}$$

Integrating this bound between 0 and t, we obtain

$$\|S_\lambda(t)x - S_\mu(t)x\| \leq M^2 t e^{2at}\|L_\mu x - L_\lambda x\|\ , \tag{4.5}$$

which converges to 0 for every $x \in \mathcal{D}(L)$ as $\lambda, \mu \to \infty$ since one then has $L_\lambda x \to Lx$. We can therefore *define* a family of linear operators $S(t)$ by $S(t)x = \lim_{\lambda\to\infty} S_\lambda(t)x$.

It is clear from (4.4) that $\|S(t)\| \leq Me^{at}$ and it follows from the semigroup property of S_λ that $S(s)S(t) = S(s+t)$. Furthermore, it follows from (4.5) and (4.3) that for every fixed $x \in \mathcal{D}(L)$, the convergence $S_\lambda(t)x \to S(t)x$ is uniform in bounded intervals of t, so that the map $t \mapsto S(t)x$ is continuous. Combining this with our *a priori* bounds on the operator norm of $S(t)$, it follows from Exercise 4.4 that S is indeed a $\mathcal{C}_0$-semigroup. It remains to show that the generator $\hat{L}$ of S coincides with L. Taking first the limit $\lambda \to \infty$ and then the limit $t \to 0$ in the identity

$$t^{-1}(S_\lambda(t)x - x) = t^{-1} \int_0^t S_\lambda(s) L_\lambda x\, ds\ ,$$

we see that $x \in \mathcal{D}(L)$ implies $x \in \mathcal{D}(\hat{L})$ and $\hat{L}x = Lx$, so that $\hat{L}$ is an extension of L. However, for $\lambda > a$, both $\lambda - L$ and $\lambda - \hat{L}$ are one-to-one between their domain and $\mathcal{B}$, so that they must coincide. □

One might think that the resolvent bound in the Hille-Yosida theorem is a consequence of the fact that the spectrum of L is assumed to be contained in the half plane $\{\lambda : \mathrm{Re}\lambda \leq a\}$. This however isn't the case, as can be seen by the following example:

Example 4.17 We take $\mathcal{B} = \bigoplus_{n\geq 1} \mathbf{C}^2$ (equipped with the usual Euclidean norms) and we define $L = \bigoplus_{n\geq 1} L_n$, where $L_n : \mathbf{C}^2 \to \mathbf{C}^2$ is given by the matrix

$$L_n = \begin{pmatrix} in & n \\ 0 & in \end{pmatrix}.$$

In particular, the resolvent $R_\lambda^{(n)}$ of L_n is given by

$$R_\lambda^{(n)} = \frac{1}{(\lambda - in)^2} \begin{pmatrix} \lambda - in & n \\ 0 & \lambda - in \end{pmatrix},$$

so that one has the upper and lower bounds

$$\frac{n}{|\lambda - in|^2} \leq \|R_\lambda^{(n)}\| \leq \frac{n}{|\lambda - in|^2} + \frac{\sqrt{2}}{|\lambda - in|}.$$

Note now that the resolvent R_λ of L satisfies $\|R_\lambda\| = \sup_{n\geq 1} \|R_\lambda^{(n)}\|$. On one hand, this shows that the spectrum of L is given by the set $\{in^2 : n \geq 1\}$, so that it does indeed lie in a half plane. On the other hand, for every fixed value $a > 0$, we have $\|R_{a+in}\| \geq \frac{n}{a^2}$, so that the resolvent bound of the Hille-Yosida theorem is certainly not satisfied.

It is therefore not surprising that L does not generate a $\mathcal{C}_0$-semigroup on $\mathcal{B}$. Even worse, trying to define $S(t) = \oplus_{n\geq 1} S_n(t)$ with $S_n(t) = e^{L_n t}$ results in $\|S_n(t)\| \geq nt$, so that $S(t)$ is an unbounded operator for every $t > 0$!

4.1.1 Adjoint semigroups

It will be very useful in the sequel to have a good understanding of the behaviour of the adjoints of strongly continuous semigroups. The reason why this is not a completely trivial topic is that, in general, it is simply not true that the adjoint semigroup $S^*(t) : \mathcal{B}^* \to \mathcal{B}^*$ of a strongly continuous semigroup is again strongly continuous. This is probably best illustrated by an example.

Take $\mathcal{B} = \mathcal{C}([0,1], \mathbf{R})$ and let $S(t)$ be the heat semigroup (with Neumann boundary conditions, say). Then $S^*(t)$ acts on finite signed measures by convolving them with the heat kernel. While it is true that $S^*(t)\mu \to \mu$ weakly as $t \to 0$, it is not true in general that this convergence is strong. For example, $S^*(t)\delta_x$ does *not* converge to δ_x in the total variation norm (which is the dual to the supremum norm on $\mathcal{C}([0,1], \mathbf{R})$). However, this difficulty can always be overcome by restricting $S^*(t)$ to a slightly smaller space than $\mathcal{B}^*$. This is the content of the following result:

Proposition 4.18 *If $S(t)$ is a $\mathcal{C}_0$-semigroup on $\mathcal{B}$, then $S^*(t)$ is a $\mathcal{C}_0$-semigroup on the closure $\mathcal{B}^\dagger$ of $\mathcal{D}(L^*)$ in $\mathcal{B}^*$ and its generator $L^\dagger$ is given by the restriction of L^* to the set $\mathcal{D}(L^\dagger) = \{x \in \mathcal{D}(L^*) : L^*x \in \mathcal{B}^\dagger\}$.*

Proof. We first show that $S^*(t)$ is strongly continuous on $\mathcal{B}^\dagger$ and we will then identify its generator. Note first that it follows from Proposition 4.10 that $S^*(t)$ maps $\mathcal{D}(L^*)$ into itself, so that it does indeed define a family of bounded operators on $\mathcal{B}^\dagger$. Since the norm of $S^*(t)$ is $\mathcal{O}(1)$ as $t \to 0$ and since $\mathcal{D}(L^*)$

is dense in $\mathcal{B}^\dagger$ by definition, it is sufficient to show that $\lim_{t\to 0} S^*(t)x = x$ for every $x \in \mathcal{D}(L^*)$. It follows immediately from Proposition 4.10 that for $x \in \mathcal{D}(L^*)$ one has the identity

$$S^*(t)x - x = \int_0^t S^*(s)L^*x\,ds\ ,$$

from which we conclude that $S^*(t)x \to x$.

It follows from Proposition 4.14 that the resolvent $R_\lambda^\dagger$ for $S^*(t)$ on $\mathcal{B}^\dagger$ is nothing but the restriction of R_λ^* to $\mathcal{B}^\dagger$. This immediately implies that $\mathcal{D}(L^\dagger)$ is given by the stated expression. □

Remark 4.19 As we saw in the example of the heat semigroup, $\mathcal{B}^\dagger$ is in general strictly smaller than $\mathcal{B}^*$. This fact was first pointed out by Phillips in [Phi55]. In our example, $\mathcal{B}^*$ consists of all finite signed Borel measures on $[0,1]$, whereas $\mathcal{B}^\dagger$ only consists of those measures that have a density with respect to Lebesgue measure.

Even though $\mathcal{B}^\dagger$ is in general a proper closed subspace of $\mathcal{B}^*$, it is large enough to be dense in $\mathcal{B}^*$, when equipped with the (much weaker) weak-* topology. This the content of our last result in the theory of strongly continuous semigroups:

Proposition 4.20 *For every $\ell \in \mathcal{B}^*$ there exists a sequence $\ell_n \in \mathcal{B}^\dagger$ such that $\ell_n(x) \to \ell(x)$ for every $x \in \mathcal{B}$.*

Proof. It suffices to choose $\ell_n = nR_n^*\ell$. Since we have $\ell_n \in \mathcal{D}(L^*)$, it is clear that $\ell_n \in \mathcal{B}^\dagger$. On the other hand, we know from the proof of the Hille-Yosida theorem that $\lim_{n\to\infty} \|nR_n x - x\| = 0$ for every $x \in \mathcal{B}$, from which the claim follows at once. □

4.2 Semigroups with selfadjoint generators

In this section, we consider the particular case of strongly continuous semigroups consisting of self-adjoint operators on a Hilbert space $\mathcal{H}$. The reason why this is an interesting case is that it immediately implies very strong smoothing properties of the operators $S(t)$ in the sense that for every $t > 0$, they map $\mathcal{H}$ into the domain of arbitrarily high powers of L. Furthermore, it is very easy to obtain explicit bounds on the norm of $S(t)$ as an operator from $\mathcal{H}$ into one of these domains. We will then see later in Section 4.3 on analytic semigroups that most of these properties still hold true for a much larger class of semigroups.

Let L be a selfadjoint operator on $\mathcal{H}$ which is bounded from above. Without loss of generality, we are going to assume that it is actually negative definite, so that $\langle x, Lx\rangle \leq 0$ for any $x \in \mathcal{H}$. In this case, we can use functional calculus (see for example [RS80], in particular chapter VIII in volume I) to define selfadjoint operators $f(L)$ for any measurable map $f\colon \mathbf{R} \to \mathbf{R}$. This is because the spectral decomposition theorem can be formulated as:

Theorem 4.21 (Spectral decomposition) *Let L be a selfadjoint operator on a separable Hilbert space $\mathcal{H}$. Then, there exists a measure space $(\mathcal{M}, \mu)$, an isomorphism $K\colon \mathcal{H} \to L^2(\mathcal{M}, \mu)$, and a function $f_L\colon \mathcal{M} \to \mathbf{R}$ such that via K, L is equivalent to the multiplication operator by f_L on $L^2(\mathcal{M}, \mu)$. In other words, one has $L = K^{-1} f_L K$ and $K\mathcal{D}(L) = \{g\ :\ f_L g \in L^2(\mathcal{M}, \mu)\}$.*

In particular, this allows one to *define* $f(L) = K^{-1}(f \circ f_L)K$, which has all the nice properties that one would expect from functional calculus, like for example $(fg)(L) = f(L)g(L)$, $\|f(L)\| = \|f\|_{L^\infty(\mathcal{M},\mu)}$, etc. Defining $S(t) = e^{Lt}$, it is an exercise to check that S is indeed a $\mathcal{C}_0$-semigroup with generator L (either use the Hille-Yosida theorem and make sure that the semigroup constructed there coincides with S or check "by hand" that $S(t)$ is indeed $\mathcal{C}_0$ with generator L).

The important property of semigroups generated by selfadjoint operators is that they do not only leave $\mathcal{D}(L)$ invariant, but they have a regularising effect in that they map $\mathcal{H}$ into the domain of any arbitrarily high power of L. More precisely, one has:

Proposition 4.22 *Let L be self-adjoint and negative definite and let $S(t)$ be the semigroup on $\mathcal{H}$ generated by L. Then, $S(t)$ maps $\mathcal{H}$ into the domain of $(1-L)^\alpha$ for any $\alpha, t > 0$ and there exist constants C_α such that $\|(1-L)^\alpha S(t)\| \leq C_\alpha(1+t^{-\alpha})$.*

Proof. By functional calculus, it suffices to show that $\sup_{\lambda\geq 0}(1+\lambda)^\alpha e^{-\lambda t} \leq C_\alpha(1+t^{-\alpha})$. One has

$$\sup_{\lambda\geq 0} \lambda^\alpha e^{-\lambda t} = t^{-\alpha} \sup_{\lambda\geq 0} (\lambda t)^\alpha e^{-\lambda t} = t^{-\alpha} \sup_{\lambda\geq 0} \lambda^\alpha e^{-\lambda} = \alpha^\alpha e^{-\alpha} t^{-\alpha} .$$

The claim now follows from the fact that there exists a constant C'_α such that $(1-\lambda)^\alpha \leq C'_\alpha(1+(-\lambda)^\alpha)$ for every $\lambda \leq 0$. □

4.3 Analytic semigroups

Obviously, the conclusion of Proposition 4.22 does not hold for arbitrary $\mathcal{C}_0$-semigroups since the group of translations from Example 4.12 does not have any smoothing properties. It does however hold for a very large class of semigroups, the so-called analytic semigroups. The study of these semigroups is the object of the remainder of this section, and the equivalent of Proposition 4.22 is going to be one of our two main results. The other result is a characterisation of generators for analytic semigroups that is analogous to the Hille-Yosida theorem for $\mathcal{C}_0$-semigroups. The difference will be that the role of the half-plane $\mathrm{Re}\lambda > a$ will be played by the complement of a sector of the complex plane with an opening angle strictly smaller than π.

Recall that a semigroup S on a Banach space $\mathcal{B}$ is *analytic* if there exists $\theta \in (0, \frac{\pi}{2})$ such that the map $t \mapsto S(t)$ (taking values in $\mathcal{L}(\mathcal{B})$) admits an analytic extension to the sector $S_\theta = \{\lambda \in \mathbf{C} : |\arg\lambda| < \theta\}$, satisfies the semigroup property there, and is such that $t \mapsto S_\varphi(t) = S(e^{i\varphi}t)$ is a strongly continuous semigroup for every $|\varphi| < \theta$. If θ is the largest angle such that the above property holds, we call S *analytic with angle θ*. The strong continuity of $t \mapsto S(e^{i\varphi}t)$ implies that there exist constants $M(\varphi)$ and $a(\varphi)$ such that

$$\|S_\varphi(t)\| \leq M(\varphi)e^{a(\varphi)t} .$$

Using the semigroup property, it is not difficult to show that M and a can be chosen bounded over compact intervals:

Proposition 4.23 *Let S be an analytic semigroup with angle θ. Then, for every $\theta' < \theta$, there exist M and a such that $\|S_\varphi(t)\| \leq Me^{at}$ for every $t > 0$ and every $|\varphi| \leq \theta'$.*

Proof. Fix $\theta' \in (0, \theta)$, so that in particular $\theta' < \pi/2$. Then there exists a constant C such that, for every $t > 0$ and every φ with $|\varphi| < \theta'$, there exist numbers $t_+, t_- \in [0, Ct]$ such that $te^{i\varphi} = t_+e^{i\theta'} + t_-e^{-i\theta'}$. It follows that one has the bound $\|S_\varphi(t)\| \leq M(\theta')M(-\theta')e^{a(\theta')Ct + a(-\theta')Ct}$, thus proving the claim. □

We next compute the generators of the semigroups S_φ obtained by evaluating S along a "ray" extending out of the origin into the complex plane:

Proposition 4.24 *Let S be an analytic semigroup with angle θ. Then, for $|\varphi| < \theta$, the generator L_φ of S_φ is given by $L_\varphi = e^{i\varphi}L$, where L is the generator of S.*

Proof. Recall Proposition 4.14 showing that for $\mathrm{Re}\lambda$ large enough the resolvent R_λ for L is given by

$$R_\lambda x = \int_0^\infty e^{-\lambda t} S(t)x \, dt \; .$$

Since the map $t \mapsto e^{-\lambda t}S(t)$ is analytic in S_θ by assumption and since, provided again that $\mathrm{Re}\lambda$ is large enough, it decays exponentially to 0 as $|t| \to \infty$, we can deform the contour of integration to obtain

$$R_\lambda x = e^{i\varphi} \int_0^\infty e^{-\lambda e^{i\varphi} t} S(e^{i\varphi}t)x \, dt \; .$$

Denoting by R^φ_λ the resolvent for the generator L_φ of S_φ, we thus have the identity $R_\lambda = e^{i\varphi} R^\varphi_{\lambda e^{i\varphi}}$, which is equivalent to $(\lambda - L)^{-1} = (\lambda - e^{-i\varphi}L_\varphi)^{-1}$, thus showing that $L_\varphi = e^{i\varphi}L$ as stated. □

We now use this to show that if S is an analytic semigroup, then the resolvent set of its generator L not only contains the right half plane, but it contains a larger sector of the complex plane. Furthermore, this characterises the generators of analytic semigroups, providing a statement similar to the Hille-Yosida theorem:

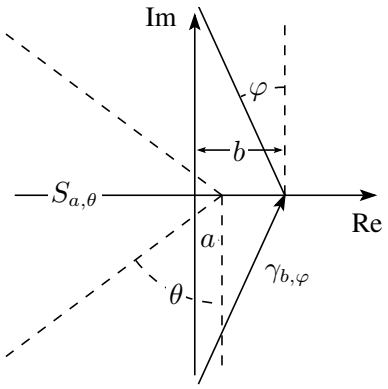
Theorem 4.25 *A closed densely defined operator L on a Banach space $\mathcal{B}$ is the generator of an analytic semigroup if and only if there exists $\theta \in (0, \frac{\pi}{2})$ and $a \geq 0$ such that the spectrum of L is contained in the sector*

$$S_{\theta,a} = \{\lambda \in \mathbf{C} \, : \, \arg(a - \lambda) \in [-\tfrac{\pi}{2} + \theta, \tfrac{\pi}{2} - \theta]\} \; ,$$

and there exists $M > 0$ such that the resolvent R_λ satisfies the bound $\|R_\lambda\| \leq M d(\lambda, S_{\theta,a})^{-1}$ for every $\lambda \notin S_{\theta,a}$.

Proof. The fact that generators of analytic semigroups are of the prescribed form is a consequence of Proposition 4.24 and the Hille-Yosida theorem.

To show the converse statement, let L be such an operator, let $\varphi \in (0, \theta)$, let $b > a$, and let $\gamma_{\varphi,b}$ be the curve in the complex plane obtained by going in a counterclockwise way around the boundary of $S_{\varphi,b}$ (see the figure on the right). For t with $|\arg t| < \varphi$, define $S(t)$ by



$$\begin{aligned} S(t) &= \frac{1}{2\pi i} \int_{\gamma_{\varphi,b}} e^{tz} R_z \, dz \qquad (4.6) \\ &= \frac{1}{2\pi i} \int_{\gamma_{\varphi,b}} e^{tz} (z - L)^{-1} \, dz \; . \end{aligned}$$

It follows from the resolvent bound that $\|R_z\|$ is uniformly bounded for $z \in \gamma_{\varphi,b}$. Furthermore, since $|\arg t| < \varphi$, it follows that e^{tz} decays exponentially as $|z| \to \infty$ along $\gamma_{\varphi,b}$, so that this expression is well-defined, does not depend on the choice of b and φ, and (by choosing φ arbitrarily close to θ) determines an analytic function $t \mapsto S(t)$ on the sector $\{t \, : \, |\arg t| < \theta\}$. As in the proof of the Hille-Yosida theorem, the function $(x, t) \mapsto S(t)x$ is jointly continuous because the convergence of the integral defining S is uniform over bounded subsets of $\{t \, : \, |\arg t| < \varphi\}$ for any $|\varphi| < \theta$.

It therefore remains to show that S satisfies the semigroup property on the sector $\{t \, : \, |\arg t| < \theta\}$ and that its generator is indeed given by L. Choosing s and t such that $|\arg s| < \theta$ and $|\arg t| < \theta$ and using the resolvent identity $R_z - R_{z'} = (z' - z)R_z R_{z'}$, we have

$$S(s)S(t) = -\frac{1}{4\pi^2} \int_{\gamma_{\varphi,b'}} \int_{\gamma_{\varphi,b}} e^{tz+sz'} R_z R_{z'} \, dz \, dz' = -\frac{1}{4\pi^2} \int_{\gamma_{\varphi,b'}} \int_{\gamma_{\varphi,b}} e^{tz+sz'} \frac{R_z - R_{z'}}{z' - z} \, dz \, dz'$$

$$= -\frac{1}{4\pi^2}\int_{\gamma_{\varphi,b}} e^{tz} R_z \int_{\gamma_{\varphi,b'}} \frac{e^{sz'}}{z'-z}\,dz'\,dz - \frac{1}{4\pi^2}\int_{\gamma_{\varphi,b'}} e^{sz} R_z \int_{\gamma_{\varphi,b}} \frac{e^{tz'}}{z'-z}\,dz'\,dz\ .$$

Here, the choice of b and b' is arbitrary, as long as $b \neq b'$ so that the inner integrals are well-defined, say $b' > b$ for definiteness. In this case, since the contour $\gamma_{\varphi,b}$ can be "closed up" to the left but not to the right, the integral $\int_{\gamma_{\varphi,b'}} \frac{e^{sz'}}{z'-z}\,dz'$ is equal to $2i\pi e^{sz}$ for every $z \in \gamma_{\varphi,b}$, whereas the integral with b and b' inverted vanishes, so that

$$S(s)S(t) = \frac{1}{2i\pi}\int_{\gamma_{\varphi,b}} e^{(t+s)z} R_z = S(s+t)\ ,$$

as required. The continuity of the map $t \mapsto S(t)x$ is a straightforward consequence of the resolvent bound, noting that it arises as a uniform limit of continuous functions. Therefore S is a strongly continuous semigroup; let us call its generator $\hat{L}$ and $\hat{R}_\lambda$ the corresponding resolvent.

To show that $L = \hat{L}$, it suffices to show that $\hat{R}_\lambda = R_\lambda$, so we make use again of Proposition 4.14. Choosing $\mathrm{Re}\lambda > b$ so that $\mathrm{Re}(z-\lambda) < 0$ for every $z \in \gamma_{\varphi,b}$, we have

$$\begin{aligned}\hat{R}_\lambda &= \int_0^\infty e^{-\lambda t} S(t)\,dt = \frac{1}{2\pi i}\int_0^\infty \int_{\gamma_{\varphi,b}} e^{t(z-\lambda)} R_z\,dz\,dt \\ &= \frac{1}{2\pi i}\int_{\gamma_{\varphi,b}} \int_0^\infty e^{t(z-\lambda)}\,dt\,R_z\,dz = \frac{1}{2\pi i}\int_{\gamma_{\varphi,b}} \frac{R_z}{z-\lambda}\,dz = R_\lambda\ .\end{aligned}$$

The last inequality was obtained by using the fact that $\|R_z\|$ decays like $1/|z|$ for large enough z with $|\arg z| \leq \frac{\pi}{2} + \varphi$, so that the contour can be "closed" to enclose the pole at $z = \lambda$. □

As a consequence of this characterisation theorem, we can study perturbations of generators of analytic semigroups. The idea is to give a constructive criterion which allows to make sure that an operator of the type $L = L_0 + B$ is the generator of an analytic semigroup, provided that L_0 is such a generator and B satisfies a type of "relative total boundedness" condition. The precise statement of this result is:

Theorem 4.26 *Let L_0 be the generator of an analytic semigroup and let $B\colon \mathcal{D}(B) \to \mathcal{B}$ be an operator such that*

- *The domain $\mathcal{D}(B)$ contains $\mathcal{D}(L_0)$.*
- *For every $\varepsilon > 0$ there exists $C > 0$ such that $\|Bx\| \leq \varepsilon\|L_0 x\| + C\|x\|$ for every $x \in \mathcal{D}(L_0)$.*

Then the operator $L = L_0 + B$ (with domain $\mathcal{D}(L) = \mathcal{D}(L_0)$) is also the generator of an analytic semigroup.

Proof. In view of Theorem 4.25 it suffices to show that there exists a sector $S_{\theta,a}$ containing the spectrum of L and such that the resolvent bound $R_\lambda \leq Md(\lambda, S_{\theta,a})^{-1}$ holds away from it.

Denote by R^0_λ the resolvent for L_0 and consider the resolvent equation for L:

$$(\lambda - L_0 - B)x = y\ , \quad x \in \mathcal{D}(L_0)\ .$$

Since (at least for λ outside of some sector) x belongs to the range of R^0_λ, we can set $x = R^0_\lambda z$ so that this equation is equivalent to

$$z - BR^0_\lambda z = y\ .$$

The claim therefore follows if we can show that there exists a sector $S_{\theta,a}$ and a constant $c < 1$ such that $\|BR^0_\lambda\| \leq c$ for $\lambda \notin S_{\theta,a}$. This is because one then has the bound

$$\|R_\lambda y\| = \|R^0_\lambda z\| \leq \frac{\|R^0_\lambda\|}{1-c}\|y\|\ .$$

Using our assumption on B, we have the bound

$$\|BR_\lambda^0 z\| \le \varepsilon \|L_0 R_\lambda^0 z\| + C\|R_\lambda^0 z\| \; . \tag{4.7}$$

Furthermore, one has the identity $L_0 R_\lambda^0 = \lambda R_\lambda^0 - 1$ and, since L_0 is the generator of an analytic semigroup by assumption, the resolvent bound $\|R_\lambda^0\| \le M d(\lambda, S_{\alpha,b})^{-1}$ for some parameters α, b. Inserting this into (4.7), we obtain the bound

$$\|BR_\lambda^0\| \le \frac{(\varepsilon|\lambda| + C)M}{d(\lambda, S_{\alpha,b})} + \varepsilon \; .$$

Note now that by choosing $\theta \in (0, \alpha)$, we can find some $\delta > 0$ such that $d(\lambda, S_{\alpha,b}) > \delta|\lambda|$ for all $\lambda \notin S_{\theta,a}$ and all $a > 1 \vee (b+1)$. We fix such a θ and we make ε sufficiently small such that one has both $\varepsilon < 1/4$ and $\varepsilon\delta^{-1} < 1/4$.

We can then make a large enough so that $d(\lambda, S_{\alpha,b}) \ge 4CM$ for $\lambda \notin S_{\theta,a}$, so that $\|BR_\lambda^0\| \le 3/4$. for these values of λ, as requested. □

Remark 4.27 As one can see from the proof, one actually needs the bound $\|Bx\| \le \varepsilon\|L_0 x\| + C\|x\|$ only for some particular value of ε that depends on the characteristics of L_0.

As a consequence, we have:

Proposition 4.28 *Let $f \in L^\infty(\mathbf{R})$. Then, the operator*

$$(Lg)(x) = g''(x) + f(x)g'(x) \; ,$$

on $L^2(\mathbf{R})$ with domain $\mathcal{D}(L) = H^2$ is the generator of an analytic semigroup.

Proof. It is well-known that the operator $(L_0 g)(x) = g''(x)$ with domain $\mathcal{D}(L) = H^2$ is self-adjoint and negative definite, so that it is the generator of an analytic semigroup with angle $\theta = \pi/2$.

Setting $Bg = fg'$, we have for $g \in H^2$ the bound

$$\|Bg\|^2 = \int_{\mathbf{R}} f^2(x)\big(g'(x)\big)^2 \, dx \le \|f\|_{L^\infty}^2 \langle g', g'\rangle = -\|f\|_{L^\infty}^2 \langle g, g''\rangle \le \|f\|_{L^\infty}\|g\|\|L_0 g\| \; .$$

It now suffices to use the fact that $2|xy| \le \varepsilon x^2 + \varepsilon^{-1} y^2$ to conclude that the assumptions of Theorem 4.26 are satisfied. □

Similarly, one can show:

Exercise 4.29 Show that the generator of an elliptic diffusion with smooth coefficients on a compact Riemannian manifold $\mathcal{M}$ generates an analytic semigroup on $L^2(\mathcal{M}, \varrho)$, where ϱ is the volume measure given by the Riemannian structure.

4.4 Interpolation spaces

The remainder of this section will be devoted to the study of the domains of fractional powers of the generator L of an analytic semigroup $S(t)$. For simplicity, we will assume *throughout this section* that there exist $M > 0$ and $w > 0$ such that $\|S(t)\| \le Me^{-wt}$, thus making sure that the resolvent set of L contains all the right half of the complex plane. The general case can be recovered easily by "shifting the generator to the left". For $\alpha > 0$, we *define* negative fractional powers of L by

$$(-L)^{-\alpha} \stackrel{\text{def}}{=} \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1} S(t) \, dt \; , \tag{4.8}$$

which is a bounded operator by the decay assumption on $\|S(t)\|$. Since $\Gamma(1) = 1$, note that if $\alpha = 1$ one does indeed recover the resolvent of L evaluated at 0. Furthermore, it is straightforward to check that one has the identity $(-L)^{-\alpha}(-L)^{-\beta} = (-L)^{-\alpha-\beta}$, which together justify the definition (4.8).

Note that it follows from this identity that $(-L)^{-\alpha}$ is injective for every $\alpha > 0$. Indeed, given some $\alpha > 0$, one can find an integer $n > 0$ such that $(-L)^{-n} = (-L)^{-n+\alpha}(-L)^{-\alpha}$. A failure for $(-L)^{-\alpha}$ to be injective would therefore result in a failure for $(-L)^{-n}$ and therefore $(-L)^{-1}$ to be injective. This is ruled out by the fact that 0 belongs to the resolvent set of L. We can therefore define $(-L)^{\alpha}$ as the unbounded operator with domain $\mathcal{D}((-L)^{\alpha}) = \text{range}(-L)^{-\alpha}$ given by the inverse of $(-L)^{-\alpha}$. This definition is again consistent with the usual definition of $(-L)^{\alpha}$ for integer values of α. This allows us to set:

Definition 4.30 For $\alpha > 0$ and given an analytic semigroup S on a Banach space $\mathcal{B}$, we define the *interpolation space* $\mathcal{B}_\alpha$ as the domain of $(-L)^{\alpha}$ endowed with the norm $\|x\|_\alpha = \|(-L)^{\alpha}x\|$. We similarly define $\mathcal{B}_{-\alpha}$ as the completion of $\mathcal{B}$ for the norm $\|x\|_{-\alpha} = \|(-L)^{-\alpha}x\|$.

Remark 4.31 If the norm of $S(t)$ grows instead of decaying with t, then we use $\lambda - L$ instead of $-L$ for some λ sufficiently large. The choice of different values of λ leads to equivalent norms on $\mathcal{B}_\alpha$.

Exercise 4.32 Show that the inclusion $\mathcal{B}_\alpha \subset \mathcal{B}_\beta$ for $\alpha \geq \beta$ hold, whatever the signs of α and β.

Exercise 4.33 Show that for $\alpha \in (0,1)$ and $x \in \mathcal{D}(L)$, one has the identity

$$(-L)^{\alpha}x = \frac{\sin \alpha\pi}{\pi} \int_0^\infty t^{\alpha-1}(t-L)^{-1}(-L)x\, dt\ . \tag{4.9}$$

Hint: Write the resolvent appearing in (4.9) in terms of the semigroup and apply the resulting expression to $(-L)^{-\alpha}x$, as defined in (4.8). The aim of the game is then to perform a smart change of variables.

Exercise 4.34 Use (4.9) to show that, for every $\alpha \in (0,1)$, there exists a constant C such that the bound $\|(-L)^{\alpha}x\| \leq C\|Lx\|^{\alpha}\|x\|^{1-\alpha}$ holds for every $x \in \mathcal{D}(L)$.
Hint: Split the integral as $\int_0^\infty = \int_0^K + \int_K^\infty$ and optimise over K. (The optimal value for K will turn out to be proportional to $\|Lx\|/\|x\|$.) In the first integral, the identity $(t-L)^{-1}(-L) = 1 - t(t-L)^{-1}$ might come in handy.

Exercise 4.35 Let L be the generator of an analytic semigroup on $\mathcal{B}$ and denote by $\mathcal{B}_\alpha$ the corresponding interpolation spaces. Let B be a (possibly unbounded) operator on $\mathcal{B}$. Using the results from the previous exercise, show that if there exists $\alpha \in [0,1)$ such that $\mathcal{B}_\alpha \subset \mathcal{D}(B)$ so that B is a bounded operator from $\mathcal{B}_\alpha$ to $\mathcal{B}$, then one has the bound

$$\|Bx\| \leq C(\varepsilon\|Lx\| + \varepsilon^{-\alpha/(1-\alpha)}\|x\|)\ ,$$

for some constant $C > 0$ and for all $\varepsilon \leq 1$. In particular, $L + B$ is also the generator of an analytic semigroup on $\mathcal{B}$.
Hint: The assumption on B implies that there exists a constant C such that $\|Bx\| \leq C\|x\|_\alpha$.

Exercise 4.36 Let L and B be as in Exercise 4.35 and denote by S_B the analytic semigroup with generator $L + B$. Use the relation $R_\lambda - R_\lambda^0 = R_\lambda^0 B R_\lambda$ to show that one has the identity

$$S_B(t)x = S(t)x + \int_0^t S(t-s)BS_B(s)x\, ds\ .$$

Hint: Start from the right hand side of the equation and use an argument similar to that of the proof of Theorem 4.25.

Exercise 4.37 Show that $(-L)^\alpha$ commutes with $S(t)$ for every $t > 0$ and every $\alpha \in \mathbf{R}$. Deduce that $S(t)$ leaves $\mathcal{B}_\alpha$ invariant for every $\alpha > 0$.

Exercise 4.38 It follows from Theorem 4.25 that the restriction $L^\dagger$ of the adjoint L^* of the generator of an analytic semigroup on $\mathcal{B}$ to the "semigroup dual" space $\mathcal{B}^\dagger$ is again the generator of an analytic semigroup on $\mathcal{B}^\dagger$. Denote by $\mathcal{B}^\dagger_\alpha$ the corresponding interpolation spaces. Show that one has $\mathcal{B}^\dagger_\alpha = \mathcal{D}((-L^\dagger)^\alpha) \subset \mathcal{D}(((-L)^\alpha)^*) = (\mathcal{B}_{-\alpha})^*$ for every $\alpha \geq 0$.

We now show that an analytic semigroup $S(t)$ always maps $\mathcal{B}$ into $\mathcal{B}_\alpha$ for $t > 0$, so that it has a "smoothing effect". Furthermore, the norm in the domains of integer powers of L can be bounded by:

Proposition 4.39 *For every $t > 0$ and every integer $k > 0$, $S(t)$ maps $\mathcal{B}$ into $\mathcal{D}(L^k)$ and there exists a constant C_k such that*

$$\|L^k S(t)\| \leq \frac{C_k}{t^k}$$

for every $t \in (0, 1]$.

Proof. In order to show that S maps $\mathcal{B}$ into the domain of every power of L, we use (4.6), together with the identity $LR_\lambda = \lambda R_\lambda - 1$ which is an immediate consequence of the definition of the resolvent R_λ of L. Since $\int_{\gamma_{\varphi,b}} e^{tz} dz = 0$ for every t such that $|\arg t| < \varphi$ and since the domain of L^k is complete under the graph norm, this shows that $S(t)x \in \mathcal{D}(L^k)$ and

$$L^k S(t) = \frac{1}{2\pi i} \int_{\gamma_{\varphi,b}} z^k e^{tz} R_z \, dz \; .$$

It follows that there exist positive constants c_i such that

$$\|L^k S(t)\| \leq \frac{1}{2\pi} \int_{\gamma_{\varphi,b}} |z|^k |e^{tz}| \|R_z\| \, d|z| \leq c_1 \int_0^\infty (1+x)^k e^{-c_2 t(x-c_3)} (1+x)^{-1} dx \; .$$

Integrating by parts $k-1$ times, we obtain

$$\|L^k S(t)\| \leq \frac{c_4}{t^{k-1}} \int_0^\infty e^{-c_2 t(x-c_4)} \, dx = \frac{c_5 e^{c_6 t}}{t^k} \; ,$$

which implies the announced bound. □

It turns out that a similar bound also holds for interpolation spaces with non-integer indices:

Proposition 4.40 *For every $t > 0$ and every $\alpha > 0$, $S(t)$ maps $\mathcal{B}$ into $\mathcal{B}_\alpha$ and there exists a constant C_α such that*

$$\|(-L)^\alpha S(t)\| \leq \frac{C_\alpha}{t^\alpha} \tag{4.10}$$

for every $t \in (0, 1]$.

Proof. The fact that $S(t)$ maps $\mathcal{B}$ into $\mathcal{B}_\alpha$ follows from Proposition 4.39 since there exists n such that $\mathcal{D}(L^n) \subset \mathcal{B}_\alpha$. We assume again that the norm of $S(t)$ decays exponentially for large t. The claim for integer values of α is known to hold by Proposition 4.39, so we fix some $\alpha > 0$ which is *not* an integer. Note first that $(-L)^\alpha = (-L)^{\alpha - [\alpha] - 1}(-L)^{[\alpha]+1}$, were we denote by $[\alpha]$ the integer part of α. We thus obtain from (4.8) the identity

$$(-L)^\alpha S(t) = \frac{(-1)^{[\alpha]+1}}{\Gamma([\alpha] - \alpha + 1)} \int_0^\infty s^{[\alpha]-\alpha} L^{[\alpha]+1} S(t+s) \, ds \; .$$

Using the previous bound for $k = [\alpha]$, we thus get for some $C > 0$ the bound

$$\|(-L)^\alpha S(t)\| \leq C \int_0^\infty s^{[\alpha]-\alpha} \frac{e^{-w(t+s)}}{(t+s)^{[\alpha]+1}}\, ds \leq Ct^{-\alpha} \int_0^\infty \frac{s^{[\alpha]-\alpha}}{(1+s)^{[\alpha]+1}}\, ds\,,$$

where we used the substitution $s \mapsto ts$. Since the last function is integrable for every $\alpha > 0$, the claim follows at once. □

Exercise 4.41 Using the fact that $S(t)$ commutes with any power of its generator, show that $S(t)$ maps $\mathcal{B}_\alpha$ into $\mathcal{B}_\beta$ for every $\alpha, \beta \in \mathbf{R}$ and that, for $\beta > \alpha$, there exists a constant $C_{\alpha,\beta}$ such that $\|S(t)x\|_{\mathcal{B}_\beta} \leq C_{\alpha,\beta}\|x\|_{\mathcal{B}_\alpha} t^{\alpha-\beta}$ for all $t \in (0, 1]$.

Exercise 4.42 Using the bound from the previous exercise and the definition of the resolvent, show that for every $\alpha \in \mathbf{R}$ and every $\beta \in [\alpha, \alpha + 1)$ there exists a constant C such that the bound $\|(t - L)^{-1}x\|_{\mathcal{B}_\beta} \leq C(1 + t)^{\beta-\alpha-1}\|x\|_{\mathcal{B}_\alpha}$ holds for all $t \geq 0$.

Exercise 4.43 Consider an analytic semigroup $S(t)$ on $\mathcal{B}$ and denote by $\mathcal{B}_\alpha$ the corresponding interpolation spaces. Fix some $\gamma \in \mathbf{R}$ and denote by $\hat{S}(t)$ the semigroup S viewed as a semigroup on $\mathcal{B}_\gamma$. Denoting by $\hat{\mathcal{B}}_\alpha$ the interpolation spaces corresponding to $\hat{S}(t)$, show that one has the identity $\hat{\mathcal{B}}_\alpha = \mathcal{B}_{\gamma+\alpha}$ for every $\alpha \in \mathbf{R}$.

Another question that can be answered in a satisfactory way with the help of interpolation spaces is the speed of convergence of $S(t)x$ to x as $t \to 0$. We know that if $x \in \mathcal{D}(L)$, then $t \mapsto S(t)x$ is differentiable at $t = 0$, so that $\|S(t)x - x\| = t\|Lx\| + o(t)$. Furthermore, one can in general find elements $x \in \mathcal{B}$ so that the convergence $S(t)x \to x$ is arbitrarily slow. This suggests that if $x \in \mathcal{D}((-L)^\alpha)$ for $\alpha \in (0, 1)$, one has $\|S(t)x - x\| = \mathcal{O}(t^\alpha)$. This is indeed the case:

Proposition 4.44 *Let S be an analytic semigroup with generator L on a Banach space $\mathcal{B}$. Then, for every $\alpha \in (0, 1)$, there exists a constant C_α, so that the bound*

$$\|S(t)x - x\| \leq C_\alpha t^\alpha \|x\|_{\mathcal{B}_\alpha} \tag{4.11}$$

holds for every $x \in \mathcal{B}_\alpha$ and every $t \in (0, 1]$.

Proof. By density, it is sufficient to show that (4.11) holds for every $x \in \mathcal{D}(L)$. For such an x, one has indeed the chain of inequalities

$$\begin{aligned}\|S(t)x - x\| &= \Big\|\int_0^t S(s)Lx\, ds\Big\| = \Big\|\int_0^t (-L)^{1-\alpha} S(s)(-L)^\alpha x\, ds\Big\| \\ &\leq C\|x\|_{\mathcal{B}_\alpha} \int_0^t \|(-L)^{1-\alpha} S(s)\|\, ds \leq C\|x\|_{\mathcal{B}_\alpha} \int_0^t s^{\alpha-1}\, ds = C\|x\|_{\mathcal{B}_\alpha} t^\alpha\,.\end{aligned}$$

Here, the constant C depends only on α and changes from one expression to the next. □

We conclude this section with a discussion on the interpolation spaces arising from a perturbed analytic semigroup. As a consequence of Exercises 4.34, 4.35, and 4.42, we have the following result:

Proposition 4.45 *Let L_0 be the generator of an analytic semigroup on $\mathcal{B}$ and denote by $\mathcal{B}_\gamma^0$ the corresponding interpolation spaces. Let B be a bounded operator from $\mathcal{B}_\alpha^0$ to $\mathcal{B}$ for some $\alpha \in [0, 1)$. Let furthermore $\mathcal{B}_\gamma$ be the interpolation spaces associated to $L = L_0 + B$. Then, one has $\mathcal{B}_\gamma = \mathcal{B}_\gamma^0$ for every $\gamma \in [0, 1]$.*

Proof. The statement is clear for $\gamma = 0$ and $\gamma = 1$. For intermediate values of γ, we will show that there exists a constant C such that $C^{-1}\|(-L_0)^\gamma x\| \le \|(-L)^\gamma x\| \le C\|(-L_0)^\gamma x\|$ for every $x \in \mathcal{D}(L_0)$.

Since the domain of L is equal to the domain of L_0, we know that the operator BR_t is bounded for every $t > 0$, where R_t is the resolvent of L. Making use of the identity

$$R_t = R_t^0 + R_t^0 B R_t \,, \tag{4.12}$$

(where we similarly denoted by R_t^0 the resolvent of L_0) it then follows from Exercise 4.42 and the assumption on B that one has for every $x \in \mathcal{B}_\gamma^0$ the bound

$$\begin{aligned}\|BR_t x\| &\le \|BR_t^0 x\| + \|BR_t^0 BR_t x\| \le C(\|R_t^0 x\|_{\mathcal{B}_\alpha^0} + \|R_t^0 BR_t x\|_{\mathcal{B}_\alpha^0}) \\ &\le C(1+t)^{\alpha-\gamma-1}\|x\|_{\mathcal{B}_\gamma^0} + C(1+t)^{\alpha-1}\|BR_t x\| \,.\end{aligned}$$

It follows that, for t sufficiently large, one has the bound

$$\|BR_t x\| \le C(1+t)^{\alpha-\gamma-1}\|x\|_{\mathcal{B}_\gamma^0} \,. \tag{4.13}$$

(Note that this bound is also valid for $\gamma = 0$.) Since one furthermore has the resolvent identity $R_s = R_t + (t-s)R_s R_t$, this bound can be extended to all $t > 0$ by possibly changing the value of the constant C.

We now show that $\|(-L)^\gamma x\|$ can be bounded by $\|(-L_0)^\gamma x\|$. We make use of Exercise 4.34 to get, for $x \in \mathcal{D}(L_0)$, the bound

$$\begin{aligned}\|x\|_{\mathcal{B}_\gamma} &= C\Big\|\int_0^\infty t^{\gamma-1} L R_t x \, dt\Big\| \\ &\le C\Big\|\int_0^\infty t^{\gamma-1} L_0 R_t^0 x \, dt\Big\| + C\int_0^\infty t^{\gamma-1}\|(L_0 R_t^0 + 1)BR_t x\| \, dt \\ &\le \|x\|_{\mathcal{B}_\gamma^0} + C\int_0^\infty t^{\gamma-1}\|BR_t x\| \, dt \\ &\le \|x\|_{\mathcal{B}_\gamma^0} + C\int_0^\infty t^{\gamma-1}(1+t)^{\alpha-\gamma-1} \, dt \|x\|_{\mathcal{B}_\gamma^0} \,.\end{aligned}$$

Here, we used again the identity (4.12) to obtain the first inequality and we used (4.13) in the last step. Since this integral converges, we have obtained the required bound.

In order to obtain the converse bound, we have similarly to before

$$\|x\|_{\mathcal{B}_\gamma^0} \le \|x\|_{\mathcal{B}_\gamma} + C\int_0^\infty t^{\gamma-1}\|BR_t x\| \, dt \,.$$

Making use of the resolvent identity, this yields for arbitrary $K > 0$ the bound

$$\begin{aligned}\|x\|_{\mathcal{B}_\gamma^0} &\le \|x\|_{\mathcal{B}_\gamma} + C\int_0^\infty t^{\gamma-1}\|BR_{t+K} x\| \, dt + CK\int_0^\infty t^{\gamma-1}\|BR_{t+K}R_t x\| \, dt \\ &\le \|x\|_{\mathcal{B}_\gamma} + C\int_0^\infty t^{\gamma-1}(t+K)^{\alpha-\gamma-1} \, dt \|x\|_{\mathcal{B}_\gamma^0} + CK\int_0^\infty t^{\gamma-1}(1+t)^{-1} \, dt \|x\| \\ &\le \|x\|_{\mathcal{B}_\gamma} + CK^{\alpha-1}\|x\|_{\mathcal{B}_\gamma^0} + CK\|x\| \,.\end{aligned}$$

By making K sufficiently large, the prefactor of the second term can be made smaller than $\frac{1}{2}$, say, so that the required bound follows by the usual trick of moving the term proportional to $\|x\|_{\mathcal{B}_\gamma^0}$ to the left hand side of the inequality. □

Exercise 4.46 Assume that $\mathcal{B} = \mathcal{H}$ is a Hilbert space and that the antisymmetric part of L is "small" in the sense that $\mathcal{D}(L^*) = \mathcal{D}(L)$ and, for every $\varepsilon > 0$ there exists a constant C such that $\|(L - L^*)x\| \le \varepsilon\|Lx\| + C\|x\|$ for every $x \in \mathcal{D}(L)$. Show that in this case the space $\mathcal{H}_{-\alpha}$ can be identified with the dual of $\mathcal{H}_\alpha$ (under the pairing given by the scalar product of $\mathcal{H}$) for $\alpha \in [0, 1]$.

It is interesting to note that the range $[0,1]$ appearing in the statement of Proposition 4.45 is not just a restriction of the technique of proof employed here. There are indeed examples of perturbations of generators of analytic semigroups of the type considered here which induce changes in the corresponding interpolation spaces $\mathcal{B}_\alpha$ for $\alpha \notin [0,1]$.

Consider for example the case $\mathcal{B} = L^2([0,1])$ and $L_0 = \Delta$, the Laplacian with periodic boundary conditions. Denote by $\mathcal{B}^0_\alpha$ the corresponding interpolation spaces. Let now $\delta \in (0,1)$ be some arbitrary index and let $g \in \mathcal{B}$ be such that $g \notin \mathcal{B}^0_\delta$. Such an element g exists since Δ is an unbounded operator. Define B as the operator with domain $\mathcal{C}^1([0,1]) \subset \mathcal{B}$ given by

$$(Bf)(x) = f'(1/2)g(x)\ . \tag{4.14}$$

It turns out that $\mathcal{B}^0_\alpha \subset \mathcal{C}^1([0,1])$ for $\alpha > 3/4$ (see for example Lemma 6.23 below), so that the assumptions of Proposition 4.45 are indeed satisfied. Consider now the interpolation spaces of index $1+\delta$. Since we know that $\mathcal{B}_\delta = \mathcal{B}^0_\delta$, we have the characterisations

$$\mathcal{B}_{1+\delta} = \{f \in \mathcal{D}(\Delta)\ :\ \Delta f + f'(1/2)g \in \mathcal{B}^0_\delta\}\ ,$$
$$\mathcal{B}^0_{1+\delta} = \{f \in \mathcal{D}(\Delta)\ :\ \Delta f \in \mathcal{B}^0_\delta\}\ .$$

Since on the other hand $g \notin \mathcal{B}^0_\delta$ by assumption, it follows that $\mathcal{B}_{1+\delta} \cap \mathcal{B}^0_{1+\delta}$ consists precisely of those functions in $\mathcal{D}(\Delta)$ that have a vanishing derivative at $1/2$. In particular, $\mathcal{B}_{1+\delta} \neq \mathcal{B}^0_{1+\delta}$.

One can also show that $\mathcal{B}_{-1/4} \neq \mathcal{B}^0_{-1/4}$ in the following way. Let $\{f_n\} \subset \mathcal{D}(L)$ be an arbitrary sequence of elements that form a Cauchy sequence in $\mathcal{B}_{3/4}$. Since we have already shown that $\mathcal{B}_{3/4} = \mathcal{B}^0_{3/4}$, this implies that $\{f_n\}$ is Cauchy in $\mathcal{B}^0_{3/4}$ as well. It then follows from the definition of the interpolation spaces that the sequence $\{\Delta f_n\}$ is Cauchy in $\mathcal{B}^0_{-1/4}$ and that the sequence $\{(\Delta + B)f_n\}$ is Cauchy in $\mathcal{B}_{-1/4}$. Assume now by contradiction that $\mathcal{B}_{-1/4} = \mathcal{B}^0_{-1/4}$.

This would entail that both $\{\Delta f_n\}$ and $\{\Delta f_n + Bf_n\}$ are Cauchy in $\mathcal{B}_{-1/4}$, so that $\{f_n'(1/2)g\}$ is Cauchy in $\mathcal{B}_{-1/4}$. This in turn immediately implies that the sequence $\{f_n'(1/2)\}$ must be Cauchy in $\mathbf{R}$. Define now f_n by

$$f_n(x) = \sum_{k=1}^n \frac{\sin(4\pi k x)}{k^2 \log k}\ .$$

It is then straightforward to check that, since $\sum_k (k \log^2 k)^{-1}$ converges, this sequence is Cauchy in $\mathcal{B}^0_{3/4}$. On the other hand, we have $f_n'(1/2) = \sum_{k=1}^n (k \log k)^{-1}$ which diverges, thus leading to the required contradiction.

Exercise 4.47 Show, again in the same setting as above, that if $g \in \mathcal{B}^0_\delta$ for some $\delta > 0$, then one has $\mathcal{B}_\alpha = \mathcal{B}^0_\alpha$ for every $\alpha \in [0, 1+\delta)$.

Remark 4.48 The operator B defined in (4.14) is not a closed operator on $\mathcal{B}$. In fact, it is not even closable! This is however of no consequence for Proposition 4.45 since the operator $L = L_0 + B$ is closed and this is all that matters.

5

Linear SPDEs / Stochastic Convolutions

We now apply the knowledge gathered in the previous sections to discuss the solution to linear stochastic PDEs. Most of the material from this section can also be found in one way or the other in the monographs [DPZ92b, DPZ96] by Da Prato and Zabczyk. Our aim is to define what we mean by the solution to a linear stochastic PDE of the form

$$dx = Lx\,dt + Q\,dW(t)\;, \quad x(0) = x_0\;, \tag{5.1}$$

where we want x to be a stochastic process taking values in a separable Banach space $\mathcal{B}$, L is the generator of a $\mathcal{C}_0$ semigroup on $\mathcal{B}$, W is a cylindrical Wiener process on some Hilbert space $\mathcal{K}$, and $Q\colon \mathcal{K} \to \mathcal{B}$ is a bounded linear operator.

We do not in general expect x to take values in $\mathcal{D}(L)$ and we do not even in general expect $QW(t)$ to be a $\mathcal{B}$-valued Wiener process, so that the usual way of defining solutions to (5.1) by simply integrating both sides of the identity does not work. However, if we perform the usual trick borrowed from PDE theory of applying some $\ell \in \mathcal{D}(L^*)$ to both sides of (5.1), then there is much more hope that the usual definition makes sense. This motivates the following definition:

Definition 5.1 A $\mathcal{B}$-valued process $x(t)$ is said to be a *weak solution* to (5.1) if, for every $t > 0$, $\int_0^t \|x(s)\|\,ds < \infty$ almost surely and the identity

$$\langle \ell, x(t)\rangle = \langle \ell, x_0\rangle + \int_0^t \langle L^*\ell, x(s)\rangle\,ds + \int_0^t \langle Q^*\ell, dW(s)\rangle\;, \tag{5.2}$$

holds almost surely for every $\ell \in \mathcal{D}(L^*)$.

Remark 5.2 **(Very important!)** The term "weak" refers to the PDE notion of a weak solution and *not* to the probabilistic notion of a weak solution to a stochastic differential equation as they appear for example in [SV79, RW94]. From a probabilistic point of view, we are always going to be dealing with strong solutions in these notes, in the sense that (5.1) can be solved pathwise for almost every realisation of the cylindrical Wiener process W.

Just as in the case of stochastic ordinary differential equations, there are examples of (nonlinear) stochastic PDEs that are sufficiently irregular so that they can only be solved in the probabilistic weak sense. We will however not consider any such example in these notes, but we refer the interested reader to the papers [WY71a, WY71b, MPS06, MP11] which explore the limits of the "pathwise" approach both in the SDE and the SPDE case.

Remark 5.3 The stochastic integral in (5.2) can be interpreted in the sense of Section 3.7 since the map $Q^*\ell\colon \mathcal{K} \to \mathbf{R}$ is Hilbert-Schmidt for every $\ell \in \mathcal{B}^*$.

Remark 5.4 Although separability of $\mathcal{B}$ was not required in the previous section on semigroup theory, it is again needed in this section, since many of the results from the section on Gaussian measure theory would not hold otherwise.

On the other hand, suppose that $f\colon \mathbf{R}_+ \to \mathcal{D}(L)$ is a continuous function and consider the function $x\colon \mathbf{R}_+ \to \mathcal{D}(L)$ given by $x(t) = S(t)x_0 + \int_0^t S(t-s)f(s)\,ds$, where S is the $\mathcal{C}_0$-semigroup generated by L. If $x_0 \in \mathcal{D}(L)$ as well, then this function is differentiable and it is easy to check, using Proposition 4.10, that it satisfies the differential equation $\partial_t x = Lx + f$. Formally replacing $f(s)\,ds$ by $Q\,dW(s)$, this suggests the following alternative definition of a solution to (5.1):

Definition 5.5 Suppose that there exists a $\mathcal{B}$-valued process $x(t)$ such that, for every $t > 0$, the identity

$$x(t) = S(t)x_0 + \int_0^t S(t-s)Q\,dW(s)\;, \tag{5.3}$$

holds almost surely (in the sense that it holds when testing against any $\ell \in \mathcal{B}^*$). Then x is said to be the *mild solution* to (5.1).

Remark 5.6 The right hand side of (5.3) is also sometimes called the *stochastic convolution* of the Wiener process QW with the semigroup S.

Remark 5.7 By the results from Section 3.7, the right hand side of (5.3) makes sense in any Hilbert space $\mathcal{H}$ containing $\mathcal{B}$ and such that $\int_0^t \operatorname{tr} \iota S(t-s)QQ^*S(t-s)^*\iota^*\,ds < \infty$, where $\iota\colon \mathcal{B} \to \mathcal{H}$ is the inclusion map. The statement can then alternatively be interpreted as saying that the right hand side belongs to $\mathcal{B} \subset \mathcal{H}$ almost surely for every t. In the case where $\mathcal{B}$ is itself a Hilbert space, (5.3) makes sense if and only if $\int_0^t \operatorname{tr} S(t-s)QQ^*S(t-s)^*\,ds < \infty$.

It turns out that these two notions of solutions are actually equivalent:

Proposition 5.8 *If the mild solution is almost surely integrable, then it is also a weak solution. Conversely, every weak solution is a mild solution.*

Proof. Note first that, by considering the process $x(t) - S(t)x_0$ and using Proposition 4.10, we can assume without loss of generality that $x_0 = 0$.

We now assume that the process $x(t)$ defined by (5.3) takes values in $\mathcal{B}$ almost surely and we show that this implies that it satisfies (5.2). Fixing an arbitrary $\ell \in \mathcal{D}(L^\dagger)$, applying $L^*\ell$ to both sides of (5.3), and integrating the result between 0 and t, we obtain:

$$\int_0^t \langle L^*\ell, x(s)\rangle\,ds = \int_0^t \int_0^s \langle L^*\ell, S(s-r)Q\,dW(r)\rangle\,ds = \int_0^t \Big\langle \int_r^t S^*(s-r)L^*\ell\,ds, Q\,dW(r)\Big\rangle\;.$$

Using Proposition 4.10 and the fact that, by Proposition 4.18, S^* is a strongly continuous semigroup on $\mathcal{B}^\dagger$, the closure of $\mathcal{D}(L^*)$ in $\mathcal{B}^*$, we obtain

$$\begin{aligned}\int_0^t \langle L^*\ell, x(s)\rangle\,ds &= \int_0^t \langle S^*(t-r)\ell, Q\,dW(r)\rangle - \int_0^t \langle \ell, Q\,dW(r)\rangle \\ &= \Big\langle \ell, \int_0^t S(t-r)Q\,dW(r)\Big\rangle - \int_0^t \langle \ell, Q\,dW(r)\rangle \\ &= \langle \ell, x(t)\rangle - \int_0^t \langle \ell, Q\,dW(r)\rangle\;,\end{aligned}$$

thus showing that (5.2) holds for every $\ell \in \mathcal{D}(L^\dagger)$. To show that x is indeed a weak solution to (5.1), we have to extend this to every $\ell \in \mathcal{D}(L^*)$. This however follows immediately from the fact that $\mathcal{B}^\dagger$ is weak-* dense in $\mathcal{B}^*$, which was the content of Proposition 4.20.

To show the converse, let now $x(t)$ be any weak solution to (5.1) (again with $x_0 = 0$). Fix an arbitrary $\ell \in \mathcal{D}(L^\dagger)$, some final time $t > 0$, and consider the function $f(s) = S^*(t-s)\ell$. Since $\ell \in \mathcal{D}(L^\dagger)$, it

follows from Proposition 4.10 that this function belongs to $\mathcal{E} \stackrel{\text{def}}{=} \mathcal{C}([0,t], \mathcal{D}(L^\dagger)) \cap \mathcal{C}^1([0,t], \mathcal{B}^\dagger)$. We are going to show that one has for such functions the almost sure identity

$$\langle f(t), x(t) \rangle = \int_0^t \langle \dot{f}(s) + L^* f(s), x(s) \rangle \, ds + \int_0^t \langle f(s), Q \, dW(s) \rangle \; . \tag{5.4}$$

Since in our case $\dot{f}(s) + L^* f(s) = 0$, this implies that the identity

$$\langle \ell, x(t) \rangle = \int_0^t \langle \ell, S(t-s) Q \, dW(s) \rangle \; , \tag{5.5}$$

holds almost surely for all $\ell \in \mathcal{D}(L^\dagger)$. By the closed graph theorem, $\mathcal{B}^\dagger$ is large enough to separate points in $\mathcal{B}$.[1] Since $\mathcal{D}(L^\dagger)$ is dense in $\mathcal{B}^\dagger$ and since $\mathcal{B}$ is separable, it follows that countably many elements of $\mathcal{D}(L^\dagger)$ are already sufficient to separate points in $\mathcal{B}$. This then immediately implies from (5.5) that x is indeed a mild solution.

It remains to show that (5.4) holds for all $f \in \mathcal{E}$. Since linear combinations of functions of the type $\varphi_\ell(s) = \ell \varphi(s)$ for $\varphi \in \mathcal{C}^1([0,t], \mathbf{R})$ and $\ell \in \mathcal{D}(L^\dagger)$ are dense in $\mathcal{E}$ (see Exercise 5.10 below) and since x is almost surely integrable, it suffices to show that (5.4) holds for $f = \varphi_\ell$. Since $\langle \ell, QW(s) \rangle$ is a one-dimensional Brownian motion, we can apply Itô's formula to $\varphi(s)\langle \ell, x(s) \rangle$, yielding

$$\varphi(t)\langle \ell, x(t) \rangle = \int_0^t \varphi(s) \langle L^* \ell, x(s) \rangle + \int_0^t \dot{\varphi}(s) \langle \ell, x(s) \rangle + \int_0^t \varphi(s) \langle \ell, Q \, dW(s) \rangle \; ,$$

which coincides with (5.4) as required. □

Remark 5.9 It is actually possible to show that if the right hand side of (5.3) makes sense for some t, then it makes sense for all t and the resulting process belongs almost surely to $L^p([0,T], \mathcal{B})$ for every p. Therefore, the concepts of mild and weak solution actually always coincide. This follows from the fact that the covariance of $x(t)$ increases with t (which is a concept that can easily be made sense of in Banach spaces as well as Hilbert spaces), see for example [DJT95].

Exercise 5.10 Consider the setting of the proof of Proposition 5.8. Let $f \in \mathcal{E} = \mathcal{C}([0,1], \mathcal{D}(L^\dagger)) \cap \mathcal{C}^1([0,1], \mathcal{B}^\dagger)$ and, for $n > 0$, define f_n on the interval $s \in [k/n, (k+1)/n]$ by cubic spline interpolation:

$$\begin{aligned} f_n(s) = {} & f(k/n)(k+1-ns)^2(1+2ns-2k) + f((k+1)/n)(ns-k)^2(3-2ns+2k) \\ & + (ns-k)(k+1-ns)^2 n (f((k+\tfrac{1}{2})/n) - f((k-\tfrac{1}{2})/n)) \\ & + (ns-k)^2(ns-k-1) n (f((k+\tfrac{3}{2})/n) - f((k+\tfrac{1}{2})/n)) \; . \end{aligned}$$

Show that f_n is a finite linear combinations of functions of the form $\ell\varphi(s)$ with $\varphi \in \mathcal{C}^1([0,1], \mathbf{R})$ and that $f_n \to f$ in $\mathcal{C}([0,1], \mathcal{D}(L^\dagger)) \cap \mathcal{C}^1([0,1], \mathcal{B}^\dagger)$.

5.1 Time and space regularity

In this subsection, we are going to study the space and time regularity of solutions to linear stochastic PDEs. For example, we are going to see how one can easily derive the fact that the solutions to the stochastic heat equation are "almost" $\frac{1}{4}$-Hölder continuous in time and "almost" $\frac{1}{2}$-Hölder continuous in space. Since we are often going to use the Hilbert-Schmidt norm of a linear operator, we introduce the notation

$$\|A\|_{\text{HS}}^2 = \text{tr}\, AA^* \; .$$

[1] Assume that, for some $x, y \in \mathcal{B}$, we have $\langle \ell, x \rangle = \langle \ell, y \rangle$ for every $\ell \in \mathcal{D}(L^*)$. We can also assume without loss of generality that the range of L is $\mathcal{B}$, so that $x = Lx'$ and $y = Ly'$, thus yielding $\langle L^*\ell, x' \rangle = \langle L^*\ell, y' \rangle$. Since L is injective and has dense domain, the closed graph theorem states that the range of L^* is all of $\mathcal{B}^*$, so that $x' = y'$ and thus also $x = y$.

For most of this section, we are going to make use of the theory of analytic semigroups. However, we start with a very weak regularity result for the solutions to stochastic PDEs whose linear operator L generates an arbitrary $\mathcal{C}_0$-semigroup:

Theorem 5.11 *Let $\mathcal{H}$ and $\mathcal{K}$ be separable Hilbert spaces, let L be the generator of a $\mathcal{C}_0$-semigroup on $\mathcal{H}$, let $Q\colon \mathcal{K} \to \mathcal{H}$ be a bounded operator and let W be a cylindrical Wiener process on $\mathcal{K}$. Assume furthermore that $\|S(t)Q\|_{\text{HS}} < \infty$ for every $t > 0$ and that there exists $\alpha \in (0, \frac{1}{2})$ such that $\int_0^1 t^{-2\alpha}\|S(t)Q\|_{\text{HS}}^2\,dt < \infty$. Then the solution x to (5.1) has almost surely continuous sample paths in $\mathcal{H}$.*

Proof. Note first that $\|S(t+s)Q\|_{\text{HS}} \le \|S(s)\|\|S(t)Q\|_{\text{HS}}$, so that the assumptions of the theorem imply that $\int_0^T t^{-2\alpha}\|S(t)Q\|_{\text{HS}}^2\,dt < \infty$ for every $T > 0$. Let us fix an arbitrary terminal time T from now on. Defining the process y by

$$y(t) = \int_0^t (t-s)^{-\alpha} S(t-s)Q\,dW(s)\;,$$

we obtain the existence of a constant C such that

$$\mathbf{E}\|y(t)\|^2 = \int_0^t (t-s)^{-2\alpha}\|S(t-s)Q\|_{\text{HS}}^2\,ds = \int_0^t s^{-2\alpha}\|S(s)Q\|_{\text{HS}}^2\,ds \le C\;,$$

uniformly for $t \in [0,T]$. It therefore follows from Fernique's theorem that for every $p > 0$ there exist a constant C_p such that

$$\mathbf{E}\int_0^T \|y(t)\|^p\,dt < C_p\;. \tag{5.6}$$

Note now that there exists a constant c_α (actually $c_\alpha = (\sin 2\pi\alpha)/\pi$) such that the identity

$$\int_s^t (t-r)^{\alpha-1}(r-s)^{-\alpha}\,dr = \frac{1}{c_\alpha}\;,$$

holds for every $t > s$. It follows that one has the identity

$$\begin{aligned}
x(t) &= S(t)x_0 + c_\alpha \int_0^t \int_s^t (t-r)^{\alpha-1}(r-s)^{-\alpha} S(t-s)\,dr\,Q\,dW(s)\\
&= S(t)x_0 + c_\alpha \int_0^t \int_0^r (t-r)^{\alpha-1}(r-s)^{-\alpha} S(t-s)Q\,dW(s)\,dr\\
&= S(t)x_0 + c_\alpha \int_0^t S(t-r)\int_0^r (r-s)^{-\alpha} S(r-s)Q\,dW(s)\,(t-r)^{\alpha-1}\,dr\\
&= S(t)x_0 + c_\alpha \int_0^t S(t-r)y(r)\,(t-r)^{\alpha-1}\,dr\;.
\end{aligned} \tag{5.7}$$

The claim thus follows from (5.6) if we can show that for every $\alpha \in (0, \frac{1}{2})$ there exists $p > 0$ such that the map

$$y \mapsto F_y\;,\quad F_y(t) = \int_0^t (t-r)^{\alpha-1} S(t-r)y(r)\,dr$$

maps $L^p([0,T],\mathcal{H})$ into $\mathcal{C}([0,T],\mathcal{H})$. Since the semigroup $t \mapsto S(t)$ is uniformly bounded (in the usual operator norm) on any bounded time interval and since $t \mapsto (t-r)^{\alpha-1}$ belongs to L^q for $q \in [1, 1/(1-\alpha))$, we deduce from Hölder's inequality that there exists a constant C_T such that one does indeed have the bound $\sup_{t\in[0,T]}\|F_y(t)\|^p \le C_T\int_0^T \|y(t)\|^p\,dt$, provided that $p > \frac{1}{\alpha}$. Since continuous functions are dense in L^p, the proof is complete if we can show that F_y is continuous for every continuous function y with $y(0) = 0$.

Fixing such a y, we first show that F_y is right-continuous and then that it is left continuous. Fixing $t > 0$, we have for $h > 0$ the bound

$$\|F_y(t+h) - F_y(t)\| \le \int_0^t \|((t+h-r)^{\alpha-1}S(h) - (t-r)^{\alpha-1})S(t-r)y(r)\|\,dr + \int_t^{t+h} (t+h-r)^{\alpha-1}\|S(t+h-r)y(r)\|\,dr$$

The second term is bounded by $\mathcal{O}(h^\delta)$ for some $\delta > 0$ by Hölder's inequality. It follows from the strong continuity of S that the integrand of the first term converges to 0 pointwise as $h \to 0$. Since on the other hand the integrand is bounded by $C(t-r)^{\alpha-1}\|y(r)\|$ for some constant C, this term also converges to 0 by the dominated convergence theorem. This shows that F_y is right continuous.

To show that F_y is also left continuous, we write

$$\|F_y(t) - F_y(t-h)\| \le \int_0^{t-h} \|((t-r)^{\alpha-1}S(h) - (t-h-r)^{\alpha-1})S(t-h-r)y(r)\|\,dr + \int_{t-h}^t (t-r)^{\alpha-1}\|S(t-r)y(r)\|\,dr\ .$$

We bound the second term by Hölder's inequality as before. The first term can be rewritten as

$$\int_0^t \|((t+h-r)^{\alpha-1}S(h) - (t-r)^{\alpha-1})S(t-r)y(r-h)\|\,dr\ ,$$

with the understanding that $y(r) = 0$ for $r < 0$. Since we assumed that y is continuous, we can again use the dominated convergence theorem to show that this term tends to 0 as $h \to 0$. □

Remark 5.12 The trick employed in (5.7) is sometimes called the "factorisation method" and was introduced in the context of stochastic convolutions by Da Prato, Kwapień, and Zabczyk [DPKZ87, DPZ92a].

This theorem is quite sharp in the sense that, without any further assumption on Q and L, it is not possible in general to deduce that $t \mapsto x(t)$ has more regularity than just continuity, even if we start with a very regular initial condition, say $x_0 = 0$. We illustrate this fact with the following exercise:

Exercise 5.13 Consider the case $\mathcal{H} = L^2(\mathbf{R})$, $\mathcal{K} = \mathbf{R}$, $L = \partial_x$ and $Q = g$ for some $g \in L^2(\mathbf{R})$ such that $g \ge 0$ and $g(x) = |x|^{-\beta}$ for some $\beta \in (0, \frac{1}{2})$ and all $|x| < 1$. This satisfies the conditions of Theorem 5.11 for any $\alpha < 1$.

Since L generates the translation group, the solution to

$$du(x,t) = \partial_x u(x,t)\,dt + g(x)\,dW(t)\ , \quad u(x,0) = 0\ ,$$

is given by

$$u(x,t) = \int_0^t g(x+t-s)\,dW(s)\ .$$

Convince yourself that for fixed t, the map $x \mapsto u(x,t)$ is in general γ-Hölder continuous for $\gamma < \frac{1}{2} - \beta$, but no better. Deduce from this that the map $t \mapsto u(\cdot,t)$ is in general also γ-Hölder continuous for $\gamma < \frac{1}{2} - \beta$ (if we consider it either as an $\mathcal{H}$-valued map or as a $\mathcal{C}_b(\mathbf{R})$-valued map), but cannot be expected to have more regularity than that. Since β can be chosen arbitrarily close to $\frac{1}{2}$, it follows that the exponent α appearing in Theorem 5.11 is in general independent of the Hölder regularity of the solution.

One of the main insights of regularity theory for parabolic PDEs (both deterministic and stochastic) is that space regularity is intimately linked to time regularity in several ways. Very often, the knowledge that a solution has a certain spatial regularity for fixed time implies that it also has a certain temporal regularity at a given spatial location.

From a slightly different point of view, if we consider time-regularity of the solution to a PDE viewed as an evolution equation in some infinite-dimensional space of functions, then the amount of regularity that one obtains depends on the functional space under consideration. As a general rule, the smaller the space (and therefore the more spatial regularity it imposes) the lower the time-regularity of the solution, viewed as a function with values in that space.

We start by giving a general result that tells us precisely in which interpolation space one can expect to find the solution to a linear SPDE associated with an analytic semigroup. This provides us with the optimal spatial regularity for a given SPDE:

Theorem 5.14 *Consider (5.1) on a Hilbert space $\mathcal{H}$, assume that L generates an analytic semigroup, and denote by $\mathcal{H}_\alpha$ the corresponding interpolation spaces. If there exists $\alpha \geq 0$ such that $Q\colon \mathcal{K} \to \mathcal{H}_\alpha$ is bounded and $\beta \in (0, \frac{1}{2} + \alpha]$ such that $\|(-L)^{-\beta}\|_{\text{HS}} < \infty$ then the solution x takes values in $\mathcal{H}_\gamma$ for every $\gamma < \gamma_0 = \frac{1}{2} + \alpha - \beta$.*

Proof. As usual, we can assume without loss of generality that 0 belongs to the resolvent set of L. It suffices to show that

$$I(T) \stackrel{\text{def}}{=} \int_0^T \|(-L)^\gamma S(t)Q\|_{\text{HS}}^2 \, dt < \infty \,, \qquad \forall T > 0 \,.$$

Since Q is assumed to be bounded from $\mathcal{K}$ to $\mathcal{H}_\alpha$, there exists a constant C such that

$$I(T) \leq C \int_0^T \|(-L)^\gamma S(t)(-L)^{-\alpha}\|_{\text{HS}}^2 \, dt = C \int_0^T \|(-L)^{\gamma-\alpha} S(t)\|_{\text{HS}}^2 \, dt \,.$$

Since $(-L)^{-\beta}$ is Hilbert-Schmidt, we have the bound

$$\|(-L)^{\gamma-\alpha} S(t)\|_{\text{HS}} \leq \|(-L)^{-\beta}\|_{\text{HS}} \|(-L)^{\beta+\gamma-\alpha} S(t)\| \leq C(1 \vee t^{\alpha-\gamma-\beta}) \,.$$

For this expression to be square integrable near $t = 0$, we need $\alpha - \gamma - \beta > -\frac{1}{2}$, which is precisely the stated condition. □

Exercise 5.15 Show that if we are in the setting of Theorem 5.14 and L is selfadjoint, then the solutions to (5.1) actually belong to $\mathcal{H}_\gamma$ for $\gamma = \gamma_0$.

Exercise 5.16 Show that the solution to the stochastic heat equation on $[0, 1]$ with periodic boundary conditions (driven by space-time white noise) has solutions in the fractional Sobolev space H^s for every $s < 1/2$. Recall that H^s is the Hilbert space with scalar product $\langle f, g \rangle_s = \sum_k \hat{f}_k \hat{g}_k (1 + k^2)^s$, where $\hat{f}_k$ denotes the kth Fourier coefficient of f.

Exercise 5.17 Consider the following modified stochastic heat equation on $[0, 1]^d$ with periodic boundary conditions:

$$dx = \Delta x \, dt + (1 - \Delta)^{-\gamma} \, dW \,,$$

where W is a cylindrical Wiener process on $L^2([0, 1]^d)$. For any given $s \geq 0$, how large does γ need to be for x to take values in H^s ?

Using this knowledge about the spatial regularity of solutions, we can now turn to the time-regularity. We have:

Theorem 5.18 *Consider the same setting as in Theorem 5.14 and fix $\gamma < \gamma_0$. Then, at all times $t > 0$, the process x is almost surely δ-Hölder continuous in $\mathcal{H}_\gamma$ for every $\delta < \frac{1}{2} \wedge (\gamma_0 - \gamma)$.*

Proof. It follows from Kolmogorov's continuity criteria, Proposition 3.43, that it suffices to check that the bound

$$\mathbf{E}\|x(t) - x(s)\|_\gamma^2 \le C|t-s|^{1\wedge 2(\tilde\gamma-\gamma)}$$

holds uniformly in $s, t \in [t_0, T]$ for every $t_0, T > 0$ and for every $\tilde\gamma < \gamma_0$. Here and below, C is an unspecified constant that changes from expression to expression. Assume that $t > s$ from now on. It follows from the semigroup property and the independence of the increments of W that the identity

$$x(t) = S(t-s)x(s) + \int_s^t S(t-r)Q\,dW(r)\;, \tag{5.8}$$

holds almost surely, where the two terms in the sum are independent. This property is also called the *Markov property*. Loosely speaking, it states that the future of x depends on its present, but not on its past. This transpires in (5.8) through the fact that the right hand side depends on $x(s)$ and on the increments of W between times s and t, but it does not depend on $x(r)$ for any $r < s$.

Furthermore, $x(s)$ is independent of the increments of W over the interval $[s, t]$, so that Proposition 4.44 allows us to get the bound

$$\begin{aligned}\mathbf{E}\|x(t) - x(s)\|_\gamma^2 &= \mathbf{E}\|S(t-s)x(s) - x(s)\|_\gamma^2 + \int_0^{t-s} \|(-L)^\gamma S(r)Q\|_{\text{HS}}^2\,dr \\ &\le C|t-s|^{2(\tilde\gamma-\gamma)\wedge 2}\mathbf{E}\|x(s)\|_{\tilde\gamma}^2 + C\int_0^{t-s}(1 \vee r^{\alpha-\gamma-\beta})^2\,dr\;.\end{aligned}$$

Here, we obtained the bound on the second term in exactly the same way as in the proof of Theorem 5.14. The claim now follows from the fact that $\alpha - \gamma - \beta = (\gamma_0 - \gamma) - \frac{1}{2}$. □

5.2 Long-time behaviour

This section is devoted to the behaviour of the solutions to (5.1) for large times. Let us again start with an example that illustrates some of the possible behaviours.

Example 5.19 Let $x \mapsto V(x)$ be some smooth "potential" and let $\mathcal{H} = L^2(\mathbf{R}, \exp(-V(x))\,dx)$. Let S denote the translation semigroup (to the right) on $\mathcal{H}$ and denote its generator by $-\partial_x$. Let us first discuss which conditions on V ensure that S is a strongly continuous semigroup on $\mathcal{H}$. It is clear that it is a semigroup and that $S(t)u \to u$ for u any smooth function with compact support. It therefore only remains to show that $\|S(t)\|$ is uniformly bounded for $t \in [0, 1]$ say. We have

$$\|S(t)u\|^2 = \int u^2(x-t)e^{-V(x)}\,dx = \int u^2(x)e^{-V(x)}e^{V(x)-V(x+t)}\,dx\;. \tag{5.9}$$

This shows that a necessary and sufficient condition for S to be a strongly continuous semigroup on $\mathcal{H}$ is that, for every $t > 0$, there exists C_t such that $\sup_{x\in\mathbf{R}}(V(x) - V(x+t)) \le C_t$ and such that C_t remains bounded as $t \to 0$. Examples of potentials leading to a $\mathcal{C}_0$-semigroup are x, $\sqrt{1+x^2}$, $\log(1+x^2)$, etc or any increasing function. Note however that the potential $V(x) = x^2$ does *not* lead to a strongly continuous semigroup. One different way of interpreting this is to consider the unitary transformation $K\colon u \mapsto \exp(\frac{1}{2}V)u$ from the "flat" space L^2 into $\mathcal{H}$. Under this transformation, the generator $-\partial_x$ is turned into

$$-(K^{-1}\partial_x K u)(x) = -\partial_x u(x) - \tfrac{1}{2}V'(x)u(x)\;.$$

Considering the characterisation of generators of $\mathcal{C}_0$-semigroups given by the Hille-Yosida theorem,

one would expect this to be the generator of a strongly continuous semigroup if V' is bounded from below, which is indeed a sufficient condition.

Let now V be such that S is a $\mathcal{C}_0$-semigroup and consider the SPDE on $\mathcal{H}$ given by

$$du(x,t) = -\partial_x u(x,t)\,dt + f(x)\,dW(t)\;, \tag{5.10}$$

where W is a one-dimensional Wiener process and f is some function in $\mathcal{H}$. The solution to (5.10) with initial condition $u_0 = 0$ is given as before by

$$u(x,t) = \int_0^t f(x+s-t)\,dW(s)\;. \tag{5.11}$$

If we fix the time t, we can make the change of variable $s \mapsto t-s$, so that $u(x,t)$ is equal in distribution to $\int_0^t f(x-s)\,dW(s)$.

We see that if f happens to be also square integrable (we will assume that this is the case in the sequel and we will also assume that f is not identically zero), then (5.11) has a limit in distribution as $t \to \infty$ given by

$$\tilde{u}(x) = \int_0^\infty f(x-s)\,dW(s)\;. \tag{5.12}$$

It is however not clear *a priori* that $\tilde{u}$ does belong to $\mathcal{H}$. On one hand, we have the bound

$$\mathbf{E}\int_{\mathbf{R}} \tilde{u}(x)^2 e^{-V(x)}\,dx = \int_{\mathbf{R}}\int_0^\infty f^2(x-t)\,dt\,e^{-V(x)}\,dx \le \int_{\mathbf{R}} f^2(t)\,dt \int_{\mathbf{R}} e^{-V(x)}\,dx\;,$$

thus showing that $\tilde{u}$ definitely belongs to $\mathcal{H}$ if e^{-V} has finite mass. On the other hand, there are examples where $\tilde{u} \in \mathcal{H}$ even though e^{-V} has infinite mass. For example, if $f(x) = 0$ for $x \le 0$, then it is necessary and sufficient to have $\int_0^\infty e^{-V(x)}\,dx < \infty$. Denote by ν the law of $\tilde{u}$ for further reference.

Furthermore, if e^{-V} is integrable, there are many measures on $\mathcal{H}$ that are invariant under the action of the semigroup S. For example, given a function $h \in \mathcal{H}$ which is periodic with period τ (that is $S(\tau)h = h$), we can check that the push-forward of the Lebesgue measure on $[0,\tau]$ under the map $t \mapsto S(t)h$ is invariant under the action of S. This is simply a consequence of the invariance of Lebesgue measure under the shift map. Given any invariant probability measure μ_h of this type, let v be an $\mathcal{H}$-valued random variable with law μ_h that is independent of W. We can then consider the solution to (5.10) with initial condition v. Since the law of $S(t)v$ is equal to the law of v by construction, it follows that the law of the solution converges to the distribution of the random variable $\tilde{u} + v$, with the understanding that $\tilde{u}$ and v are independent.

This shows that in the case $\int e^{-V(x)}\,dx < \infty$, it is possible to construct solutions u to (5.10) such that the law of $u(\cdot\,,t)$ converges to $\mu_h \star \nu$ for any periodic function h.

Exercise 5.20 Construct an example of a potential V such that the semigroup S from the previous example is *not* strongly continuous. You can do this by choosing V in such a way that $\lim_{t\to 0}\|S(t)\| = +\infty$, even though each of the operators $S(t)$ for $t > 0$ is bounded! **Hint:** Choose V of the form $V(x) = x^3 - \sum_{n>0} nW(\frac{x-c_n}{n})$, where W is an isolated "spike" and c_n are suitably chosen constants.

Example 5.19 shows that in general, the long-time behaviour of solutions to (5.1) may depend on the choice of initial condition, even if the solutions are stable in the sense that they remain bounded in probability. It also shows that depending on the behaviour of $\mathcal{H}$, L and Q, the law of the solutions may or may not converge to a limiting distribution in the space in which solutions are considered.

In order to formalise the concept of "long-time behaviour of solutions" for (5.1), it is convenient to introduce the *Markov semigroup* associated to (5.1). Given a linear SPDE with solutions in $\mathcal{B}$, we can define a family $\mathcal{P}_t$ of bounded linear operators on $\mathscr{B}_b(\mathcal{B})$, the space of Borel measurable bounded functions from $\mathcal{B}$ to $\mathbf{R}$ by

$$(\mathcal{P}_t\varphi)(x) = \mathbf{E}\varphi\Big(S(t)x + \int_0^t S(t-s)Q\,dW(s)\Big)\;. \tag{5.13}$$

The operators $\mathcal{P}_t$ are *Markov operators* in the sense that the map $A \mapsto \mathcal{P}_t \mathbf{1}_A(x)$ is a probability measure on $\mathcal{B}$ for every fixed x. In particular, one has $\mathcal{P}_t \mathbf{1} = \mathbf{1}$ and $\mathcal{P}_t \varphi \geq 0$ if $\varphi \geq 0$, that is the operators $\mathcal{P}_t$ preserve positivity. It follows furthermore from (5.8) and the independence of the increments of W over disjoint time intervals that $\mathcal{P}_t$ satisfies the semigroup property $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s$ for any two times $s, t \geq 0$.

Exercise 5.21 Show that $\mathcal{P}_t$ maps the space $\mathcal{C}_b(\mathcal{B})$ of continuous bounded functions from $\mathcal{B}$ to $\mathbf{R}$ into itself. (Recall that we assumed $\mathcal{B}$ to be separable.)

If we denote by $\mathcal{P}_t(x, \cdot\,)$ the law of $S(t)x + \int_0^t S(t-s)Q\, dW(s)$, then $\mathcal{P}_t$ can alternatively be represented as

$$(\mathcal{P}_t \varphi)(x) = \int_{\mathcal{B}} \varphi(y)\, \mathcal{P}_t(x, dy) \; .$$

It follows that its dual $\mathcal{P}_t^*$ acts on measures with finite total variation by

$$(\mathcal{P}_t^* \mu)(A) = \int_{\mathcal{B}} \mathcal{P}_t(x, A)\, \mu(dx) \; .$$

Since it preserves the mass of positive measures, $\mathcal{P}_t^*$ is a continuous map from the space $\mathscr{P}(\mathcal{B})$ of Borel probability measures on $\mathcal{B}$ (endowed with the total variation topology) into itself. It follows from (5.13) and the definition of the dual that $\mathcal{P}_t^* \mu$ is nothing but the law at time t of the solution to (5.1) with its initial condition u_0 distributed according to μ, independently of the increments of W over $[0, t]$. With these notations in place, we define:

Definition 5.22 A Borel probability measure μ on $\mathcal{B}$ is an *invariant measure* for (5.1) if $\mathcal{P}_t^* \mu = \mu$ for every $t > 0$, where $\mathcal{P}_t$ is the Markov semigroup associated to solutions of (5.1) via (5.13).

In the case $\mathcal{B} = \mathcal{H}$ where we consider (5.1) on a Hilbert space $\mathcal{H}$, the situations in which such an invariant measure exists are characterised in the following theorem:

Theorem 5.23 *Consider (5.1) with solutions in a Hilbert space $\mathcal{H}$ and define the self-adjoint operator $Q_t \colon \mathcal{H} \to \mathcal{H}$ by*

$$Q_t = \int_0^t S(s) Q Q^* S^*(s)\, ds \; .$$

Then there exists an invariant measure μ for (5.1) if and only if one of the following two equivalent conditions are satisfied:

1. *There exists a positive definite trace class operator $Q_\infty \colon \mathcal{H} \to \mathcal{H}$ such that the identity $2\mathrm{Re}\langle Q_\infty L^* x, x\rangle + \|Q^* x\|^2 = 0$ holds for every $x \in \mathcal{D}(L^*)$.*
2. *One has $\sup_{t>0} \mathrm{tr}\, Q_t < \infty$.*

Furthermore, any invariant measure is of the form $\nu \star \mu_\infty$, where ν is a measure on $\mathcal{H}$ that is invariant under the action of the semigroup S and μ_∞ is the centred Gaussian measure with covariance Q_∞.

Proof. The proof goes as follows. We first show that μ being invariant implies that 2. holds. Then we show that 2. implies 1., and we conclude the first part by showing that 1. implies the existence of an invariant measure.

Let us start by showing that if μ is an invariant measure for (5.1), then 2. is satisfied. By choosing $\varphi(x) = e^{i\langle h, x\rangle}$ for arbitrary $h \in \mathcal{H}$, it follows from (5.13) that the Fourier transform of $\mathcal{P}_t^* \mu$ satisfies the equation

$$\widehat{\mathcal{P}_t^* \mu}(x) = \hat{\mu}(S^*(t)x) e^{-\frac{1}{2}\langle x, Q_t x\rangle} \; . \tag{5.14}$$

Taking logarithms and using the fact that $|\hat{\mu}(x)| \leq 1$ for every $x \in \mathcal{H}$ and every probability measure μ, It follows that if μ is invariant, then

$$\langle x, Q_t x\rangle \leq -2\log|\hat{\mu}(x)|\;, \quad \forall x \in \mathcal{H}\;, \quad \forall t > 0\;. \tag{5.15}$$

Choose now a sufficiently large value of $R > 0$ so that $\mu(\|x\| > R) < 1/8$ (say) and define a symmetric positive definite operator $A_R\colon \mathcal{H} \to \mathcal{H}$ by

$$\langle h, A_R h\rangle = \int_{\|x\|\leq R} |\langle x, h\rangle|^2\, \mu(dx)\;.$$

Since, for any orthonormal basis, one has $\|x\|^2 = \sum_n |\langle x, e_n\rangle|^2$, it follows that A_R is trace class and that $\operatorname{tr} A_R \leq R^2$. Furthermore, one has the bound

$$|1 - \hat{\mu}(h)| \leq \int_{\mathcal{H}} |1 - e^{i\langle h,x\rangle}|\, \mu(dx) \leq \sqrt{\langle h, A_R h\rangle} + \frac{1}{4}\;.$$

Combining this with (5.15), it follows that $\langle x, Q_t x\rangle$ is bounded by $2\log 4$ for every $x \in \mathcal{H}$ such that $\langle x, A_R x\rangle \leq 1/4$ so that, by homogeneity,

$$\langle x, Q_t x\rangle \leq (8\log 4)\langle x, A_R x\rangle\;.$$

It follows that $\operatorname{tr} Q_t \leq (8\log 4)R^2$, so that 2. is satisfied. To show that 2. implies 1., note that $\sup \operatorname{tr} Q_t < \infty$ implies that

$$Q_\infty = \int_0^\infty S(t) Q Q^* S^*(t)\, dt\;,$$

is a well-defined positive definite trace class operator (since $t \mapsto Q_t^{1/2}$ forms a Cauchy sequence in the space of Hilbert-Schmidt operators). Furthermore, one has the identity

$$\langle x, Q_\infty x\rangle = \langle S^*(t)x, Q_\infty S^*(t)x\rangle + \int_0^t \|Q^* S^*(s)x\|^2\, ds\;.$$

for $x \in \mathcal{D}(L^*)$, both terms on the right hand side of this expression are differentiable. Taking the derivative at $t = 0$, we get

$$0 = 2\mathrm{Re}\langle Q_\infty L^* x, x\rangle + \|Q^* x\|^2\;,$$

which is precisely the identity in 1.

Let now Q_∞ be a given operator as in 1., we want to show that the centred Gaussian measure μ_∞ with covariance Q_∞ is indeed invariant for $\mathcal{P}_t$. For $x \in \mathcal{D}(L^*)$, it follows from Proposition 4.10 that the map $F_x\colon t \mapsto \langle Q_\infty S^*(t)x, S^*(t)x\rangle$ is differentiable with derivative given by $\partial_t F_x(t) = 2\mathrm{Re}\langle Q_\infty L^* S^*(t)x, S^*(t)x\rangle$. It follows that

$$F_x(t) - F_x(0) = 2\int_0^t \mathrm{Re}\langle Q_\infty L^* S^*(s)x, S^*(s)x\rangle\, ds = -\int_0^t \|Q^* S^*(s)x\|^2\, ds\;,$$

so that one has the identity

$$Q_\infty = S(t) Q_\infty S^*(t) + \int_0^t S(s) Q Q^* S^*(s)\, ds = S(t) Q_\infty S^*(t) + Q_t\;.$$

Inserting this into (5.14), the claim follows. Here, we used the fact that $\mathcal{D}(L^*)$ is dense in $\mathcal{H}$, which is always the case on a Hilbert space, see [Yos95, p. 196].

Since it is obvious from (5.14) that every measure of the type $\nu \star \mu_\infty$ with ν invariant for S is also invariant for $\mathcal{P}_t$, it remains to show that the converse also holds. Let μ be invariant for $\mathcal{P}_t$ and define μ_t as the push-forward of μ under the map $S(t)$. Since $\hat{\mu}_t(x) = \hat{\mu}(S^*(t)x)$, it follows from (5.14) and the invariance of μ that there exists a function $\psi\colon \mathcal{H} \to \mathbf{R}$ such that $\hat{\mu}_t(x) \to \psi(x)$ uniformly on bounded sets, $\psi \circ S(t)^* = \psi$, and such that $\hat{\mu}(x) = \psi(x)\exp(-\frac{1}{2}\langle x, Q_\infty x\rangle)$. It therefore only remains to show that there exists a probability measure ν on $\mathcal{H}$ such that $\psi = \hat{\nu}$.

In order to show this, it suffices to show that the family of measures $\{\mu_t\}$ is tight, that is for every $\varepsilon > 0$ there exists a compact set K such that $\mu_t(K) \geq 1 - \varepsilon$ for every t. Prohorov's theorem [Bil68, p. 37] then ensures the existence of a sequence t_n increasing to ∞ and a measure ν such that $\mu_{t_n} \to \nu$ weakly. In particular, $\hat{\mu}_{t_n}(x) \to \hat{\nu}(x)$ for every $x \in \mathcal{H}$, thus concluding the proof.

To show tightness, denote by ν_t the centred Gaussian measure on $\mathcal{H}$ with covariance Q_t and note that one can find a sequence of bounded linear operators $A_n \colon \mathcal{H} \to \mathcal{H}$ with the following properties:

a. One has $\|A_{n+1}x\| \geq \|A_n x\|$ for every $x \in \mathcal{H}$ and every $n \geq 0$.

b. The set $B_R = \{x \,:\, \sup_n \|A_n x\| \leq R\}$ is compact for every $R > 0$.

c. One has $\sup_n \operatorname{tr} A_n Q_\infty A_n^* < \infty$.

(By diagonalising Q_∞, the construction of such a family of operators is similar to the construction, given a positive sequence $\{\lambda_n\}$ with $\sum_n \lambda_n < \infty$, of a positive sequence a_n with $\lim_{n\to\infty} a_n = +\infty$ and $\sum_n a_n \lambda_n < \infty$.) Let now $\varepsilon > 0$ be arbitrary. It follows from Prohorov's theorem that there exists a compact set $\hat{K} \subset \mathcal{H}$ such that $\mu(\mathcal{H} \setminus \hat{K}) \leq \frac{\varepsilon}{2}$. Furthermore, it follows from property *c.* above and the fact that $Q_\infty \geq Q_t$ that there exists $R > 0$ such that $\nu_t(\mathcal{H} \setminus B_R) \leq \frac{\varepsilon}{2}$. Define a set $K \subset \mathcal{H}$ by

$$K = \{z - y \,:\, z \in \hat{K}\,, y \in B_R\}\,.$$

It is straightforward to check, using the Heine-Borel theorem, that K is precompact.

If we now take X and Y to be independent $\mathcal{H}$-valued random variables with laws μ_t and ν_t respectively, then it follows from the definition of a mild solution and the invariance of μ that $Z = X + Y$ has law μ. Since one has the obvious implication $\{Z \in \hat{K}\} \& \{Y \in B_R\} \Rightarrow \{X \in K\}$, it follows that

$$\mu_t(\mathcal{H} \setminus K) = \mathbf{P}(X \notin K) \leq \mathbf{P}(Z \notin \hat{K}) + \mathbf{P}(Y \notin B_R) \leq \varepsilon\,,$$

thus showing that the sequence $\{\mu_t\}$ is tight as requested. □

It is clear from Theorem 5.23 that if (5.1) does have a solution in some Hilbert space $\mathcal{H}$ and if $\|S(t)\| \to 0$ as $t \to \infty$ in that same Hilbert space, then it also possesses a unique invariant measure on $\mathcal{H}$. It turns out that as far as the "uniqueness" part of this statement is concerned, it is sufficient to have $\lim_{t\to\infty} \|S(t)x\| = 0$ for every $x \in \mathcal{H}$:

Proposition 5.24 *If $\lim_{t\to\infty} \|S(t)x\| = 0$ for every $x \in \mathcal{H}$, then (5.1) can have at most one invariant measure. Furthermore, if an invariant measure μ_∞ exists in this situation, then one has $\mathcal{P}_t^*\nu \to \mu_\infty$ weakly for every probability measure ν on $\mathcal{H}$.*

Proof. In view of Theorem 5.23, the first claim follows if we show that δ_0 is the only measure that is invariant under the action of the semigroup S. Let ν be an arbitrary probability measure on $\mathcal{H}$ such that $S(t)^\sharp \nu = \nu$ for every $t > 0$ and let $\varphi \colon \mathcal{H} \to \mathbf{R}$ be a bounded continuous function. One then has indeed

$$\int_{\mathcal{H}} \varphi(x)\nu(dx) = \lim_{t\to\infty} \int_{\mathcal{H}} \varphi(S(t)x)\nu(dx) = \varphi(0)\,, \tag{5.16}$$

where we first used the invariance of ν and then the dominated convergence theorem.

To show that $\mathcal{P}_t^*\nu \to \mu_\infty$ whenever an invariant measure exists we use the fact that in this case, by Theorem 5.23, one has $Q_t \uparrow Q_\infty$ in the trace class topology. Denoting by μ_t the centred Gaussian measure with covariance Q_t, the fact that L^2 convergence implies weak convergence then implies that there exists a measure $\hat{\mu}_\infty$ such that $\mu_t \to \hat{\mu}_\infty$ weakly. Furthermore, the same reasoning as in (5.16) shows that $S(t)^\sharp \nu \to \delta_0$ weakly as $t \to \infty$. The claim then follows from the fact that $\mathcal{P}_t^*\nu = (S(t)^\sharp \nu) \star \mu_t$ and from the fact that convolving two probability measures is a continuous operation in the topology of weak convergence. □

Note that the condition $\lim_{t\to\infty} \|S(t)x\| = 0$ for every x is *not* sufficient in general to guarantee the existence of an invariant measure for (5.1). This can be seen again with the aid of Example 5.19. Take an increasing function V with $\lim_{x\to\infty} V(x) = \infty$, but such that $\int_0^\infty e^{-V(x)}\,dx = \infty$. Then, since

$\exp(V(x)-V(x+t)) \le 1$ and $\lim_{t\to\infty} \exp(V(x)-V(x+t)) = 0$ for every $x \in \mathbf{R}$, it follows from (5.9) and the dominated convergence theorem that $\lim_{t\to\infty} \|S(t)u\| = 0$ for every $u \in \mathcal{H}$. However, the fact that $\int_0^\infty e^{-V(x)}\,dx = \infty$ prevents the random process $\tilde{u}$ defined in (5.12) from belonging to $\mathcal{H}$, so that (5.10) has no invariant measure in this particular situation.

Exercise 5.25 Show that if (5.1) has an invariant measure μ_∞ but there exists $x \in \mathcal{H}$ such that $\limsup_{t\to\infty} \|S(t)x\| > 0$, then one cannot have $\mathcal{P}_t^*\delta_x \to \mu_\infty$ weakly. In this sense, the statement of Proposition 5.24 is sharp.

5.3 Convergence in stronger topologies

Proposition 5.24 shows that if (5.1) has an invariant measure μ_∞, one can in many cases expect to have $\mathcal{P}_t^*\nu \to \mu_\infty$ weakly for every initial measure ν. It is however not clear *a priori* whether such a convergence also holds in some stronger topologies on the space of probability measures. If we consider the finite-dimensional case (that is $\mathcal{H} = \mathbf{R}^n$ for some $n > 0$), the situation is clear: the condition $\lim_{t\to\infty} \|S(t)x\| = 0$ for every $x \in \mathcal{H}$ then implies that $\lim_{t\to\infty} \|S(t)\| = 0$, so that L has to be a matrix whose eigenvalues all have strictly negative real parts. One then has:

Proposition 5.26 *In the finite-dimensional case, assume that all eigenvalues of L have strictly negative real parts and that Q_∞ has full rank. Then, there exists $T > 0$ such that $\mathcal{P}_t^*\delta_x$ has a smooth density $p_{t,x}$ with respect to Lebesgue measure for every $t > T$. Furthermore, μ_∞ has a smooth density p_∞ with respect to Lebesgue measure and there exists $c > 0$ such that, for every $\lambda > 0$, one has*

$$\lim_{t\to\infty} e^{ct} \sup_{y\in\mathbf{R}^n} e^{\lambda|y|} |p_\infty(y) - p_{t,x}(y)| = 0 \,.$$

In other words, $p_{t,x}$ converges to p_∞ exponentially fast in any weighted norm with exponentially increasing weight.

The proof of Proposition 5.26 is left as an exercise. It follows in a straightforward way from the explicit expression for the density of a Gaussian measure.

In the infinite-dimensional case, the situation is much less straightforward. The reason is that there exists no natural reference measure (the equivalent of the Lebesgue measure) with respect to which one could form densities.

In particular, even though one always has $\|\mu_t - \mu_\infty\|_{\mathrm{TV}} \to 0$ in the finite-dimensional case (provided that μ_∞ exists and that all eigenvalues of L have strictly negative real part), one cannot expect this to be true in general. Consider for example the SPDE

$$dx = -x\,dt + Q\,dW(t) \,, \quad x(t) \in \mathcal{H} \,,$$

where W is a cylindrical process on $\mathcal{H}$ and $Q\colon \mathcal{H} \to \mathcal{H}$ is a Hilbert-Schmidt operator. One then has

$$Q_t = \frac{1-e^{-2t}}{2} QQ^* \,, \qquad Q_\infty = \frac{1}{2} QQ^* \,.$$

Combining this with Proposition 3.65 (dilates of an infinite-dimensional Gaussian measure are mutually singular) shows that if QQ^* has infinitely many non-zero eigenvalues, then μ_t and μ_∞ are mutually singular in this case.

One question that one may ask oneself is under which conditions the convergence $\mathcal{P}_t^\nu \to \mu_\infty$ takes place in the total variation norm. There is a straightforward interpretation to the total variation convergence $\mathcal{P}_t^\nu \to \mu_\infty$: for large times, a sample drawn from the invariant distribution is with high probability indistinguishable from a sample drawn from the Markov process at time t. Compare this with the notion of weak convergence which relies on the topology of the underlying space and only asserts that the two samples are close with high probability in the sense determined by the topology in question. For example, $\|\delta_x - \delta_y\|$ is always equal to 2 if $x \neq y$, whereas $\delta_x \to \delta_y$ weakly if $x \to y$.

Exercise 5.27 Show that the two definitions of the total variation distance given above are indeed equivalent by constructing a coupling that realises the infimum in (3.3). It is useful for this to consider the measure $\mu \wedge \nu$ which, in μ and ν have densities $\mathcal{D}_\mu$ and $\mathcal{D}_\nu$ with respect to some common reference measure π, is given by $(\mathcal{D}_\mu(x) \wedge \mathcal{D}_\nu(x))\pi(dx)$.

An alternative characterisation of the total variation norm is as the dual norm to the supremum norm on the space $\mathscr{B}_b(\mathcal{B})$ of bounded Borel measurable functions on $\mathcal{B}$:

$$\|\mu - \nu\|_{\text{TV}} = \sup\Big\{ \int \varphi(x)\mu(dx) - \int \varphi(x)\nu(dx) \;:\; \sup_{x\in\mathcal{B}} |\varphi(x)| \le 1 \Big\} .$$

It turns out that, instead of showing directly that $\mathcal{P}_t^*\nu \to \mu_\infty$ in the total variation norm, it is somewhat easier to show that one has $\mathcal{P}_t^*\nu \to \mu_\infty$ in a type of "weighted total variation norm", which is slightly stronger than the usual total variation norm. Given a weight function $V\colon \mathcal{B} \to \mathbf{R}_+$, we define a weighted supremum norm on measurable functions by

$$\|\varphi\|_V = \sup_{x\in\mathcal{B}} \frac{|\varphi(x)|}{1 + V(x)} ,$$

as well as the dual norm on measures by

$$\|\mu - \nu\|_{\text{TV},V} = \sup\Big\{ \int \varphi(x)\mu(dx) - \int \varphi(x)\nu(dx) \;:\; \|\varphi\|_V \le 1 \Big\} . \tag{5.17}$$

Since we assumed that $V > 0$, it is obvious that one has the relation $\|\mu - \nu\|_{\text{TV}} \le \|\mu - \nu\|_{\text{TV},V}$, so that convergence in the weighted norm immediately implies convergence in the usual total variation norm. By considering the Jordan decomposition of $\mu - \nu = \varrho_+ - \varrho_-$, it is clear that the supremum in (5.17) is attained at functions φ such that $\varphi(x) = 1 + V(x)$ for ϱ_+-almost every x and $\varphi(x) = -1 - V(x)$ for ϱ_--almost every x. In other words, an alternative expression for the weighted total variation norm is given by

$$\|\mu - \nu\|_{\text{TV},V} = \int_{\mathcal{X}} (1 + V(x))\, |\mu - \nu|(dx) , \tag{5.18}$$

just like the total variation norm is given by $\|\mu - \nu\|_{\text{TV}} = |\mu - \nu|(\mathcal{X})$.

The reason why it turns out to be easier to work in a weighted norm is the following: For a suitable choice of V, we are going to see that in a large class of examples, one can construct a weight function V and find constants $c < 1$ and $T > 0$ such that

$$\|\mathcal{P}_T^*\mu - \mathcal{P}_T^*\nu\|_{\text{TV},V} \le c\|\mu - \nu\|_{\text{TV},V} , \tag{5.19}$$

for any two probability measures μ and ν. This implies that the map $\mathcal{P}_T$ is a contraction on the space of probability measures, which must therefore have exactly one fixed point, yielding both the existence of an invariant measure μ_∞ and the exponential convergence of $\mathcal{P}_t^*\nu$ to μ_∞ for every initial probability measure ν which integrates V.

This argument is based on the following abstract result that can be applied to arbitrary Markov semigroups on Polish (that is separable, complete, metric) spaces:

Theorem 5.28 (Harris) *Let $\mathcal{P}_t$ be a Markov semigroup on a Polish space $\mathcal{X}$ such that there exists a time $T_0 > 0$ and a function $V\colon \mathcal{X} \to \mathbf{R}_+$ such that:*

- *The exist constants $\gamma < 1$ and $K > 0$ such that $\mathcal{P}_{T_0}V(x) \le \gamma V(x) + K$ for every $x \in \mathcal{X}$.*
- *There exist $R > 2K/(1-\gamma)$ and $\delta > 0$ such that $\|\mathcal{P}_{T_0}^*\delta_x - \mathcal{P}_{T_0}^*\delta_y\|_{\text{TV}} \le 2(1-\delta)$ for every pair x, y such that $V(x) + V(y) \le R$.*

Then, there exists $T > 0$ such that (5.19) holds for some $c < 1$.

In a nutshell, the argument for the proof of Theorem 5.28 is the following. There are two mechanisms that allow to decrease the weighted total variation distance between two probability measures:

1. The mass of the two measures moves into regions where the weight $V(x)$ becomes smaller.
2. The two measures "spread out" in such a way that there is an increase in the overlap between them.

The two conditions of Theorem 5.28 are tailored such as to combine these two effects in order to obtain an exponential convergence of $\mathcal{P}_t^*\mu$ to the unique invariant measure for $\mathcal{P}_t$ as $t \to \infty$.

Remark 5.29 The condition that there exists $\delta > 0$ such that $\|\mathcal{P}_{T_0}^*\delta_x - \mathcal{P}_{T_0}^*\delta_y\|_{\mathrm{TV}} \leq 2(1-\delta)$ for any $x, y \in A$ is sometimes referred to in the literature as the set A being a *small set*.

Remark 5.30 Traditional proofs of Theorem 5.28 as given for example in [MT93] tend to make use of coupling arguments and estimates of return times of the Markov process described by $\mathcal{P}_t$ to level sets of V. The basic idea is to make use of (3.3) to get a bound on the total variation between $\mathcal{P}_T^*\mu$ and $\mathcal{P}_T^*\nu$ by constructing an explicit coupling between two instances x_t and y_t of a Markov process with transition semigroup $\{\mathcal{P}_t\}$. Because of the second assumption in Theorem 5.28, one can construct this coupling in such a way that every time the process (x_t, y_t) returns to some sufficiently large level set of V, there is a probability δ that $x_{t'} = y_{t'}$ for $t' \geq t + T_0$. The first assumption then guarantees that these return times have exponential tails and a renewal-type argument allows to conclude.

Such proofs are quite involved at a technical level and are by consequence not so easy to follow, especially if one wishes to get a spectral gap bound like (5.19) and not "just" an exponential decay bound like

$$\|\mathcal{P}_T^*\delta_x - \mathcal{P}_T^*\delta_y\|_{\mathrm{TV}} \leq Ce^{-\gamma T} ,$$

with a constant C depending on x and y. Furthermore, they require more background in advanced probability theory (in particular renewal theory) than what is assumed for the scope of these notes.

The elementary proof given here is taken from [HM08b] and is based on the arguments first exposed in [HM08a]. It has the disadvantage of being less intuitively appealing than proofs based on coupling arguments, but this is more than offset by the advantage of fitting into less than two pages without having to appeal to advanced mathematical concepts. It also has the advantage of being generalisable to situations where level sets of the Lyapunov function are not small sets, see [HMS09].

Before we turn to the proof of Theorem 5.28, we define for every $\beta > 0$ the distance function

$$d_\beta(x,y) = \begin{cases} 0 & \text{if } x = y \\ 2 + \beta V(x) + \beta V(y) & \text{if } x \neq y. \end{cases}$$

One can check that the positivity of V implies that this is indeed a distance function, albeit a rather strange one. We define the corresponding "Lipschitz" seminorm on functions $\varphi\colon \mathcal{X} \to \mathbf{R}$ by

$$\|\varphi\|_{\mathrm{Lip}_\beta} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_\beta(x,y)} .$$

We are going to make use of the following lemma:

Lemma 5.31 *With the above notations, one has* $\|\varphi\|_{\mathrm{Lip}_\beta} = \inf_{c \in \mathbf{R}} \|\varphi + c\|_{\beta V}$.

Proof. It is obvious that $\|\varphi\|_{\mathrm{Lip}_\beta} \leq \|\varphi + c\|_{\beta V}$ for every $c \in \mathbf{R}$. On the other hand, if x_0 is any fixed point in $\mathcal{X}$, one has

$$|\varphi(x)| \leq |\varphi(x_0)| + \|\varphi\|_{\mathrm{Lip}_\beta}(2 + \beta V(x) + \beta V(x_0)) , \tag{5.20}$$

for all $x \in \mathcal{X}$. Set now

$$c = -\sup_{x \in \mathcal{X}} (\varphi(x) - \|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(x))) .$$

It follows from (5.20) that c is finite. Furthermore, one has

$$\varphi(y) + c \le \varphi(y) - (\varphi(y) - \|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(y))) = \|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(y)) \,,$$

and

$$\begin{aligned}\varphi(y) + c &= \inf_{x\in\mathcal{X}} (\varphi(y) - \varphi(x) + \|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(x))) \\ &\ge \inf_{x\in\mathcal{X}} \|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(x) - d_\beta(x,y)) = -\|\varphi\|_{\mathrm{Lip}_\beta}(1 + \beta V(y)) \,,\end{aligned}$$

which implies that $\|\varphi + c\|_{\beta V} \le \|\varphi\|_{\mathrm{Lip}_\beta}$. □

With this result in hand, we are ready to tackle the proof of Harris's theorem:

Proof of Theorem 5.28. During this proof, we use the notation $\mathcal{P} \stackrel{\text{def}}{=} \mathcal{P}_{T_0}$ for simplicity. We are going to show that there exists a choice of $\beta \in (0,1)$ such that there is $\alpha < 1$ satisfying the bound

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \le \alpha d_\beta(x,y)\|\varphi\|_{\mathrm{Lip}_\beta} \,, \tag{5.21}$$

uniformly over all measurable functions $\varphi\colon \mathcal{X} \to \mathbf{R}$ and all pairs $x, y \in \mathcal{X}$. Note that this is equivalent to the bound $\|\mathcal{P}\varphi\|_{\mathrm{Lip}_\beta} \le \alpha\|\varphi\|_{\mathrm{Lip}_\beta}$. Combining this with Lemma 5.31 and (5.18), we obtain that, for $T = nT_0$, one has the bound

$$\begin{aligned}\|\mathcal{P}_T^*\mu - \mathcal{P}_T^*\nu\|_{\mathrm{TV},V} &= \inf_{\|\varphi\|_V\le 1} \int_{\mathcal{X}} (\mathcal{P}_T\varphi)(x)\,(\mu-\nu)(dx) \\ &= \inf_{\|\varphi\|_V\le 1} \inf_{c\in\mathbf{R}} \int_{\mathcal{X}} ((\mathcal{P}_T\varphi)(x) + c)\,(\mu-\nu)(dx) \\ &\le \inf_{\|\varphi\|_V\le 1} \inf_{c\in\mathbf{R}} \|\mathcal{P}_T\varphi + c\|_V \int_{\mathcal{X}} (1 + V(x))\,|\mu-\nu|(dx) \\ &\le \inf_{\|\varphi\|_V\le 1} \inf_{c\in\mathbf{R}} \|\mathcal{P}_T\varphi + c\|_{\beta V}\|\mu-\nu\|_{\mathrm{TV},V} \\ &= \inf_{\|\varphi\|_V\le 1} \|\mathcal{P}_T\varphi\|_{\mathrm{Lip}_\beta}\|\mu-\nu\|_{\mathrm{TV},V} \\ &\le \alpha^n \inf_{\|\varphi\|_V\le 1} \|\varphi\|_{\mathrm{Lip}_\beta}\|\mu-\nu\|_{\mathrm{TV},V} \le \frac{\alpha^n}{\beta}\|\mu-\nu\|_{\mathrm{TV},V} \,.\end{aligned}$$

Since $\alpha < 1$, the result (5.19) then follows at once by choosing n sufficiently large.

Let us turn now to (5.21). If $x = y$, there is nothing to prove, so we assume that $x \ne y$. Fix an arbitrary non-constant function φ and assume without loss of generality that $\|\varphi\|_{\mathrm{Lip}_\beta} = 1$. It follows from Lemma 5.31 that, by adding a constant to it if necessary, we can assume that $|\varphi(x)+c| \le (1+\beta V(x))$.

This immediately implies the bound

$$\begin{aligned}|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| &\le (2 + \beta\mathcal{P}V(x) + \beta\mathcal{P}V(y)) \\ &\le 2 + 2\beta K + \beta\gamma V(x) + \beta\gamma V(y) \,.\end{aligned} \tag{5.22}$$

Suppose now that x and y are such that $V(x) + V(y) \ge R$ and note that (5.23) then immediately implies that, for every $\varepsilon > 0$, we have

$$|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \le 2(1-\varepsilon) + \beta(V(x) + V(y))\Big(\gamma + \frac{2\beta K + 2\varepsilon}{\beta R}\Big) \,. \tag{5.23}$$

At this stage, we see that if $\varepsilon = 0$, then due to the assumption $R < 2K/(1-\gamma)$ the constant multiplying $\beta(V(x) + V(y))$ is strictly less than 1. This shows that, for every $\beta > 0$, there exists $\alpha > 0$ such that (5.21) holds for $V(x) + V(y) \ge R$. Actually, the optimal choice is given by

$$\alpha = 1 - \beta\frac{R(1-\gamma) - 2K}{2 + \beta R} \,,$$

which is positive by the assumption on R.

Take now a pair x, y such that $V(x) + V(y) \leq R$. Note that we can write $\varphi = \varphi_1 + \varphi_2$ with $|\varphi_1(x)| \leq 1$ and $|\varphi_2(x)| \leq \beta V(x)$. (Set $\varphi_1(x) = (\varphi(x) \wedge 1) \vee (-1)$.) It follows from the assumptions on V that there exists some $\delta > 0$ such that

$$\begin{aligned}|\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| &\leq |\mathcal{P}\varphi_1(x) - \mathcal{P}\varphi_1(y)| + |\mathcal{P}\varphi_2(x) - \mathcal{P}\varphi_2(y)| \\ &\leq \|\mathcal{P}^*\delta_x - \mathcal{P}^*\delta_y\|_{\text{TV}} + \beta(\mathcal{P}V)(x) + \beta(\mathcal{P}V)(y) \\ &\leq 2(1-\delta) + \beta(2K + \gamma V(x) + \gamma V(y)) \\ &\leq ((1 - \delta + \beta K) \vee \gamma) d_\beta(x, y) \; .\end{aligned}$$

If we now choose $\beta = \frac{\delta}{2K}$, say, (5.21) holds with $\alpha = (1 - \frac{1}{2}\delta) \vee \gamma < 1$. Combining this estimate with the one obtained previously shows that one can indeed find α and β such that (5.21) holds for all x and y in $\mathcal{X}$, thus concluding the proof of Theorem 5.28. □

Remark 5.32 The typical situation in practice is that γ and K are of order 1, but δ is very small. In this case, the proof shows that the spectral gap $1 - \alpha$ for the operator $\mathcal{P}$ is of $\mathcal{O}(\delta)$. Furthermore, one also needs to take $\beta = \mathcal{O}(\delta)$, so that one typically obtains bounds of the type

$$\|\mathcal{P}_t^*\mu - \mathcal{P}_t^*\nu\|_{\text{TV},V} \leq \frac{Ce^{-c\delta t}}{\delta}\|\mu - \nu\|_{\text{TV},V} \; ,$$

for $\delta \ll 1$.

One could argue that while it guarantees its uniqueness, this theorem does not guarantee the existence of an invariant measure since the fact that $\mathcal{P}_T^*\mu = \mu$ for some fixed T does not imply that $\mathcal{P}_t^*\mu = \mu$ for every t. However, one has:

Lemma 5.33 *If there exists a probability measure such that $\mathcal{P}_T^*\mu = \mu$ for some fixed time $T > 0$, then there also exists a probability measure μ_∞ such that $\mathcal{P}_t^*\mu_\infty = \mu_\infty$ for all $t > 0$.*

Proof. Define the measure μ_∞ by

$$\mu_\infty(A) = \frac{1}{T}\int_0^T \mathcal{P}_t^*\mu(A)\, dt \; .$$

It is then a straightforward exercise to check that it does have the requested property. □

We are now able to use this theorem to obtain the following result on the convergence of the solutions to (5.1) to an invariant measure in the total variation topology:

Theorem 5.34 *Assume that (5.1) has a solution in some Hilbert space $\mathcal{H}$ and that there exists a time T such that $\|S(T)\| < 1$ and such that $S(T)$ maps $\mathcal{H}$ into the image of $Q_T^{1/2}$. Then (5.1) admits a unique invariant measure μ_∞ and there exists $\gamma > 0$ such that*

$$\|\mathcal{P}_t^*\nu - \mu_\infty\|_{\text{TV}} \leq C(\nu)e^{-\gamma t} \; ,$$

for every probability measure ν on $\mathcal{H}$ with finite second moment.

Proof. Let $V(x) = \|x\|$ and denote by μ_t the centred Gaussian measure with covariance Q_t. We then have

$$\mathcal{P}_t V(x) \leq \|S(t)x\| + \int_{\mathcal{H}} \|x\|\, \mu_t(dx) \; ,$$

which shows that the first assumption of Theorem 5.28 is satisfied. A simple variation of Exercise 3.59 (use the decomposition $\mathcal{H} = \tilde{\mathcal{H}} \oplus \ker K$) shows that the Cameron-Martin space of μ_T is given by $\operatorname{Im} Q_T^{1/2}$ endowed with the norm

$$\|h\|_T = \inf\{\|x\| \, : \, h = Q_T^{1/2}x\} \; .$$

Since we assumed that $S(T)$ maps $\mathcal{H}$ into the image of $Q_T^{1/2}$, it follows from the closed graph theorem that there exists a constant C such that $\|S(T)x\|_T \le C\|x\|$ for every $x \in \mathcal{H}$. It follows from the Cameron-Martin formula that the total variation distance between $\mathcal{P}_T^*\delta_x$ and $\mathcal{P}_T^*\delta_y$ is equal to the total variation distance between $\mathcal{N}(0,1)$ and $\mathcal{N}(\|S(T)x - S(T)y\|_T, 1)$, so that the second assumption of Theorem 5.28 is also satisfied.

Both the existence of μ_∞ and the exponential convergence of $\mathcal{P}_t^*\nu$ towards it is then a consequence of Banach's fixed point theorem in the dual to the space of measurable functions with $\|\varphi\|_V < \infty$. □

Remark 5.35 The proof of Theorem 5.34 shows that if its assumptions are satisfied, then the map $x \mapsto \mathcal{P}_t^*\delta_x$ is continuous in the total variation distance for $t \ge T$.

Remark 5.36 Since $\operatorname{Im} S(t)$ decreases with t and $\operatorname{Im} Q_t^{1/2}$ increases with t, it follows that if $\operatorname{Im} S(t) \subset \operatorname{Im} Q_t^{1/2}$ for some t, then this also holds for any subsequent time. This is consistent with the fact that Markov operators are contraction operators in the supremum norm, so that if $x \mapsto \mathcal{P}_t^*\delta_x$ is continuous in the total variation distance for some t, the same must be true for all subsequent times.

While Theorem 5.34 is very general, it is sometimes not straightforward at all to verify its assumptions for arbitrary linear SPDEs. In particular, it might be complicated *a priori* to determine the image of $Q_t^{1/2}$. The task of identifying this subspace can be made easier by the following result:

Proposition 5.37 *The image of $Q_t^{1/2}$ is equal to the image of the map A_t given by*

$$A_t\colon L^2([0,t],\mathcal{K}) \to \mathcal{H}\,, \qquad A_t\colon h \mapsto \int_0^t S(s)Qh(s)\,ds\,.$$

Proof. Since $Q_t = A_tA_t^*$, we can use polar decomposition [RS80, Thm VI.10] to find an isometry J_t of $(\ker A_t)^\perp \subset \mathcal{H}$ (which extends to $\mathcal{H}$ by setting $J_tx = 0$ for $x \in \ker A_t$) such that $Q_t^{1/2} = A_tJ_t$. Alternatively, it follows from Proposition 3.71 that the Cameron-Martin space of the measure $\mathcal{P}_t^*\delta_0$ is given by the image of A_t. Since on the other hand this Cameron-Martin space is equal to the image of $Q_t^{1/2}$ by Exercise 3.59, the claim follows. □

One case where it is straightforward to check whether $S(t)$ maps $\mathcal{H}$ into the image of $Q_t^{1/2}$ is the following:

Example 5.38 Consider the case where $\mathcal{K} = \mathcal{H}$, L is selfadjoint, and there exists a function $f\colon \mathbf{R} \to \mathbf{R}_+$ such that $Q = f(L)$. (This identity should be interpreted in the sense of the functional calculus already mentioned in Theorem 4.21.)

If we assume furthermore that $f(\lambda) > 0$ for every $\lambda \in \mathbf{R}$, then the existence of an invariant measure is equivalent to the existence of a constant $c > 0$ such that $\langle x, Lx\rangle \le -c\|x\|^2$ for every $x \in \mathcal{H}$. Using functional calculus, we see that the operator Q_T is then given by

$$Q_T = \frac{f^2(L)}{2L}(1 - e^{-2LT})\,,$$

and, for every $T > 0$, the Cameron-Martin norm for μ_T is equivalent to the norm

$$\|x\|_f = \Big\|\frac{\sqrt{L}}{f(L)}x\Big\|\,.$$

In order to obtain convergence $\mathcal{P}_t^*\nu \to \mu_\infty$ in the total variation topology, it is therefore sufficient that there exist constants $c, C > 0$ such that $f(\lambda) \ge Ce^{-c\lambda}$ for $\lambda \ge 0$.

This shows that one cannot expect convergence in the total variation topology to take place under similarly weak conditions as in Proposition 5.24. In particular, convergence in the total variation topology requires some non-degeneracy of the driving noise which was not the case for weak convergence.

Exercise 5.39 Consider again the case $\mathcal{K} = \mathcal{H}$ and L selfadjoint with $\langle x, Lx\rangle \leq -c\|x\|^2$ for some $c > 0$. Assume furthermore that Q is selfadjoint and that Q and L commute, so that there exists a space $L^2(\mathcal{M}, \mu)$ isometric to $\mathcal{H}$ and such that both Q and L can be realised as multiplication operators (say f and g respectively) on that space. Show that:

- In order for there to exist solutions in $\mathcal{H}$, the set $A_Q \stackrel{\text{def}}{=} \{\lambda \in \mathcal{M} : f(\lambda) \neq 0\}$ must be "essentially countable" in the sense that it can be written as the union of a countable set and a set of μ-measure 0.
- If there exists $T > 0$ such that $\operatorname{Im} S(T) \subset \operatorname{Im} Q_T^{1/2}$, then μ is purely atomic and there exists some possibly different time $t > 0$ such that $S(t)$ is trace class.

Exercise 5.38 suggests that there are many cases where, if $S(t)$ maps $\mathcal{H}$ to $\operatorname{Im} Q_t^{1/2}$ for some $t > 0$, then it does so for all $t > 0$. It also shows that, in the case where L and Q are selfadjoint and commute, Q must have an orthnormal basis of eigenvectors with all eigenvalues non-zero. Both statements are certainly not true in general. We see from the following example that there can be infinite-dimensional situations where $S(t)$ maps $\mathcal{H}$ to $\operatorname{Im} Q_t^{1/2}$ even though Q is of rank one!

Example 5.40 Consider the space $\mathcal{H} = \mathbf{R} \oplus L^2([0,1], \mathbf{R})$ and denote elements of $\mathcal{H}$ by (a, u) with $a \in \mathbf{R}$. Consider the semigroup S on $\mathcal{H}$ given by

$$S(t)(a,u) = (a, \tilde{u}) \,, \quad \tilde{u}(x) = \begin{cases} a & \text{for } x \leq t \\ u(x-t) & \text{for } x > t. \end{cases}$$

It is easy to check that S is indeed a strongly continuous semigroup on $\mathcal{H}$ and we denote its generator by $(0, \partial_x)$. We drive this equation by adding noise only on the first component of $\mathcal{H}$. In other words, we set $\mathcal{K} = \mathbf{R}$ and $Q1 = (1, 0)$ so that, formally, we are considering the equation

$$da = dW(t) \,, \qquad du = \partial_x u \, dt \,.$$

Even though, at a formal level, the equations for a and u look decoupled, they are actually coupled via the domain of the generator of S. In order to check whether $S(t)$ maps $\mathcal{H}$ into $\mathcal{H}_t \stackrel{\text{def}}{=} \operatorname{Im} Q_t^{1/2}$, we make use of Proposition 5.37. This shows that $\mathcal{H}_t$ consists of elements of the form

$$\int_0^t h(s) \chi_s \, ds \,,$$

where $h \in L^2([0,t])$ and χ_s is the image of $(1,0)$ under $S(s)$, which is given by $(1, \mathbf{1}_{[0, s\wedge 1]})$. On the other hand, the image of $S(t)$ consists of all elements (a, u) such that $u(x) = a$ for $x \leq t$. Since one has $\chi_s(x) = 0$ for $x > s$, it is obvious that $\operatorname{Im} S(t) \not\subset \mathcal{H}_t$ for $t < 1$.

On the other hand, for $t > 1$, given any $a > 0$, we can find a function $h \in L^2([0,t])$ such that $h(x) = 0$ for $x \leq 1$ and $\int_0^t h(x)\, dx = a$. Since, for $s \geq 1$, one has $\chi_s(x) = 1$ for every $x \in [0,1]$, it follows that one does have $\operatorname{Im} S(t) \subset \mathcal{H}_t$ for $t > 1$.

6

Semilinear SPDEs

Now that we have a good working knowledge of the behaviour of solutions to linear stochastic PDEs, we are prepared to turn to nonlinear SPDEs. In these notes, we will restrict ourselves to the study of *semilinear* SPDEs with *additive* noise.

In this context, a *semilinear* SPDE is one such that the nonlinearity can be treated as a perturbation of the linear part of the equation. The word *additive* for the noise refers to the fact that, as in (5.1), we will only consider noises described by a fixed operator $Q\colon \mathcal{K} \to \mathcal{B}$, rather than by an operator-valued function of the solution. We will therefore consider equations of the type

$$dx = Lx\,dt + F(x)\,dt + Q\,dW(t)\;, \quad x(0) = x_0 \in \mathcal{B}\;, \tag{6.1}$$

where L is the generator of a strongly continuous semigroup S on a separable Banach space $\mathcal{B}$, W is a cylindrical Wiener process on some separable Hilbert space $\mathcal{K}$, and $Q\colon \mathcal{K} \to \mathcal{B}$ is bounded. Furthermore, F is a measurable function from some linear subspace $\mathcal{D}(F) \subset \mathcal{B}$ into $\mathcal{B}$. We will say that a process $t \mapsto x(t) \in \mathcal{D}(F)$ is a *mild solution* to (6.1) if the identity

$$x(t) = S(t)x_0 + \int_0^t S(t-s)F(x(s))\,ds + \int_0^t S(t-s)Q\,dW(s)\;, \tag{6.2}$$

holds almost surely for every $t > 0$.

6.1 Local solutions

Throughout this section, we will make the standing assumption that the linear version of (6.1) (that is the corresponding equation with $F = 0$) does have a continuous solution with values in $\mathcal{B}$. In order to simplify notations, we borrow the notation from [DPZ92b] and write

$$W_L(t) \stackrel{\text{def}}{=} \int_0^t S(t-s)Q\,dW(s)\;.$$

In the nonlinear case, there are situations where solutions explode after a finite, but possibly random, time interval. (This statement is of course not specific to SPDEs, just think of the ordinary differential equation $\dot{x} = x^2$ on $\mathbf{R}$.) In order to be able to account for such a situation, we introduce the notion of a local solution. Recall first that, given a cylindrical Wiener process W defined on some probability space $(\Omega, \mathbf{P})$, we can associate to it the natural filtration $\{\mathscr{F}_t\}_{t\geq 0}$ generated by the increments of W. In other words, for every $t > 0$, $\mathscr{F}_t$ is the smallest σ-algebra such that the random variables $W(s) - W(r)$ for $s, r \leq t$ are all $\mathscr{F}_t$-measurable.

In this context, a *stopping time* is a positive random variable τ such that the event $\{\tau \leq T\}$ is $\mathscr{F}_T$-measurable for every $T \geq 0$. With this definition at hand, we have:

Definition 6.1 A *local mild solution* to (6.1) is a $\mathcal{D}(F)$-valued stochastic process x together with a stopping time τ such that $\tau > 0$ almost surely and such that the identity

$$x(t) = S(t)x_0 + \int_0^t S(t-s)F(x(s))\,ds + W_L(t)\;, \tag{6.3}$$

holds almost surely for every stopping time t such that $t < \tau$ almost surely.

Remark 6.2 In some situations, it might be of advantage to allow F to be a map from $\mathcal{D}(F)$ to $\mathcal{B}'$ for some superspace $\mathcal{B}'$ such that $\mathcal{B} \subset \mathcal{B}'$ densely and such that $S(t)$ extends to a continuous linear map from $\mathcal{B}'$ to $\mathcal{B}$. The prime example of such a space $\mathcal{B}'$ is an interpolation space with negative index in the case where the semigroup S is analytic. The definition of a mild solution carries over to this situation without any change.

A local mild solution (x, τ) is called *maximal* if, for every mild solution $(\tilde{x}, \tilde{\tau})$, one has $\tilde{\tau} \leq \tau$ almost surely.

Exercise 6.3 Show that local mild solutions to (6.1) coincide with local mild solutions to (6.1) with L replaced by $\tilde{L} = L - c$ and F replaced by $\tilde{F} = F + c$ for any constant $c \in \mathbf{R}$.

Our first result on the existence and uniqueness of mild solutions to nonlinear SPDEs makes the rather restrictive assumption that the nonlinearity F is defined on the whole space $\mathcal{B}$ and that it is locally Lipschitz there:

Theorem 6.4 *Consider (6.1) on a Banach space $\mathcal{B}$ and assume that W_L is a continuous $\mathcal{B}$-valued process. Assume furthermore that $F\colon \mathcal{B} \to \mathcal{B}$ is such that its restriction to every bounded set is Lipschitz continuous. Then, there exists a unique maximal mild solution (x, τ) to (6.1). Furthermore, this solution has continuous sample paths and one has $\lim_{t\uparrow\tau} \|x(t)\| = \infty$ almost surely on the set $\{\tau < \infty\}$.*

If F is globally Lipschitz continuous, then $\tau = \infty$ almost surely.

Proof. Given any realisation $W_L \in \mathcal{C}(\mathbf{R}_+, \mathcal{B})$ of the stochastic convolution, we are going to show that there exists a time $\tau > 0$ depending only on W_L up to time τ and a unique continuous function $x\colon [0, \tau) \to \mathcal{B}$ such that (6.3) holds for every $t < \tau$. Furthermore, the construction will be such that either $\tau = \infty$, or one has $\lim_{t\uparrow\tau} \|x(t)\| = \infty$, thus showing that (x, τ) is maximal.

The proof relies on the Banach fixed point theorem. Given a terminal time $T > 0$ and a continuous function $g\colon \mathbf{R}_+ \to \mathcal{B}$, we define the map $M_{g,T}\colon \mathcal{C}([0, T], \mathcal{B}) \to \mathcal{C}([0, T], \mathcal{B})$ by

$$(M_{g,T}u)(t) = \int_0^t S(t - s)F(u(s))\, ds + g(t)\ . \tag{6.4}$$

The proof then works in almost exactly the same way as the usual proof of uniqueness of a maximal solution for ordinary differential equations with Lipschitz coefficients. Note that we can assume without loss of generality that the semigroup S is bounded, since we can always subtract a constant to L and add it back to F. Using the fact that $\|S(t)\| \leq M$ for some constant M, one has for any $T > 0$ the bound

$$\sup_{t\in[0,T]} \|M_{g,T}u(t) - M_{g,T}v(t)\| \leq MT \sup_{t\in[0,T]} \|F(u(t)) - F(v(t))\|\ . \tag{6.5}$$

Furthermore, one has

$$\sup_{t\in[0,T]} \|M_{g,T}u(t) - g(t)\| \leq MT \sup_{t\in[0,T]} \|F(u(t))\|\ . \tag{6.6}$$

Fix now an arbitrary constant $R > 0$. Since F is locally Lipschitz, it follows from (6.5) and (6.6) that there exists a maximal $T > 0$ such that $M_{g,T}$ maps the ball of radius R around g in $\mathcal{C}([0, T], \mathcal{B})$ into itself and is a contraction with contraction constant $\frac{1}{2}$ there. This shows that $M_{g,T}$ has a unique fixed point for T small enough and the choice of T was obviously performed by using knowledge of g only up to time T. Setting $g(t) = S(t)x_0 + W_L(t)$, the pair (x, T), where T is as just constructed and x is the unique fixed point of $M_{g,T}$ thus yields a local mild solution to (6.1).

In order to construct the maximal solution, we iterate this construction in the same way as in the finite-dimensional case. Uniqueness and continuity in time also follows as in the finite-dimensional case. In the case where F is globally Lipschitz continuous, denote its Lipschitz constant by K. We then see from (6.5) that $M_{g,T}$ is a contraction on the whole space for $T < 1/(KM)$, so that the choice of

T can be made independently of the initial condition, thus showing that the solution exists for all times. □

While this setting is very straightforward and did not make use of any PDE theory, it nevertheless allows to construct solutions for an important class of examples, since every composition operator of the form $(N(u))(\xi) = (f \circ u)(\xi)$ is locally Lipschitz on $\mathcal{C}(K, \mathbf{R}^d)$ (for K a compact subset of $\mathbf{R}^n$, say), provided that $f\colon \mathbf{R}^d \to \mathbf{R}^d$ is locally Lipschitz continuous.

A much larger class of examples can be found if we restrict the regularity properties of F, but assume that L generates an analytic semigroup:

Theorem 6.5 *Let L generate an analytic semigroup on $\mathcal{B}$ (denote by $\mathcal{B}_\alpha$, $\alpha \in \mathbf{R}$ the corresponding interpolation spaces) and assume that Q is such that the stochastic convolution W_L has almost surely continuous sample paths in $\mathcal{B}_\alpha$ for some $\alpha \geq 0$. Assume furthermore that there exists $\gamma \geq 0$ and $\delta \in [0, 1)$ such that, for every $\beta \in [0, \gamma]$, the map F extends to a locally Lipschitz continuous map from $\mathcal{B}_\beta$ to $\mathcal{B}_{\beta-\delta}$ that, together with its local Lipschitz constant, grows at most polynomially.*

Then, (6.1) has a unique maximal mild solution (x, τ) with x taking values in $\mathcal{B}_\beta$ for every $\beta < \beta_\star \stackrel{\text{def}}{=} \alpha \wedge (\gamma + 1 - \delta)$.

Proof. In order to show that (6.1) has a unique mild solution, we proceed in a way similar to the proof of Theorem 6.4 and we make use of Exercise 4.41 to bound $\|S(t-s)F(u(s))\|$ in terms of $\|F(u(s))\|_{-\delta}$. This yields instead of (6.5) the bound

$$\sup_{t\in[0,T]} \|M_{g,T}u(t) - M_{g,T}v(t)\| \leq MT^{1-\delta} \sup_{t\in[0,T]} \|F(u(t)) - F(v(t))\| , \tag{6.7}$$

and similarly for (6.6), thus showing that (6.1) has a unique $\mathcal{B}$-valued maximal mild solution (x, τ). In order to show that $x(t)$ actually belongs to $\mathcal{B}_\beta$ for $t < \tau$ and $\beta \leq \alpha \wedge \gamma$, we make use of a "bootstrapping" argument, which is essentially an induction on β.

For notational convenience, we introduce the family of processes $W_L^a(t) = \int_{at}^t S(t-r)Q\, dW(r)$, where $a \in [0, 1)$ is a parameter. Note that one has the identity

$$W_L^a(t) = W_L(t) - S((1-a)t)W_L(at)$$

so that if W_L is continuous with values in $\mathcal{B}_\alpha$, then the same is true for W_L^a.

We are actually going to show the following stronger statement. Fix an arbitrary time $T > 0$. Then, for every $\beta \in [0, \beta_\star)$ there exist exponents $p_\beta \geq 1$, $q_\beta \geq 0$, and constants $a \in (0, 1)$, $C > 0$ such that the bound

$$\|x_t\|_\beta \leq Ct^{-q_\beta}\Big(1 + \sup_{s\in[at,t]} \|x_s\| + \sup_{0\leq s\leq t} \|W_L^a(s)\|_\beta\Big)^{p_\beta} , \tag{6.8}$$

holds almost surely for all $t \in (0, T]$.

The bound (6.8) is obviously true for $\beta = 0$ with $p_\beta = 1$ and $q_\beta = 0$. Assume now that, for some $\beta = \beta_0 \in [0, \gamma]$, the bound (6.8) holds. We will then argue that, for any $\varepsilon \in (0, 1 - \delta)$, the statement (6.8) also holds for $\beta = \beta_0 + \varepsilon$ (and therefore also for all intermediate values), provided that we adjust the constants appearing in the expression. Since it is possible to go from $\beta = 0$ to any value of $\beta < \gamma + 1 - \delta$ in a finite number of such steps, the claim then follows at once.

From the definition of a mild solution, we have the identity

$$x_t = S((1-a)t)x_{at} + \int_{at}^t S(t-s)F(x(s))\, ds + W_L^a(t) .$$

Since $\beta \leq \gamma$, it follows from our polynomial growth assumption on F that there exists $n > 0$ such that, for $t \in (0, T]$,

$$\|x_t\|_{\beta+\varepsilon} \leq Ct^{-\varepsilon}\|x_{at}\|_\beta + \|W_L^a(t)\|_{\beta+\varepsilon} + C\int_{at}^t (t-s)^{-(\varepsilon+\delta)}(1 + \|x_s\|_\beta)^n\, ds$$

$$\leq C(t^{-\varepsilon} + t^{1-\varepsilon-\delta}) \sup_{at\leq s\leq t} (1 + \|x_s\|_\beta^n) + \|W_L^a(t)\|_{\beta+\varepsilon}$$
$$\leq Ct^{-\varepsilon} \sup_{at\leq s\leq t} (1 + \|x_s\|_\beta^n) + \|W_L^a(t)\|_{\beta+\varepsilon} \, .$$

Here, the constant C depends on everything but t and x_0. Using the induction hypothesis, this yields the bound

$$\|x_t\|_{\beta+\varepsilon} \leq Ct^{-\varepsilon-nq_\beta}(1 + \sup_{s\in[a^2t,t]} \|x_s\| + \sup_{0\leq s\leq t} \|W_L^a(s)\|_\beta)^{np_\beta} + \|W_L^a(t)\|_{\beta+\varepsilon} \, ,$$

thus showing that (6.8) does indeed hold for $\beta = \beta_0 + \varepsilon$, provided that we replace a by a^2 and set $p_{\beta+\varepsilon} = np_\beta$ and $q_{\beta+\varepsilon} = \varepsilon + nq_\beta$. This concludes the proof of Theorem 6.5. □

6.2 Multiplicative noise

In tis section, we will see how some of the results in this chapter can be extended to situations when the intensity of the noise depends on the current state of the solution. In other words, we consider equations of the type

$$dx = Lx\,dt + F(x)\,dt + Q(x)\,dW \, , \tag{6.9}$$

In this section, we will always consider solutions taking values in a Hilbert space $\mathcal{H}$ and we denote by $\mathcal{H}_\alpha$ the corresponding scale of interpolation spaces with respect to L, which we assume to be the generator of an analytic semigroup on $\mathcal{H}$.

As before, we assume that F is a locally Lipschitz continuous map from $\mathcal{H}$ to $\mathcal{H}_{-\delta}$ for some $\delta \in [0, 1)$. We also assume that W is a cylindrical Wiener process on some Hilbert space $\mathcal{K}$ and that Q is a locally Lipschitz continuous map from $\mathcal{H}$ to $\mathcal{L}_2(\mathcal{K}, \mathcal{H}_\alpha)$ (the space of Hilbert-Schmidt operators from $\mathcal{K}$ to $\mathcal{H}_{-\alpha}$) for some $\alpha \in [0, \frac{1}{2})$. It will be clear from the sequel why the index $-\frac{1}{2}$ is a reasonable cut-off point here.

Similarly to above, we have

Definition 6.6 A local mild solution to (6.9) is a continuous $\mathcal{H}$-valued stochastic process x, together with a stopping time τ such that $\tau > 0$ almost surely and such that, for every $t > 0$, the identity

$$x(t) = S(t)x_0 + \int_0^{t\wedge\tau} S(t-s)F(x(s))\,ds + \int_0^{t\wedge\tau} S(t-s)Q(x(s))\,dW(s) \, , \tag{6.10}$$

holds almost surely on the event $\{t < \tau\}$. We furthermore impose that there exists $R > 0$ such that $\sup_{s\leq\tau} \|x(s)\| < R$ almost surely.

Remark 6.7 For this definition to make sense, we have to make sure that both terms on the right hand side of (6.10) make sense separately, the first as a Bochner integral and the second as a stochastic integral against W as in Corollary 3.93.

This follows from the boundedness condition on x, since the local Lipschitz continuity of F and Q then ensures that, for every finite time horizon, there exists a constant C (depending on R) such that $\|S(t-s)F(x(s))\| \leq C(t-s)^{-\delta}$ and $\|S(t-s)Q(x(s))\|_{\mathcal{L}_2(\mathcal{K},\mathcal{H})} \leq C(t-s)^{-\alpha}$. Since the first is integrable and the second is square integrable, the claim follows. In the multiplicative noise case, the situation becomes significantly more involved if we try to define Banach-space valued solutions since the definition of the stochastic integral is then much more subtle.

Remark 6.8 The reason why we cannot use the exact same definition as in the additive noise case is that for t an arbitrary stopping time, there is no reason why $s \mapsto S(t-s)Q(x(s))$ would be adapted, so that it is no longer clear what the meaning of the last term appearing in the right hand side actually is.

Remark 6.9 If (x, τ) is a local mild solution and y is a stochastic process such that, for every t, $y(t) = x(t)$ almost surely on the event $\{t < \tau\}$, then (y, τ) is again a local mild solution, so that it is natural to consider these as equivalent.

Furthermore, if $\bar\tau$ is a stopping time such that $\bar\tau \le \tau$ almost surely and $\bar x$ is a process such that $\bar x(t) = x(t)$ almost surely on $\{t < \bar\tau\}$ (whenever this is the case, we say that (x, τ) *extends* $(\bar x, \bar\tau)$), then $(\bar x, \bar\tau)$ is also a local mild solution. This is not a completely obvious fact, but follows from the local property of the Itô integral: if two integrands agree almost surely on some event A then their Itô integrals also agree almost surely on A.

This remark suggests the following definition.

Definition 6.10 We say that $(\bar x, \bar\tau)$ is a *maximal mild solution* if, on the set $\{\bar\tau < \infty\}$, one has $\lim_{t\to\bar\tau} \|\bar x(t)\| = \infty$ almost surely and there exists a sequence (x_n, τ_n) of local mild solutions with τ_n increasing, $\bar\tau = \sup_n \tau_n$ almost surely, and such that $(\bar x, \bar\tau)$ extends each of the (x_n, τ_n).

Remark 6.11 The pair $(\bar x, \bar\tau)$ itself is in general *not* a local mild solution in the sense of Definition 6.6 since $\bar x$ is not bounded up to $\bar\tau$. Even if we relax that condition, it may simply happen that the stochastic integral in (6.10) makes no sense because of the divergence of the integrand near $t = \tau$, which is a problem that did not arise in the previous subsection.

Similarly, we can define a concept of a *local weak solution* by adapting Definition 5.1:

Definition 6.12 A local weak solution to (6.9) is an $\mathcal{H}$-valued stochastic process x together with a stopping time τ such that $\tau > 0$ almost surely and such that the identity

$$\langle \ell, x(t) - x_0\rangle = \int_0^t \langle L^*\ell, x(s)\rangle\, ds + \int_0^{t\wedge\tau} \langle \ell, F(x(s))\rangle\, ds + \int_0^{t\wedge\tau} \langle \ell, Q(x(s))\, dW(s)\rangle\,, \quad (6.11)$$

holds almost surely for every time t and for every $\ell \in \mathcal{D}(L^*)$. As before, we furthermore impose that there exists $R > 0$ such that $\sup_{s\le\tau} \|x(s)\| < R$ almost surely.

Remark 6.13 Since $\ell \in \mathcal{D}(L^*)$ by assumption, ℓ belongs to the dual of $\mathcal{H}_{-1}$. As a consequence, it also belongs to the dual of $\mathcal{H}_{-\delta}$ for $\delta \le 1$, so that the right hand side of (6.11) makes sense thanks to our boundedness assumption, just as in Remark 6.7.

As in the case of linear equations, we start by showing that these two notions of a solution actually do coincide:

Proposition 6.14 *Let L, F and Q be as above. Then weak solutions are mild solutions and vice-versa.*

Proof. The proof that mild solutions are weak solutions is virtually identical to the proof of Proposition 5.8.

Only "problem": have to use stochastic Fubini.... □ Finish this up

The main result of this section is the following well-posedness result:

Theorem 6.15 *Let L, F and Q be as above. Then (6.9) admits a unique maximal solution.*

Proof. The proof is quite similar to the proof of Theorems 6.4 and 6.5, the main difference being that we cannot subtract the "stochastic convolution" since Q is now allowed to depend on the solution itself. As a consequence, we cannot construct solutions in a pathwise manner as before, but instead we have to perform a fixed point argument in a space of stochastic processes. The problem with this is that a stochastic process can typically take arbitrarily large values, which will be a problem if F and / or Q grow too fast at infinity. There are two ways to circumvent this: considering stochastic processes that

are stopped when they reach values that are too large or approximating F and Q by globally Lipschitz functions. We choose the latter approach

However, for every □

6.3 Interpolation inequalities and Sobolev embeddings

The kind of bootstrapping arguments used in the proof of Theorem 6.5 above are extremely useful to obtain regularity properties of the solutions to semilinear parabolic stochastic PDEs. However, they rely on obtaining bounds on the regularity of F from one interpolation space into another. In many important situations, the interpolation spaces turn out to be given by fractional Sobolev spaces. For the purpose of these notes, we are going to restrict ourselves to the analytically easier situation where the space variable of the stochastic PDE under consideration takes values in the d-dimensional torus $\mathbf{T}^d$. In other words, we restrict ourselves to situations where the operator describing the linearised evolution is endowed with periodic boundary conditions.

This will make the proof of the embedding theorems presented in these notes technically more straightforward. For the corresponding embeddings with more general boundary conditions or even on more general manifolds or unbounded domains, we refer for example to the comprehensive series of monographs [Tri83, Tri92, Tri06].

Recall that, given a distribution $u \in L^2(\mathbf{T}^d)$, we can decompose it as a Fourier series:

$$u(x) = \sum_{k \in \mathbf{Z}^d} u_k e^{i\langle k, x\rangle} ,$$

where the identity holds for (Lebesgue) almost every $x \in \mathbf{T}^d$. Furthermore, the L^2 norm of u is given by Parseval's identity $\|u\|^2 = \sum |u_k|^2$. We have

Definition 6.16 The fractional Sobolev space $H^s(\mathbf{T}^d)$ for $s \geq 0$ is given by the subspace of functions $u \in L^2(\mathbf{T}^d)$ such that

$$\|u\|_{H^s}^2 \stackrel{\text{def}}{=} \sum_{k \in \mathbf{Z}^d} (1 + |k|^2)^s |u_k|^2 < \infty . \tag{6.12}$$

Note that this is a separable Hilbert space and that $H^0 = L^2$. For $s < 0$, we define H^s as the closure of L^2 under the norm (6.12).

Remark 6.17 By the spectral decomposition theorem, H^s for $s > 0$ is the domain of $(1 - \Delta)^{s/2}$ and we have $\|u\|_{H^s} = \|(1 - \Delta)^{s/2} u\|_{L^2}$.

A very important situation is the case where L is a differential operator with constant coefficients (formally $L = P(\partial_x)$ for some polynomial $P\colon \mathbf{R}^d \to \mathbf{R}$) and $\mathcal{H}$ is either an L^2 space or some Sobolev space. In this case, one has

Lemma 6.18 *Assume that $P\colon \mathbf{R}^d \to \mathbf{R}$ is a polynomial of degree $2m$ such that there exist positive constants c, C such that the bound*

$$(-1)^{m+1} c|k|^{2m} \leq P(k) \leq (-1)^{m+1} C|k|^{2m} ,$$

holds for all k outside of some compact set. Then, the operator $P(\partial_x)$ generates an analytic semigroup on $\mathcal{H} = H^s$ for every $s \in \mathbf{R}$ and the corresponding interpolation spaces are given by $\mathcal{H}_\alpha = H^{s+2m\alpha}$.

Proof. By inspection, noting that $P(\partial_x)$ is conjugate to the multiplication operator by $P(ik)$ via the Fourier decomposition. □

Note first that for any two positive real numbers a and b and any pair of positive conjugate exponents p and q, one has Young's inequality

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}\,, \qquad \frac{1}{p} + \frac{1}{q} = 1\,. \tag{6.13}$$

As a corollary of this elementary bound, we obtain Hölder's inequality, which can be viewed as a generalisation of the Cauchy-Schwartz inequality:

Proposition 6.19 (Hölder's inequality) *Let $(\mathcal{M}, \mu)$ be a measure space and let p and q be a pair of positive conjugate exponents. Then, for any pair of measurable functions $u, v\colon \mathcal{M} \to \mathbf{R}$, one has*

$$\int_{\mathcal{M}} |u(x)v(x)|\,\mu(dx) \le \|u\|_p\,\|v\|_q\,,$$

for any pair (p, q) of conjugate exponents.

Proof. It follows from (6.13) that, for every $\varepsilon > 0$, one has the bound

$$\int_{\mathcal{M}} |u(x)v(x)|\,\mu(dx) \le \frac{\varepsilon^p \|u\|_p^p}{p} + \frac{\|v\|_q^q}{q\varepsilon^q}\,,$$

Setting $\varepsilon = \|v\|_q^{\frac{1}{p}}\|u\|_p^{\frac{1}{p}-1}$ concludes the proof. □

One interesting consequence of Hölder's inequality is the following interpolation inequality for powers of selfadjoint operators:

Proposition 6.20 *Let A be a positive definite selfadjoint operator on a separable Hilbert space $\mathcal{H}$ and let $\alpha \in [0, 1]$. Then, the bound $\|A^\alpha u\| \le \|Au\|^\alpha \|u\|^{1-\alpha}$ holds for every $u \in \mathcal{D}(A^\alpha) \subset \mathcal{H}$.*

Proof. The extreme cases $\alpha \in \{0, 1\}$ are obvious, so we assume $\alpha \in (0, 1)$. By the spectral theorem, we can assume that $\mathcal{H} = L^2(\mathcal{M}, \mu)$ and that A is the multiplication operator by some positive function f. Applying Hölder's inequality with $p = 1/\alpha$ and $q = 1/(1-\alpha)$, one then has

$$\begin{aligned}\|A^\alpha u\|^2 &= \int f^{2\alpha}(x)u^2(x)\,\mu(dx) = \int |fu|^{2\alpha}(x)\,|u|^{2-2\alpha}(x)\,\mu(dx)\\ &\le \Big(\int f^2(x)u^2(x)\,\mu(dx)\Big)^\alpha \Big(\int u^2(x)\,\mu(dx)\Big)^{1-\alpha}\,,\end{aligned}$$

which is exactly the bound we wanted to show. □

An immediate corollary is:

Corollary 6.21 *For any $t > s$ and any $r \in [s, t]$, the bound*

$$\|u\|_{H^r}^{t-s} \le \|u\|_{H^t}^{r-s}\|u\|_{H^s}^{t-r} \tag{6.14}$$

is valid for every $u \in H^t$.

Proof. Apply Proposition 6.20 with $\mathcal{H} = H^s$, $A = (1-\Delta)^{\frac{t-s}{2}}$, and $\alpha = (r-s)/(t-s)$. □

Exercise 6.22 As a consequence of Hölder's inequality, show that for any collection of n measurable functions and any exponents $p_i > 1$ such that $\sum_{i=1}^n p_i^{-1} = 1$, one has the bound

$$\int_{\mathcal{M}} |u_1(x)\cdots u_n(x)|\,\mu(dx) \le \|u_1\|_{p_1}\cdots\|u_n\|_{p_n}\,.$$

Following our earlier discussion regarding fractional Sobolev spaces, it would be convenient to be able to bound the L^p norm of a function in terms of one of the fractional Sobolev norms. It turns out that bounding the L^∞ norm is rather straightforward:

Lemma 6.23 *For every $s > \frac{d}{2}$, the space $H^s(\mathbf{T}^d)$ is contained in the space of continuous functions and there exists a constant C such that $\|u\|_{L^\infty} \le C\|u\|_{H^s}$.*

Proof. It follows from Cauchy-Schwarz that

$$\sum_{k\in\mathbf{Z}^d} |u_k| \le \Big(\sum_{k\in\mathbf{Z}^d} (1+|k|^2)^s |u_k|^2\Big)^{1/2} \Big(\sum_{k\in\mathbf{Z}^d} (1+|k|^2)^{-s}\Big)^{1/2} .$$

Since the sum in the second factor converges if and only if $s > \frac{d}{2}$, the claim follows. □

Exercise 6.24 In dimension $d = 2$, find an example of an unbounded function u such that $\|u\|_{H^1} < \infty$.

Exercise 6.25 Show that for $s > \frac{d}{2}$, H^s is contained in the space $\mathcal{C}^\alpha(\mathbf{T}^d)$ for every $\alpha < s - \frac{d}{2}$.

As a consequence of Lemma 6.23, we are able to obtain a more general Sobolev embedding for all L^p spaces:

Theorem 6.26 (Sobolev embeddings) *Let $p \in [2,\infty]$. Then, for every $s > \frac{d}{2} - \frac{d}{p}$, the space $H^s(\mathbf{T}^d)$ is contained in the space $L^p(\mathbf{T}^d)$ and there exists a constant C such that $\|u\|_{L^p} \le C\|u\|_{H^s}$.*

Proof. The case $p = 2$ is obvious and the case $p = \infty$ has already been shown, so it remains to show the claim for $p \in (2,\infty)$. The idea is to divide Fourier space into "blocks" corresponding to different length scales and to estimate separately the L^p norm of every block. More precisely, we define a sequence of functions $u^{(n)}$ by

$$u^{-1}(x) = u_0 , \quad u^{(n)}(x) = \sum_{2^n \le |k| < 2^{n+1}} u_k e^{i\langle k,x\rangle} ,$$

so that one has $u = \sum_{n\ge -1} u^{(n)}$. For $n \ge 0$, one has

$$\|u^{(n)}\|_{L^p}^p \le \|u^{(n)}\|_{L^2}^2 \|u^{(n)}\|_{L^\infty}^{p-2} . \tag{6.15}$$

Choose now $s' = \frac{d}{2} + \varepsilon$ and note that the construction of $u^{(n)}$, together with Lemma 6.23, guarantees that one has the bounds

$$\|u^{(n)}\|_{L^2} \le 2^{-ns}\|u^{(n)}\|_{H^s} , \quad \|u^{(n)}\|_{L^\infty} \le C\|u^{(n)}\|_{H^{s'}} \le C2^{n(s'-s)}\|u^{(n)}\|_{H^s} .$$

Inserting this into (6.15), we obtain

$$\|u^{(n)}\|_{L^p} \le C\|u^{(n)}\|_{H^s} 2^{n\left((s'-s)\frac{p-2}{p} - \frac{2s}{p}\right)} = C\|u^{(n)}\|_{H^s} 2^{n\left(\varepsilon\frac{p-2}{p} + \frac{d}{2} - \frac{d}{p} - s\right)} \le C\|u\|_{H^s} 2^{n\left(\varepsilon + \frac{d}{2} - \frac{d}{p} - s\right)} .$$

It follows that $\|u\|_{L^p} \le |u_0| + \sum_{n\ge 0} \|u^{(n)}\|_{L^p} \le C\|u\|_{H^s}$, provided that the exponent appearing in this expression is negative which, since ε can be chosen arbitrarily small, is precisely the case whenever $s > \frac{d}{2} - \frac{d}{p}$. □

Remark 6.27 For $p \ne \infty$, one actually has $H^s(\mathbf{T}^d) \subset L^p(\mathbf{T}^d)$ with $s = \frac{d}{2} - \frac{d}{p}$, but this borderline case is more difficult to obtain.

Combining the Sobolev embedding theorem and Hölder's inequality, it is eventually possible to estimate in a similar way the fractional Sobolev norm of a product of two functions:

Theorem 6.28 *Let r, s and t be positive exponents such that $s \wedge r \geq t$ and $s + r > t + \frac{d}{2}$. Then, if $u \in H^r$ and $v \in H^s$, the product $w = uv$ belongs to H^t.*

Proof. Define $u^{(n)}$ and $v^{(m)}$ as in the proof of the Sobolev embedding theorem and set $w^{(m,n)} = u^{(m)}v^{(n)}$. Note that one has $w_k^{(m,n)} = 0$ if $|k| > 2^{3+(m\vee n)}$. It then follows from Hölder's inequality that if $p, q \geq 2$ are such that $p^{-1} + q^{-1} = \frac{1}{2}$, one has the bound

$$\|w^{(m,n)}\|_{H^t} \leq C2^{t(m\vee n)}\|w^{(m,n)}\|_{L^2} \leq C2^{t(m\vee n)}\|u^{(m)}\|_{L^p}\|v^{(n)}\|_{L^q} \, .$$

Assume now that $m > n$. The conditions on r, s and t are precisely such that there exists a pair (p, q) as above with

$$r > t + \frac{d}{2} - \frac{d}{p} \, , \qquad s > \frac{d}{2} - \frac{d}{q} \, .$$

In particular, we can find some $\varepsilon > 0$ such that

$$\|u^{(m)}\|_{L^p} \leq C\|u^{(m)}\|_{H^{r-t-\varepsilon}} \leq C2^{-m(t+\varepsilon)}\|u\|_{H^r} \, , \quad \|v^{(n)}\|_{L^p} \leq C\|v^{(n)}\|_{H^{s-\varepsilon}} \leq C2^{-n\varepsilon}\|v\|_{H^s} \, .$$

Inserting this into the previous expression, we find that

$$\|w^{(m,n)}\|_{H^t} \leq C2^{-m\varepsilon - n\varepsilon}\|u\|_{H^r}\|u\|_{H^s} \, .$$

Since our assumptions are symmetric in u and v, we obtain a similar bound for the case $m \leq n$, so that

$$\|w\|_{H^t} \leq \sum_{m,n>0} \|w^{(m,n)}\|_{H^t} \leq C\|u\|_{H^r}\|u\|_{H^s} \sum_{m,n>0} 2^{-m\varepsilon - n\varepsilon} \leq C\|u\|_{H^r}\|u\|_{H^s} \, ,$$

as requested. □

Exercise 6.29 Show that the conclusion of Theorem 6.28 still holds if $s = t = r$ is a positive integer, provided that $s > \frac{d}{2}$.

Exercise 6.30 Similarly to Exercise 6.22, show that one can iterate this bound so that if $s_i > s \geq 0$ are exponents such that $\sum_i s_i > s + \frac{(n-1)d}{2}$, then one has the bound

$$\|u_1 \cdots u_n\|_s \leq C\|u_1\|_{s_1} \cdots \|u_n\|_{s_n} \, .$$

Hint: The case $s \geq \frac{d}{2}$ is simple, so it suffices to consider the case $s < \frac{d}{2}$.

The functional inequalities from the previous section allow to check that the assumptions of Theorems 6.4 and 6.5 are verified by a number of interesting equations.

6.4 Reaction-diffusion equations

This is a class of (stochastic) partial differential equations that model the evolution of reactants in a gel, described by a spatial domain D. They are of the type

$$du = \Delta u \, dt + f \circ u \, dt + Q \, dW(t) \, , \tag{6.16}$$

where $u(x,t) \in \mathbf{R}^d$, $x \in D \subset \mathbf{R}^n$, describes the density of the various components of the reaction at time t and location x. The nonlinearity $f\colon \mathbf{R}^d \to \mathbf{R}^d$ describes the reaction itself and the term Δu describes the diffusion of the reactants in the gel. The noise term $Q\,dW$ should be interpreted as a crude attempt to describe the fluctuations in the quantities of reactant due both to the discrete nature of the underlying particle system and the interaction with the environment[1].

[1] A more realistic description of these fluctuations would result in a covariance Q that depends on the solution u. Since we have not developed the tools necessary to treat this type of equations, we restrict ourselves to the simple case of a constant covariance operator Q.

Equations of the type (6.16) also appear in the theory of amplitude equations, where they appear as a kind of "normal form" near a change of linear instability. In this particular case, one often has $d = 2$ and $f(u) = \kappa u - u|u|^2$ for some $\kappa \in \mathbf{R}$, see [BHP05]. A natural choice for the Banach space $\mathcal{B}$ in which to consider solutions to (6.16) is the space of bounded continuous functions $\mathcal{B} = \mathcal{C}(D, \mathbf{R}^d)$ since the composition operator $u \mapsto f \circ u$ (also sometimes called Nemitskii operator) then maps $\mathcal{B}$ into itself and inherits the regularity properties of f. If the domain D is sufficiently regular then the semigroup generated by the Laplacian Δ is the Markov semigroup for a Brownian motion in D. The precise description of the domain of Δ is related to the behaviour of the corresponding Brownian motion when it hits the boundary of D. In order to avoid technicalities, let us assume from now on that D consists of the torus $\mathbf{T}^n$, so that there is no boundary to consider.

Exercise 6.31 Show that in this case, Δ generates an analytic semigroup on $\mathcal{B} = \mathcal{C}(\mathbf{T}^n, \mathbf{R}^d)$ and that for $\alpha \in \mathbf{N}$, the interpolation space $\mathcal{B}_\alpha$ is given by $\mathcal{B}_\alpha = \mathcal{C}^{2\alpha}(\mathbf{T}^n, \mathbf{R}^d)$.

If Q is such that the stochastic convolution has continuous sample paths in $\mathcal{B}$ almost surely and f is locally Lipschitz continuous, we can directly apply Theorem 6.4 to obtain the existence of a unique local solution to (6.16) in $\mathcal{C}(\mathbf{T}^n, \mathbf{R}^d)$. We would like to obtain conditions on f that ensure that this local solution is also a global solution, that is the blow-up time τ is equal to infinity almost surely.

If f happens to be a globally Lipschitz continuous function, then the existence and uniqueness of global solutions follows from Theorem 6.4. Obtaining global solutions when f is not globally Lipschitz continuous is slightly more tricky. The idea is to obtain some *a priori* estimate on some functional of the solution which dominates the supremum norm and ensures that it cannot blow up in finite time.

Let us first consider the deterministic part of the equation alone. The structure we are going to exploit is the fact that the Laplacian generates a Markovian semigroup. Actually, we can even consider sub-Markovian semigroups, namely semigroups $\mathcal{P}_t$ of positive measures over some fixed Polish space $\mathcal{X}$ such that $\mathcal{P}_t(x, \mathcal{X}) \leq 1$ for every x. This then allows to cover for example the case of the Laplacian with Dirichlet boundary conditions. The following general fact will turn out to be useful:

Lemma 6.32 *Let $\mathcal{P}_t$ be a sub-Markovian Feller[2] semigroup over a Polish space $\mathcal{X}$. Extend it to $\mathcal{C}_b(\mathcal{X}, \mathbf{R}^d)$ by applying it to each component independently. Let $V : \mathbf{R}^d \to \mathbf{R}_+$ be convex (that is $V(\alpha x + (1-\alpha)y) \leq \alpha V(x) + (1-\alpha)V(y)$ for all $x, y \in \mathbf{R}^d$ and $\alpha \in [0,1]$) with $V(0) = 0$ and define $\tilde{V} : \mathcal{C}_b(\mathcal{X}, \mathbf{R}^d) \to \mathbf{R}_+$ by $\tilde{V}(u) = \sup_{x \in \mathcal{X}} V(u(x))$. Then $\tilde{V}(\mathcal{P}_t u) \leq \tilde{V}(u)$ for every $t \geq 0$ and every $u \in \mathcal{C}_b(\mathcal{X}, \mathbf{R}^d)$.*

Remark 6.33 One should keep in mind the following example: take $\mathcal{X} = \mathbf{T}^n$, the n-dimensional torus and take $\mathcal{P}_t = S(t)$ to be the heat semigroup, which is indeed a Feller semigroup on $\mathcal{C}_b(\mathbf{T}^n, \mathbf{R})$. This can be extended to $\mathbf{R}^n$-valued functions by evolving each component independently under the heat equation. In particular, this result then shows that $\tilde{V}(S(t)u) \leq \tilde{V}(u)$ for every convex V and every $u \in \mathcal{C}(\mathbf{T}^n, \mathbf{R}^d)$. Another example to keep in mind is that of $\mathcal{X} = D \subset \mathbf{R}^n$ with $\mathcal{P}_t(x, \cdot)$ given by the law of Brownian motion starting at x and killed when it reaches ∂D. In this case, $\mathcal{P}_t$ is not Markovian, but it is still sub-Markovian.

Proof. Note first that if V is convex, then it is continuous and, for every probability measure μ on $\mathbf{R}^d$, one has the inequality

$$V\Big(\int_{\mathbf{R}^d} x\, \mu(dx)\Big) \leq \int_{\mathbf{R}^d} V(x)\, \mu(dx)\;. \tag{6.17}$$

One can indeed check by induction that (6.17) holds if μ is a "simple" measure consisting of a convex combination of finitely many Dirac measures. The general case then follows from the continuity of

[2] A Markov semigroup is Feller if it maps continuous functions into continuous functions.

V and the fact that every probability measure on $\mathbf{R}^d$ can be approximated (in the topology of weak convergence) by a sequence of simple measures.

Denote now by $\mathcal{P}_t(x,\cdot)$ the transition probabilities for $\mathcal{P}_t$, so that $\mathcal{P}_t u$ is given by the formula $(\mathcal{P}_t u)(x) = \int_{\mathcal{X}} u(y)\,\mathcal{P}_t(x,dy)$. In order to sidestep the problem that $\mathcal{P}_t(x,\cdot)$ is not a probability measure, we extend it to a probability measure $\hat{\mathcal{P}}(x,\cdot)$ on $\hat{\mathcal{X}} = \mathcal{X} \cup \{\star\}$, where $\star$ denotes a "cemetery state". This extension is performed by setting

$$\hat{\mathcal{P}}(x,\cdot) = \mathcal{P}(x,\cdot) + (1 - \mathcal{P}(x,\mathcal{X}))\delta_\star(\cdot)\ .$$

With this notation at hand, we also extend any function $u \in \mathcal{C}(\mathcal{X},\mathbf{R}^d)$ into a function $\hat{u} \in \mathcal{C}(\hat{\mathcal{X}},\mathbf{R}^d)$ by setting $\hat{u}(\star) = 0$. One then has

$$\begin{aligned}
\tilde{V}(\mathcal{P}_t u) &= \sup_{x\in\mathcal{X}} V\Big(\int_{\mathcal{X}} u(y)\,\mathcal{P}_t(x,dy)\Big) = \sup_{x\in\mathcal{X}} V\Big(\int_{\hat{\mathcal{X}}} \hat{u}(y)\,\hat{\mathcal{P}}_t(x,dy)\Big)\\
&= \sup_{x\in\mathcal{X}} V\Big(\int_{\mathbf{R}^d} v\,(\hat{u}^\sharp\hat{\mathcal{P}}_t(x,\cdot))(dv)\Big) \le \sup_{x\in\mathcal{X}} \int_{\mathbf{R}^d} V(v)\,(\hat{u}^\sharp\hat{\mathcal{P}}_t(x,\cdot))(dv)\\
&= \sup_{x\in\mathcal{X}} \int_{\mathcal{X}} V(\hat{u}(y))\,\hat{\mathcal{P}}_t(x,dy) = \sup_{x\in\mathcal{X}} \int_{\mathcal{X}} V(u(y))\,\mathcal{P}_t(x,dy)\\
&\le \sup_{y\in\mathcal{X}} V(u(y)) = \tilde{V}(u)\ ,
\end{aligned} \tag{6.18}$$

where we made use of the fact that $V(0) = 0$ to obtain the identities on the first and third lines. We also made use of (6.17) on the second line. This is precisely the required inequality. □

This is the main ingredient allowing us to obtain a priori estimates for the solution to (6.16):

Proposition 6.34 *Consider the setting for equation (6.16) described above. Assume that Q is such that W_Δ has continuous sample paths in $\mathcal{B} = \mathcal{C}(\mathbf{T}^n,\mathbf{R}^d)$ and that there exists a convex twice differentiable function $V\colon \mathbf{R}^d \to \mathbf{R}_+$ such that $\lim_{|x|\to\infty} V(x) = \infty$ and such that, for every $R > 0$, there exists a constant C such that $\langle \nabla V(x), f(x+y)\rangle \le CV(x)$ for every $x \in \mathbf{R}^d$ and every y with $|y| \le R$. Then (6.16) has a global solution in $\mathcal{B}$.*

Proof. We denote by $u(t)$ the local mild solution to (6.16). Our aim is to obtain *a priori* bounds on $\tilde{V}(u(t))$ that are sufficiently good to show that one cannot have $\lim_{t\to\tau} \|u(t)\| = \infty$ for any finite (stopping) time τ.

Setting $v(t) = u(t) - W_\Delta(t)$, the definition of a mild solution shows that v satisfies the equation

$$v(t) = e^{\Delta t}v(0) + \int_0^t e^{\Delta(t-s)}(f \circ (v(s) + W_\Delta(s)))\,ds \stackrel{\text{def}}{=} e^{\Delta t}v(0) + \int_0^t e^{\Delta(t-s)}F(s)\,ds\ .$$

Since $t \mapsto v(t)$ is continuous by Theorem 6.4 and the same holds for W_Δ by assumption, the function $t \mapsto F(t)$ is continuous in $\mathcal{B}$. Therefore, one has

$$\lim_{h\to 0} \frac{1}{h}\Big(\int_0^h e^{\Delta(h-s)}F(s)\,ds - he^{\Delta h}F(0)\Big) = 0\ .$$

Combining this with (6.18), we therefore obtain for $\tilde{V}(v)$ the bound

$$\limsup_{h\to 0} h^{-1}(\tilde{V}(v(t+h)) - \tilde{V}(v(t))) \le \limsup_{h\to 0} h^{-1}(\tilde{V}(v(t) + hF(t)) - \tilde{V}(v(t)))\ .$$

Since V furthermore belongs to $\mathcal{C}^2$ by assumption, we also have the estimate

$$\tilde{V}(v(t) + hF(t)) = \sup_{x\in\mathbf{T}^n} (V(v(x,t)) + h\langle \nabla V(v(x,t)), F(x,t)\rangle) + \mathcal{O}(h^2)\ .$$

Using the definition of F and the assumptions on V, it follows that for every $R > 0$ there exists a constant C such that, provided that $\|W_\Delta(t)\| \le R$, one has

$$\limsup_{h\to 0} h^{-1}(\tilde{V}(v(t+h)) - \tilde{V}(v(t))) \le C\tilde{V}(v(t)) \, .$$

A standard comparison argument for ODEs then shows that $\tilde{V}(v(t))$ cannot blow up as long as $\|W_\Delta(t)\|$ does not blow up, thus concluding the proof. □

Example 6.35 A prime example of equation that falls into the category of reaction-diffusion equations considered here is the stochastic Allen-Cahn equation given by

$$du = \Delta u \, dt + \frac{1}{\varepsilon}(u - u^3) \, dt + Q \, dW \, , \tag{6.19}$$

viewed for example as an equation on the space $\mathcal{B} = \mathcal{C}([-L, L]^n, \mathbf{R})$, endowed with periodic boundary conditions. Here, a typical example for Q would be a Fourier symbol of the type

$$\widehat{Qu}(k) = q_k \hat{u}(k) \, ,$$

where $\hat{u}$ denotes the Fourier series of u. In this case, one can check that $V(u) = u^2$ is

This is a model for phase separation: for small values of $\varepsilon > 0$, solutions to (6.19) tend to stay close to the two stable fixed points for the ordinary differential equation $\dot{u} = u - u^3$, namely $u = 1$ and $u = -1$. The regions where $u \approx \pm 1$ are separated by boundaries that model the interface between two phases of a ??, see for example [].

Exercise 6.36 In the case $d = 1$, show that the assumptions of Proposition 6.34 are satisfied for $V(u) = u^2$ if f is any polynomial of odd degree with negative leading coefficient.

Exercise 6.37 Show that in the case $d = 3$, (6.16) has a unique global solution when we take for f the right-hand side of the Lorentz attractor:

$$f(u) = \begin{pmatrix} \sigma(u_2 - u_1) \\ u_1(\varrho - u_3) - u_2 \\ u_1 u_2 - \beta u_3 \end{pmatrix} ,$$

where ϱ, σ and β are three arbitrary positive constants.

Exercise 6.38 Show that a proposition similar to Proposition **??** holds if we replace Δ

6.5 The stochastic Navier-Stokes equations

The Navier-Stokes equations govern the motion of an idealised incompressible fluid and are one of the most studied models in the theory of partial differential equations, as well as in theoretical and mathematical physics. We are going to use the symbol $u(x, t)$ to denote the instantaneous velocity of the fluid at position $x \in \mathbf{R}^d$ and time t, so that $u(x, t) \in \mathbf{R}^d$. With these notations, the deterministic Navier-Stokes equations are given by

$$\partial_t u = \nu \Delta u - (u \cdot \nabla)u - \nabla p \, , \qquad \operatorname{div} u = 0 \, , \tag{6.20}$$

where the (scalar) pressure p is determined implicitly by the incompressibility condition $\operatorname{div} u = 0$ and $\nu > 0$ denotes the kinematic viscosity of the fluid. In principle, these equations make sense for any value of the dimension d. However, even the deterministic equations (6.20) are known to have global smooth solutions for arbitrary smooth initial data only in dimension $d = 2$. We are therefore going to restrict ourselves to the two-dimensional case in the sequel. As we saw already in the introduction,

solutions to (6.20) tend to 0 as time goes to ∞, so that an external forcing is required in order to obtain an interesting stationary regime.

One natural way of adding an external forcing is given by a stochastic force that is white in time and admits a translation invariant correlation function in space. In this way, it is possible to maintain the translation invariance of the equations (in a statistical sense), even though the forcing is not constant in space. We are furthermore going to restrict ourselves to solutions that are periodic in space in order to avoid the difficulties arising from partial differential equations in unbounded domains. The incompressible stochastic Navier-Stokes equations on the torus $\mathbf{R}^2$ are given by

$$du = \nu\Delta u\,dt - (u\cdot\nabla)u\,dt - \nabla p\,dt + Q\,dW(t)\;, \qquad \operatorname{div} u = 0\;, \tag{6.21}$$

where p and $\nu > 0$ are as above. In order to put these equations into the more familiar form (6.1), we denote by Π the orthogonal projection onto the space of divergence-free vector fields. In Fourier components, Π is given by

$$(\Pi u)_k = u_k - \frac{k\langle k, u_k\rangle}{|k|^2}\;. \tag{6.22}$$

(Note here that the Fourier coefficients of a vector field are themselves vectors.) With this notation, one has

$$du = \nu\Delta u\,dt + \Pi(u\cdot\nabla)u\,dt + Q\,dW(t) \stackrel{\text{def}}{=} \Delta u\,dt + F(u)\,dt + Q\,dW(t)\;.$$

It is clear from (6.22) that Π is a contraction in any fractional Sobolev space. For $t \geq 0$, it therefore follows from Theorem 6.28 that

$$\|F(u)\|_{H^t} \leq \|u\|_{H^s}\|\nabla u\|_{H^{s-1}} \leq C\|u\|_{H^s}^2\;, \tag{6.23}$$

provided that $s \geq t+1$ and $s > \frac{t}{2} + \frac{1}{2} + \frac{d}{4}$. In particular, this bound holds for $s = t+1$, provided that $t > 0$.

Furthermore, in this setting, since L is just the Laplacian, if we choose $\mathcal{H} = H^s$, then the interpolation spaces $\mathcal{H}_\alpha$ are given by $\mathcal{H}_\alpha = H^{s+2\alpha}$. This allows us to apply Theorem 6.5 to show that the stochastic Navier-Stokes equations admit local solutions for any initial condition in H^s, provided that $s > 1$, and that the stochastic convolution takes values in that space. Furthermore, these solutions will immediately lie in any higher order Sobolev space, all the way up to the space in which the stochastic convolution lies.

This line of reasoning does however not yield any *a priori* bounds on the solution, so that it may blow up in finite time. The Navier-Stokes nonlinearity satisfies $\langle u, F(u)\rangle = 0$ (the scalar product is the L^2 scalar product), so one may expect bounds in L^2, but we do not know at this stage whether initial conditions in L^2 do indeed lead to local solutions. We would therefore like to obtain bounds on $F(u)$ in negative Sobolev spaces. In order to do this, we exploit the fact that H^{-s} can naturally be identified with the dual of H^s, so that

$$\|F(u)\|_{H^{-s}} = \sup\Big\{\int F(u)(x)\,v(x)\,dx\;,\quad v\in\mathcal{C}^\infty\;,\quad \|v\|_{H^s}\leq 1\Big\}\;.$$

Making use of the fact that we are working with divergence-free vector fields, one has (using Einstein's convention of summation over repeated indices):

$$\int F(u)\,v\,dx = -\int v_j u_i \partial_i u_j\,dx \leq \|v\|_{L^p}\|\nabla u\|_{L^2}\|u\|_{L^q}\;,$$

provided that $p, q > 2$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{2}$. We now make use of the fact that $\|u\|_{L^q} \leq C_q\|\nabla u\|_2$ for every $q \in [2,\infty)$ (but $q = \infty$ is excluded) to conclude that for every $s > 0$ there exists a constant C such that

$$\|F(u)\|_{-s} \leq C\|\nabla u\|_{L^2}^2\;. \tag{6.24}$$

In order to get *a priori* bounds for the solution to the 2D stochastic Navier-Stokes equations, one

can then make use of the following trick: introduce the vorticity $w = \nabla \wedge u = \partial_1 u_2 - \partial_2 u_1$. Then, provided that $\int u\,dx = 0$ (which, provided that the range of Q consists of vector fields with mean 0, is a condition that is preserved under the solutions to (6.21)), the vorticity is sufficient to describe u completely by making use of the incompressibility assumption $\operatorname{div} u = 0$. Actually, the map $w \mapsto u$ can be described explicitly by

$$u_k = (Kw)_k = \frac{k^\perp w_k}{|k|^2}\,, \qquad (k_1, k_2)^\perp = (-k_2, k_1)\,.$$

This shows in particular that K is actually a bounded operator from H^s into H^{s+1} for every s. It follows that one can rewrite (6.21) as

$$dw = \nu\Delta w\,dt + (Kw\cdot\nabla)w\,dt + \tilde{Q}\,dW(t) \stackrel{\text{def}}{=} \Delta w\,dt + \tilde{F}(w)\,dt + \tilde{Q}\,dW(t)\,. \tag{6.25}$$

Since $\tilde{F}(w) = \nabla \wedge F(Kw)$, it follows from (6.24) that one has the bounds

$$\|\tilde{F}(w)\|_{-1-s} \le C\|w\|_{L^2}^2\,,$$

so that $\tilde{F}$ is a locally Lipschitz continuous map from L^2 into H^s for every $s < -1$. This shows that (6.25) has unique local solutions for every initial condition in L^2 and that these solutions immediately become as regular as the corresponding stochastic convolution.

Denote now by $\tilde{W}_L$ the stochastic convolution

$$\tilde{W}_L(t) = \int_0^t e^{\Delta(t-s)}\tilde{Q}\,dW(s)\,,$$

and define the process $v(t) = w(t) - W_L(t)$. With this notation, v is the unique solution to the random PDE

$$\partial_t v = \nu\Delta v + \tilde{F}(v + \tilde{W}_L)\,.$$

It follows from (6.23) that $\|\tilde{F}(w)\|_{H^{-s}} \le C\|w\|_{H^s}^2$, provided that $s > 1/3$. For the sake of simplicity, assume from now on that $\tilde{W}_L$ takes values in $H^{1/2}$ almost surely. Using the fact that $\langle v, \tilde{F}(v)\rangle = 0$, we then obtain for the L^2-norm of v the following *a priori* bound:

$$\begin{aligned}
\partial_t\|v\|^2 &= -2\nu\|\nabla v\|^2 - 2\langle \tilde{W}_L, \tilde{F}(v+\tilde{W}_L)\rangle \\
&\le -2\nu\|\nabla v\|^2 + 2\|\tilde{W}_L\|_{H^{1/2}}\|v+\tilde{W}_L\|_{H^{1/2}}^2 \\
&\le -2\nu\|\nabla v\|^2 + 4\|\tilde{W}_L\|_{H^{1/2}}(\|v\|_{H^{1/2}}^2 + \|\tilde{W}_L\|_{H^{1/2}}^2) \\
&\le -2\nu\|\nabla v\|^2 + 4\|\tilde{W}_L\|_{H^{1/2}}(\|v\|\|\nabla v\| + \|\tilde{W}_L\|_{H^{1/2}}^2) \\
&\le \frac{8}{\nu}\|\tilde{W}_L\|_{H^{1/2}}^2\|v\|^2 + 2\|\tilde{W}_L\|_{H^{1/2}}^3\,,
\end{aligned} \tag{6.26}$$

so that global existence of solutions then follows from Gronwall's inequality.

This calculation is only formal, since it is not known in general whether the L^2-norm of v is differentiable as a function of time. The bound that one obtains from (6.26) can however be made rigorous in a very similar way as for the example of the stochastic reaction-diffusion equation, so that we will not reproduce this argument here.

7

Variation of solutions

Our aim in this chapter is to show that the solutions to semilinear stochastic PDEs as studied in the previous section are differentiable both with respect to their initial condition and with respect to the driving noise. Knowing that this is indeed the case is of interest in a number of situations. For example,

Add some examples: Funaki / Brassesco & Co; ergodicity

7.1 Differentiation with respect to the initial condition

In this section, we study how the solutions to (6.1) depend on their initial conditions. Since the map $M_{g,T}$ that was used in (6.4) to construct the solutions to (6.1) is Fréchet differentiable (it is actually differentiable arbitrarily often, provided that F is) and since it is a contraction for sufficiently small values of t, we will be able to apply the classical implicit functions theorem (see for example [RR04] for a Banach space version) to deduce that for every realisation of the driving noise, the map $u_s \mapsto u_t$ is Fréchet differentiable, provided that $t > s$ is sufficiently close to s. Here, the precise meaning of "sufficiently close" depends in general both on the value of u_s and on the size of the stochastic convolution W_L for times greater than s.

Iterating this argument and using the fact that the composition of Fréchet differentiable functions is again Fréchet differentiable, we conclude that, for any $s \le t < \tau$, the map $u_s \mapsto u_t$ given by the solutions to (6.1) is Fréchet differentiable in $\mathcal{B}$. Inspecting the expression for the derivative given by the implicit functions theorem, we conclude that the derivative $J_{s,t}\varphi$ in the direction $\varphi \in \mathcal{B}$ satisfies the following random linear equation in its mild formulation:

$$\partial_t J_{s,t}\varphi = -LJ_{s,t}\varphi + DF(u_t)J_{s,t}\varphi\,, \quad J_{s,s}\varphi = \varphi\,. \tag{7.1}$$

Note that, by the properties of monomials, it follows from Assumption **AN** that

set up assumption...

$$\|DN(u)v\|_\gamma \le C(1 + \|u\|_{\gamma+a})^{n-1}\|v\|_{\gamma+a}\,,$$

for every $\gamma \in [-a, \gamma_\star)$. A fixed point argument similar to the one in Theorem 6.4 shows that the solution to (7.1) is unique, but note that it does not allow us to obtain bounds on its moments. We only have that for any T smaller than the explosion time to the solutions of (6.1), there exists a (random) constant C such that

$$\sup_{0\le s<t<T}\sup_{\|\varphi\|\le 1}\|J_{s,t}\varphi\| \le C\,. \tag{7.2}$$

The constant C depends exponentially on the size of the solution u in the interval $[0, T]$. However, if we obtain better control on $J_{s,t}$ by some means, we can then use the following bootstrapping argument:

Proposition 7.1 *For every $\gamma < \gamma_\star + 1$, there exist exponents $\tilde p_\gamma, \tilde q_\gamma \ge 0$, and constants $C > 0$ and $\gamma_0 < |\gamma|$ such that we have the bound*

$$\|J_{t,t+s}\varphi\|_\gamma \le Cs^{-\tilde p_\gamma}\sup_{r\in[\frac{s}{2},s]}(1 + \|u_{t+r}\|_{\gamma_0})^{\tilde q_\gamma}\|J_{t,t+r}\varphi\|\,, \tag{7.3}$$

for every $\varphi \in \mathcal{H}$ and every $t, s > 0$. If $\gamma < 1 - a$, then one can choose $\gamma_0 = \tilde p_\gamma = 0$ and $\tilde q_\gamma = n - 1$.

Since an almost identical argument will be used in the proof of Proposition 7.3 below, we refer the reader there for details. We chose to present that proof instead of this one because the presence of an adjoint causes slight additional complications.

For $s \le t$, let us define operators $K_{s,t}$ via the solution to the (random) PDE

$$\partial_s K_{s,t}\varphi = LK_{s,t}\varphi - DN^*(u_s)K_{s,t}\varphi\ , \quad K_{t,t}\varphi = \varphi\ , \quad \varphi \in \mathcal{H}\ . \tag{7.4}$$

Note that this equation runs *backwards* in time and is random through the solution u_t of (6.1). Here, $DN^*(u)$ denotes the adjoint in $\mathcal{H}$ of the operator $DN(u)$ defined earlier. Fixing the terminal time t and setting $\varphi_s = K_{t-s,t}\varphi$, we obtain a more usual representation for φ_s:

$$\partial_s \varphi_s = -L\varphi_s + DN^*(u_{t-s})\varphi_s\ . \tag{7.5}$$

The remainder of this subsection will be devoted to obtaining regularity bounds on the solutions to (7.4) and to the proof that $K_{s,t}$ is actually the adjoint of $J_{s,t}$. We start by showing that, for γ sufficiently close to (but less than) $\gamma_\star + 1$, (7.4) has a unique solution for every path $u \in \mathcal{C}(\mathbf{R}, \mathcal{H}_\gamma)$ and $\varphi \in \mathcal{H}$.

Proposition 7.2 *There exists $\gamma < \gamma_\star + 1$ such that, for every $\varphi \in \mathcal{H}$, equation (7.4) has a unique continuous $\mathcal{H}$-valued solution for every $s < t$ and every $u \in \mathcal{C}(\mathbf{R}, \mathcal{H}_\gamma)$. Furthermore, $K_{s,t}$ depends only on u_r for $r \in [s,t]$ and the map $\varphi \mapsto K_{s,t}\varphi$ is linear and bounded.*

Proof. As in Proposition **??**, we define a map $\Phi_{T,u}\colon \mathcal{H} \times \mathcal{C}([0,T],\mathcal{H}) \to \mathcal{C}([0,T],\mathcal{H})$ by

$$\big(\Phi_{T,u}(\varphi_0,\varphi)\big)_t = e^{-Lt}\varphi_0 + \int_0^t e^{-L(t-s)}(DN^*(u_s))\varphi_s\, ds\ .$$

It follows from Assumption **ADN** with $\beta = -a$ that there exists $\gamma < \gamma_\star + 1$ such that $DN^*(u)\colon \mathcal{H} \to \mathcal{H}_{-a}$ is a bounded linear operator for every $u \in \mathcal{H}_\gamma$. Proceeding as in the proof of Proposition **??**, we see that Φ is a contraction for sufficiently small T. □

Similarly to before, we can use a bootstrapping argument to show that $K_{s,t}\varphi$ actually has more regularity than stated in Proposition 7.2.

Proposition 7.3 *For every $\beta < \beta_\star + 1$, there exists $\gamma < \gamma_\star + 1$, exponents $\bar p_\beta, \bar q_\beta > 0$, and a constant C such that*

$$\|K_{t-s,t}\varphi\|_\beta \le Cs^{-\bar p_\beta} \sup_{r\in[\frac{s}{2},s]} (1 + \|u_{t-r}\|_\gamma)^{\bar q_\beta} \|K_{t-r,t}\varphi\|\ , \tag{7.6}$$

for every $\varphi \in \mathcal{H}$, every $t, s > 0$, and every $u \in \mathcal{C}(\mathbf{R}, \mathcal{H}_\gamma)$.

Proof. Fix $\beta < \beta_\star + a$ and $\delta \in (0, 1-a)$ and assume that the bound (7.6) holds for $\|K_{s,t}\varphi\|_\beta$. Since we run s "backwards in time" from $s = t$, we consider again t as fixed and use the notation $\varphi_s = K_{t-s,t}\varphi$. We then have, for arbitrary $\alpha \in (0,1)$,

$$\|\varphi_s\|_{\beta+\delta} \le Cs^{-\delta}\|\varphi_{\alpha s}\|_\beta + C\int_{\alpha s}^s (s-r)^{-(\delta+a)}\|DN^*(u_{t-r})\varphi_r\|_{\beta-a}\, dr\ ,$$

provided that γ is sufficiently close to $\gamma_\star + 1$ such that $DN^*\colon \mathcal{H}_\gamma \to \mathcal{L}(\mathcal{H}_\beta, \mathcal{H}_{\beta-a})$ by Assumption **ADN**. Furthermore, the operator norm of $DN^*(v)$ is bounded by $C(1 + \|v\|_\gamma)^{n-1}$, yielding

$$\begin{aligned}\|\varphi_s\|_{\beta+\delta} &\le Cs^{-\delta}\|\varphi_{\alpha s}\|_\beta + Cs^{-(\delta+a)} \sup_{r\in[\alpha s,s]} (1 + \|u_r\|_\gamma)^{n-1}\|\varphi_r\|_\beta \\ &\le Cs^{-(\delta+a)} \sup_{r\in[\alpha s,s]} (1 + \|u_r\|_\gamma)^{n-1}\|\varphi_r\|_\beta\ .\end{aligned}$$

Iterating these bounds as in Proposition **??** concludes the proof. □

The following lemma appears also in [MP06, BM07]. It plays a central role in establishing the

representation of the Malliavin matrix given in (**??**) on which this article as well as [MP06, BM07] rely heavily.

Proposition 7.4 *For every $0 \le s < t$, $K_{s,t}$ is the adjoint of $J_{s,t}$ in $\mathcal{H}$, that is $K_{s,t} = J_{s,t}^*$.*

Proof. Fixing $0 \le s < t$ and $\varphi, \psi \in \mathcal{H}_\infty$, we claim that the expression

$$\langle J_{s,r}\varphi, K_{r,t}\psi\rangle \,, \tag{7.7}$$

is independent of $r \in [s,t]$. Evaluating (7.7) at both $r = s$ and $r = t$ then concludes the proof.

We now prove that (7.7) is independent of r as claimed. It follows from (7.4) and Proposition **??** that, with probability one, the map $r \mapsto K_{r,t}\varphi$ is continuous with values in $\mathcal{H}_{\beta+1}$ and differentiable with values in $\mathcal{H}_\beta$, provided that $\beta < \beta_\star$. Similarly, the map $r \mapsto J_{s,r}\psi$ is continuous with values in $\mathcal{H}_{\gamma+1}$ and differentiable with values in $\mathcal{H}_\gamma$, provided that $\gamma < \gamma_\star$. Since $\gamma_\star + \beta_\star > -1$ by assumption, it thus follows that (7.7) is differentiable in r for $r \in (s,t)$ with

$$\begin{aligned}\partial_r\langle J_{s,r}\varphi, K_{r,t}\psi\rangle &= \langle (L + DN(u_r))J_{s,r}\varphi, K_{r,t}\psi\rangle \\ &\quad - \langle J_{s,r}\varphi, (L + DN^*(u_r))K_{r,t}\psi\rangle = 0 \,.\end{aligned}$$

Since furthermore both $r \mapsto K_{r,t}\varphi$ and $r \mapsto J_{s,r}\psi$ are continuous in r on the closed interval, the proof is complete. See for example [DL92, p. 477] for more details. □

7.2 Higher order derivatives

Let us derive now a formula for the higher-order variations of the solution with respect to its initial condition.

For integer $n \ge 2$, let $\varphi = (\varphi_1, \cdots, \varphi_n) \in \mathcal{H}^{\otimes n}$ and $s = (s_1, \cdots, s_n) \in [0,\infty)^n$ and define $\vee s = s_1 \vee \cdots \vee s_n$. We will now define the n-th variation of the equation $J_{s,t}^{(n)}\varphi$ which intuitively is the cumulative effect on u_t of varying the value of u_{s_k} in the direction φ_k.

If $I = \{n_1 < \ldots < n_{|I|}\}$ is an ordered subset of $\{1,\ldots,n\}$ (here $|I|$ means the number of elements in I), we introduce the notation $s_I = (s_{n_1}, \ldots, s_{n_{|I|}})$ and $\varphi_I = (\varphi_{n_1}, \ldots, \varphi_{n_{|I|}})$. Now the n-th variation of the equation $J_{s,t}^{(n)}\varphi$ solves

$$\begin{aligned}\partial_t J_{s,t}^{(n)}\varphi &= -LJ_{s,t}^{(n)}\varphi + DN(u(t))J_{s,t}^{(n)}\varphi + \mathcal{G}_{s,t}^{(n)}(u(t),\varphi), \quad t > \vee s, \\ J_{s,t}^{(n)}\varphi &= 0, \quad t \le \vee s,\end{aligned} \tag{7.8}$$

where

$$\mathcal{G}_{s,t}^{(n)}(u,\varphi) = \sum_{\nu=2}^{m\wedge n} \sum_{I_1,\ldots,I_\nu} D^{(\nu)}N(u)\Big(J_{s_{I_1},t}^{(|I_1|)}\varphi_{I_1}, \ldots, J_{s_{I_\nu},t}^{(|I_\nu|)}\varphi_{I_\nu}\Big) \,, \tag{7.9}$$

and the second sum runs over all partitions of $\{1,\ldots,n\}$ into disjoint, ordered non-empty sets $I_1, \ldots, I_\nu$.

The variations of constants formula then implies that

$$J_{s,t}^{(n)}\varphi = \int_0^t J_{r,t}\mathcal{G}_{s,r}^{(n)}(u_r,\varphi)dr \,, \tag{7.10}$$

see also [BM07]. We obtain the following bound on the higher-order variations:

Proposition 7.5 *If $\beta_\star > a - 1$ then there exists $\gamma < \gamma_\star + 1$ such that, for every $n > 0$, there exist exponents N_n and M_n such that*

$$\|J_{s,t}^{(n)}\varphi\| \le C \sup_{r\in[0,t]} (1 + \|u_r\|_\gamma)^{N_n} \sup_{0\le u<v\le t} (1 + \|J_{u,v}\|)^{M_n} \,,$$

uniformly over all n-uples φ with $\|\varphi_k\| \le 1$ for every k.

Proof. We proceed by induction. As a shorthand, we set

$$\mathcal{E}(M,N) = \sup_{r\in[0,t]} (1+\|u_r\|_\gamma)^N \sup_{0\le u<v\le t} (1+\|J_{u,v}\|)^M \ .$$

The result is trivially true for $n=1$ with $M_1=1$ and $N_1=0$. For $n>1$, we combine (7.10) and (7.9), and we use Assumption **??**, part 2., to obtain

$$\begin{aligned}\|J^{(n)}_{s,t}\varphi\| &\le C\int_0^t \|J_{r,t}\|_{-a\to 0}\Big(1+\|u_r\|^n + \sum_I \|J^{|I|}_{s_I,r}\varphi_I\|^n\Big)\,dr \\ &\le C\mathcal{E}(nM_{n-1}, n(N_{n-1}+1))\int_0^t \|K_{r,t}\|_{0\to a}\,dr \ .\end{aligned}$$

To go from the first to the second line, we used the induction hypothesis, the fact that $K_{r,t}=J^*_{r,t}$, and the duality between $\mathcal{H}_a$ and $\mathcal{H}_{-a}$.

It remains to apply Proposition 7.3 with $\beta=a$ to obtain the required bound. □

7.3 Differentiation with respect to the noise

Show that solutions are differentiable as functions of the initial condition and of the noise. Derive formulae for the derivatives (to any order).

7.4 Integration by parts formula

One very useful tool in the study of the long-time behaviour of solutions to stochastic PDEs is an "integration by parts formula" on Wiener space.

8

Approximations to SPDEs

This chapter is devoted to the theory of approximation of stochastic PDEs. The typical situation that we have in mind is a numerical approximation, but one could also think of a sequence of stochastic PDEs depending on a parameter, for which one wants to show that the solutions converge to the solutions to some limiting equation. In the context of numerical approximation, there are two conceptually distinct components to the approximation procedure: *spatial approximation* and *temporal approximation*. Spatial approximation consists in approximating the "real" Banach space of functions $\mathcal{B}$ in which the solutions take values by some (typically finite-dimensional, but not necessarily) space $\mathcal{B}_n$ in which we have an approximating SPDE which is close in some sense to the original equation. Temporal approximation on the other hand consists in constructing a step by step approximation to the solution of the equation.

Let us illustrate the difference between these two concepts by the simplest example we know: the stochastic heat equation. Recall that this is the equation

$$du = \partial_x^2 u \, dt + dW(t) \, , \tag{8.1}$$

where we take $u \in \mathcal{H} = L^2([0,\pi])$, W a cylindrical Wiener process on $\mathcal{H}$, and we endow the linear operator ∂_x^2 with Dirichlet boundary conditions. This operator can of course be diagonalised explicitly. Its eigenfunctions are given by $e_n(x) = \sqrt{2/\pi}\sin(nx)$ (the normalisation constant is chosen in such a way that $\|e_n\| = 1$) and the corresponding eigenvalues are $\lambda_n = n^2$.

One of the simplest possible *spatial* approximation for (8.1) can be obtained by writing

$$u(x,t) \approx U_N(x,t) \stackrel{\text{def}}{=} \sum_{n=1}^{N} u_n(t) e_n(x) \, ,$$

and by taking for u_n the solutions to

$$du_n = -n^2 u_n \, dt + \langle e_n, dW(t) \rangle \, . \tag{8.2}$$

Since the e_n's form an orthonormal basis of $\mathcal{H}$, the processes $\langle e_n W(t) \rangle$ are in turn nothing but a sequence of independent standard Wiener processes. In this particular case, denoting by Π_N the orthogonal projection in $\mathcal{H}$ onto the first N eigenvectors e_n, it is very easy to see that if we set $u_n(0) = \langle e_n, u(0) \rangle$, then we have $U_N(t) = \Pi_N u(t)$ for every $t \geq 0$. Since we know from Exercise 5.16 that $u(t)$ belongs to the interpolation space $\mathcal{H}_\alpha$ for every $\alpha < 1/4$, this allows us to obtain an error estimate on $\|U_N(t) - u(t)\|$ by using the fact that

$$\|U_N(t) - u(t)\| = \|(\Pi_N - 1)u(t)\| \leq \lambda_N^{-\alpha} \|u(t)\|_\alpha \, .$$

Since $\lambda_N \approx N^2$ and we can take α arbitrarily close to $1/4$, it follows that we have the approximate bound $\|U_N(t) - u(t)\| \approx \mathcal{O}(N^{-1/2})$ for large values of N.

Temporal approximations for stochastic PDEs are slightly more tricky to study because the individual terms of the equation do not necessarily make sense in the space in which solutions take values, so that the usual explicit Euler scheme does not work. A temporal approximation to (8.1) could be obtained for example by fixing a small parameter $h > 0$, setting $t_n = nh$ and defining a sequence u_n recursively by

$$u_{n+1} = u_n + h\partial_x^2 u_{n+1} + \sqrt{h}\, \Delta W_n \, , \tag{8.3}$$

where the ΔW_n's are i.i.d. copies of a Gaussian random variable with Cameron-Martin space $\mathcal{H}$. Note that even though the increments ΔW_n of the cylindrical Wiener process W are not $\mathcal{H}$-valued, the sequence of approximations u_n is. This can be seen by rewriting (8.3) as

$$u_{n+1} = (1 - h\partial_x^2)^{-1}(u_n + \sqrt{h}\,\Delta W_n)\;,$$

and noting that the operator $(1 - h\partial_x^2)^{-1}$ is a Hilbert-Schmidt operator on $\mathcal{H}$.

8.1 Spatial approximation

This section is devoted to the spatial approximation of stochastic PDEs. A very simple example of spatial approximation was given by (8.3), for which we were furthermore able to estimate a rate of convergence. This simple example however relied on very explicit calculations (we used the fact that the linear operator ∂_x^2 can be diagonalised explicitly and that the noise acts independently on each eigenspace), so we would like to obtain a more abstract result that can be applied to a larger class of examples.

The typical example of spatial approximation that one should keep in mind is given by the finite difference approximation to a one-dimensional reaction-diffusion equation. Consider the stochastic PDE given by

$$du = \partial_x^2 u\,dt + f(u)\,dt + dW(t)\;, \tag{8.4}$$

where ∂_x^2 denotes the second derivative operator on $\mathcal{H} = L^2([0,1], \mathbf{R})$ with periodic boundary conditions, $f\colon \mathbf{R} \to \mathbf{R}$ is a globally Lipschitz continuous nonlinearity, and W is a cylindrical Wiener process on $\mathcal{H}$. It is then straightforward to check that the assumptions of Theorem 6.5 are satisfied. Actually, there are several possible choices of $\mathcal{B}$ that work. As we have seen in Section 6.4, we could choose $\mathcal{B}$ to be the space of continuous functions, but in the case when f is globally Lipschitz continuous, the choice $\mathcal{B} = \mathcal{H}$ would also do. In both cases, one can take $\alpha = \delta = 0$.

Let us now try to construct a finite-dimensional approximation to (8.4). For any given $n > 0$, we subdivide the interval $[0,1]$ into n subintervals of equal size and we denote by $u_n^i(t)$ an approximation to $u(i/n, t)$ for $i = \{0, \dots, n\}$ with the convention that $u_n^0 = u_n^n$. Our aim is to derive a system of n coupled ordinary stochastic differential equations for the u_n^i such that their solution yields a good approximation to the solution to (8.4) evaluated at the gridpoints.

If $v\colon [0,1] \to \mathbf{R}$ is a smooth function, then a simple second-order Taylor expansion around $x = i/n$ shows that

$$\partial_x^2 v(i/n) = n^2(v((i+1)/n) + v((i-1)/n) - 2v(i/n)) + \mathcal{O}(n^{-2})\;. \tag{8.5}$$

As a consequence, it would seem that a reasonable approximation to ∂_x^2 is given by this second-order difference operator so that, when deriving an equation for u_n^i, we would like to replace the term $\partial_x^2 u$ in the right hand side of (8.4) by $n^2(u_n^{i+1} + u_n^{i-1} - 2u_n^i)$.

Remark 8.1 Note that (8.5) is only valid for *smooth* functions v. (Actually, v needs to be $\mathcal{C}^4$ for this bound to hold and it certainly needs to be $\mathcal{C}^2$ for this approximation to converge at all.) This could potentially be a problem since the solutions to many interesting stochastic PDEs are not spatially smooth. It turns out that this problem is not severe, as long as the finite difference approximation appears in the linear part of the equation. If it appears in the nonlinear part, the lack of regularity may cause serious problems, as we will se in Section **??** below.

It is less immediate to see how the noise term dW should be approximated. Since we have seen in Section 3.7 that a cylindrical Wiener process on L^2 corresponds to space-time white noise, it seems intuitively clear that we should consider a sequence $\{w_i\}_{i=1}^n$ of independent Wiener processes and use w_i to drive the equation for u_n^i. The only question that remains is that of the correct amplitude that we should choose for w_i. This question can be answered by the following heuristic argument. Denote by

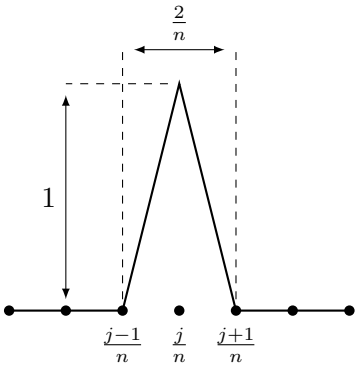
$\iota_n\colon \mathbf{R}^n \to \mathcal{H}$ the map that returns the piecewise linear function that agrees on the gridpoints with the given values. Formally, we have

$$(\iota_n u)(x) = (1-\delta)u^i + \delta u^{i+1}\,, \qquad x = \frac{i+\delta}{n}\,, \quad \delta \in [0,1]\,.$$

Again, we make the convention $u^0 = u^n$. We also note that while the choice of i and δ is ambiguous if x is an integer multiple of $1/n$, the resulting value for $(\iota_n u)(x)$ is not. Let now ξ_n be a normal Gaussian random variable on $\mathbf{R}^n$, namely the ξ_n^i are all independent and normally distributed with variance 1. In this case, the covariance for $\iota_n\xi_n$ is given by

$$\mathbf{E}(\iota_n\xi_n)(i/n)(\iota_n\xi_n)(j/n) = \delta_{ij}\,, \tag{8.6}$$

Furthermore, for any given x, the function defined by $y \mapsto \mathbf{E}(\iota_n\xi_n)(x)(\iota_n\xi_n)(y)$ is affine on every interval of the form $[j/n, (j+1)/n]$. This, together with (8.6) is in principle sufficient to completely determine the covariance of $\iota_n\xi_n$. Note now that, for $x = j/n$, the function $y \mapsto C^n(x,y) \stackrel{\text{def}}{=} \mathbf{E}(\iota_n\xi_n)(x)(\iota_n\xi_n)(y)$ has the shape depicted in the figure shown to the right. We thus see that, for every x of the form j/n with $j \in \{0,\dots,n\}$, we have the identity



$$\int_0^1 C_n(x,y)\,dy = \frac{1}{n}\,. \tag{8.7}$$

On the other hand, we would like to obtain an approximation to spatial white noise, which has a Dirac delta-function as its covariance. As a consequence of (8.7), the function $nC_n(x,y)$ is a good approximation to $\delta(x-y)$. In view of this argument, it seems natural to approximate dW on the gridpoints by $\sqrt{n}\,dw_n^i$, where the w_n^i are i.i.d. standard Wiener processes.

Collecting all of these arguments, we finally obtain the following approximation to (8.4):

$$du_n^i = n^2(u_n^{i+1} + u_n^{i-1} - 2u_n^i)\,dt + f(u_n^i)\,dt + \sqrt{n}\,dw_n^i\,. \tag{8.8}$$

At this stage, guessing the speed and type of convergence of (8.8) towards (8.4) is not straightforward anymore. The aim of this section is to give an abstract approximation result that includes this case and allows to obtain explicit bounds on the convergence rate. We will then show how this abstract result can be applied to approximations like (8.8), as well as several other examples. We will also see when these arguments break down, mainly due to the lack of spatial regularity, as already anticipated in Remark 8.1.

8.1.1 An abstract result

Throughout this section, we put ourselves in the following context. We consider a semilinear stochastic PDE of the form

$$du = Lu\,dt + F(u)\,dt + Q\,dW(t)\,, \tag{8.9}$$

where W is a cylindrical Wiener process on some separable Hilbert space $\mathcal{K}$, and the operators $L\colon \mathcal{D}(L) \to \mathcal{B}$, $F\colon \mathcal{D}(F) \to \mathcal{B}$, $Q\colon \mathcal{K} \to \mathcal{B}$ are as in Theorem 6.5, so that we have a good local existence theory for its solutions.

The solutions to this equation will then be approximated by a sequence of equations taking values in a sequence of (possibly finite-dimensional, but no necessarily so) Banach spaces $\mathcal{B}^{(n)}$ which are identified with subspaces of $\mathcal{B}$ via an injection map $\iota_n\colon \mathcal{B}^{(n)} \to \mathcal{B}$ and a projection map $\pi_n\colon \mathcal{B} \to \mathcal{B}^{(n)}$ such that $\pi_n \circ \iota_n$ is the identity on $\mathcal{B}^{(n)}$. The main application of the abstract theorem given in this section will be the numerical approximation of stochastic PDEs. However, it is more general than that, and also covers various homogenisation results where a "smooth" linear operator is approximated by

operators with highly oscillating coefficients. A number of such applications will be given in the next subsections.

Example 8.2 A standard example of a pair of operators ι_n and π_n satisfying our assumptions is given by the projection of a square integrable function on $[0, 2\pi]$ onto its first n Fourier coefficients. In this case, one would take for example $\mathcal{B} = L^2([0, 2\pi])$ and $\mathcal{B}^{(n)} = \mathbf{R} \oplus \mathbf{C}^n$ (since the '0' Fourier coefficient is always real for a real-valued function, but the other Fourier coefficients can be complex). In this case, the operator ι_n simply outputs the trigonometric polynomial with the given Fourier coefficients.

Another typical example would be given by $\mathcal{B} = \mathcal{C}([0, 1])$, $\mathcal{B}^{(n)} = \mathbf{R}^{n+1}$, and π_n the operator that evaluates its arguments at the points k/n for $k \in \{0, \dots, n\}$. In this case, a natural "reconstruction" operator ι_n would be given by simply performing a linear interpolation between the gridpoints.

However, the setup considered here is much more flexible and allows for a wide range of situations. For example, if $\mathcal{B} = \mathcal{C}^1([0, 1])$, one could keep not only the values of the function at gridpoints, but also the values of its derivative. In this case, a natural reconstruction operator would be given by interpolating with cubic splines.

Our sequence of approximating SPDEs is of the form

$$du_n = L_n u_n \, dt + F_n(u_n) \, dt + Q_n \, dW_n(t) \, , \tag{8.10}$$

where we assume that $L_n \colon \mathcal{D}(L_n) \to \mathcal{B}^{(n)}$, $F_n \colon \mathcal{D}(F_n) \to \mathcal{B}^{(n)}$ and $Q_n \colon \mathcal{K}^{(n)} \to \mathcal{B}^{(n)}$ again satisfy the assumptions of Theorem 6.5 for every fixed value of n, but with constants that are *independent* of n. If $\mathcal{B}^{(n)}$ happens to be finite-dimensional, as is usually the case in practice, then these assumptions are all trivially satisfied, provided that F_n is locally Lipschitz continuous, but the independence of n is less trivial.

More precisely, denoting by S_n the analytic semigroup generated by L_n and by $\mathcal{B}_\alpha^{(n)}$ the corresponding interpolation spaces, we are going to make throughout this section the following assumptions on the approximations L_n and F_n. First of all, we assume that the stochastic convolution given by the solutions to the approximating equation converges to the one given by the original equation and that we have some control on the convergence rate:

Assumption 8.3 Denote by Z the solution to the linearised equation and by Z_n its approximation:

$$Z(t) = \int_0^t S(t-s) Q \, dW(s) \, , \quad Z_n(t) = \int_0^t S_n(t-s) Q_n \, dW_n(s) \, .$$

Then, we assume that Z has a continuous $\mathcal{B}$-valued version and that there exists a joint realisation of the cylindrical Wiener processes W and W_n and an exponent γ_0 such that the bound

$$\mathbf{E} \sup_{t \in [0,T]} \|\iota_n Z_n(t) - Z(t)\| \leq C(T) n^{-\gamma_0} \, , \tag{8.11}$$

holds for every $T > 0$ for some constant $C(T)$.

The reason why we do not restate this assumption as a consequence of number of conditions on Q_n and L_n is that there are many instances where it is relatively easy to check (8.11) directly. This is because both processes are Gaussian and there are explicit formulae available for their covariances in a number of interesting situations.

Our next assumption basically states that the semigroup generated by L_n approximates the one generated by L, when applied to elements in $\mathcal{B}_{-\delta}$:

Assumption 8.4 There exist exponents γ_2 and γ_3 such that

$$\|\iota_n S_n(t) \pi_n x - S(t) x\| \leq C t^{-\alpha} (1 \wedge t^{-\gamma_2} n^{-\gamma_3}) \|x\|_{-\alpha} \, , \tag{8.12}$$

holds for every $\alpha \in [0, \delta]$ and every $x \in \mathcal{B}_{-\alpha}$.

Note that the factor $t^{-\delta}$ is natural since this is the bound that one has on $S(t)x$. The reason why we allow the error term to become large for small times is somewhat related: in the case of a numerical approximation, the reason why one expects S_n to be a good approximation of S is that S has some smoothing properties, so that the small-scale behaviour of the initial condition x gets "washed out", thus allowing it to be approximated by a finite-dimensional object. This argument however breaks down for very small times, so that we have to allow for the approximation to become bad as $t \to 0$.

Finally, we have of course to assume that the approximating nonlinearity F_n is a reasonable approximation to the "true" nonlinearity F:

Assumption 8.5 There exists an exponent γ_4 such that, for every $R > 0$, there exists a constant C such that

$$\|\iota_n F_n(\pi_n x) - F(x)\|_{-\delta} \le C n^{-\gamma_4} ,$$

holds for every $x \in \mathcal{B}$ with $\|x\| \le R$.

With these assumptions at hand, the main abstract theorem of this section is the following:

Theorem 8.6 *Let $U_0 \in \mathcal{B}$, $R \ge \|U_0\| + 1$ and $T > 0$. Let T_R be the stopping time given by $T_R = T \wedge \inf\{t > 0 : \|u(t)\| \ge R\}$, where u denotes the solution to (8.9) with initial condition U_0. Assume furthermore that, for every $n > 0$, there exists a continuous $\mathcal{B}$-valued solution u_n to (8.10) with initial condition $\pi_n U_0$.*

Then, for every $\varepsilon > 0$ and every $R \ge \|U_0\| + 1$, there exists a constant $C > 0$ and a value $n_0 > 0$ such that

$$\mathbf{P}\Big(\sup_{s \le T_R} s^\alpha \|\iota_n u_n(s) - u(s)\| > C n^{-\gamma} \Big) \le \varepsilon , \tag{8.13}$$

for all $n \ge n_0$. In this expression, the exponent γ is given by $\gamma = \gamma_0 \wedge \gamma_3 \wedge \gamma_4 \wedge (1-\delta)\frac{\gamma_2}{\gamma_3}$, and α is given by $\alpha = (1-\delta) \wedge \gamma_2$.

Remark 8.7 One feature of this statement is that it requires no a priori bounds on the solutions of either (8.9) or (8.10). The main reason is that we only consider solutions up to the stopping time T_R, which takes care of possible blow-ups in the solution u to (8.9). However, even if u does not blow up, the weak assumptions that we have on F_n do not necessarily prevent a finite-time blow-up of u_n. This however, is taken care of *a posteriori* by the fact that u_n converges to u, so that it will be less than $R+1$ (say) with very high probability. This is also why we need to include n_0 in the statement: for small values of n, it may well happen that u_n blows up before time T_R with a fixed probability, so that (8.13) cannot in general be expected to hold for all n.

Proof. Note first that the although the statement of Theorem 8.6 is probabilistic, its proof is completely deterministic. The only point where probabilities enter the game is in the bounds on the stochastic convolution. Indeed, it follows form (8.11) and Chebychev's inequality that

$$\mathbf{P}\Big(\sup_{t \in [0,T]} \|\iota_n Z_n(t) - Z(t)\| \ge L \Big) \le L^{-1} C(T) n^{-\gamma_0} .$$

In particular, there exists a constant C depending on ε and T such that the bound

$$\mathbf{P}\Big(\sup_{t \in [0,T]} \|\iota_n Z_n(t) - Z(t)\| \le C n^{-\gamma_0} \,\&\, \sup_{t \in [0,T]} \|Z(t)\| \le R \Big) \ge 1 - \varepsilon , \tag{8.14}$$

holds for every $n > 0$. The theorem will thus be proven if we can find a constant C depending on the initial condition such that the "deterministic" bound

$$\sup_{s \le T_R} s^\alpha \|\iota_n u_n(s) - u(s)\| \le C n^{-\gamma} , \tag{8.15}$$

holds for all realisations of the stochastic convolutions such that the bounds in the left hand side of (8.14) hold. As a shorthand, it will be convenient to write $v(t) = u(t) - Z(t)$ and similarly for v_n. With this notation, and recalling the assumed bound on $Z - \iota_n Z_n$, we see that (8.15) is implied by the same bound with u (resp. u_n) replaced by v (resp. v_n).

In order to prove the result, we will first show that a bound of the type (8.15) holds for a sufficiently short time t and we will then patch such bounds together in order to get the desired bound up to time T_R. We therefore try first to obtain a bound of the type (8.15), but with slightly different initial conditions for u and u_n. Fix initial conditions $U_0 \in \mathcal{B}$ and $\pi_n U_0 \in \mathcal{B}^{(n)}$ and denote by u and u_n the solutions to (8.9) with initial condition U_0 and (8.10) with initial condition $\pi_n U_0$ respectively.

These notations will be convenient due to the fact that we will Using the mild formulation of the solution and the triangle inequality, we then obtain for every $t \geq r$ the bound

$$\begin{aligned}\|v(t) - \iota_n v_n(t)\| &\leq \|S(t)v(r) - S(t)\iota_n v_n(r)\| + \|S(t)\iota_n v_n(r) - \iota_n S_n(t) v_n(r)\| \\ &\quad + \Big\| \int_r^t \Big(S(t-s)\iota_n - \iota_n S_n(t-s) \Big) F_n(u_n(s))\, ds \Big\| \\ &\quad + \Big\| \int_r^t S(t-s) \Big(\iota_n F_n(u_n(s)) - F(\iota_n u_n(s)) \Big)\, ds \Big\| \\ &\quad + \Big\| \int_r^t S(t-s) \Big(F(\iota_n u_n(s)) - F(u(s)) \Big)\, ds \Big\| \, .\end{aligned}$$

Assume for the moment that t is furthermore sufficiently small so that $t \leq \bar{T}_R$, where

$$\bar{T}_R = T_R \wedge \inf\{t \geq 0 \,:\, \|\iota_n v_n(t)\| \geq KR\} \, ,$$

for a constant K that is yet to be determined.

Using our assumptions on each term then yields the (pathwise) bound

$$\begin{aligned}\|v(t) - \iota_n v_n(t)\| &\leq C\|v(r) - \iota_n v_n(r)\| + C(1 \wedge t^{-\gamma_2} n^{-\gamma_3})\|\iota_n v_n(r)\| + C n^{-\gamma_0} \\ &\quad + C_R \int_0^t (t-s)^{-\delta} (1 \wedge (t-s)^{-\gamma_2} n^{-\gamma_3})\, ds \\ &\quad + C_R n^{-\gamma_4} + C_R \int_0^t (t-s)^{-\delta} \|u(s) - \iota_n u_n(s)\|\, ds \, ,\end{aligned} \tag{8.16}$$

where the symbol C_R denotes as before a generic constant depending on R. Note that in the particular case $r = 0$ when $v_n(r) = \pi_n v(r)$, the first term vanishes and the term $\|\iota_n v_n(r)\|$ should be replaced by $\|U_0\|$ in the second term. Note furthermore that, by (8.11), the term $\|u(s) - \iota_n u_n(s)\|$ on the last line can be replaced by $\|v(s) - \iota_n v_n(s)\|$ without any change, except possibly for the values of the constants C_R.

Before we proceed, we also remark that the term on the second line is bounded by $C n^{-\gamma_{23}}$ with

$$\gamma_{23} = \gamma_3 \Big(\frac{1-\delta}{\gamma_2} \wedge 1 \Big) \, ,$$

which can be seen by splitting the integral into two contributions, according to which term dominates in the bound (8.12).

With $\alpha = (1-\delta) \wedge \gamma_2$ as in the statement of the theorem, we now set

$$\|u\|_{\alpha;r,t} \stackrel{\text{def}}{=} \sup_{s \in [r,t]} (s-r)^\alpha \|u(s)\| \, .$$

With this notation at hand, we can take the supremum over t on both sides of (8.16), so that

$$\|v - \iota_n v_n\|_{\alpha;r,t} \leq C\|v(r) - \iota_n v_n(r)\| + C_R n^{-\gamma} + C_R t^{1-\delta} \|v - \iota_n v_n\|_{\alpha;r,t} \, .$$

Here, we used the fact that $\sup_{s \leq t} t^\alpha (1 \wedge t^{-\gamma_2} n^{-\gamma_3}) \leq n^{-\gamma_{23}}$, which can easily be seen by an explicit

calculation. Similarly, if we do not multiply by the weight function t^α, we obtain

$$\|v - \iota_n v_n\|_{0;t} \le C(\|v(r)\| + \|\iota_n v_n(r)\|) + C_R n^{-\gamma} + C_R t^{1-\delta}\|v - \iota_n v_n\|_{0;t} \,. \tag{8.17}$$

Note that, unlike the previous bound, this bound does *not* become small if $v(r) - \iota_n v_n(r)$ is small. This is due to the fact that the bound (8.12) is badly behaved near $t = 0$. However, the bound (8.19) will turn out to be useful in the sequel since it provides a uniform bound on the approximate solution near $t = 0$.

We now choose t sufficiently small so that $C_R t^{1-\delta} \le \frac{1}{2}$, which then finally yields the bounds

$$\|u - \iota_n u_n\|_{\alpha;t} \le C_R n^{-\gamma} \,, \tag{8.18}$$

if $\tilde{U}_0 = \pi_n U_0$ (recall the remark after (8.16)) and

$$\|u - \iota_n u_n\|_{\alpha;t} \le \bar{C}\|U_0 - \iota_n \tilde{U}_0\| + C_R n^{-\gamma} \,,$$

otherwise, for some constant $\bar{C}$. We furthermore have the uniform bound

$$\|u - \iota_n u_n\|_{0;t} \le \bar{C}(\|U_0\| + \|\iota_n \tilde{U}_0\|) + C_R n^{-\gamma} \,, \tag{8.19}$$

where we can use the same constant $\bar{C}$ as in the previous bound. Denote by τ the length of the time interval such that these bounds hold. Note that we still haven't fixed K and that the constants C_R appearing in these bounds do depend on the precise value of K. The constant $\bar{C}$ however only depends on (8.12) and is independent of R and K.

We now make the choice $K = 4\bar{C} + 1$, which then fixes the values of the constants C_R appearing in the previous bounds. With this choice of K, we see that whenever we start with initial conditions U_0 and $\tilde{U}_0$ such that $\|U_0\| \le R$ and $\|\iota_n \tilde{U}_0\| \le R$, we have $\bar{T}_R \ge \tau \wedge T_R$ for a sufficiently large value of n. Indeed, assume by contradiction that this is not the case. We can then apply the bound (8.19) with $t = \bar{T}_R$, thus yielding the bound

$$\|\iota_n u_n(T_R)\| \le R + 2R\bar{C} + C_R n^{-\gamma} \,,$$

which is indeed a contradiction to the fact that, if $\bar{T}_R < T_R$, $\|\iota_n u_n(T_R)\| \ge KR$ by definition.

We now have all the ingredients ready for the final run. The argument goes as follows: by the previous argument, the bound (8.18) holds for $t \le \tau \wedge T_R$. In particular, for any $\kappa > 0$, we can choose n large enough so that

$$\sup_{s \in [\tau/2, \tau]} \|u(s) - \iota_n u_n(s)\| \mathbf{1}_{s \le T_R} \le \kappa \,. \tag{8.20}$$

Similarly, provided that $\|\iota_n \tilde{U}_0\| \le R$, we have

$$\sup_{s \in [\tau/2, \tau]} \|u(s) - \iota_n u_n(s)\| \mathbf{1}_{s \le T_R} \le \kappa + \bar{C}\|U_0 - \iota_n \tilde{U}_0\| \,. \tag{8.21}$$

Let now N be the smallest integer such that $N\tau/2 \ge T$ and choose n large enough so that the bounds (8.20) and (8.21) hold for a sufficiently small value of κ so that $\kappa(1 + \bar{C})^N \le 1$. With this choice, we can now iterate our bounds in the following way. □

It is sometimes useful to obtain convergence results in expectation, rather than in probability. At the level of generality of Theorem 8.6, it is completely unrealistic to expect any convergence in expectation. Indeed, it may happen that the approximate equation explodes before time T_R with finite probability, which would immediately make any moment infinite.

However, it is not too difficult to show that this scenario is the only obstruction. Indeed, if, besides the assumptions of Theorem 8.6, we assume that both the limiting equation and the approximations satisfy a uniform bound on moments of all orders, then we can strengthen our conclusion significantly.

Theorem 8.8 ....

Proof. .... □

8.1.2 Finite elements approximation

8.1.3 Finite differences approximation

8.2 Temporal approximation

Explain why Euler-Maruyama is useless, just like any explicit method.

8.2.1 The θ-method

We consider approximations of the type

$$u_{n+1} = u_n + A(\theta u_n + (1-\theta)u_{n+1})\,h + F(u_n)\,h + Q\delta W_n\ .$$

Advantages: easy to implement. In the case $\theta = \frac{1}{2}$ and $F = 0$, the invariant measure of the approximate dynamic is equal to that of the original dynamic for every h.

8.2.2 Exponential integrators

We consider approximations of the type

$$u_{n+1} = u_n + e^{Ah}u_n + F(u_n)\,h + Q\delta W_n\ .$$

Advantages: In the case $F = 0$, the law is equal to the law of the exact solution, subsampled at multiples of h.

Disadvantages: it requires the computation of e^{Ah} which can be expensive in both time and storage.

8.3 What can go wrong?

As we have seen many times throughout these notes, the main difference between solutions to deterministic PDEs and solutions to SPDEs is that the latter can show very low spatial and temporal regularity. This can have far-reaching consequences for numerical approximations. As a general rule, when implementing a numerical approximation for the nonlinear term of a stochastic PDE, one should always ask oneself if this term remains meaningful when applied to a function that has the kind of regularity that one expects from the solutions to the SPDE in question. Very often, this would result in elements taking value in Sobolev spaces with negative index. In this case, one should check whether $\langle \ell, F_n(\pi x)\rangle \to \langle \ell, F(x)\rangle$ for sufficiently "regular" functionals ℓ.

This section is devoted to a case study that illustrates that relatively straightforward looking numerical approximations to SPDEs with rough solutions can go dramatically wrong if these guidelines are not respected. We consider the stochastic Burgers' equation

$$du = \nu\partial_x^2 u\,dt + u\,\partial_x u\,dt + dW(t)\ , \tag{8.22}$$

where W is a cylindrical process on $\mathcal{H} = L^2([0,2\pi])$ and ∂_x^2 denotes the generator of the heat semigroup on $L^2([0,2\pi])$ with periodic boundary conditions. If we view (8.22) as a process in $\mathcal{C}(S^1,\mathbf{R})$, the space of continuous functions on the unit circle, then it is relatively straightforward to see that (8.22) has a unique global solution. The existence of a unique local solution follows indeed from Theorem 6.5 with $\alpha = \beta = \gamma = 0$ and $\delta = 1/2$. The fact that this solution is global then follows by an argument similar to those performed previously for the stochastic reaction-diffusion and Navier-Stokes equations. One good candidate for a Lyapunov function in the case of the stochastic Burgers' equation considered here is the supremum norm of the solution, since the deterministic equation yields a contraction for this norm.

Consider now the nonlinearity $F(u) = u\,\partial_x u$ to this equation on $\mathcal{H} = L^2([0,2\pi])$. If we identify the endpoints of the interval $[0,2\pi]$, working in effect with periodic boundary conditions, then this

nonlinearity makes sense even for functions u that are not differentiable but only belong to L^2, since one can set

$$\langle \varphi, F(u) \rangle \stackrel{\text{def}}{=} -\frac{1}{2}\langle \varphi', u^2 \rangle \ .$$

This expression makes sense for every function $\varphi \in \mathcal{C}^1([0,\pi])$. In particular, this yields the *a priori* bound

$$|\langle \varphi, F(u) \rangle| \le \frac{1}{2}\|u\|^2 \|\varphi'\|_{L^\infty} \ . \tag{8.23}$$

On the other hand, it would appear at first sight quite natural to approximate F by setting

$$F_\varepsilon(u)(x) = u(x)\frac{u(x+\varepsilon) - u(x)}{\varepsilon} \ . \tag{8.24}$$

While this approximation is good if u is sufficiently regular, it can be very bad when u gets rougher. In particular, it is very far from satisfying a bound of the type (8.23). Assuming that π/ε is an integer, this can easily be seen with the example $\varphi = 1$ and $u = \sin(\pi x/\varepsilon)$. In this case, we have $\langle \varphi, F(u) \rangle = 0$, but $\langle \varphi, F_\varepsilon(u) \rangle = -1$. This example is of course slightly artificial since we expect our solutions to be more regular than that. More precisely, one would expect solutions to (8.22) to have the same regularity as solutions to the stochastic heat equation, which are Hölder continuous for every exponent smaller than $\frac{1}{2}$. We are going to see in a moment that even for u's with this level of regularity, one expects (8.24) to be a very bad approximation for F.

A much more suitable approximation is given by

$$\tilde{F}_\varepsilon(u)(x) = \frac{u^2(x+\varepsilon) - u^2(x)}{2\varepsilon} \ ,$$

which relies on the fact that F can be written as $F(u) = \frac{1}{2}\partial_x(u^2)$. In this case, a simple change of variables yields the identity

$$\langle \varphi, \tilde{F}_\varepsilon(u) \rangle = \int_0^{2\pi} \frac{\varphi(x-\varepsilon) - \varphi(x)}{2\varepsilon} u^2(x)\, dx \ ,$$

so that the bound (8.23) still holds with F replaced by $\tilde{F}_\varepsilon$. Let us now compare $\tilde{F}_\varepsilon$ to F_ε. It is easy to check that one has the identity

$$\tilde{F}_\varepsilon(u)(x) - F_\varepsilon(u)(x) = \frac{\big(u(x+\varepsilon) - u(x)\big)^2}{2\varepsilon} \ .$$

It follows that $\tilde{F}_\varepsilon$ and F_ε are close to each other for small values of ε only if u is Hölder continuous with exponent greater than $\frac{1}{2}$. On the other hand, one expects solutions to (8.22) to be Hölder continuous only for exponents *less* than $\frac{1}{2}$ so that (8.24) may well yield an approximation to the wrong equation in the limit $\varepsilon \to 0$ (if indeed it converges at all).

The rest of this section is devoted to an argument showing that if we consider the stochastic Burgers' equation, which has a nonlinearity $u\partial_x u$, then a numerical scheme based on the second approximation is expected to give a good approximation to the solution, while a numerical scheme based on the first approximation will converge, but to the wrong solution!

Divide the interval $[0, 2\pi]$ into $2N$ identical subintervals of size $\Delta x = \frac{\pi}{N}$ and consider the following finite difference scheme:

$$\begin{aligned} dv(x) &= \frac{\nu}{\Delta x^2}(v(x+\Delta x) + v(x-\Delta x) - 2v(x))\, dt + \frac{v(x)}{\Delta x}(v(x+\Delta x) - v(x))\, dt + \frac{dW_x(t)}{\sqrt{\Delta x}} \\ &\stackrel{\text{def}}{=} \nu(\Delta_N v)(x)\, dt + (F_N(v))(x)\, dt + \frac{dW_x(t)}{\sqrt{\Delta x}} \ . \end{aligned} \tag{8.25}$$

where the W_x's are independent standard Wiener processes and x takes values in $[0, 2\pi] \cap \mathbf{N}\Delta x$ with

periodic boundary conditions (that is we impose $v(x+2\pi) = v(x)$). It follows from ???? that the solutions to the linearised equation do converge to the solutions to the stochastic heat equation, and so it seems reasonable to expect that the above scheme approximates (8.22). However, one has

Proposition 8.9 *As $N \to \infty$, the solution v to (8.25) converges in law to the solution u to*

$$du = \nu\partial_x^2 u\,dt + u\,\partial_x u\,dt - \frac{dt}{4\nu} + dW(t)\;.$$

In particular, it does not *converge to the solution to (8.22).*

Proof (idea). A complete proof is beyond the scope of these notes, since it is rather long and technical. We therefore only reproduce its main steps here. The idea is to decompose the solution into a regular part and a rapidly oscillating part by projecting onto spectral subspaces of the linear operator Δ_N. Since the nonlinearity is a lower order perturbation (in the sense that it contains only one spatial derivative, while the linear part contains two derivatives), one then expects the rapidly oscillating part to be well-described by the solution to the linearised equation.

The approximation is then well-behaved on the regular part. However, it turns out that through the nonlinearity, the rapidly oscillating part feeds a constant term back into the regular part of the equation. Let us show how this mechanism works in slightly more detail. It is straightforward to check that the eigenvectors for the discretised Laplacian Δ_N are given by e^{inx} with $n = -N+1, \dots, N$, with eigenvalues

$$\lambda_n = \frac{2}{\Delta x^2}(\cos n\Delta x - 1) = -\Big(\frac{2}{\Delta x}\sin(\frac{n\Delta x}{2})\Big)^2 \stackrel{\text{def}}{=} -\eta_n^2\;.$$

(Note that for fixed n and small Δx, one has indeed $\lambda n \approx -n^2$ as one would expect.) It then follows that the solution to the disctretised linearised equation is given by

$$\tilde{v}(x,t) = \sum_{n\neq 0} \frac{e^{inx}\xi_n(t)}{2\sqrt{\nu\pi}i\eta_n}\;,$$

where the ξ_n's are an i.i.d. sequence of Ornstein-Uhlenbeck processes with variance 1 and time-constants η_n^2. Since the nonlinearity is of lower order compared to the linear part of the equation, one expects the high-frequency components of v to be well approximated by the solution $\tilde{v}$ to the linearised equation. We therefore introduce a cutoff scale N^α and we write $v \approx \bar{v}^{(\alpha)} + \tilde{v}^{(\alpha)}$, where $\tilde{v}^{(\alpha)}$ is defined by

$$\tilde{v}^{(\alpha)}(x,t) = \sum_{|n|\in[N^\alpha,N]} \frac{e^{inx}\xi_n(t)}{2\sqrt{\nu\pi}i\eta_n}\;,$$

and $\bar{v}^{(\alpha)}$ is a linear combination of e^{inx} with $|n| < N^\alpha$. If α is large enough, then it should be possible to obtain a good approximation to the solution of (8.25) in this way. We then argue that if α is small enough, then the choice of space discretisation is not very important for $\bar{v}^{(\alpha)}$, since it is always a very smooth function. Furthermore, products of the form $\bar{v}^{(\alpha)}\tilde{v}^{(\alpha)}$ have only high-frequency components. Denoting by $W_L^{(\alpha)}$ the projection of the solution to the linearised equation onto the subspace spanned by e^{inx} with $|n| < N^\alpha$, this suggests that the smooth component $\bar{v}^{(\alpha)}$ is an approximate solution to

$$\bar{v}^{(\alpha)}(t) \approx S_N(t)v_0 + W_L^{(\alpha)}(t) + \int_0^t S(t-s)\bar{v}^{(\alpha)}(t)\partial_x\bar{v}^{(\alpha)}(t)\,ds + \int_0^t S_N(t-s)F_\varepsilon(\tilde{v}^{(\alpha)}(t))\,ds\;. \tag{8.26}$$

It remains to argue that the factor $F_N(\tilde{v}^{(\alpha)}(t))$ appearing in the last term can be approximately replaced by a constant. The discrete derivative $\tilde{v}$ of $\tilde{v}$ is given by

$$\frac{\tilde{v}^{(\alpha)}(x+\Delta x) - \tilde{v}^{(\alpha)}(x)}{\Delta x} = \sum_{|n|\in[N^\alpha,N]} \frac{e^{inx}\xi_n(t)}{2\sqrt{\nu\pi}i\eta_n}\frac{e^{in\Delta x}-1}{\Delta x}\;.$$

The expectation of $F_N(\tilde{v}^{(\alpha)}(t))$ is therefore given by

$$\mathbf{E}F_N(\tilde{v}^{(\alpha)}(t)) = \sum_{n=N^\alpha}^{N} \frac{1}{2\pi\nu\eta_n^2} \frac{\cos n\Delta x - 1}{\Delta x} = -\sum_{n=N^\alpha}^{N} \frac{\Delta x}{4\pi\nu} \approx -\frac{1}{4\nu} ,$$

independently of x. Compare this with the fact that the "true" derivative $\partial_x \tilde{v}^{(\alpha)}$ of $\tilde{v}^{(\alpha)}$ is given by

$$\partial_x \tilde{v}^{(\alpha)}(x) = \sum_{n\neq 0} \frac{n e^{inx} \xi_n(t)}{2\sqrt{\nu\pi}\eta_n} ,$$

so that it is possible to check that $\mathbf{E}(\tilde{v}^{(\alpha)}(x)\partial_x \tilde{v}^{(\alpha)}(x)) = 0$, as one would expect for a total derivative. The argument is completed by arguing that:

- The Fourier coefficients of $F_N(\tilde{v}^{(\alpha)}(t))$ are approximately of magnitude $\frac{1}{\sqrt{N}}$, because they arise as a sum of independent terms of mean 0.
- The Fourier modes with high wavenumber are suppressed by the action of the linear semigroup S_N.

This shows that the term $F_N(\tilde{v}^{(\alpha)}(t))$ in (8.26) can be replaced by its expectation $-\frac{1}{4\nu}$, so that one expects the numerical approximation to converge, as $\Delta x \to 0$, to the solution to

$$du = \nu\partial_x^2 u\, dt + u\partial_x u\, dt - \frac{1}{4\nu}\, dt + dW(t) .$$

The reader interested in a rigorous proof of this fact, as well as in related questions, is referred to the articles [Hai12, HV11, Hai11]. □

Exercise 8.10 Consider the finite difference scheme

$$dv(x) = \frac{\nu}{\Delta x^2}(v(x+\Delta x)+v(x-\Delta x)-2v(x))\, dt + \frac{g(v(x))}{\Delta x}(v(x+\Delta x)-v(x))\, dt + \frac{dW_x(t)}{\sqrt{\Delta x}} ,$$

for a "nice" function g. (Say smooth with bounded derivatives of all orders.) Modify the argument given above to argue that in the limit $N \to \infty$, its solutions should converge to the solutions to the SPDE

$$du = \nu\partial_x^2 u\, dt + g(u)\, \partial_x u\, dt - \frac{g'(u)}{4\nu}\, dt + dW(t) .$$

Hint: Convince yourself first that one can decompose the solution as $u = \bar{u} + \tilde{u}$ with $\bar{u}$ a component that is large, but smooth and $\tilde{u}$ an oscillatory part which is small in the supremum norm, but large in the H^1 norm. For large N, one then has $g(u) \approx g(\bar{u}) + g'(\bar{u})\tilde{u}$. This second term will then produce a correction term that doesn't vanish in the limit in a similar way as before.

9

Ergodic Theory for Semilinear SPDEs

The aim of this section is to extend some of the results from Section 5.2 to the setting of semilinear SPDEs.

9.1 Structure of the set of invariant measures

Ergodic decomposition theorem as a consequence of Birkhoff's ergodic theorem.

9.2 Strong convergence results

Aim is to show how Theorem 5.28 can be applied to semilinear SPDEs.

9.2.1 Harris chains

Explain the theory of Harris chains. Introduce the strong Feller property. Show that SF implies continuity in total variation of TP's.

9.2.2 The Bismut-Elworthy-Li formula

Give proof of BEL formula and show how to use it to verify SF property. Do reaction-diffusion in detail, as in Sandra's paper. As an application, give a short proof of regularity of densities of finite-dimensional projections.

9.3 Weak convergence results

Introduce the asymptotic strong Feller property. Show that ASF + Topological irreducibility implies uniqueness of IM. Show how to check ASF in a couple of examples where noise acts on "dominating modes".

9.4 Spectral gap results

Explain what a spectral gap is. Maybe mention spectral mapping theorem to relate it to generator.

9.4.1 Spectral gaps in total variation distance

Basically show ho to apply Theorem 5.28 to nonlinear equations.

9.4.2 Spectral gaps in Wasserstein distance

Explain what spectral gap is. We consider norms of the type

$$\|\varphi\|_{V,W} = \sup_{u\in\mathcal{B}}\Big(\frac{|\varphi(u)|}{V(u)} + \frac{\|D\varphi(u)\|}{W(u)}\Big) \tag{9.1}$$

The conditions presented in this section are taken from [HM09], but they are strongly inspired by those in [HM08a, HMS09]. One of our aims is to be applicable to equations with conservative quadratic nonlinearities, as arising in many applications. The main ingredient is a gradient bound of the following type, which was shown in Section 9.3 to hold for a large class of stochastic PDEs:

Assumption 9.1 There exist continuous functions $U_1, U_2\colon \mathcal{B} \to \mathbf{R}_+$ such that, for every $\varepsilon > 0$ there is a constant $C_\varepsilon > 0$ such that the bound

$$\|D\mathcal{P}\varphi(x)\|^2 \le \varepsilon U_1^2(x)(\mathcal{P}\|D\varphi\|^2)(x) + C_\varepsilon U_2^2(x)(\mathcal{P}\varphi^2)(x)\;, \tag{9.2}$$

holds for every $x \in \mathcal{B}$ and every $\varphi \in \mathcal{C}_b^1(\mathcal{B})$.

Remark 9.2 In the case of the stochastic Navier-Stokes equations, this bound can be shown to hold with $U(x) = V(x) = \exp(\eta\|x\|^2)$ for arbitrarily small $\eta > 0$, with x the vorticity of the vector field and $\|x\|$ its L^2-norm.

The aim of this section is to show that if a Markov operator $\mathcal{P}$ satisfies Assumption 9.1 and has sufficiently good contraction properties, then it is possible to deduce a spectral gap in the norm (9.1) (for a suitable choice of V and W), provided that the transition probabilities satisfy a kind of topological irreducibility condition. This statement can be formulated precisely in the following way:

Theorem 9.3 *Let $\mathcal{P}$ be a Markov operator over a separable Banach space $\mathcal{B}$ mapping $\mathcal{C}_0^\infty(\mathcal{B})$ into $\mathcal{C}_{V,W}^1$ and satisfying Assumption 9.1. Suppose furthermore that there exist continuous functions $V, W\colon \mathcal{B} \to \mathbf{R}_+$ satisfying*

$$U_1^2\mathcal{P}W^2 + U_2^2\mathcal{P}V^2 \le CW^2\;, \qquad \mathcal{P}V \le \frac{1}{2}V + K\;, \tag{9.3}$$

for some constants C and K.

Finally, assume that there exists a point $x_\star \in \mathcal{B}$ such that, for every $\varepsilon > 0$ and every $C > 0$ there exists $\alpha > 0$ such that

$$\inf_{x\,:\,V(x)\le C} \mathcal{P}(x, B_\varepsilon(x_\star)) \ge \alpha\;. \tag{9.4}$$

Then, $\mathcal{P}_t$ has exactly one invariant probability measure $\mu_\star$. Furthermore, there exist constants C and $\gamma > 0$ such that the bound

$$\|\mathcal{P}^n\varphi\|_{1;V,W} \le Ce^{-\gamma n}\|\mathcal{P}^n\varphi\|_{1;V,W}\;, \tag{9.5}$$

holds for every $\varphi \in \mathcal{C}_{V,W}^1$. Finally, there exist constants $\delta > 0$ and $\beta > 0$ and $\varrho < 1$ such that the bound

$$\|\mathcal{P}\varphi\|_{1;1+\beta V,\delta^{-1}W} \le \varrho\|\mathcal{P}\varphi\|_{1;1+\beta V,\delta^{-1}W}\;, \tag{9.6}$$

holds for every $\varphi \in \mathcal{C}_{V,W}^1$.

Remark 9.4 If the second inequality in (9.3) holds with $\frac{1}{2}$ replaced by any other constant smaller than 1, then one can always reduce it to this case by considering an iterate of $\mathcal{P}$ instead of $\mathcal{P}$ itself.

Also, it will follow from the proof that we do not need (9.4) to hold for every $C > 0$, but only for some $C > 2K$.

Proof. An important role will be played by the distance function $d_{V,W}$ given by

$$d_{V,W}(x,y) = \varrho_W(x,y) \wedge (V(x) + V(y)) \; . \tag{9.7}$$

Here, for any positive continuous function W bounded away from 0, we defined the metric

$$\varrho_W(x,y) = \inf_{\gamma(0)=x;\gamma(1)=y} \int_0^1 W(x)\, \|\dot\gamma(s)\|\, ds \; ,$$

where the infimum runs over all smooth curves $\gamma\colon [0,1] \to \mathcal{B}$ with the prescribed boundary conditions. (Note that if $W \equiv 1$, then one simply has $\varrho_W(x,y) = \|x-y\|$.)

We now make use the following trick, which is reminiscent of the trick already used in the proof of Theorem 5.28. For $\delta > 0$ and $\beta > 0$, we introduce the distance

$$\hat d_{\beta,\delta}(x,y) = \delta^{-1}\varrho_W(x,y) \wedge (2 + \beta V(x) + \beta V(y)) \; .$$

This distance is of course equivalent to the distance $d_{V,W}$ introduced in (9.7), but it turns out that allowing the freedom of choosing both δ and β sufficiently small will considerably simplify the proofs. With this definition, a $\mathcal{C}^1$ function φ is Lipschitz continuous with Lipschitz constant 1 with respect to $\hat d_{\beta,\delta}$ if and only if one has

$$\|D\varphi(x)\| \le \delta^{-1} W(x) \; , \qquad |\varphi(x)| \le 1 + \beta V(x) \; . \tag{9.8}$$

(For the second inequality, one might have to add a suitable constant to φ.) Denoting by L the set of all such functions, we show that it is possible to choose δ and β in such a way that the bound

$$\hat d_{\beta,\delta}(\mathcal{P}(x,\cdot\,),\mathcal{P}(y,\cdot\,)) \le \alpha d_{\beta,\delta}(x,y) \; , \tag{9.9}$$

holds for some $\alpha < 1$ uniformly over all pairs $x, y \in \mathcal{B}$. Similarly to the proof of Theorem 5.28, we will now show (9.9) separately in three different cases and we use separately the three ingredients of the theorem in each of these cases.

The case $\boldsymbol{\varrho_W(x,y) \le \delta(2 + \beta V(x) + \beta V(y))}$. In this case, we make use of the gradient bound (9.2), together with the "super-Lyapunov" structure (9.3) to deduce that if φ satisfies (9.8), then for every $\varepsilon > 0$ there exists C_ε such that the bound

$$\|D\mathcal{P}\varphi(x)\| \le \varepsilon\delta^{-1} W(x) + C_\varepsilon W(x) \; ,$$

holds uniformly for all such φ and for all $\beta \le 1$, say. It follows that by first choosing $\varepsilon = \frac{1}{4}$ and then choosing δ small enough so that $C_\varepsilon \le 1/(4\delta)$, one has

$$\|D\mathcal{P}\varphi(x)\| \le \frac{1}{2\delta} W(x) \; ,$$

which immediately implies that

$$\begin{aligned}
\hat d_{\beta,\delta}(\mathcal{P}(x,\cdot\,),\mathcal{P}(y,\cdot\,)) &\le \sup_{\varphi\in \mathrm{L}} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \\
&\le \sup_{\varphi\in \mathrm{L}} \inf_\gamma \int_0^1 \|D\mathcal{P}\varphi(\gamma(s))\|\, |\dot\gamma(s)|\, ds \\
&\le \frac{1}{2\delta} \sup_\varphi \inf_\gamma \int_0^1 W(\gamma(s))\, |\dot\gamma(s)|\, ds \\
&\le \frac{1}{2\delta} \varrho_W(x,y) \le \frac{1}{2} \hat d_{\beta,\delta}(x,y) \; ,
\end{aligned}$$

as requested.

The case $\boldsymbol{\varrho_W(x,y) > \delta(2 + \beta V(x) + \beta V(y))}$ and $\boldsymbol{V(x) + V(y) \ge 4(K+2)}$. In this case, we simply make use of the fact that V is assumed to be a Lyapunov function. We have indeed

$$\hat d_{\beta,\delta}(\mathcal{P}(x,\cdot\,),\mathcal{P}(y,\cdot\,)) \le \sup_{\varphi\in \mathrm{L}} |\mathcal{P}\varphi(x) - \mathcal{P}\varphi(y)| \le \sup_{\varphi\in \mathrm{L}} (|\mathcal{P}\varphi(x)| + |\mathcal{P}\varphi(y)|)$$

$$\leq 2 + \beta \mathcal{P} V(x) + \beta \mathcal{P} V(y) \leq 2 + \frac{\beta}{2}(V(x) + V(y)) + \beta K$$
$$\leq (2 - 2\beta) + \frac{3\beta}{4}(V(x) + V(y)) \leq (1 - (\beta \wedge \tfrac{1}{4}))\hat{d}_{\beta,\delta}(x, y) \,,$$

which again yields a contraction, but with a strength that depends this time on the parameter β. Finally, we have

The case $\varrho_W(x, y) > \delta(2 + \beta V(x) + \beta V(y))$ and $V(x) + V(y) < 4(K + 2)$. In this case, we make use of our final assumption, namely (9.4). At this stage we assume that $\delta > 0$ is fixed, sufficiently small so that our first step goes through. We can then find some sufficiently small $\varepsilon > 0$ so that $\hat{d}_{\beta,\delta}(x_\star, y) \leq \frac{1}{2}$ for all $y \in B_\varepsilon(x_\star)$, uniformly over $\beta \leq 1$. In this case, we can decompose the Markov operator $\mathcal{P}$ into a combination $\mathcal{P} = \alpha\mathcal{P}_1 + (1 - \alpha)\mathcal{P}_2$ of Markov operators such that $\mathcal{P}_1(x, B_\varepsilon(x_\star)) = 1$ for every x such that $V(x) \leq 4(K + 2)$. We conclude that

$$\begin{aligned}\hat{d}_{\beta,\delta}(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) &\leq \alpha \hat{d}_{\beta,\delta}(\mathcal{P}_1(x, \cdot), \mathcal{P}_1(y, \cdot)) + (1 - \alpha)\hat{d}_{\beta,\delta}(\mathcal{P}_2(x, \cdot), \mathcal{P}_2(y, \cdot)) \\ &\leq \alpha + (1 - \alpha)(2 + \beta \mathcal{P}_2 V(x) + \beta \mathcal{P}_2 V(x)) \\ &\leq \alpha + 2(1 - \alpha) + \beta \mathcal{P} V(x) + \beta \mathcal{P} V(x) \\ &\leq 2 - \alpha + \frac{\beta}{2}(V(x) + V(y) + 2K) \leq 2 - \alpha + \beta(3K + 4) \,.\end{aligned}$$

We can now choose β sufficiently small so that this constant is strictly smaller than 2. Since on the other hand one has $\hat{d}_{\beta,\delta}(x, y) \geq 2$, the claim now follows. □

9.4.3 Application to numerical methods

If numerical method approximates dynamic in the same norm as SG, then one obtains stability of the IM under discretisation.

9.4.4 Application to Poisson equations

Can solve Poisson equation in space in which one has a SG result.

Index of Notations

Symbol	Meaning
$\mathcal{B}$	Arbitrary separable Banach space
$\mathcal{B}^*$	Dual space of $\mathcal{B}$
$\mathcal{B}^\dagger$	Semigroup dual to $\mathcal{B}$ with respect to a given semigroup S
$\mathcal{B}_\alpha$	Interpolation space of index α for an analytic semigroup on $\mathcal{B}$
$\mathscr{B}_b$	Space of bounded Borel measurable functions
$\mathcal{C}$	Space of all continuous functions
$\mathcal{C}^\alpha$	Space of Hölder continuous functions with Hölder exponent α
$\mathcal{C}_b$	Space of bounded continuous functions
C_μ	Covariance operator of μ as a bilinear map $\mathcal{B}^* \times \mathcal{B}^* \to \mathbf{R}$
$\hat{C}_\mu$	Covariance operator of μ as a linear map $\mathcal{B}^* \to \mathcal{B}$
δ_x	Dirac measure at x
$d(\mu, \nu)$	Wasserstein-1 distance between μ and ν with respect to the metric d
$\mathcal{D}_\mu$	Density of the measure μ with respect to a given reference measure
$\mathcal{D}(A)$	Domain of the linear operator A
$f^\sharp \mu$	Push-forward of the measure μ under the map f
$\mathcal{H}$	Arbitrary separable Hilbert space
H^s	Fractional Sobolev space of index s
$\mathcal{H}_\mu$	Cameron-Martin space of the Gaussian measure μ
$\mathcal{H}_\alpha$	Interpolation space of index α for an analytic semigroup on $\mathcal{H}$
$J_{s,t}$	Jacobian of the solution to an SPDE between times s and t
$\mathcal{L}_2(\mathcal{H}, \mathcal{K})$	Space of Hilbert-Schmidt operators from $\mathcal{H}$ to $\mathcal{K}$
L	Generator of a strongly continuous or analytic semigroup S
L^*	Adjoint of L
$L^\dagger$	Generator of the adjoint semigroup S^*
L^p	Space of functions with integrable pth power
$\hat{\mu}$	Fourier transform of the measure μ
$\mathscr{M}(\mathcal{X})$	Space of finite signed measures on $\mathcal{X}$
$\mathcal{N}(0, 1)$	Normal distribution on $\mathbf{R}$ with mean 0 and variance 1
$\mathscr{P}(\mathcal{X})$	Space of probability measures on $\mathcal{X}$
$\mathcal{P}_t$	Markov semigroup acting on observables
Q_t	Covariance of the solution to a linear SPDE at time t
$S(t)$	Semigroup generated by L
$\mathbf{T}^d$	d-dimensional torus.
$\mathcal{X}$	Arbitrary Polish space.
$\|\cdot\|_\alpha$	Norm of the interpolation space of index α
$\|\cdot\|_{\text{HS}}$	Hilbert-Schmidt norm of a linear operator
$\|\cdot\|_{\text{TV}}$	Total variation norm of a signed measure

References

[BHP05] D. BLÖMKER, M. HAIRER, and G. A. PAVLIOTIS. Modulation equations: stochastic bifurcation in large domains. *Comm. Math. Phys.* **258**, no. 2, (2005), 479–512.

[Bil68] P. BILLINGSLEY. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

[BM07] Y. BAKHTIN and J. C. MATTINGLY. Malliavin calculus for infinite-dimensional systems with additive noise. *J. Funct. Anal.* **249**, no. 2, (2007), 307–353.

[Boc33] S. BOCHNER. Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind. *Fund. Math.* **20**, (1933), 262–276.

[Bog98] V. I. BOGACHEV. *Gaussian measures*, vol. 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[Bog07] V. I. BOGACHEV. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.

[Bor75] C. BORELL. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30**, no. 2, (1975), 207–216.

[BS27] S. BANACH and H. STEINHAUS. Sur le principe de la condensation de singularités. *Fundamenta Mathematicae* **9**, (1927), 50–61.

[Dav80] E. B. DAVIES. *One-parameter semigroups*, vol. 15 of *London Mathematical Society Monographs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1980.

[DJT95] J. DIESTEL, H. JARCHOW, and A. TONGE. *Absolutely summing operators*, vol. 43 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[DL92] R. DAUTRAY and J.-L. LIONS. *Mathematical analysis and numerical methods for science and technology. Vol. 5*. Springer-Verlag, Berlin, 1992. Evolution problems. I, With the collaboration of Michel Artola, Michel Cessenat and Hélène Lanchon, Translated from the French by Alan Craig.

[DPKZ87] G. DA PRATO, S. KWAPIEŃ, and J. ZABCZYK. Regularity of solutions of linear stochastic equations in Hilbert spaces. *Stochastics* **23**, no. 1, (1987), 1–23.

[DPZ92a] G. DA PRATO and J. ZABCZYK. A note on stochastic convolution. *Stochastic Anal. Appl.* **10**, no. 2, (1992), 143–153.

[DPZ92b] G. DA PRATO and J. ZABCZYK. *Stochastic equations in infinite dimensions*, vol. 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[DPZ96] G. DA PRATO and J. ZABCZYK. *Ergodicity for infinite-dimensional systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[Fer70] X. FERNIQUE. Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Sér. A-B* **270**, (1970), A1698–A1699.

[FHH+01] M. FABIAN, P. HABALA, P. HÁJEK, V. MONTESINOS SANTALUCÍA, J. PELANT, and V. ZIZLER. *Functional analysis and infinite-dimensional geometry*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 8. Springer-Verlag, New York, 2001.

[Gar02] R. J. GARDNER. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)* **39**, no. 3, (2002), 355–405 (electronic).

[GIP15] M. GUBINELLI, P. IMKELLER, and N. PERKOWSKI. Paracontrolled distributions and singular PDEs. *Forum of Mathematics, Pi* **3**(2015). doi:10.1017/fmp.2015.2.

[Hai11] M. HAIRER. Rough stochastic PDEs. *Comm. Pure Appl. Math.* **64**, no. 11, (2011), 1547–1585. doi:10.1002/cpa.20383.

[Hai12] M. HAIRER. Singular perturbations to semilinear stochastic heat equations. *Probab. Theory Related Fields* **152**, no. 1-2, (2012), 265–297. doi:10.1007/s00440-010-0322-7.

[Hai14] M. HAIRER. A theory of regularity structures. *Inventiones mathematicae* **198**, no. 2, (2014), 269–504. arXiv:1303.5113. doi:10.1007/s00222-014-0505-4.

[Hai18] M. HAIRER. Renormalisation of parabolic stochastic PDEs. *Jpn. J. Math.* **13**, no. 2, (2018), 187–233. arXiv:1803.03044. doi:10.1007/s11537-018-1742-x.

[Hil53] T. H. HILDEBRANDT. Integration in abstract spaces. *Bull. Amer. Math. Soc.* **59**, (1953), 111–139.

[HM08a] M. HAIRER and J. C. MATTINGLY. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36**, no. 6, (2008), 2050–2091.

[HM08b] M. HAIRER and J. C. MATTINGLY. Yet another look at Harris' ergodic theorem for Markov chains, 2008. Preprint.

[HM09] M. HAIRER and A. J. MAJDA. A simple framework to justify linear response theory, 2009. Preprint.

[HMS09] M. HAIRER, J. C. MATTINGLY, and M. SCHEUTZOW. A weak form of Harris's theorem with applications to stochastic delay equations, 2009. Preprint.

[HV11] M. HAIRER and J. VOSS. Approximations to the stochastic Burgers equation. *J. Nonlinear Sci.* **21**, no. 6, (2011), 897–920. doi:10.1007/s00332-011-9104-3.

[Kat82] M. P. KATS. Continuity of universally measurable linear mappings. *Sibirsk. Mat. Zh.* **23**, no. 3, (1982), 83–90, 221.

[Law04] G. F. LAWLER. An introduction to the stochastic Loewner evolution. In *Random walks and geometry*, 261–293. Walter de Gruyter GmbH & Co. KG, Berlin, 2004.

[Led96] M. LEDOUX. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, vol. 1648 of *Lecture Notes in Math.*, 165–294. Springer, Berlin, 1996.

[LT91] M. LEDOUX and M. TALAGRAND. *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[Lun95] A. LUNARDI. *Analytic semigroups and optimal regularity in parabolic problems*. Progress in Nonlinear Differential Equations and their Applications, 16. Birkhäuser Verlag, Basel, 1995.

[MP06] J. C. MATTINGLY and É. PARDOUX. Malliavin calculus for the stochastic 2D Navier-Stokes equation. *Comm. Pure Appl. Math.* **59**, no. 12, (2006), 1742–1790.

[MP11] L. MYTNIK and E. PERKINS. Pathwise uniqueness for stochastic heat equations with Hölder continuous coefficients: the white noise case. *Probab. Theory Related Fields* **149**, no. 1-2, (2011), 1–96. doi:10.1007/s00440-009-0241-7.

[MPS06] L. MYTNIK, E. PERKINS, and A. STURM. On pathwise uniqueness for stochastic heat equations with non-Lipschitz coefficients. *Ann. Probab.* **34**, no. 5, (2006), 1910–1959.

[MT93] S. P. MEYN and R. L. TWEEDIE. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[Paz83] A. PAZY. *Semigroups of linear operators and applications to partial differential equations*, vol. 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[Phi55] R. S. PHILLIPS. The adjoint semi-group. *Pacific J. Math.* **5**, (1955), 269–283.

[PR07] C. PRÉVÔT and M. RÖCKNER. *A concise course on stochastic partial differential equations*, vol. 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[Pro56] Y. V. PROHOROV. Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.* **1**, (1956), 177–238.

[PZ07] S. PESZAT and J. ZABCZYK. *Stochastic partial differential equations with Lévy noise*, vol. 113 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2007. An evolution equation approach.

[RR04] M. RENARDY and R. C. ROGERS. *An introduction to partial differential equations*, vol. 13 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second ed., 2004.

[RS80] M. REED and B. SIMON. *Methods of Modern Mathematical Physics I–IV*. Academic Press, San Diego, California, 1980.

[Rud91] W. RUDIN. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second ed., 1991.

[RW94] L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov processes, and martingales. Vol. 1*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, second ed., 1994. Foundations.

[RY94] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second ed., 1994.

[SC74] V. N. SUDAKOV and B. S. CIREL′SON. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **41**, (1974), 14–24, 165. Problems in the theory of probability distributions, II.

[Sch48] E. SCHMIDT. Die Brunn-Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. I. *Math. Nachr.* **1**, (1948), 81–157.

[Sch66] L. SCHWARTZ. Sur le théorème du graphe fermé. *C. R. Acad. Sci. Paris Sér. A-B* **263**, (1966), A602–A605.

[She07] S. SHEFFIELD. Gaussian free fields for mathematicians. *Probab. Theory Related Fields* **139**, no. 3-4, (2007), 521–541.

[SS05] M. SANZ-SOLÉ. *Malliavin calculus*. Fundamental Sciences. EPFL Press, Lausanne, 2005. With applications to stochastic partial differential equations.

[SS06] O. SCHRAMM and S. SHEFFIELD. Contour lines of the two-dimensional discrete Gaussian free field, 2006. Preprint.

[SV79] D. W. STROOCK and S. R. S. VARADHAN. *Multidimensional diffusion processes*, vol. 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.

[Tri83] H. TRIEBEL. *Theory of function spaces*, vol. 38 of *Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its Applications in Physics and Technology]*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1983.

[Tri92] H. TRIEBEL. *Theory of function spaces. II*, vol. 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992.

[Tri06] H. TRIEBEL. *Theory of function spaces. III*, vol. 100 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 2006.

[Vil03] C. VILLANI. *Topics in optimal transportation*, vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[Vil09] C. VILLANI. *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[WY71a] S. WATANABE and T. YAMADA. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11**, (1971), 155–167.

[WY71b] S. WATANABE and T. YAMADA. On the uniqueness of solutions of stochastic differential equations. II. *J. Math. Kyoto Univ.* **11**, (1971), 553–563.

[Yos95] K. YOSIDA. *Functional analysis*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.

[Ziz03] V. ZIZLER. Nonseparable Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. 2*, 1743–1816. North-Holland, Amsterdam, 2003.