Stochastic PDEs and their asymptotic behaviour

Alexandre Boritchev

March 4, 2021

Contents

Notation

Many of the objects used here are very classical/well-known for most of you...but maybe not for everybody and this is normal. Therefore, the notes are reasonably self-contained; see the sections below for the definitions of relevant objects.
Although this course is mostly intended for students with an analysis and PDE background, it is very important that it is followable if your prior knowledge is mainly probabilistic...but some notation may seem strange. It is similar for some notation which is standard in probability if you are not used to it.

Basics

We denote by $\mathbb{N}$ the set of natural numbers $\{1, 2, \dots\}$ (the "non-French" notation), by $\mathbb{N}_0$ – the set $\mathbb{N} \cup \{0\}$ (the "French $\mathbb{N}$") and by $\mathbb{R}^+$ – the set of non-negative real numbers.
The expressions l.h.s., r.h.s. stand, respectively, for "left-hand side" and "right-hand side". The abbreviations r.v., a.e. and a.s. stand, respectively, for "random variable", "almost everywhere" and "almost surely". Finally, i.i.d. means "independent identically distributed".
Subindices t, x stand for (partial) derivatives. For example, something called a_t does (also) depend on the parameter t, but is considered as the (sometimes weak) derivative of a function a with respect to t.
The m-th derivative of a function f of a scalar argument is denoted $f^{(m)}$.
The law of a r.v. ξ is denoted by $\mathcal{D}(\xi)$. If ξ is a Gaussian r.v. with mean value m and dispersion σ^2, we write $\xi \sim N(m, \sigma^2)$.
All random processes which we consider in this text have continuous trajectories. In the notation of random variables and processes we often drop the random parameter ω.
For the Wiener process, w^ω is the trajectory (i.e. element of the set of continuous functions) corresponding to an element ω of the probability space Ω on which the process is defined.
For a metric space M, $C_b(M)$ indicates the space of bounded real-valued

continuous functions on M, equipped with the sup-norm.
A mapping between metric spaces whose Lipschitz constant is bounded by K is called K–**Lipschitz.** A map between Banach spaces B_1 and B_2 is said to be **locally Lipschitz** if its restriction to any ball in B_1 is Lipschitz.

Measure theory

For $0 < T \leq \infty$ we provide the interval $[0, T]$ with the Borel sigma-algebra $\mathcal{B}$ and the Lebesgue measure dt. Then for $1 \leq p \leq \infty$ and a **separable** Banach space B we denote by $L_p(0, T; B)$ the L_p–space of measurable mappings $([0, T], \mathcal{L}, dt) \to (B, \mathcal{B}_B)$. This is a Banach space. When for a function $u(t, x)$, $t \geq 0$, $x \in S^1$, we consider $|u(t, \cdot)|_\infty$ (recall that the space $L_\infty(S^1)$ is not separable), we regard this object as a function of t and not as an element of some space $L_p([0, T], L_\infty)$.
For a measurable space (which will always be Polish in this book) $(X, \mathcal{F})$ we denote by $\mathcal{P}(X, \mathcal{F})$ the space of probability measures on $(X, \mathcal{F})$. A Polish (i.e. metric separable) space M is equipped with the Borel sigma–algebra $\mathcal{B} = \mathcal{B}_M$; we abbreviate the **space of probability measures** $\mathcal{P}(M, \mathcal{B}_M)$ to $\mathcal{P}(M)$.
For a measure μ on a measurable space and an integrable function f the standard pairing is written

$$\langle \mu, f \rangle = \langle f, \mu \rangle = \int f \, d\mu \, .$$

The "upper half-arrow" $\rightharpoonup$ denotes weak convergence of measures.

Functional spaces

In this course, we always work with functions of $x \in S^1$ with zero mean value. Accordingly, we denote by L_p the subspace of $L_p(S^1)$ formed by functions of zero mean value, and denote by H^m, $m \in \mathbb{R}$, the L_2–Sobolev space of functions with zero mean value, equipped with the homogeneous norm $\| \cdot \|_m$ (which would be only a seminorm without the zero-mean restriction). For more information on Sobolev spaces, see the corresponding section.
In particular, if $m \in \mathbb{N}$, then $\|u\|_m^2 = \int_{S^1} u^{(m)}(x)^2 dx$. Elements of spaces L_p and spaces $H^m(S^1)$ with $m \leq 1/2$ are equivalence classes, where functions coinciding a.e. are considered as the same object. If $m > 1/2$, then $H^m(S^1)$ is embedded in the space of continuous functions, and in this case a function

in $H^m(S^1)$ is identified with the only continuous representative of the class of equivalence.
We usually regard functions of t, x as curves of t, valued in a certain space of functions of x, and we work with various functional spaces corresponding to the situations when these curves are continuous, or smooth, or when they only are integrable.
If $m = 0$, we write $\|u\| := \|u\|_0 = |u|_2$.
For $T_2 > T_1 \geq 0$ and a Banach space B endowed with a norm $\|\cdot\|_B$, $C(T_1, T_2; B)$ stands for the space of continuous B-valued curves defined for $t \in [T_1, T_2]$: it is a Banach space with the supremum norm $\sup_{t\in[T_1,T_2]} \|u(t)\|_B$.
We use the abbreviations:

$$X_T^m = C(0, T; H^m)\,, \quad X_T^0 = X_T\,,$$

and denote $\dot{X}_T^m = \{\xi \in X_T^m : \xi(0) = 0\}$, and similarly for $\dot{X}_T$.
We consider the following norm for $C^n(T_1, T_2; B)$, $n \in \mathbb{N}_0$:

$$\|u\|_{C^n(T_1,T_2;B)} := \sum_{i=0}^{n} \left\| \frac{d^i u}{dt^i} \right\|_{C(T_1,T_2;B)}.$$

For a Polish space X we denote by $\|\cdot\|_{L(X)}$ the norm (B.1.1) in the Banach space of bounded Lipschitz functions on X, and by $\|\cdot\|^*_{L(X)}$ – the norm in the dual Banach space.
The notation $X \Subset Y$, where X, Y are Banach spaces, means that the former is compactly embedded in the latter.

Motivation: PDEs and the Wiener process. How can they be combined? And why Markov processes and stationary measures?

Trajectories of the Wiener process are only (almost surely) $(1/2 - \varepsilon)$–Hölder; in particular, they are not derivable. However, in some contexts their weak derivative can give birth to something meaningful (without using the Itô calculus, which will however be needed later for the Ito formula). We begin by a (not very interesting) example of a **SDE (stochastic differential equation)**.

Example. *Clearly, for any given random trajectory w, the SDE*

$$u_t - w_t = 0$$

has a unique (in the sense of distributions) solution $u(t) = w(t) + u(0)$.

The general method of considering the equation for $u - w$ will be OK in the whole course (sometimes in a generalised form, taking the solution of the heat equation $e^{t\Delta}w$ instead of w)...but for SPDEs in general this is not always the case.

Considering more involved features of the noise is anyhow necessary for the Ito formula, which is useful to study the stationary measure (and more generally to study the evolution in time of averaged quantities).
Before passing to Markovian concept, let us look at another standard "nice" example.

Example. *The Ornstein-Uhlenbeck/Langevin SDE (damping+diffusion):*

$$u_t = -\alpha u + \beta w_t.$$

We make the change of variables

$$v(t) := u(t) - \beta w(t),$$

and then

$$\tilde{v}(t) := e^{\alpha t} v(t).$$

We find that

$$\tilde{v}_t(t) = -\alpha \beta e^{\alpha t} w(t).$$

To study long-term behaviour of the solutions, it is relevant to look at dynamical system-type concepts 'of Markov type'.

Definition (INFORMAL). *A Markov chain for discrete space and time is a process u such that for all a and all $x_1, \dots, x_{k-1}, b$,*

$$\begin{aligned} &\mathbf{P}(u(k+1) = a | u(k) = b) \\ &= \mathbf{P}(u(k+1) = a | u(k) = b, u(k-1) = x_{k-1}, \dots, u(1) = x_1) \end{aligned}$$

(in other words: to know the law of the future, it suffices to know the present, no need to know the past).
Heuristically: it is sufficient to observe what happens now, history books are useless to help predict the future!

Remark .0.1. *The equality above implies the one with $u(k+1)$ replaced with $u(k+n)$ for any $n \geq 1$. In other words: if history books are useless to predict the next step, they are also useless to predict the situation at any future step.*

Definition (INFORMAL). *Given a process u in discrete space and time, a stationary measure is a probability measure μ on the space state such that for all k, x, y,*

$$\mu(x) = \sum_y \mathbf{P}(u(k+1) = x | u(k) = y)\ \mu(y).$$

Physicists call this a 'non equilibrium steady state'. Heuristically: particles do move, but their distribution does not change. So if one does not care about their individual properties/temporal trajectories, nothing changes!

Remark. *The two definitions above can of course be simplified if the process is **homogeneous in time** (i.e., transition probabilities from the step k to the step $k+1$ do not depend on k). In this case, the transition probabilities above can be indexed by the length of the intervals over which we observe the evolution – we do not care about the starting time.*

Our goal will be to define these objects in continuous time and space, introducing **Markov processes**, and eventually to generalise them even more to the context of **SPDEs (stochastic PDEs)**. Three potential difficulties (only the first one is important) and one nice thing:

- We need to use a continuous analogue to express the infinitesimal transition rates. Two ways out: work with expected values of functions or work directly with measures. For simplicity, in this course we will choose the latter. However, the former is more empirical/physical since it is concerned with the evolution of averaged quantities and not with the abstract underlying measure structure.
- For SPDEs, the state space is not $\mathbb{R}$ (or $\mathbb{R}^d$), but a subset of a separable Banach space (non-separability makes it more difficult to work with Sobolev spaces based on L^∞).
- For continuous time, there is a subtle measure-theoretical difficulty: we need to be careful when we define 'the events up to the past' (adapted vs progressively measurable). In our context of almost surely continuous paths, this will not be an issue and we will not talk about it anymore.

- If we deal with an autonomous equation and a random additive time-decorrelated force, the continuous analogues of transition rates are time-independent, so we have homogeneity in time. This will simplify our definitions a lot.

A Tools from functional analysis, ODEs and PDEs

We recall that in our course, we work with **Banach** (complete normed) spaces, endowed with their Borel sigma-algebras, which we denote as $(X, \mathcal{B})$. They are most often **Polish** (i.e. complete separable metric). Note that the most well-known example of a non-Polish space which appears in PDEs is L_∞ (and the corresponding Sobolev spaces $W^{m,\infty}$), but we only define those in order to use the corresponding norms, and we never work on them directly.

A.1 ODEs: local and global well-posedness

Definition A.1.1. *Consider a Banach space X and a continuous function $f : X \to X$. Then we refer to the* ***autonomous*** *ODE*

$$u'(t) = f(u),\ u(t = T_0) = u_0 \in X,\ t \in [T_0,\ T_1].$$

as ***a Cauchy problem.*** *We say that this problem is* ***well-posed*** *if there exists a unique solution $u \in C^1([T_0,\ T_1]; X)$ satisfying the equality above and depending continuously on u_0.*

Remark A.1.2. *One can also consider a* ***non-autonomous*** *version of the Cauchy problem, allowing the function f to also depend on the time t, and not only on the state variable $x(t)$. We also have nicer properties of the solution map (more regular dependence on t, u_0) if f is regular.*

Remark A.1.3. *We will not make precise statements from the classical theory of ODEs: everything which we need will be stated later for the more general setting of SDEs. Nevertheless, here are the main statements/ideas:*

- *If f is globally Lipschitz, there exists a solution for T as large as we want (i.e. a* ***global solution****) and by the* ***Gronwall lemma****, we have an exponential bound for this solution.*

- *If f is only locally Lipschitz, the statement above is still true but only for a finite u_0-dependent time.*
- *An important concept is that of a* ***Lyapunov function****, defined on the phase space and which does not grow too fast along the flow. If it has good properties (essentially, if it goes to infinity at infinity), this guarantees the existence of a global solution and gives a bound on it.*

A.2 Spaces of functions

First let us state the elementary Riesz–Thorin inequality, also referred to as the log–convexity of Lebesgue norms. It follows from Hölder's inequality.

THEOREM A.2.1. *Consider the spaces $L_p := L_p(X, \mathcal{F}, \mu)$. Let $1 \leq p_0 < p_1 \leq \infty$, and for $0 < \theta < 1$ define p_θ by: $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$. Then*

$$|f|_{p_\theta} \leq |f|_{p_0}^{1-\theta} \, |f|_{p_1}^{\theta}. \tag{A.2.1}$$

Sobolev spaces are generalisations of Lebesgue spaces, taking into account the integrability of derivatives. For generality, we give definitions which are valid in $\mathbb{R}^d$ as well as on the (flat) torus $\mathbb{T}^d$; however, in this course we will only work with the one-dimensional torus=circle S^1. Moreover, since we only work with zero-mean functions, we will actually work with homogeneous semi-norms only, since they are norms on the corresponding subspaces: see Remark A.2.5.

DEFINITION A.2.2. *For $m \in \mathbb{Z}_+$, $p \in [1, +\infty[$, the* ***homogeneous Sobolev seminorm*** *in $W^{m,p}$ of a function u is given by*

$$\|u\|_{\dot{W}^{m,p}} := \Big(\int_x \sum_{|\alpha|=m} \Big| \frac{\partial^m u}{\partial x^\alpha}(x) \Big|^p dx \Big)^{1/p}.$$

We use the usual multi–index conventions

$$|\alpha| = |(\alpha_1, \dots, \alpha_d)| = \alpha_1 + \dots + \alpha_d; \ \partial x^\alpha = \partial x_1^{\alpha_1} \cdot \dots \cdot \partial x_d^{\alpha_d},$$

and for $p = +\infty$ the definition above is modified as usual with a supremum in x and over the multi–indices.

Definition A.2.3. *For $m \in \mathbb{Z}_+,\ p \in [1, +\infty]$, the* ***Sobolev norm*** *in $W^{m,p}$ of a function u is given by*

$$\|u\|_{W^{m,p}} := \sum_{k=0}^{m} \|u\|_{\dot{W}^{k,p}}.$$

Note that on S^1, the Sobolev norms are totally ordered. Namely, for $u \in W^{m,p}$ we have:

$$|u|_{n,q} \leq |u|_{m,p},\ \text{for } n < m,\ q \in [1, +\infty] \text{ or } n = m,\ q \in [1, p]. \qquad \text{(A.2.2)}$$

Moreover, we have the classical interpolation inequality

$$\|u\|_n \leq \|u\|_k^{\theta} \|u\|_m^{(1-\theta)},\ \text{for } 0 \leq k \leq n \leq m, \qquad \text{(A.2.3)}$$

where $\theta = (m - n)/(m - k)$.

As usual, we set $H^m = W^{m,2}$, $m \in \mathbb{N}$, and denote the corresponding homogeneous Sobolev seminorms by $\|\cdot\|_{\dot{H}^m}$. These seminorms can be defined alternatively (and equivalently) using Fourier series (on the torus) and the Fourier transform (on the whole space).

Definition A.2.4. *A function u is said to belong to the space H^m on the torus if the corresponding norm*

$$\|u\|_m^2 = \sum_{k \in \mathbb{Z}^d} (1 + (2\pi|k|))^{2m} |\hat{u}|_k^2$$

is finite, and similarly on the whole space with the norm

$$\|u\|_m^2 = \int_{\xi \in \mathbb{R}^d} (1 + 2\pi|\xi|)^{2m} |\hat{u}(\xi)|^2.$$

Here, $\hat{u}$ denotes respectively the Fourier coefficients/transform.

Remark A.2.5. *If we restrict ourselves to the subspace of zero mean value functions on the torus, one can check (using Poincaré's inequality) that seminorms and norms are equivalent.*

Theorem A.2.6. *For $m > n$, the injection of H^m into H^n is compact.*

We start with a classical result. The inequalities on $\mathbb{R}$ imply the ones on S^1 by a localisation argument, multiplying by a bump function. Analogous inequalities exist in the multidimensional setting, but with different restrictions on the norms and different powers of Lebesgue/Sobolev norms.

LEMMA A.2.7. *(the Gagliardo–Nirenberg inequality). Let $m \in \mathbb{N}$ and $\beta \in \mathbb{N}_0$, $\beta \leq m-1$. Let also $p, q, r \in [1, \infty]$ and*

$$\theta \in [\beta/m, 1).$$

Then there exists $C = C(\beta, q, r, m) > 0$ such that

$$|h|_{\beta,r} \leq C|h|_{m,p}^{\theta}|h|_q^{1-\theta}, \tag{A.2.4}$$

provided that

$$\beta - \frac{1}{r} = \theta\Big(m - \frac{1}{p}\Big) - (1-\theta)\,\frac{1}{q}\,,$$

and, in addition,

$$\theta = \beta/m \quad \text{ if } p = 1 \text{ or } p = \infty.$$

In particular,

$$|h|_{k,r} \leq C\|h\|_m^{\theta}\|h\|^{1-\theta}, \qquad \theta = \frac{2rk + r - 2}{2rm}, \tag{A.2.5}$$

if $r \geq 2$ and $0 \leq k \leq m-1$. If $k \geq 1$, then applying (A.2.4) to h_x we get another useful for us inequality:

$$|h|_{k,r} \leq C\|h\|_m^{\theta}|h|_{1,1}^{1-\theta}, \qquad \theta = \frac{2}{r}\,\frac{rk-1}{2m-1}, \tag{A.2.6}$$

if $2m \geq k + 1 + (2m-2)/r$ and $1 \leq k \leq m-1$.

A.3 Semilinear parabolic PDEs: classical and mild solutions

We will not give any results in this section: the definitions and results are standard enough and if you have trouble finding a good source, I can provide an example. Moreover, we will give a detailed presentation later in the setting in which we need it. However, let us recall a few key ideas in a general setting of a PDE of the type:

$$u_t + f(u) = \Delta u, \; 0 \leq t \leq T.$$

- A **classical solution** is a function u which solves the equation above for $t > 0$. Usually it is required to belong to $C^1((0,T];X)$ for some Banach space X (else the equation makes no sense) and to $C([0,T];Y)$ for some other (larger) Banach space Y (else the initial condition makes no sense). The regularity of X is higher than that of Y, so the Laplacian is well-defined.

- A **mild solution** is defined using the heat semigroup $S(t)$ generated by $-\Delta$ and a Duhamel formulation which allows us to rewrite the PDE in a time-integrated form.

- We will only work in a nice setting where these two notions are uniquely defined and equivalent. The easy part is to observe that a classical solution is also necessarily mild. Then, one proves the existence and uniqueness of a mild solution by a fixed-point theorem in the space $C^1((0,T];X) \cap C([0,T];Y)$ (the hard part is to find an appropriate norm: a power of t is needed in front of the X norm), and after some calculations one proves that a mild solution is also classical.

B Tools from measure theory and probability

For more details on the results below, see [1, Chapter 11] and [2, Section 1.3.1]. More specifically, these references contain in a short and (as much as possible) self-contained manner some concepts which we do not have the time to treat. However, they play an important role when studying SPDEs; this is the case for example for the Kolmogorov–Chapman relations and the (strong) Markov property.

B.1 Measure theory

The results below apply, e.g.,

- When $(X, \mathcal{F}, \mu)$ is a Banach space with a probability Borel measure.

- When we work with the Lebesgue measure (eventually generalised to work with integrals on separable Banach spaces, i.e. the so-called **Bochner integral**. We never work directly with integrals in time on non-separable spaces such as L_∞; however, it is sometimes useful to consider the corresponding norms as functionals on separable spaces such as L_1, and to integrate these norms in time.

- When X is the set $\mathbb{N}$, given the counting measure, which assigns the measure one to each natural number (the setting we implicitly use when working with infinite series).

THEOREM B.1.1. ***The Lebesgue dominated convergence theorem:*** *Assume that there exists a nonnegative μ-integrable function g such that μ-a.e. $|f_n| \leq g$ for all n and $\lim_{n\to+\infty} f_n = f$. Then f is μ-integrable and*

$$\lim_{n\to+\infty} \int_X f_n(x)d\mu(x) = \int_X f(x)d\mu(x).$$

THEOREM B.1.2. ***Fatou's lemma:*** *Assume that $(f_n)_{n\geq 0}$ is a sequence of non-negative functions. Then we have:*

$$\int_X \liminf_{n\to+\infty} f_n(x)d\mu(x) \leq \liminf_{n\to+\infty} \int_X f_n(x)d\mu(x) \leq \infty.$$

THEOREM B.1.3. ***The monotone convergence theorem:*** *Assume that for μ-a.e. x the sequence of positive numbers $(f_n(x))_{n\geq 0}$ is monotone increasing. Then*

$$\lim_{n\to+\infty} \int_X f_n(x)d\mu(x) = \int_X \lim_{n\to+\infty} f_n(x)d\mu(x)$$

(where both sides of the equality can be infinite).

Now we consider a Polish space $\mathcal{O}$.

DEFINITION B.1.4. *A sequence $\{\mu_n\} \subset \mathcal{P}(\mathcal{O})$ is said to be weakly converging to $\mu \in \mathcal{P}(\mathcal{O})$ (denoted by $\mu_n \rightharpoonup \mu$) if*

$$\forall f \in \mathcal{C}_b(\mathcal{O}),\ \langle \mu_n, f\rangle \to \langle \mu, f\rangle.$$

For a measure μ on a measurable space X and an integrable function f the standard pairing is written

$$(\mu, f) \to \langle \mu, f\rangle = \langle f, \mu\rangle = \int f\, d\mu\,.$$

This mapping is non-degenerate: if $\langle \mu, f\rangle = 0$ for all measures $\mu \in \mathcal{P}(X)$, then $f = 0$; if $\langle \mu, f\rangle = \langle \nu, f\rangle$ for all functions $f \in C_b(X)$, then $\mu = \nu$.

Now we denote

$$L(\mathcal{O}) := \{f \in \mathcal{C}_b(\mathcal{O}) : \mathrm{Lip}(f) < \infty\}, \tag{B.1.1}$$

where $\mathrm{Lip}(f)$ denotes the Lipschitz constant of f, and endow the space $L(\mathcal{O})$ with the norm

$$\|f\|_{L(\mathcal{O})} := \|f\|_{\mathcal{C}_b(\mathcal{O})} + \mathrm{Lip}(f)\,. \tag{B.1.2}$$

Remark B.1.5. *The space $L(\mathcal{O})$ with the norm* (B.1.2) *is Banach.*

Definition B.1.6. *For $\mu, \nu \in \mathcal{P}(\mathcal{O})$ the Lipschitz-dual distance between μ and ν (defined on the dual Banach space $L(\mathcal{O})^*$, restricted to the space of measures $\mathcal{P}(\mathcal{O})$, naturally embedded in $L(\mathcal{O})^*$) is given by:*

$$\|\mu - \nu\|^*_{L(\mathcal{O})} := \sup_{\bar{B}}(\langle f, \mu\rangle - \langle f, \nu\rangle) = \sup_{\bar{B}} |\langle f, \mu\rangle - \langle f, \nu\rangle| \leq 2, \tag{B.1.3}$$

where $\bar{B} = \bar{B}^1_{L(\mathcal{O})}$ is the closed ball $\{\|f\|_{L(\mathcal{O})} \leq 1\}$.

Remark B.1.7. *If the metric space is clear from the context, we abbreviate $\|f\|_{L(\mathcal{O})}$ to $\|f\|_L$ and $\|\mu\|^*_{L(\mathcal{O})}$ to $\|\mu\|^*_L$.*

The following celebrated theorem, essentially due to Kantorovich, characterises the weak convergence of measures in terms of the Lipschitz-dual distance:

Theorem B.1.8. *Let $\mathcal{O}$ be a Polish space. Then:*

1. *$(\mathcal{P}(\mathcal{O}), \|\cdot\|^*_{L(\mathcal{O})})$ is a complete metric space.*
2. *If $\{\mu_n\} \subset \mathcal{P}(\mathcal{O})$ and $\mu \in \mathcal{P}(\mathcal{O})$, then $\mu_n \rightharpoonup \mu$ if and only if $\|\mu_n - \mu\|^*_{L(\mathcal{O})} \to 0$.*

Let $\mathcal{O}$ be a Polish space and $\mathcal{P}_1(\mathcal{O})$ be the space of probability measures on it with finite first moment:

$$\mathcal{P}_1(\mathcal{O}) = \{\mu \in \mathcal{P}(\mathcal{O}) : \int_{\mathcal{O}} \mathrm{dist}_{\mathcal{O}}(x, x_0)\mu(dx) < \infty\},$$

where x_0 is a fixed point in $\mathcal{O}$ (the space will not change if we replace x_0 by another point in $\mathcal{O}$). In his celebrated study of the optimal transport problem, Kantorovich introduced in $\mathcal{P}_1(\mathcal{O})$ the distance

$$\|\mu - \nu\|_{Kant} = \|\mu - \nu\|_{Kant,\mathcal{O}} := \sup_{\mathrm{Lip} f \leq 1} \big|\langle f, \mu\rangle - \langle f, \nu\rangle\big| \leq \infty \tag{B.1.4}$$

(obviously, we may assume that $f(x_0) = 0$). This distance is greater or equal than the Lipschitz–dual distance (B.1.3), and it is easy to see that if $\mathcal{O}$ is bounded then the two are equivalent. However if $\mathcal{O}$ is unbounded, then the convergence in the Kantorovich distance is stronger than the weak convergence in $\mathcal{P}(\mathcal{O})$.

EXAMPLE B.1.9. *In $\mathcal{P}(\mathbb{R})$ consider the sequence of measures $\mu_n = (1 - n^{-1/2})\delta_0 + n^{-3/2}\chi_{[0,n]}dx$. Obviously $\mu_n \rightharpoonup \delta_0$ as $n \to \infty$, but choosing in* (B.1.4) *$f(x) = x$ we see that $\|\mu_n - \delta_0\|_{Kant} \geq \sqrt{n}/2$.*

To bound (B.1.4) in terms of the Lipschitz-dual distance between μ and ν one has to control "the behaviour of the measures at infinity In particular, the following is true (for the proof, see [1, Appendix 11.G]):

LEMMA B.1.10. *Let μ and ν be two measures in $\mathcal{P}_1(\mathcal{O})$ such that*

$$\langle dist_{\mathcal{O}}(x,x_0)^\gamma, \mu\rangle,\ \langle(dist_{\mathcal{O}}(x,x_0)^\gamma, \nu\rangle \leq K, \quad K \geq 1, \tag{B.1.5}$$

for some $\gamma > 1$. Then

$$\|\mu - \nu\|_{Kant} \leq C_\gamma K^{1/\gamma}\big(\|\mu-\nu\|_L^*\big)^{1-1/\gamma}\big). \tag{B.1.6}$$

B.2 Basics of probability

If $(\Omega, \mathcal{F})$ is a measurable space and $\mathbb{P}$ is any element of $\mathcal{P}(\Omega, \mathcal{F})$, then the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**. Elements of $\mathcal{F}$ are called **events**. Events of zero measure are called **null sets**. A measurable mapping defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **random variable**.

Theorems B.2.1 and B.2.4, given below, are corollaries of the monotone convergence theorem.

THEOREM B.2.1. *Let $m_1 \geq m_2$ and $\{\xi_n, n \geq 1\}$ be a sequence of r.v.'s in H^{m_1} such that $\mathbb{E}\|\xi_n\|_{m_1}^\alpha \leq K$ for all n and some $1 \leq \alpha < \infty$. Assume also that a.s. ξ_n converges in the space H^{m_2} to a random variable ξ. Then a.s. $\xi = \xi'$, where ξ' is a random variable in H^{m_1} and*

$$\mathbb{E}\|\xi'\|_{m_1}^\alpha \leq K. \tag{B.2.1}$$

EXERCICE B.2.1. *Prove this result (the proof is in [1, Appendix A]).*

Applying this theorem we will not make a difference between the r.v. ξ and its modification ξ', and will regard ξ as a r.v. in H^{m_1}, taking infinite value on a null-set.

THEOREM B.2.2. *Let a sequence of Borel measures $\{\mu_n, n \geq 1\}$ on a Polish space $\mathcal{O}$ weakly converge to a measure $\mu_0 \in \mathcal{P}(\mathcal{O})$. Then there exists a collection of random variables $\{\xi_n, n \geq 0\}$, defined on the same probability space and valued in $\mathcal{O}$, such that $\mathcal{D}\xi_n = \mu_n$ for each $n \geq 0$, and $\xi_n \xrightarrow[n\to\infty]{} \xi_0$, a.s.*

COROLLARY B.2.3. *Let $m_1 \geq m_2$ and $\{\mu_n, n \geq 1\}$ be a sequence of measures in $\mathcal{P}(H^{m_1})$ such that $\langle \|u\|_{m_1}^p, \mu_n\rangle \leq K$ for all n and some $1 \leq p < \infty$. Assume also that $\mu_n \rightharpoonup \mu$ in $\mathcal{P}(H^{m_2})$. Then $\mu(H^{m_1}) = 1$, so μ may be regarded as a measure on H^{m_1}, and $\langle \|u\|_{m_1}^p, \mu\rangle \leq K$.*

Proof. The result follows from the theorem above by applying Skorokhod's Theorem B.2.2. □

THEOREM B.2.4. *Let $1 \leq p, \alpha < \infty$ and $\{\xi_n, n \geq 1\}$ be a sequence of r.v.'s in L_p such that $\mathbb{E}|\xi_n|_p^\alpha \leq K$ for all n. Assume that a.s. $\xi_n^\omega \to \xi^\omega$ in L_1, where ξ is a r.v. in L_1. Then $\xi \in L_p$ a.s. and*

$$\mathbb{E}|\xi|_p^\alpha \leq K. \tag{B.2.2}$$

EXERCICE B.2.2. *Prove this result (the proof is in [1, Appendix A]).*

EXERCICE B.2.3. *Formulate and prove a corollary, proceeding as above.*

B.3 Measure theory on product spaces and the Markov property

Let $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$ be measurable spaces. Consider the space $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ (if X_1, X_2 are Polish spaces, endowed with their Borel sigma-algebras $\mathcal{B}_{X_1}, \mathcal{B}_{X_2}$, then $\mathcal{B}_1 \otimes \mathcal{B}_2 = \mathcal{B}_{X_1 \times X_2}$).
Then for any measurable function f on $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ and each $x_2 \in X_2$, the function $x_1 \mapsto f(x_1, x_2)$ is measurable on $(X_1, \mathcal{B}_1)$, and if in addition f is integrable, then **Fubini's theorem** holds:

THEOREM B.3.1. *Let μ_1 and μ_2 be probability measures on $(X_1, \mathcal{B}_1)$ and $(X_2, \mathcal{B}_2)$. Then a measurable function f on $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ is integrable with respect to $\mu_1 \times \mu_2$ if and only if*

i) $\int |f(x_1, x_2)| \, \mu_2(dx_2) < \infty, \ \ \mu_1$*–a.s;*
ii) $\int \Big(\int |f(x_1, x_2)| \, \mu_2(dx_2) \Big) \mu_1(dx_1) < \infty$.
If this holds, then

$$\int f(x_1, x_2)(\mu_1 \times \mu_2)(dx_1 \, dx_2) = \int \Big(\int f(x_1, x_2) \, \mu_2(dx_2) \Big) \mu_1(dx_1) \, .$$

Definition B.3.2. *For two Polish spaces* $(X_1, \mathcal{B}_1), (X_2, \mathcal{B}_2)$*, a* ***transition probability*** *is defined as a mapping*

$$\lambda : (x_1, B_2) \in (X_1 \times \mathcal{B}_2) \mapsto \lambda(x_1, B_2) \in \mathbb{R}_+$$

such that

- *For every* x_1*,* $\lambda(x_1, \cdot)$ *is a probability measure.*
- *For every* B_2*,* $\lambda(\cdot, B_2)$ *is measurable.*

We will admit the following result.

Lemma B.3.3. *Let* μ_1 *be a positive Borel measure on* $\mathcal{B}_1$*. Then the mapping*

$$\mu_1 \mapsto \int_{\mathcal{B}_1} \lambda(x_1, \cdot) d\mu_1(x_1)$$

induces a positive Borel measure μ_2 *on* $\mathcal{B}_2$*. Moreover, this mapping is linear, and images of probability measures are probability measures.*

Definition B.3.4. *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, a* ***(continuous) random process*** *valued on a Polish space* X *is a random variable valued in* $C(I, X)$ *for some time interval* I*. In other words, it is a measurable mapping*

$$a : \omega \mapsto (a^\omega(t))_{t \in I} \in C(I, X).$$

Definition B.3.5. *Consider a random variable* h *defined on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and valued on a Polish space* X*. For a* σ*-algebra* $\mathcal{G} \subset \mathcal{F}$ *the* ***conditional expectation*** *of* h *with respect to* $\mathcal{G}$*, denoted* $\mathbf{E}(h \mid \mathcal{G})$*, is the unique* $\mathcal{G}$*-measurable random variable such that for all* $\mathcal{G}$*-measurable* $\mathbb{R}$*-valued random variables* ϕ*,*

$$\int_X h(\omega)\phi(\omega) \ d\omega = \int_X \mathbf{E}(h \mid \mathcal{G})(\omega)\phi(\omega) \ d\omega.$$

For a measurable subset $A \subset \mathcal{F}$, the ***conditional probability*** $\mathbf{P}(A \mid \mathcal{G})(\omega)$ *is defined in the natural way by*

$$\mathbf{P}(A \mid \mathcal{G}) := \mathbf{E}(1_A \mid \mathcal{G}).$$

In particular, if A is $\mathcal{G}$-measurable, $\mathbf{P}(A \mid \mathcal{G}) = 1_A$.

Definition B.3.6. *A* ***filtered probability space*** *is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and an increasing sequence of σ-algebras $(\mathcal{F}_t)_{t\in I} \subset \mathcal{F}$, where I is an interval.*

Definition B.3.7. *A process $(h(t))_{t\in I}$ defined on a filtered probability space is said to be* ***adapted*** *if for every $t \in I$, $h(t)$ is $\mathcal{F}_t$-measurable.*

Definition B.3.8. *A (homogeneous)* ***Markov process*** *consists of:*

- *A (continuous) random process $(u(t))_{t\geq 0}$, adapted with respect to a filtered probability space*

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

  and valued on a separable Banach space X.

- *A family of probability measures $\mathbf{P}_v,\ v \in X$, such that for each $Q \in \mathcal{F}$, $v \mapsto \mathbf{P}_v(Q)$ is measurable.*

- *A family of transition probabilities $P_t,\ t \geq 0$ defined on $X \times \mathcal{B}_X$.*

such that

- *For all $v \in X$, $\mathbf{P}_v(u(0) = v) = 1$.*

- *For all $v \in X$, $s, t \geq 0$ and $\Gamma \in B_X$ we have the property*

  $$\mathbf{P}_v(u(t+s) \in \Gamma \mid \mathcal{F}_s) = \mathbf{E}_v(1_{u(t+s)\in\Gamma} \mid \mathcal{F}_s) = P_t(u(s), \Gamma),\ \mathbf{P}_v - a.s.$$

Definition B.3.9. *For a Markov process v, the* ***Markov semigroup for measures*** *is the family of linear operators $(S_t^*)_{t\geq 0}$ acting on the space of probability measures on X and defined by*

$$S_t^*(\mu) = \int_X P_t(x, \cdot)\ d\mu(x)$$

(see Lemma B.3.3).

Note that $(S_t^*)_{t\geq 0}$ is indeed a semigroup (i.e. $S_0^* = Id$ and $S_{s+t}^* = S_s^* \circ S_t^*$); see [1, Section 1.6].

B.4 The Wiener process, SDEs and Ito's formula

For more details and references to proofs, see [1, Appendix H].

Definition B.4.1. *A Wiener process is:*

- *A* ***probability space*** $(\Omega, \mathcal{F}, \mathbf{P})$ *(i.e. space+σ-algebra+probability measure).*
- *A measurable function $\omega \mapsto w^\omega(\cdot)$ from Ω onto the space of continuous functions $C(\mathbb{R}, \mathbb{R})$.*

with the following properties:

- *We have $w^\omega(0) \equiv 0$.*
- *For every s, t, the law of $w(t) - w(s)$ is $\mathcal{N}(0, |t-s|)$ (Gaussian, mean 0, variance $|t-s|$).*
- *For $t_1 < \cdots < t_k$, the increments $w(t_k) - w(t_{k-1}), \ldots, w(t_2) - w(t_1)$ are independent. Equivalently, one can require that for $t_2 \geq t_1$, the increment $w(t_2) - w(t_1)$ is independent with respect to the σ–algebra generated by $w(t)$, $t \leq t_1$.*

Most importantly: we know (by a Donsker central-limit type construction) that such a process exists. Once we know it, we do not care who Ω is. As usual, only the law of the process is important.

Now let us say a few words about Ito's formula for a solution of finite-dimensional SDEs, which is essentially a simplified version of that in [1, Chapter 11]. We will often also use it for SPDEs; to do so rigorously, a finite-dimensional **Galerkin projection** and a limiting procedure are necessary.

We consider equations of the form

$$\dot{x}_j(t) = f_j(t, x(t)) + \sum_{j=1}^n b_j \dot{\beta}_j(t), \quad 1 \leq j \leq n, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{B.4.1}$$

with initial data

$$x_j(0) = x_{0j}, \qquad 1 \leq j \leq n. \tag{B.4.2}$$

Here $t \geq 0$, $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ is a continuous locally Lipschitz mapping, b_j's are any real numbers (we do not exclude the deterministic case, i.e. that all of them vanish), $\beta_j(t)$'s are standard independent Wiener processes, and x_0 is a r.v. in $\mathbb{R}^n$, independent from those processes. The derivatives $\dot{\beta}_j(t)$ are defined only in the sense of distributions, so we say that $x^\omega(t)$, $t \in [0,T]$, $0 < T \leq \infty$, is a solution of the problem (B.4.1), (B.4.2) if

$$x_j^\omega(t) = x_{0j}^\omega + \int_0^t f_j(s, x^\omega(s))\, ds + \sum_{j=1}^n b_j \beta_j^\omega(t), \qquad 0 \leq t \leq T, \tag{B.4.3}$$

for a.e. ω.
A solution of (B.4.1), (B.4.2), if it exists, is unique in the sense that any two solutions $x^1(t)$ and $x^2(t)$ must coincide a.s. Indeed, for a.e. ω the difference $x^1 - x^2$ satisfies a deterministic equation with zero initial data, so it vanishes identically for the same reason as in the theory of ordinary differential equations. As in the deterministic case, a solution for (B.4.1), (B.4.2) may not exist for all times, but it does exist if the vector field f is globally Lipschitz:

Theorem B.4.2. *If the mapping $x \mapsto f(t,x)$ is Lipschitz on $\mathbb{R}^n$ with a Lipschitz constant, independent from $0 \leq t \leq T$, then the problem is globally well-posed.*

The global Lipschitz property restriction is often too strong. But if f is only locally Lipschitz in x, then still the problem (B.4.1), (B.4.2) is soluble if eq. (B.4.1) has a **Lyapunov function**.

Theorem B.4.3. *Assume that there exist a C^2–smooth function $V(x)$ and a constant $c > 0$ such that*

$$\frac{1}{2}\sum_j b_j^2 \frac{\partial^2}{\partial x_j^2} V(x) + \sum_j f_j(t,x) \frac{\partial}{\partial x_j} V(x) \leq cV(x) \quad \forall\, x, \ \ \forall\, t \geq 0, \tag{B.4.4}$$

and

$$\inf_{|x| \geq R} V(x) \to \infty \quad \text{as} \quad R \to \infty. \tag{B.4.5}$$

Then the problem (B.4.1), (B.4.2) *has a unique solution $x(t)$, and*

$$\mathbb{E} V(x(t)) \leq e^{ct} \mathbb{E} V(x_0) \tag{B.4.6}$$

(the r.h.s. may be infinite).

Example B.4.4. *Assume that the vector field f satisfies*

$$\langle f(t,x)\cdot x\rangle \le -\alpha\|x\|^2 \qquad \forall\, t\ge 0,\ \forall\, x, \tag{B.4.7}$$

with some $\alpha>0$. Let us take $V(x)=e^{\sigma\|x\|^2}$, where $0<\sigma\le\alpha/B^2$, where $B=\max_{1\le j\le n}|b_j|$. Then

$$\sum_j f_j(t,x)\frac{\partial}{\partial x_j}V(x)=e^{\sigma\|x\|^2}2\sigma x\cdot f(x)\le -V(x)2\sigma\alpha\|x\|^2,$$

and

$$\frac{1}{2}\sum_j b_j^2\frac{\partial^2}{\partial x_j^2}V(x)\le \sigma V(x)(\sum b_j^2+2\sigma B^2\|x\|^2).$$

So the l.h.s. of (B.4.4) *is bounded by*

$$\sigma V(x)\big(-2\alpha\|x\|^2+2\sigma B^2\|x\|^2+\sum b_j^2\big).$$

Since $\sigma B^2\le\alpha$, then (B.4.4) *holds with $c=\sigma\sum b_j^2$. Relation* (B.4.5) *obviously is valid, so the last theorem implies*

$$\mathbb{E}e^{\sigma\|x(t)\|^2}\le e^{t\sigma\sum b_j^2}\,\mathbb{E}e^{\sigma\|x_0\|^2}. \tag{B.4.8}$$

A powerful tool to study equation (B.4.1) is given by **Ito's formula**. For us, it suffices to consider its expected value.

Theorem B.4.5. *Let $x^\omega(t)$ be a solution of* (B.4.1) *for $0\le t\le T$ and $F(x)$ be a C^2–smooth function. Assume that*

$$\int_0^T \mathbb{E}\left[|\nabla F(x^\omega(t))\cdot f(t,x^\omega(t))|+\|\nabla F(x^\omega(t))\|+|\sum_j b_j^2\,\frac{\partial^2}{\partial x_j^2}F(x^\omega(t))|\right]^2 dt<\infty. \tag{B.4.9}$$

Then for $0\le t\le T$ we have

$$\frac{\partial}{\partial t}\mathbb{E}F(x(t))=\mathbb{E}\left(\nabla F(x(t))\cdot f(t,x(t))\right)+\tfrac{1}{2}\,\mathbb{E}\sum_j b_j^2\,\frac{\partial^2}{\partial x_j^2}F(x(t)),\quad a.s.\ in\ t. \tag{B.4.10}$$

As in similar situations above, (B.4.10) should be understood in the integrated sense: $\mathbb{E}F(x(t)) - \mathbb{E}F(x(0))$ equals to the integral of the term in the r.h.s., for all $t \geq 0$. Since in view of (B.4.9) the expectation $\mathbb{E}F(x(t))$ is an absolutely continuous function of t, then by the latter relation (B.4.10) holds for a.e. t.

Lemma B.4.6. *Consider a vector field $f(t,x)$ satisfying* (B.4.7) *and a C^2–function $F(x)$. Assume that $\mathbb{E}e^{\sigma\|x_0\|^2} < \infty$ for some $0 < \sigma \leq \alpha/B^2$ and*

$$|\nabla F(x) \cdot f(t,x)| + \|\nabla F(x)\| + |\sum_j b_j^2 \frac{\partial^2}{\partial x_j^2} F(x)| \leq C_T e^{\sigma\|x\|^2} \quad \forall\, x,\ \forall\, t \in [0,T], \tag{B.4.11}$$

with a suitable constant $C_T > 0$. Then a solution $x(t)$ of the problem (B.4.1), (B.4.2) *satisfies* (B.4.9) *for $0 \leq t \leq T$. So $x(t)$ also meets* (B.4.10).

C SDEs: well-posedness and stationary measure

C.1 Well-posedness of SDEs

For well-posedness of ODEs, one studies the operator

$$u_0 \mapsto (u(t))_{t\in[T_1,T_2]}.$$

For SDEs (or for ODEs with an additional time-dependent parameter), we are concerned with the mapping

$$\Big(u_0, (w^\omega(t))_{t\in[T_1,T_2]}\Big) \mapsto (u^\omega(t))_{t\in[T_1,T_2]},\ \mathbb{R} \times C(T_1,T_2;\mathbb{R}) \to C(T_1,T_2;\mathbb{R}).$$

In other words, we want our solution to be continuous not only with respect to the initial value, but also with respect to the (random) noise.
This is a natural requirement, which also makes sense when studying, for example, control theory (it is NOT a stochastic concept, we just quantify how the solution depends on the force). It is also reminiscent of the dependence with respect to the parameters in the context of the classical Cauchy-Lipschitz theorem for ODEs. As we have seen before, for the Ornstein-Uhlenbeck SDE:

$$u_t = -\alpha u + \beta w_t.$$

everything is OK because we have an explicit formula. Indeed,

$$\tilde{v}_t(t) = -\alpha\beta e^{\alpha t} w(t),$$

where

$$\tilde{v}(t) := e^{\alpha t}(u(t) - \beta w(t)).$$

In general (for example when replacing u with a nice enough $f(u)$), there are two alternative strategies to prove continuity:

- The abstract one with the Banach fixed point theorem (considering $w(t)$ as an external parameter).
- A more pragmatic one: we work directly on the difference between solutions. Namely, we decompose it into two parts: the difference of solutions with the same u_0 and different w (dependence with respect to the forcing) and that of solutions with different u_0 and the same w (dependence with respect to the initial condition).

C.2 Stationary measure for SDEs

We will admit the following result. For the lemma's proof, follow the lines of [1, Sections 1.6, 11.5] (and similarly for the proof of other such statements in similar situations).

Below, we use the notation $\mathbf{P}_{u_0=x}$. Its meaning is that we consider a deterministic initial condition $u_0 = x$ and then the quantity $u(t)$ becomes a random process as described below, so that the quantities given below are measurable.

Lemma C.2.1. *The solutions $u(t)$ of the Ornstein-Uhlenbeck SDE described above form a Markov process, endowed with the filtered probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where the filtration $(\mathcal{F}_t)_{t\geq 0}$ is generated by the random variables $(w(s))_{0\leq s\leq t}$ and the transition probabilities P_t are defined by*

$$P_t(x, \Gamma) := \mathbf{P}_{u_0=x}(u(t) \in \Gamma)$$

and the measures $\mathbf{P}_v$ by

$$\mathbf{P}_v := \mathbf{P}_{u_0=v}.$$

Definition C.2.2. *A* ***stationary measure*** *is a probability measure which is a fixed point for the Markov operator S_t^* for all $t > 0$.*

To establish the existence of a stationary measure for the Ornstein-Uhlenbeck SPDE

$$u_t = -\alpha u + \beta w_t,$$

we use a **Bogolyubov-Krylov argument** (tightness+extraction of a subsequence). Namely, we consider averages

$$m_k := \frac{1}{k}\int_0^k \mu_t,\ \ \mu_t := S_t^*(\delta_0).$$

Applying the Itô formula to the expected value $\mathbf{E}u^2$ (one of the few circumstances in which we use it in our whole course!), we prove that it remains finite, so for any ε all the mass except at most ε remains concentrated on a fixed k**-independent** compact (by Chebyshev's inequality): in other words, we have a tight sequence. Therefore (by Prokhorov's theorem) we can substract a converging subsequence from μ_k, and its limit will be a stationary measure.

Given the form of the solution, for a deterministic u_0, the law of u at a time t is necessarily Gaussian. Since the limit (in the sense of weak convergence) of Gaussian laws is necessarily Gaussian (see exercice below), the same is true for the stationary measure.

Applying the Itô formula to the expected values $\mathbf{E}u$ and $\mathbf{E}u^2$, one proves that a stationary measure for the PDE above necessarily has the law $\mathcal{N}(0, \beta^2/2\alpha)$ (therefore it is unique, and one can estimate the exponential rate of convergence to it).

Exercice C.2.1. *Fill in the missing details (in particular about the Ito formula). In order to prove that the Bogolyubov-Krylov weak extracted limit is a stationary measure, we admit that for a well-posed SDE, the Markov operator S_t^* is continuous with respect to the weak convergence; see [1, Section 1.6.]. We also use the result of the exercice below on the limit of Gaussian real-valued laws*

Solution. *First we apply the Ito formula to $f(u) = u$ (the domination verifications are standard and follow from the fact that all polynomial moments of a Wiener process are finite:*

$$\frac{d\mathbb{E}u}{dt} = -\alpha\mathbb{E}u + 0.$$

Similarly:

$$\frac{d\mathbb{E}u^2}{dt} = -2\alpha\mathbb{E}u^2 + \beta^2.$$

Therefore, $\mathbb{E}u^2(t)$ is uniformly bounded by $\beta^2/2\alpha$. By linearity, the same is true for $\int_{\mathbb{R}} |x|^2 dm_k$, uniformly with respect to k.
Consequently by Chebyshev's inequality:

$$\forall N \in \mathbb{N}, \forall t > 0,\ m_k(x :\ |x| > N) \leq \beta^2/2\alpha N^2.$$

Therefore:

$$\forall \varepsilon > 0,\ \exists N > 0 :\ m_k([-N, N]) > 1 - \varepsilon,\ \forall k \geq 1.$$

By Prokhorov's theorem, the Bogolyubov-Krylov sequence converges to some limit μ. It remains to prove that the corresponding limit is a stationary measure.
For $\theta > 0$ we have:

$$\begin{aligned} S_\theta^*(m_{j_k}) &= \frac{1}{j_k}\int_0^{j_k} S_\theta^*(\mu(t))dt = \frac{1}{j_k}\int_0^{j_k} \mu(t+\theta)dt = \frac{1}{j_k}\int_\theta^{j_k+\theta} \mu(t)dt \\ &= \underbrace{-\frac{1}{j_k}\int_0^{\theta} \mu(t)dt}_{:=I} + \underbrace{\frac{1}{j_k}\int_0^{j_k} \mu(t)dt}_{=m_{j_k}} + \underbrace{\frac{1}{j_k}\int_{j_k}^{j_k+\theta} \mu(t)dt}_{:=K}. \end{aligned}$$

For any $f \in \mathcal{C}_b(H^1)$ the terms $|\langle K, f\rangle|$ and $|\langle I, f\rangle|$ are bounded from above by $\frac{\theta}{j_k}|f|_\infty$, so they converge to zero when $k \to \infty$. Since $m_{j_k} \rightharpoonup \mu$ and the operator S_θ^* is weakly continuous, then passing to the limit in the equality above, we get $S_\theta^*\mu = \mu$.

Exercice C.2.2. *Prove that a (weak) limit of Gaussian real-valued laws is necessarily Gaussian.*

Solution (Exercice C.2.2). *Consider the limit of Gaussian laws $N(\alpha_n, \beta_n)$. Convergence of the mean values α_n follows by duality with a constant function. Now we look for a converging subsequence of β_n. It necessarily exists, except in the case $\beta_n \to \infty$, which we can rule out because of the converse of Prokhorov's lemma (we escape from any compact). On a more intuitive level, consider the duality with any function with compact support and then (by localisation) with any continuous bounded function: if the variance explodes, the laws converge weakly to 0.*

It remains to prove that if $\alpha_n \to \alpha$ and $\beta_n \to \beta$, then $N(\alpha_n, \beta_n) \rightharpoonup N(\alpha, \beta)$: this is not hard. For a quantitative version, see the next exercice.
Finally, we conclude by uniqueness of a weak limit. Note that we might converge to a Dirac measure: this simply corresponds to $\beta = 0$.

Exercice C.2.3. *Prove the exponential rate of (weak) convergence to the stationary measure.*

Solution (Exercice C.2.3). *By the calculations above, the mean value and variance of the laws of $u(t)$* ***(which are Gaussian by the explicit formula we have!)*** *converge exponentially. It remains to prove that this yields exponential convergence for the laws in the dual-Lipschitz metric. We consider $\phi \in \mathcal{L}(\mathbb{R})$ of norm smaller than 1 and we get that if $|\alpha - \alpha'|, |\beta - \beta'| \leq \varepsilon$, then:*

$$
\begin{aligned}
&\Big| \int_{\mathbb{R}} \phi(x) \exp(-|x-\alpha'|^2/2\beta')dx - \int_{\mathbb{R}} \phi(x) \exp(-|x-\alpha|^2/2\beta)dx \Big| \\
&\leq \Big| \int_{\mathbb{R}} |\phi(x)| |\exp(-|x-\alpha'|^2/2\beta') - \exp(-|x-\alpha|^2/2\beta)| dx \Big| \\
&\leq \Big| \int_{\mathbb{R}} C(\alpha,\beta)\varepsilon |\phi(x)| dx \Big| \leq C(\alpha,\beta)\varepsilon.
\end{aligned}
$$

In the last line, we simply used that $(\gamma, \delta) \mapsto exp(-|x-\gamma|^2/2\delta)$ is locally Lipschitz.

D SPDEs: well-posedness and stationary measure.

D.1 Well-posedness

For SPDEs, we are concerned with

$$\Big(u_0, (\xi^\omega(t))_{t\in[T_1,T_2]}\Big) \mapsto (u^\omega(t))_{t\in[T_1,T_2]}.$$

The only difference w.r.t. SDEs is that we need to study the operator on an appropriate functional space, both for the initial condition and for the force. In particular, we need to build an analogue of the Wiener process on a separable Hilbert space: we will only see the easiest way to do it, using Fourier analysis.

We will not dwell on the different concepts of weak/strong/mild solutions **in the sense of PDEs** since in our setting (linear and semilinear parabolic equations) all these definitions coincide. We will also not talk about the concepts of weak/strong solutions **in the sense of probability** (solutions for some forcing satisfying a law vs solutions for a given forcing); however, the latter is implicitly important when studying stationary measures (see [1, Chapter 1]). Indeed, in our setting we begin by studying solutions with a deterministic initial conditions, and then when considering the transition probabilities and the semigroup of Markov operators there is a (hidden) structure of laws of solutions considered at intermediate time (i.e., when we consider the transition semigroup, we care about the law of an initial condition and not about the initial condition itself). In fact, time-homogeneity of the forcing allows for the convenient semigroup notation, which allows us to manipulate solutions directly on the measure level, which sometimes makes us forget how the whole thing works pathwise!
We remind that we always work with Lebesgue/Sobolev spaces of zero-mean functions on the torus, which we will denote as above by $L_p/H^m/W^{m,p}$. This simplifies our life a lot notation-wise; in particular, the Sobolev seminorms are now actual norms, equivalent to the ones we formerly defined.

Definition (INFORMAL) D.1.1. *Consider a separable Hilbert space. It necessarily has a countable Fourier basis denoted $(e_k)_{k\in\mathbb{Z}}$.*
The easiest way to define a (generalised) Wiener process is to consider ***independent*** *Wiener processes w_k and then to (formally) define the sum:*

$$\sum_{k\in\mathbb{Z}^*} a_k w_k(t) e_k. \tag{D.1.1}$$

Depending on the asymptotics of the coefficients a_k, the sum converges almost surely in an appropriate space.

We consider the canonical basis of zero mean value functions in $L_2(S^1)$:

$$\begin{cases} e_k = \sqrt{2}\cos(2\pi kx), \\ e_{-k} = \sqrt{2}\sin(2\pi kx), \end{cases} \quad k\in\mathbb{N}. \tag{D.1.2}$$

Now each Sobolev space $H^m(S^1),\ m\in\mathbb{N}\cup\{0\}$ is a Hilbert space endowed with the canonical Fourier basis given by

$$\frac{1}{(2\pi|k|)^m} e_k,\ k\in\mathbb{Z}^*$$

We will define for each m:

$$B_m := \sum_{k\in\mathbb{Z}^*} (2\pi k)^{2m} a_k^2. \tag{D.1.3}$$

LEMMA D.1.2. *Provided $B_m < \infty$, the sum given by* (D.1.1) *is almost surely in X_T^m, and moreover*

$$\mathbf{E}\|\xi(t)\|_{H^m}^2 = B_m t, \ \forall t.$$

Proof. See [1, Section 1.2.]: first Theorem 1.2.1., then Corollary 1.2.4. The proof is not very technical but uses many ingredients: Doob's inequality, an extraction procedure and monotonicity to conclude on the convergence of the whole sequence. □

In this case, we will say that the sum (D.1.1) is **a Wiener process in** H^m.

DEFINITION D.1.3. *Consider $u_0 \in H^1$ and a Wiener process $\xi(t) \in H^2$. The* ***stochastic heat equation*** *is given by:*

$$u_t - \nu u_{xx} = \xi_t,$$

where $\nu > 0$ is a constant.

THEOREM D.1.4. *The stochastic heat equation is well-posed in H^m for a Wiener process $\xi(t) \in H^{m^*}$, for any $m, m^* \in \mathbb{R}$, $m < m^*$.*

Proof. The main idea is straightforward: decoupling of the different Fourier modes, each of them satisfying an Ornstein-Uhlenbeck equation (in particular, uniqueness is straightforward).
Consequently, a solution exists formally as a Fourier series and it remains to prove its convergence and the continuity of the solution operator (which is linear so the verification is simpler).
Without loss of generality we may assume that $m_* \leq m + 2$. Suppose first that $u_0 \in H^m, \xi, \xi_t \in X_T^{m_*}$ and consider the curve

$$v(t) = e^{\nu\Delta t}u_0 + \int_0^t e^{\nu\Delta(t-\tau)}\xi_\tau(\tau)\,d\tau = e^{\nu\Delta t}u_0 + \xi(t) - \nu\int_0^t \Delta e^{\nu\Delta(t-\tau)}\xi(\tau)\,d\tau, \tag{D.1.4}$$

where $\Delta = \frac{\partial^2}{\partial x^2}$ and $\{e^{\nu t\Delta}, t \geq 0\}$, is the semigroup of operators generated by $\nu\Delta$. That is, $\big(e^{\nu\tau\Delta}u_0\big)(x) = v(\tau, x)$, where $v(t, x)$ is the solution of the unforced heat equation. Denote

$$\Delta e^{\nu\Delta(t-\tau)}\xi(\tau) =: \Theta(\tau) = \sum \Theta_s(\tau)e_s\,.$$

Then $\Theta_s(\tau) = |2\pi s|^2 e^{-\nu|2\pi s|^2(t-\tau)}\xi_s(\tau)$, so

$$\begin{aligned}|\Theta_s(\tau)|\,|2\pi s|^m &= (\nu(t-\tau))^{-1+\frac{m_*-m}{2}}\Big(\nu(|2\pi s|^2(t-\tau))^{\frac{2+m-m_*}{2}}e^{-\nu|2\pi s|^2(t-\tau)}\Big)|2\pi s|^{m_*}|\xi_s(\tau)|\\ &\leq C(t-\tau)^{-1+\frac{m_*-m}{2}}|2\pi s|^{m_*}|\xi_s(\tau)|\,,\quad C = C(\nu, m, m_*),\end{aligned}$$

since $2+m-m_* \geq 0$ and $\sup_{r\geq 0} r^a e^{-r} < \infty$ if $a \geq 0$. Therefore we have

$$\|\Theta(\tau)\|_m^2 \leq C(t-\tau)^{-2+m_*-m}\sum |2\pi s|^{2m_*}|\xi_s(\tau)|^2 = C(t-\tau)^{-2+m_*-m}\|\xi(\tau)\|_{m_*}^2\,.$$

Hence,

$$\begin{aligned}\|v(t)\|_m &\leq \|u_0\|_m + \|\xi(t)\|_m + C_1\int_0^t \|\Theta(\tau)\|_m\,d\tau\\ &\leq \|u_0\|_m + C_2\|\xi\|_{\dot X_T^{m_*}}\big(1+\int_0^t \theta^{-1+\frac{m_*-m}{2}}\,d\theta\big) \leq \|u_0\|_m + C_3\|\xi\|_{\dot X_T^{m_*}}\,,\end{aligned}$$

since $m_* > m$, where C_1, C_2, C_3 depend on ν, m, m_*, T. That is,

$$\|v\|_{X_T^m} \leq \|u_0\|_m + C_3\|\xi\|_{\dot X_T^{m_*}}\,. \tag{D.1.5}$$

By continuity the solution mapping $(u_0, \xi) \mapsto v$ extends to a bounded linear operator, and $v(u_0, \xi)$ solves the stochastic heat equation for any $(u_0, \xi) \in H^m \times \dot X_T^{m_*}$ (see [1, Section 1.3] for details). □

D.2 Stationary measure

The only thing which we need to change in the definitions is that now our state space is a functional space.

In particular, the Markov semigroups are defined in the same way as in the SDE setting.

Existence of a stationary measure for the stochastic heat equation is again obtained by a Krylov-Bogolyubov argument. However, one needs to be careful

about the functional space in which one works, in order to use a compactity/tightness argument.
The uniqueness (and the explicit expression as a measure with Gaussian Fourier coefficients) follows from decoupling into Ornstein-Uhlenbeck processes on the different Fourier projections.

EXERCICE D.2.1. *Calculate this stationary measure. In which Sobolev spaces H^s, $s \in \mathbb{R}$, is it supported?*

E The stochastic Burgers equation

E.1 Well-posedness for the stochastic Burgers equation

DEFINITION E.1.1. *Consider $u_0 \in H^1$ and a Wiener process $\xi(t) \in H^2$. The* ***stochastic Burgers equation*** *is given by:*

$$u_t + uu_x - \nu u_{xx} = \xi_t,$$

where $\nu > 0$ is a constant.

The next objective is to prove the well-posedness of the Burgers equation in the space H^1, i.e. the fact that in the same sense as the stochastic equation considered above, it has a unique solution, and the corresponding mapping $(u_0, \xi(\cdot)) \mapsto u(\cdot)$ is continuous from $H^1 \times C(0, T; H^2)$ to $C(0, T; H^1)$ for all $T > 0$.

We will make the splitting $u = v + h$, where h solves the heat equation with forcing ξ_t. Note that it suffices to prove well-posedness for the equation satisfied by v. The proof is quite technical and can be found in [1, Chapter 1.3]: it consists of two parts, Theorem 1.3.6. for existence and uniqueness of solutions, and Lemma 1.3.11. for continuous dependence on the noise and the initial condition.
The fundamental idea is that of **Galerkin approximations**. It consists in writing the projection on the $2k$ lower Fourier modes of the Burgers equation, considering its solutions and passing to the limit $k \to \infty$ in an appropriate functional space.
One fundamental use of this projection is to allow us to apply Ito's formula, as we will see next.

Remark E.1.2. *One may generalise to smoother noise/initial conditions, following [1, Section 1.3]. The structure on the proof remains the same, but some calculations are longer.*

Theorem E.1.3. *Assume that $B_M < \infty$ for all M and u_0 is deterministic and C^∞-smooth. Then we have the energy balance:*

$$\frac{d}{dt}\mathbb{E}\tfrac{1}{2}\|u(t)\|^2 = -\nu\mathbb{E}\|u\|_1^2 + \tfrac{1}{2}B_0. \tag{E.1.1}$$

Moreover, we have:

$$\frac{d}{dt}\mathbb{E}\|u(t)\|_m^2 = -2\nu\mathbb{E}\|u\|_{m+1}^2 - \mathbb{E}\langle u, \partial_x u^2\rangle_m + B_m. \tag{E.1.2}$$

Proof. The Itô formula for finite-dimensional approximations of the Burgers equation gives us, after passing to the limit:

$$\mathbb{E}\|u(t)\|^2 - \mathbb{E}\|u(0)\|^2 = -\int_0^t \mathbb{E}\big(2\nu\|u\|_1^2 + \underbrace{\langle u, \partial_x u^2\rangle}_{0}\big)d\tau + tB_0,$$

and similarly

$$\mathbb{E}\|u(t)\|_m^2 - \mathbb{E}\|u(0)\|_m^2 = -\int_0^t \mathbb{E}\big(2\nu\|u\|_{m+1}^2 + \langle u, \partial_x u^2\rangle_m\big)d\tau + tB_m.$$

For details, see [1, Prop. 1.4.6]. □

E.2 Stationary measure for the stochastic Burgers equation

Lemma E.2.1. *For any $u_0 \in H^1$, the quantity*

$$\frac{1}{T}\int_0^T \|u(t)\|_2^2 dt$$

is bounded, uniformly in T.

Proof. See [2, Theorem 1.4.4.]. □

Using the estimate above and the compact injection of H^2 into H^1, one proves:

THEOREM E.2.2. *The Burgers equation defined as above admits a unique stationary measure on H^1.*

REMARK E.2.3. *If the noise is infinitely smooth, then the stationary measure also is, i.e. it is supported on the space $C^\infty = \cap_{m\geq 0} H^m$. This is due to the parabolic smoothing effect of the semigroup associated to the Burgers equation.*

For the uniqueness and the speed of convergence, one uses a pathwise contraction property in L_1 - an (almost) unique feature of the equation.

LEMMA E.2.4. *Consider two solutions of the Burgers equation with the same forcing and different initial values $u_{0,1}, u_{0,2}$. Then the difference between the corresponding solutions in L_1, $|u_1(t) - u_2(t)|_1$, is nonincreasing.*

Proof. See [2, Section 3.2.]. $\square$

This result is very important since it allows us to prove uniqueness of the stationary measure (under one additional technical assumption: the noise is assumed to be in H^4). We want to prove that for any measures μ_1, μ_2, the distance between $S_t^* \mu_1$ and $S_t^* \mu_2$ tends to 0. Since the semigroup S_t^* is linear, by Fubini's theorem it suffices to work with deterministic initial conditions, i.e. to consider Dirac measures $\mu_1 = \delta_{u_{0,1}}, \mu_2 = \delta_{u_{0,2}}$. Using once again Fubini's theorem, we observe that it suffices to prove that the distance between $u_1(t)$ and $u_2(t)$ with the same forcing tends to 0, almost surely. This follows from the definition of the dual-Lipschitz metric which quantifies weak convergence: indeed, the dominated convergence theorem would imply the convergence to 0 of

$$\begin{aligned} \sup_{\|f\|_{L(L_1)}\leq 1} |\langle f, S_t^* \delta_{u_{0,1}} \rangle - \langle f, S_t^* \delta_{u_{0,2}} \rangle| &= \sup_{\|f\|_{L(L_1)}\leq 1} \left| \mathbb{E}\Big[f(u_1(t)) - f(u_2(t)) \Big] \right| \\ &\leq \sup_{\|f\|_{L(\mathcal{O})}\leq 1} \mathbb{E} |f(u_1(t)) - f(u_2(t))|. \end{aligned}$$

To prove that the distance above tends to 0, we need to use a "small-noise argument". The idea is that almost surely, there is a large enough time interval on which the noise is small. Smallness of the noise implies that of solutions in L_1 - which are then stuck together forever because of the nonincreasing Lemma E.2.4. It remains to interpolate between L_1 and H^2 in order to have smallness in H^1 of the distance between solutions.
For details, see [2, Section 3.3.].

F Generalisations, miscellanea

F.1 More around the Burgers equation

Depending on the viscosity coefficient ν, one can estimate the rate of convergence to the stationary measure: if we work on the space L_1, which requires more careful functional analysis, there is a ν-independent bound for this rate. Moreover, one can measure more precisely the expected value of Sobolev norms, Fourier coefficients... and quantify sharply the dependence on ν.

REMARK F.1.1. *It is also relevant to study kicked processes, i.e. to consider the unforced PDE between integer time moments and adding independent identically distributed random impulsions (which do not have to be Gaussian!). The results are similar and parallel, but technically easier, compared to the ones described above. Indeed, one considers Markov chains instead of Markov processes.*
Moreover, there is a Donsker-type result of convergence of the solutions to the corresponding solutions for ***white noise*** *(i.e. the solutions we considered above, white noise being the weak derivative in time of a Wiener process). In other words, one takes kicks with frequency $1/N$, and then one rescales with the factor $1/\sqrt{N}$, like in the central limit theorem. Then the resulting process converges to the white-forced one, in a variety of ways. For more details, see [1, Chapter 10].*

REMARK F.1.2. *It is also very relevant to study the Burgers equation for $\nu = 0$...which is not at all parabolic/smoothing so its properties are very different. Nevertheless, some very useful properties (including convergence to a unique stationary measure) "pass to the limit".*

References

[1] A. Boritchev and S. Kuksin. *One-dimensional turbulence and the stochastic Burgers equation.* American Mathematical Society, 2021.

[2] S. Kuksin and A. Shirikyan. *Mathematics of Two-Dimensional Turbulence*, volume 194 of *Cambridge tracts in mathematics.* Cambridge University Press, 2012.