# SPDEs, classical and new

BY NICOLAS PERKOWSKI

Freie Universität Berlin

*July 15, 2020*

## Table of contents

# Introduction

Stochastic partial differential equations (SPDEs) are simply PDEs with "noise". Just as there are many different types of PDEs, there are at least as many different types of SPDEs. In this lecture we will only be interested in *semilinear parabolic SPDEs*, which are of the form

$$\partial_t u = \Delta u + f(u, \nabla u) + g(u, \nabla u)\xi, \qquad u(0) = u_0, \tag{1}$$

where $u: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ with $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ and $\xi$ is the noise. The equation is called *parabolic* because if $d = 1$ and if we interpret the highest order differential operators as polynomials and ignore the nonlinearities (i.e. $\partial_t \to t$ and $\Delta = \partial_{xx} \to x^2$), then the identity $t = x^2$ describes a parabola. It is called *semilinear* because it depends linearly on the highest order derivatives and the nonlinearities only involve lower order derivatives.

We will learn two approaches to SPDEs. First we discuss the classical approach from Walsh's Saint Flour notes [51]. This is based on an extension of Itô calculus to multiple parameters. We construct Itô integrals $\int_0^t \int_{\mathbb{R}^d} \varphi(s, x)\xi(ds, dx)$ using the same ideas as for the construction of the Itô integral $\int_0^t H_s dB_s$ for a Brownian motion $B$. This will allow us to solve SPDEs with *space-time white noise* (which we will define) in dimension $d = 1$ as long as $f$ and $g$ do not depend on $\nabla u$. We can also solve linear equations (linear $f$ and $g = 1$) in $d \geqslant 1$. If the noise is less singular than white noise (sufficiently "correlated" and not "white"), then we can treat equations in any dimension with Walsh's approach.

Then we will be interested in certain equations with white noise and $f$ depending on $\nabla u$ in $d = 1$, or nonlinear $f$ in $d > 1$, or in equations with a noise which only depends on space but which is constant in time. For these equations our previous tools break down. Therefore we will learn a pathwise approach to SPDEs which overcomes this problem. To understand the philosophy of the pathwise approach we will first learn some basic *rough path* theory, a pathwise approach to stochastic ordinary differential equations. Then we will extend the ideas of rough paths to higher dimensions, in the framework of *paracontrolled distributions* from [22], which combine techniques from Fourier analysis with the rough path philosophy and apply this to SPDEs.

## 1 Some motivating examples

**Example 1. (Interface growth)** SPDEs arise naturally as scaling limits of microscopic systems, similarly as Brownian motion arising as scaling limit of random walks. Consider for example an *exclusion process* on $\mathbb{Z}$, i.e. a system of particles which perform continuous time independent random walks with rate $p$ (resp. $1-p$) of jumping to the right (resp. the left), but which are not allowed to jump on top of each other; each site has at most one particle.

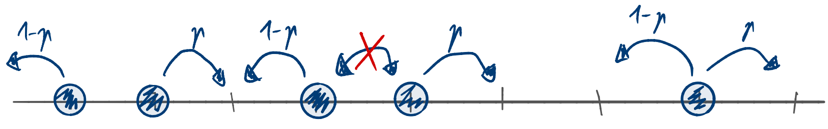


**Figure 1.** Simple exclusion process

If the random walks are symmetric, i.e. $p=\frac{1}{2}$, then on large scales this particle system resembles (converges to) the SPDE

$$\partial_t u = \frac{1}{2}\partial_{xx} u + \partial_x \xi,$$

on $\mathbb{R}_+ \times \mathbb{R}$, where $\xi$ is a space-time white noise (roughly speaking $\xi(t,x)$ is independent of $\xi(s,y)$ whenever $(t,x) \neq (s,y)$) and where $\partial_x$ is the derivative in the sense of distributions (we will learn how to interpret that). This equation is called the *infinite-dimensional Ornstein-Uhlenbeck process*. If we take a small perturbation around the symmetric jump rates and consider $p=\frac{1}{2}+\lambda\varepsilon$, with $\lambda \in \mathbb{R}$ and $\varepsilon \rightarrow 0$ as we scale out, then the scaling limit is given by the *stochastic Burgers equation*

$$\partial_t u = \frac{1}{2}\partial_{xx} u + \lambda \partial_x u^2 + \partial_x \xi.$$

Since formally $\partial_x u^2 = 2u\partial_x u$, this equation is of the type (1), with a very singular noise $\tilde{\xi} = \partial_x \xi$.

One motivation for studying this particle system is that it corresponds to a simple interface model: We can imagine a piecewise linear curve $(h(t,k): t \geqslant 0, k \in \mathbb{Z})$ over $\mathbb{Z}$, such that $h(t,k+1)-h(t,k)=1$ if there is a particle at site $k$, and $h(t,k+1)-h(t,k)=-1$ otherwise. We can convince ourselves that then local maxima become local minima with rate $p$, and local minima become local maxima with rate $1-p$.

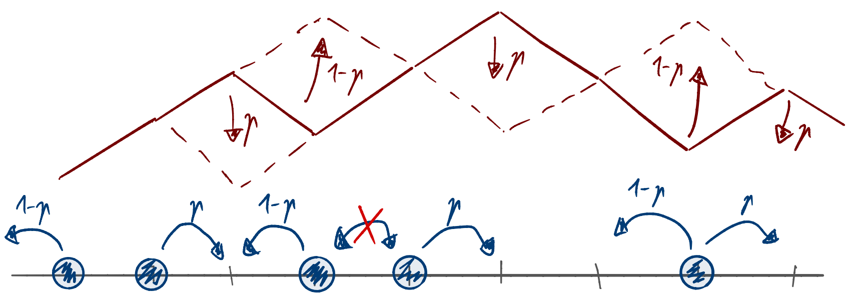


**Figure 2.** Simple exlusion as interface model

The oscillations of this random interface could for example be a toy model for snow crystals piling up on a windshield (see Figure 3). Except in that case up and down are not symmetric, so we would not expect to have $p=\frac{1}{2}$ but rather $p<\frac{1}{2}$. On the other hand if there is some melting, we also would not expect to have $p=0$. The large scale behavior for $p\in\left(0,\frac{1}{2}\right)$ is very difficult and not well understood yet. But if we take $p=\frac{1}{2}-\varepsilon$ ("slow growth"), then the scaling limit for the interface is the *KPZ equation*

$$\partial_t h=\frac{1}{2}\partial_{xx}h-(\partial_x h)^2+\xi.$$

By formally differentiating $h$ in $x$ we see that its derivative should solve the stochastic Burgers equation with $\lambda=\frac{1}{2}$, which is not surprising given how we constructed the interface.

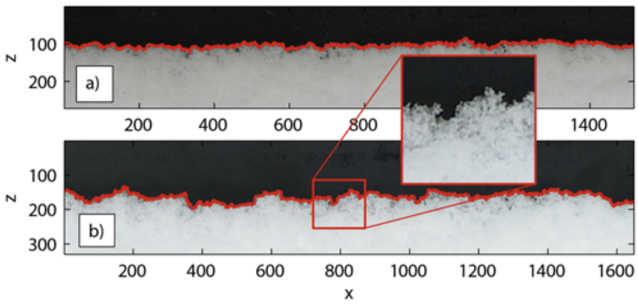


**Figure 3.** Snow crystals. Löwe et al., Geophys. Res. Letters, Vol. 34, L21507, 2007.

**Example 2. (Filtering)** Let $B,W$ be independent Brownian motions and consider a signal given by the solution to the SDE

$$dX_t=b(X_t)dt+dB_t,\qquad X_0\sim p(x)dx,$$

where $p$ is a probability density on $\mathbb{R}$. We would like to know the current state of $X_t$, but unfortunately we cannot observe $X$ directly, but only

$$Y_t=\int_0^t f(X_s)ds+W_t.$$

Our aim is to compute the conditional probability distribution of $X_t$ given the entire observation until now, $(Y_s)_{s\in[0,t]}$. That is, we want to compute $\mathbb{E}[\varphi(X_t)|(Y_s)_{s\in[0,t]}]$ for all continuous and bounded test functions $\varphi$. One can show that this conditional expectation is given by

$$\mathbb{E}[\varphi(X_t)|(Y_s)_{s\in[0,t]}]=\frac{\int_{\mathbb{R}}u(t,x)\varphi(x)dx}{\int_{\mathbb{R}}u(t,x)dx},$$

where $u$ solves the *Zakai equation*

$$\partial_t u=\frac{1}{2}\partial_{xx}u-\partial_x(bu)+fu\partial_t Y,\qquad u(0)=p.$$

By a Girsanov transform we can construct an equivalent probability measure under which $Y$ is a Brownian motion, we write $\tilde{W}$, and then the equation becomes

$$\partial_t u=\frac{1}{2}\partial_{xx}u-\partial_x(bu)+u\xi,\qquad u(0)=\varphi,$$

where formally $\xi(t,x)=f(x)\partial_t\tilde{W}(t)$.

**Example 3.** SPDEs also arise as scaling limits of population models: Consider independent continuous time random walkers on $\mathbb{Z}^d$, which can branch into new particles or die, according to a random landscape. More precisely, let $(\eta(x))_{x\in\mathbb{Z}^d}$ be an i.i.d. family of centered random variables with sufficiently many moments. If a particle is at site $x$ and $\eta(x)$ is positive, then we interpret this as a favorable environment and the particle can reproduce with rate $\eta(x)^+ = \max\{\eta(x), 0\}$. If the reproduction event happens, then the particle splits into two new independent particles, which both follow the same dynamics as the other particles. If however $\eta(x) < 0$, then the environment is unfavorable and the particle is being killed with rate $\eta(x)^-$. On large scales and for large numbers of particles this system approaches the *parabolic Anderson model*

$$\partial_t u = \frac{1}{2}\Delta u + u\xi,$$

on $\mathbb{R}_+ \times \mathbb{R}^d$, where $\xi$ is a *space white noise* and independent of time.

**Conventions and notation** Throughout these notes, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ is a filtered probability space satisfying the usual conditions. We write $a \lesssim b$ if there exists a constant $C > 0$, independent of $a$ and $b$, such that $a \leqslant Cb$. Similarly for $\gtrsim$ and $\simeq$. For example, it follows from Hölder's inequality that $(x+y)^p \leqslant 2^{p-1}(x^p + y^p)$ for $p \geqslant 1$ and $x, y \geqslant 0$, so we would write $(x+y)^p \lesssim x^p + y^p$. If we want to stress that the implicit constant depends on one of the (unimportant) variables, we denote it with a subscript. For example $(x+y)^p \lesssim_p x^p + y^p$.

# 1 Classical approach to SPDEs

Here we present the approach of Walsh [51]. It is based on a multi-parameter extension of Itô calculus, in which the time variable plays a distinguished role because it gives rise to the flow of information (filtration) and to the martingale structure.

## 1.1 Gaussian martingale measures and stochastic integration

Our first task is to mathematically model the noise appearing in the SPDEs. For simplicity we restrict our attention to noise on $\mathbb{R}_+ \times \mathbb{R}^d$, but we could replace $\mathbb{R}^d$ by any Polish space $(S, \mathcal{S})$ (for example if we want to solve an equation on a domain $D \subset \mathbb{R}^d$ we could take $(D, \mathcal{B}(D))$.

Recall that a signed measure $K$ is the difference of two (positive) measures $K^+, K^-$, that is $K = K^+ - K^-$, where at least one of the two is finite (we cannot have both $K^+(\mathbb{R}^d) = K^-(\mathbb{R}^d) = \infty$, because then $K(\mathbb{R}^d)$ would not be defined).

**Definition 1.1.** *Let $K$ be a signed measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$. We say that $K$ is* symmetric *if $K(A \times B) = K(B \times A)$ for all $A, B \in \mathcal{B}(\mathbb{R}^d)$. $K$ is* positive definite *if for each bounded measurable function $f$ for which the integral makes sense we have*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x) f(y) K(dx, dy) \geqslant 0.$$

*If $K$ is symmetric, positive definite, and $\sigma$-finite, and if there exists a symmetric, positive definite, and $\sigma$-finite positive measure $|K|$ with $|K(A \times B)| \leqslant |K|(A \times B)$, then we call $K$ a* covariance measure *and we define*

$$(f, g)_K := \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x) g(y) K(dx, dy), \qquad \|f\|_K := \sqrt{(f, f)_K}.$$

*We define $\mathcal{A}_K = \{A \in \mathcal{B}(\mathbb{R}^d) : K(A \times A) < \infty\}$. $\mathcal{A}_K$ is a ring of sets (closed under finite unions and $A, B \in \mathcal{A}_K \Rightarrow A \setminus B \in \mathcal{A}_K$), but this will not be important for us.*

**Remark 1.2.**

i. If $K$ is a positive measure, we can take $|K| = K$. Similarly, if $-K$ is a positive measure (so $K^+ = 0$), then we can take $|K| = -K$.

ii. In general $|K|$ will not be the total variation of $K$, because the total variation is not necessarily positive definite.

iii. The measure $|K|$ will be needed for the construction of the stochastic integral below. If we were not interested in stochastic integration, we could remove the assumption that $|K|$ exists.

iv. For $A, B \in \mathcal{A}_K$ we have

$$K(A \times B) = (\mathbb{1}_A, \mathbb{1}_B)_K$$

by definition.

v. Since $(f, g) \mapsto (f, g)_K$ is a symmetric and positive definite bilinear form, it satisfies the Cauchy-Schwarz inequality and the triangle inequality for the (semi-)norm:

$$(f, g)_K \leqslant \|f\|_K \|g\|_K, \qquad \|f + g\|_K \leqslant \|f\|_K + \|g\|_K.$$

If this is not clear to you, you should prove both of these. You could use the same proof as in $\mathbb{R}^n$.

Question: Can you interpret a covariance matrix in $\mathbb{R}^{n \times n}$ as a covariance measure, if we take a different measurable space than $\mathbb{R}^d$ in the definition of covariance measures?

Recall that a real valued stochastic process $(X_i)_{i \in I}$ is called *Gaussian* if for all $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $i_1, \ldots, i_n \in I$ the random variable $\alpha_1 X_{i_1} + \cdots + \alpha_n X_{i_n}$ is normally distributed.

**Definition 1.3.** *Let $K$ be a covariance measure. A centered Gaussian process*

$$(W_t(A) : (t, A) \in \mathbb{R}_+ \times \mathcal{A}_K)$$

*is called a* Gaussian martingale measure with covariance $K$ *if for all $A, B \in \mathcal{A}_K$:*

*i. $W_0(A) = 0$;*

*ii. $(W_t(A))_{t \geqslant 0}$ is a continuous martingale in the usual augmentation $(\mathcal{F}_t)_{t \geqslant 0}$ of*

$$\mathcal{F}_t^\circ = \sigma(W_s(B) : s \leqslant t, B \in \mathcal{A}_K);$$

*iii. $\mathbb{E}[W_s(A) W_t(B)] = (s \wedge t) K(A \times B)$ for all $s, t \geqslant 0$.*

**Lemma 1.4.** *Let $K$ be a covariance measure. Then there exists a Gaussian martingale measure with covariance $K$ and its law is unique.*

**Proof.** We know from Stochastics II or III that if $I$ is any index set and $\Gamma : I \times I \to \mathbb{R}$ is any positive definite covariance function, then there exists a Gaussian process with covariance $\Gamma$ and its law is unique. Here we have $I = \mathbb{R}_+ \times \mathcal{A}_K$ and

$$\Gamma((s, A), (t, B)) = s \wedge t K(A \times B) = \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{[0,t]}(r) \mathbb{1}_{[0,s]}(r) \mathbb{1}_A(x) \mathbb{1}_B(y) K(dx, dy) \right) dr.$$

Therefore,

$$
\begin{aligned}
&\sum_{i,j=1}^{n} \alpha_i \alpha_j \Gamma((t_i, A_i), (t_j, B_j)) \\
&= \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} \sum_{i,j=1}^{n} \alpha_i \alpha_j \mathbb{1}_{[0,t_i]}(r) \mathbb{1}_{[0,t_j]}(r) \mathbb{1}_{A_i}(x) \mathbb{1}_{A_j}(y) K(dx, dy) \right) dr \\
&= \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} \left( \sum_{i=1}^{n} \alpha_i \mathbb{1}_{[0,t_i]}(r) \mathbb{1}_{A_i}(x) \right) \left( \sum_{i=1}^{n} \alpha_i \mathbb{1}_{[0,t_i]}(r) \mathbb{1}_{A_i}(y) \right) K(dx, dy) \right) dr \geqslant 0.
\end{aligned}
$$

So let $(\tilde{W}_t(A))_{t \geqslant 0, A \in \mathcal{A}_K}$ be the Gaussian process with covariance $\Gamma$ which we obtain from Kolmogorov's extension theorem. Then we have for $s \leqslant t$ and for all $p \geqslant 1$, using the Gaussianity of $\tilde{W}_t(A) - \tilde{W}_s(A)$:

$$
\mathbb{E}[|\tilde{W}_t(A) - \tilde{W}_s(A)|^p] = \mathbb{E}[|X|^p] \mathbb{E}[|\tilde{W}_t(A) - \tilde{W}_s(A)|^2]^{p/2} = \mathbb{E}[|X|^p] |t - s|^{p/2} \|\mathbb{I}_A\|_K^p
$$

for a standard normal random variable $X$. So by Kolmogorov's continuity criterion the process $(\tilde{W}_t(A))_{t \geqslant 0}$ has a continuous modification $(W_t(A))_{t \geqslant 0}$. Moreover, for any $r_1, \ldots, r_n \in [0, s]$ and $B_1, \ldots, B_n \in \mathcal{A}_K$ we have

$$
\mathbb{E}[(W_t(A) - W_s(A)) W_{r_i}(B_i)] = (t \wedge r_i - s \wedge r_i) K(A \times B_i) = 0,
$$

so by Gaussianity $W_t(A) - W_s(A)$ is independent of $(W_{r_1}(B_1), \ldots, W_{r_n}(B_n))$. A monotone class argument shows that $W_t(A) - W_s(A)$ is even independent of $\mathcal{F}_s$ and therefore the martingale property follows. □

Question: Which $p$ could we use for Kolmogorov's continuity criterion in the proof? Which Hölder regularity does the process $(W_t(A))_{t \geqslant 0}$ have?

Note that point iii. of the definition implies that $W(A)$ is a multiple of a Brownian motion, more precisely $W(A)$ has the same distribution as $\sqrt{K(A \times A)} B$ for a Brownian motion $B$.

**Lemma 1.5.** *Let $W$ be a Gaussian martingale measure with covariance $K$. Let $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}_K$ be disjoint sets with $\bigcup_n A_n \in \mathcal{A}_K$. Then for all $t \geqslant 0$:*

$$
W_t\left( \bigcup_n A_n \right) = \sum_n W_t(A_n),
$$

*where the series on the right hand side converges in $L^2(\mathbb{P})$.*

**Proof.** Exercise 1.1. □

**Remark 1.6.** We could call $W_t$ an *$L^2$-valued measure*. But in general it is not true that $W_t(\omega, \cdot)$ is a measure for fixed $\omega$. Indeed, even if $\sum_n W_t(A_n)$ converges almost surely and not only in $L^2(\mathbb{P})$ (which we did not show), the identity

$$
W_t\left( \bigcup_n A_n \right) = \sum_n W_t(A_n)
$$

is only true outside of a null set. And the null set may depend on the sequence $(A_n)$, of which there are uncountably many. In most cases of interest there exists no modification of $W$ which is a measure.

**Remark 1.7.** We formally have

$$
\mathbb{E}[W_t(x) W_s(y)] = t \wedge s f(x, y),
$$

where $f(x, y)$ is the density of $K$ with respect to the Lebesgue measure. Of course, $K$ does not have to be absolutely continuous and $W_t(x)$ does not have to be defined for single points $x$. But if we ignore that and proceed formally, assuming that $W_t(x)$ is continuous in $x$, then we get with the ball $B(x, r) = \{y: |x - y| < r\}$:

$$\begin{aligned}\mathbb{E}[W_t(x)W_s(y)] &= \lim_{r\to 0}\frac{1}{|B(x,r)|^2}\mathbb{E}[W_t(B(x,r))W_s(B(y,r))]\\ &= \lim_{r\to 0}\frac{1}{|B(x,r)|^2} t\wedge s K(B(x,r)\times B(y,r)).\end{aligned}$$

If $K$ has a continuous density $f$ with respect to the Lebesgue measure, then the right hand side indeed converges to $t \wedge s f(x, y)$.

**Example 1.8.**

i. Consider $n$ independent one-dimensional Brownian motions $B^1, \ldots, B^n$ and let $\sigma_1, \ldots, \sigma_n \in C_b(\mathbb{R}^d)$ be such that for each $k$ either $\sigma_k(x) \geqslant 0$ for all $x \in \mathbb{R}^d$, or $\sigma_k \in L^1(\mathbb{R}^d)$. Then

$$W_t(A) := \int_{\mathbb{R}^d} \mathbb{1}_A(x) W_t(x) dx, \qquad W_t(x) := \sum_{k=1}^n \sigma_k(x) B_t^k,$$

is a Gaussian martingale measure (Exercise 1.2).

ii. The most important example for us is $K(dx, dy) = |K|(dx, dy) = \delta_x(dy) dx dy$ for the Dirac delta

$$\delta_x(A) := \begin{cases} 1, & x \in A, \\ 0, & \text{else.} \end{cases}$$

We mostly write

$$K(dx, dy) = \delta(x - y) dx dy$$

and interpret $\delta(x - y)$ as the "density" of $K$ (which of course does not exist, but formally it is $\infty$ if $x = y$ and 0 otherwise). In that case we call $W$ a *space-time white noise* (or rather the formal derivative $\partial_t W_t$ is called space-time white noise). Heuristically, we have

$$\mathbb{E}[\partial_t W_t(x) \partial_s W_s(y)] = \partial_t \partial_s (s \wedge t) \delta(x - y) = \delta(t - s) \delta(x - y),$$

i.e. $\partial_t W_t(x)$ and $\partial_s W_s(y)$ are independent unless $s = t$ and $x = y$ (and in that case the variance is infinite). So space-time white noise is completely decorrelated, the noise is independent in any two distinct space-time points. More rigorously, whenever

$$[s_1, t_1] \times A_1 \cap [s_2, t_2] \times A_2 = \emptyset,$$

then $W_{t_1}(A_1) - W_{s_1}(A_1)$ is independent of $W_{t_2}(A_1) - W_{s_2}(A_1)$.

iii. We will see later that we can define $W_t(f)$ for $f$ with $|||f|||_{|K|} < \infty$ by an extension argument. For now we take this for granted and use it to define a mollified space-time white noise: Let $W$ be a space-time white noise and let $\rho \in C_c^\infty(\mathbb{R}^d, \mathbb{R})$ be a positive function with $\int_{\mathbb{R}^d} \rho(x) dx = 1$. Then we define the mollification of $W$ as follows: Let $W$ be a space-time white noise and let for $\lambda(A) < \infty$ ($\lambda$ is Lebesgue measure)

$$V_t(A) := W_t\bigg(\int_A \rho(x - \cdot) dx\bigg),$$

which should be formally interpreted as

$$\int_A \bigg(\int_{\mathbb{R}^d} \rho(x - y) W_t(y) dy\bigg) dx = \int_A (\rho * W_t)(x) dx.$$

$V_t(A)$ is well defined because

$$\left\|\left|\int_A \rho(x-\cdot)dx\right|\right\|_{|K|}^2 = \int_{\mathbb{R}^d\times\mathbb{R}^d}\left|\int_A \rho(x-y_1)dx\right|\left|\int_A \rho(x-y_2)dx\right|\delta(y_1-y_2)dy_1\,dy_2$$
$$= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\mathbb{1}_A(x)\rho(x-y_1)dx\right)^2 dy_1$$
$$\leqslant \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbb{1}_A(x)^2\rho(x-y_1)dx\,dy_1 = \lambda(A),$$

where we applied Jensen's inequality which is allowed because $\rho(x-y_1)dx$ is a probability measure. This new Gaussian martingale measure has the covariance

$$(f,g)_K = \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(x)\rho(x-z)dx\right)\left(\int_{\mathbb{R}^d} g(y)\rho(y-z)dy\right)dz$$
$$= \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)g(y)(\rho*\tilde{\rho})(x-y)dx\,dy,$$

where $\tilde{\rho}(x)=\rho(-x)$ and $a*b(x)=\int_{\mathbb{R}^d}a(x-z)b(z)dz$. Since

$$(f,f)_K = \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(x)\rho(x-z)dx\right)^2 dz \geqslant 0,$$

$K$ is positive definite. We can take $|K|=K$, since $K$ is positive.

iv. For $\alpha\in[0,d)$ the fractional kernel

$$|K|(dx,dy) = K(dx,dy) = |x-y|^{-\alpha}dx\,dy$$

is positive definite (Exercise 1.2).

Next we want to construct stochastic integrals against Gaussian random measures. We take the same approach as in the construction of the usual Itô integral: We first define the integral for simple integrands, then prove Itô's isometry, and then extend the integral to general integrands by a continuity argument.

**Definition 1.9.** *The* bounded elementary processes $b\mathcal{E}$ *are all processes of the form*

$$H(s,x) = \sum_{k=0}^{n-1}\sum_{\ell=0}^{L_k} h_{k,\ell}\mathbb{I}_{(t_k,t_{k+1}]\times A_{k,\ell}}(s,x),$$

*where* $n\in\mathbb{N}$, $L_k\in\mathbb{N}$, $0\leqslant t_0<\ldots<t_n$, $h_{k,\ell}\in L^\infty(\mathcal{F}_{t_k})$, $A_{k,\ell}\in\mathcal{B}(\mathbb{R}^d)$.

*The* predictable $\sigma$-algebra $\mathcal{P}$ *is the smallest* $\sigma$*-algebra on* $\Omega\times\mathbb{R}_+\times\mathbb{R}^d$ *with respect to which all* $H\in b\mathcal{E}$ *are measurable. A stochastic process is* predictable *if it is measurable with respect to* $\mathcal{P}$.

**Definition 1.10.** *Let* $W$ *be a Gaussian martingale measure with covariance* $K$*. We define for a predictable process* $H$:

$$\|H\|_W^2 = \mathbb{E}\left[\int_0^\infty \||H(s,\cdot)|\|_{|K|}^2 ds\right] = \mathbb{E}\left[\int_0^\infty\int_{\mathbb{R}^d\times\mathbb{R}^d}|H(s,x)||H(s,y)||K|(dx,dy)ds\right],$$

*and we write*

$$L^2(W) = \{H \textit{ predictable s.t. } \|H\|_W<\infty\}.$$

One can show that after identifying predictable $H,\tilde{H}\in L^2(W)$ with $\|H-\tilde{H}\|_W=0$ the space $L^2(W)$ is a Banach space; see Exercise 2.5 in [51]. We will see in the Itô isometry that the norm

$$\mathbb{E}\Big[\int_0^\infty\int_{\mathbb{R}^d\times\mathbb{R}^d} H(s,x)H(s,y)K(dx,dy)ds\Big]$$

would be more canonical, but for proving the Banach space property and for showing the following density statement it is convenient to replace $K\to|K|$ and $H\to|H|$.

Since $H\in b\mathcal{E}$ could involve sets $A_k$ with $K(A_k\times A_k)=\infty$ it is not necessarily true that $b\mathcal{E}\subset L^2(W)$. But we can simply intersect with $L^2(W)$:

**Lemma 1.11.** *The set $b\mathcal{E}\cap L^2(W)$ is dense in $L^2(W)$.*

**Proof.** See Walsh [51], Proposition 2.3. □

**Definition 1.12.** *For $H\in b\mathcal{E}\cap L^2(W)$ we define*

$$\int_{[0,t]\times\mathbb{R}^d} H(s,x)W(ds,dx):=\sum_{k=0}^{n-1}\sum_{\ell=0}^{L_k} h_{k,\ell}(W_{t_{k+1}\wedge t}(A_{k,\ell})-W_{t_k\wedge t}(A_{k,\ell})),\qquad t\geqslant 0.$$

Since $W_t(A\cup B)=W_t(A)+W_t(B)$ for disjoint $A,B$ (by Lemma 1.5), it is not hard to check that this definition is independent of the specific representation

$$H(s,x)=\sum_{k=0}^{n-1}\sum_{\ell=0}^{L_k} h_{k,\ell}\mathbb{I}_{(t_k,t_{k+1}]\times A_{k,\ell}}(s,x)$$

of $H$.

**Proposition 1.13. (Itô isometry)** *Let $H\in b\mathcal{E}\cap L^2(W)$. Then*

$$M_t=\int_{[0,t]\times\mathbb{R}^d} H(s,x)W(ds,dx),\qquad t\geqslant 0,$$

*is a continuous square-integrable martingale with quadratic variation*

$$\langle M\rangle_t=\int_0^t\|H(s,\cdot)\|_K^2ds=\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}H(s,x)H(s,y)K(dx,dy)ds.$$

*In particular, Itô's isometry holds:*

$$\mathbb{E}\Big[\Big(\int_{\mathbb{R}_+\times\mathbb{R}^d}H(s,x)W(ds,dx)\Big)^2\Big]=\mathbb{E}\Big[\int_0^\infty\|H(s,\cdot)\|_K^2ds\Big]\leqslant\|H\|_W^2.$$

**Proof.** To simplify notation let us write $W^{k,\ell}=W(A_{k,\ell})$. $M$ is a finite sum of usual Itô integrals with respect to the martingales $W^{k,\ell}$. Therefore, it is a continuous local martingale and its quadratic variation is

$$\begin{aligned}\langle M\rangle_t=&\sum_{k_1,k_2=0}^{n-1}\sum_{\ell_1=0}^{L_{k_1}}\sum_{\ell_2=0}^{L_{k_2}}\Big\langle\int_0^\cdot h_{k_1,\ell_1}\mathbb{I}_{(t_{k_1},t_{k_1+1}]}(s)dW_s^{k_1,\ell_1},\int_0^\cdot h_{k_2,\ell_2}\mathbb{I}_{(t_{k_2},t_{k_2+1}]}(s)dW_s^{k_2,\ell_2}\Big\rangle_t\\=&\sum_{k=0}^{n-1}\sum_{\ell_1,\ell_2=0}^{L_k}\int_0^t h_{k,\ell_1}h_{k,\ell_2}\mathbb{I}_{(t_k,t_{k+1}]}(s)K(A_{k,\ell_1}\times A_{k,\ell_2})ds\\=&\int_0^t\|H(s,\cdot)\|_K^2ds,\end{aligned}$$

where we used that the intervals $(t_{k_1}, t_{k_1+1}]$ and $(t_{k_2}, t_{k_2+1}]$ are disjoint unless $k_1 = k_2$. Since $\|H\|_W < \infty$ by assumption, we get that $\mathbb{E}[\langle M\rangle_t]$ is bounded in $t \geqslant 0$ and therefore $M$ is a uniformly integrable and square-integrable martingale and Itô's isometry holds. □

Let $\mathcal{M}_c^2$ be the family of uniformly integrable continuous martingales $M$ such that

$$M_0 = 0 \qquad \text{and} \qquad \mathbb{E}[M_\infty^2] < \infty.$$

Here we identify two martingales if they are indistinguishable. We know from the stochastic analysis lecture (based on Doob's $L^2$-inequality) that $\mathcal{M}_c^2$ is a Hilbert space with inner product

$$(M, N)_{\mathcal{M}_c^2} = \mathbb{E}[M_\infty N_\infty] = \mathbb{E}[\langle M, N\rangle_\infty].$$

We have just shown that for $H \in b\mathcal{E} \cap L^2(W)$

$$\left\| \int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx) \right\|_{\mathcal{M}_c^2} \leqslant \|H\|_{L^2(W)}.$$

Hence, the map

$$\begin{aligned} J_W \colon b\mathcal{E} \cap L^2(W) &\longrightarrow \mathcal{M}_c^2, \\ H &\longmapsto \int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx), \end{aligned}$$

is Lipschitz continuous, and it is linear by construction

**Theorem 1.14. (Itô integral)** *There is a unique continuous linear map*

$$L^2(W) \ni H \mapsto \int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx) \in \mathcal{M}_c^2$$

*which extends $J_W$ from $b\mathcal{E} \cap L^2(W)$ to all of $L^2(W)$. We call $\int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx)$ the* stochastic integral *or* Itô integral *of $H$ with respect to $W$.*

**Proof.** By Lemma 1.11 we know that $b\mathcal{E} \cap L^2(W)$ is dense in $L^2(W)$. So let $H \in L^2(W)$ and let $(H^n) \subset b\mathcal{E} \cap L^2(W)$ be such that $\|H - H^n\|_{L^2(W)} \to 0$. Then by the Itô isometry $(\int_{[0,\cdot]\times\mathbb{R}^d} H^n(s,x) W(ds,dx))_n$ is a Cauchy sequence in $\mathcal{M}_c^2$ and therefore it converges to a limit $\int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx) \in \mathcal{M}_c^2$. Using once more the isometry we see that the definition of $\int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx)$ does not depend on the specific sequence $(H^n)$, and any other approximating sequence gives the same limit. Linearity, Itô isometry and Lipschitz continuity of $H \mapsto \int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx)$ are inherited from the linearity, Itô isometry and Lipschitz continuity of the map on $b\mathcal{E} \cap L^2(W)$. □

Let us formulate more explicitly what the previous theorem says:

**Corollary 1.15.** *For all $H \in L^2(W)$, there is a sequence $(H^n) \subset b\mathcal{E} \cap L^2(W)$ such that $H^n \to H$ in $L^2(W)$. And for any approximating sequence $H^n \to H$ in $L^2(W)$, we have $\int_{[0,\cdot]\times\mathbb{R}^d} H^n(s,x) W(ds,dx) \to \int_{[0,\cdot]\times\mathbb{R}^d} H(s,x) W(ds,dx)$ in $\mathcal{M}_c^2$. Moreover, Itô's isometry holds: For all $t \in [0,\infty]$ we have*

$$\mathbb{E}\left[\left(\int_{[0,t]\times\mathbb{R}^d} H(s,x) W(ds,dx)\right)^2\right] = \mathbb{E}\left[\int_0^t \|H(s,\cdot)\|_K^2 ds\right] \leqslant \|\mathbb{1}_{[0,t]} H\|_W^2.$$

It is possible to localize the integral using stopping times, but we will not need this.

**Theorem 1.16. (Burkholder-Davis-Gundy inequality)** *For all $p>0$ there exists a constant $C_p>0$ such that for all $H\in L^2(W)$ and for all $t\in[0,\infty]$:*

$$\frac{1}{C_p}\mathbb{E}\left[\left(\int_0^t \|H(s,\cdot)\|_K^2 ds\right)^{p/2}\right]\leqslant\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_{[0,s]\times\mathbb{R}^d} H(r,x)W(dr,dx)\right|^p\right]$$
$$\leqslant C_p\mathbb{E}\left[\left(\int_0^t \|H(s,\cdot)\|_K^2 ds\right)^{p/2}\right],$$

*where it may happen that one (and then both) of the terms is infinite.*

**Proof.** This is just the usual Burkholder-Davis-Gundy inequality for the continuous martingale $M_t=\int_{[0,t]\times\mathbb{R}^d}H(s,x)W(ds,dx)$ with quadratic variation $\langle M\rangle_t=\int_0^t\|H(s,\cdot)\|_K^2 ds$. $\square$

Question: Assume you wanted to localize the stochastic integral to extend it to more general $H$. How would you go about it (without working out the details)?

## 1.2 Weak and mild solutions

Let $W$ be a Gaussian martingale measure. Our aim is to solve the SPDE

$$\partial_t u=\Delta u+f(u)+g(u)\partial_t W,\qquad u(0)=u_0, \tag{1.1}$$

where purely formally we interpret $\partial_t W(t,x)$ as the density of $W$ with respect to Lebesgue measure, i.e.

$$\int_{\mathbb{R}_+\times\mathbb{R}^d} H(s,x)W(ds,dx)=\int_{\mathbb{R}_+\times\mathbb{R}^d} H(s,x)\partial_s W(s,x)ds dx.$$

If this is confusing, you can just see (1.1) as notation and our aim is to find a rigorous interpretation now.

As for stochastic (ordinary) differential equations we could try to interpret (1.1) as an integral equation:

$$u(t,x)=u(0,x)+\int_0^t \Delta u(s,x)ds+\int_0^t f(u(s,x))ds+\int_0^t g(u(s,x))W(ds,x),$$

for all $x\in\mathbb{R}^d$. But there are two problems with this: First of all, the stochastic integral on the right hand side looks weird, it is not of the form

$$\int_{[0,t]\times\mathbb{R}^d} H(s,x)W(ds,dx)$$

like the integrals which we constructed. And as mentioned before, in general $W$ is not actually of the form $W(\cdot,x)dx$ (absolutely continuous in $x$). Moreover, we will see later that in most interesting examples the solution $u$ to our equation is not actually differentiable, so $\Delta u$ makes no sense. The solution to both these problems is to consider weak solutions:

**Definition 1.17.** *A predictable stochastic process $u$ is called a* weak solution *to (1.1) if for all $\varphi\in C_c^\infty(\mathbb{R}^d)$ all of the following integrals are well defined and*

$$\begin{aligned}\int_{\mathbb{R}^d} u(t,x)\varphi(x)dx &=\int_{\mathbb{R}^d} u(0,x)\varphi(x)dx+\int_0^t\int_{\mathbb{R}^d} u(s,x)\Delta\varphi(x)dx ds\\ &\quad+\int_0^t\int_{\mathbb{R}^d} f(u(s,x))\varphi(x)dx ds\\ &\quad+\int_{[0,t]\times\mathbb{R}^d} g(u(s,x))\varphi(x)W(ds,dx).\end{aligned} \tag{1.2}$$

Note that if $u(s,\cdot)\in C^2$, then integration by parts gives

$$\int_{\mathbb{R}^d}\Delta u(s,x)\varphi(x)dx=\int_{\mathbb{R}^d}u(s,x)\Delta\varphi(x)dx,$$

and this motivates why we moved the Laplacian to $\Delta$ in the definition. Indeed, if we formally assume that $W(dt,dx)=\partial_t W(t,x)dx$ and that $u$ and $\partial_t W$ are smooth in space and time, then the definition becomes

$$\int_0^t\int_{\mathbb{R}^d}(\partial_s u(s,x)-\Delta u(s,x)-f(u(s,x))-g(u(s,x))\partial_s W(s,x))\varphi(x)dx\,ds=0$$

for all $\varphi\in C_c^\infty(\mathbb{R}^d)$, and this would then imply that

$$\partial_s u(s,x)-\Delta u(s,x)-f(u(s,x))-g(u(s,x))\partial_s W(s,x)=0$$

for Lebesgue-almost all $(s,x)$ (and thus for all $(s,x)$ if $u\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^d)$). So weak solutions (formally) generalize classical solutions.

The intution behind the definition is that we "test" the solution $u$ against the test function $\varphi$, and the result has to show the right behavior. For example we could imagine $u$ as a temperature and $\varphi$ as a thermometer which picks up the average temperature in some region.

Weak solutions make intuitive sense and the definition is more permissive than the classical definition and in particular we will see that, under appropriate conditions, weak solutions exist and are unique. But to show this, we will use an equivalent charaterization of weak solutions.

**Example 1.18.** Consider the equation

$$\partial_t v=\Delta v,\qquad v(0)=v_0.$$

By differentiating the following expression we see that

$$v(t,x)=P_t v_0(x)=\mathbb{E}[v_0(x+\sqrt{2}B_t)]=\int_{\mathbb{R}^d}p(t,x-y)v_0(y)dy=p_t*v_0(x),$$

where

$$p(t,x)=(4\pi t)^{-\frac{d}{2}}e^{-\frac{|x|^2}{4t}}$$

satisfies $\partial_t p=\Delta p$ and

$$p(0,\cdot)*v_0(x):=\lim_{t\to 0}p(t,\cdot)*v_0(x)=v_0(x)$$

if $v_0$ is continuous. Therefore, we formally get for "nice" $f$:

$$\begin{aligned}\partial_t\int_0^t\int_{\mathbb{R}^d}p(t-s,x-y)f(y)dy&=\int_{\mathbb{R}^d}p(0,x-y)f(y)dy+\int_0^t\int_{\mathbb{R}^d}\partial_t p(t-s,x-y)f(y)dy\\&=f(x)+\Delta\left(\int_0^t\int_{\mathbb{R}^d}p(t-s,x-y)f(y)dy\right).\end{aligned}$$

In other words, $w(t,x)=\int_0^t\int_{\mathbb{R}^d}p(t-s,x-y)f(y)dy$ solves

$$\partial_t w=\Delta w+f,\qquad w(0)=0.$$

Moreover, $u=v+w$ solves

$$\partial_t u=\Delta v+\Delta w+f=\Delta u+f,\qquad u(0)=v_0.$$

This is essentially the *variation of constants* formula from ODE theory, only applied in a more complicated context.

For the remainder of this section we use the notation

$$p(t,x) = (4\pi t)^{-\frac{d}{2}} e^{-\frac{|x|^2}{4t}}$$

and

$$P_t f(x) = p_t * f(x) = \int_{\mathbb{R}^d} f(x-y) p(t,y) dy,$$

and we call $(P_t)_{t \geqslant 0}$ the semigroup generated by $\Delta$.

**Definition 1.19.** *A predictable stochastic process $u$ is called a* mild solution *to (1.1) if all of the following integrals are well defined and*

$$\begin{aligned} u(t,x) &= \int_{\mathbb{R}^d} p(t,x-y) u_0(y) dy + \int_0^t \int_{\mathbb{R}^d} p(t-s,x-y) f(u(s,y)) dy ds \\ &\quad + \int_{[0,t]\times\mathbb{R}^d} p(t-s,x-y) g(u(s,y)) W(ds,dy). \end{aligned} \tag{1.3}$$

**Remark 1.20.** Under appropriate assumptions one can show that mild solutions and weak solutions are equivalent: $u$ is a weak solution if and only if it is a mild solution. See for example the proof of Theorem 3.2 in Walsh [51], or Proposition 3.2.2 in Pardoux [44].

Question: We formally have $P_t = e^{t\Delta}$. Replace $\Delta$ by a bounded linear operator (say multiplication with $a \in \mathbb{R}$) and convince yourself, that then

$$\partial_t u = a u + f, \qquad u(0) = u_0$$

is equivalent to

$$u(t) = e^{ta} u_0 + \int_0^t e^{(t-s)a} f(s) ds.$$

This is just the usual variation of constants formula.

## 1.3 A general existence-uniqueness result

Based on the mild formulation (1.3) we can now prove the existence and uniqueness of solutions to (1.1), under appropriate conditions.

**Definition 1.21.** *A predictable stochastic process $u$ is called* locally $L^2$-bounded *if for all $T \geqslant 0$*

$$\sup_{t\in[0,T], x\in\mathbb{R}^d} \mathbb{E}[u(t,x)^2] < \infty.$$

*Similarly, $(v(x))_{x\in\mathbb{R}^d}$ is called $L^2$-bounded if $\sup_x \mathbb{E}[v(x)^2] < \infty$.*

**Definition 1.22.** *Let $K$ be covariance measure. For $\alpha \geqslant 0$ we write $K \in \mathcal{K}^\alpha$ if for all $T > 0$ there exists $C(T) > 0$ with*

$$\sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d\times\mathbb{R}^d} p(t,x-y_1) p(t,x-y_2) |K|(dy_1,dy_2) \leqslant C(T) t^{-\alpha}, \qquad t \in [0,T].$$

We will see that we can solve (1.1) as long as $K \in \mathcal{K}^\alpha$ for some $\alpha \in [0,1)$ (and of course under appropriate conditions on $f$ and $g$ and $u_0$).

**Example 1.23.**

i. Let $K(dy_1, dy_2) = \delta_{y_1}(dy_2) dy_1$ (i.e. the covariance measure of the space-time white noise). Then $K = |K|$ and

$$\begin{aligned}\int_{\mathbb{R}^d \times \mathbb{R}^d} p(t, x - y_1) p(t, x - y_2) |K|(dy_1, dy_2) &= \int_{\mathbb{R}^d} p(t, y)^2 dy\\ &= \int_{\mathbb{R}^d} (4\pi t)^{-d} e^{-\frac{|y|^2}{2t}} dy\\ &= (2\pi t)^{d/2} (4\pi t)^{-d} \simeq t^{-d/2}.\end{aligned}$$

So $K \in \mathcal{K}^{d/2}$, and $d/2 < 1$ only for $d = 1$.

ii. For the mollified space-time white noise we have $|K| = K = (\rho * \tilde{\rho})(y_1 - y_2) dy_1 dy_2$ and

$$\begin{aligned}&\int_{\mathbb{R}^d \times \mathbb{R}^d} p(t, x - y_1) p(t, x - y_2) |K|(dy_1, dy_2)\\ &\leqslant \|\rho * \tilde{\rho}\|_\infty \int_{\mathbb{R}^d \times \mathbb{R}^d} p(t, y_1) p(t, y_2) dy_1 dy_2\\ &= \|\rho * \tilde{\rho}\|_\infty,\end{aligned}$$

so $K \in \mathcal{K}^0$.

iii. For the "finite-dimensional noise"

$$W_t(A) = \sum_{k=1}^n \int_{\mathbb{R}^d} \mathbb{1}_A(x) \sigma_k(x) dx B_t^k$$

we have $|K|(dy_1, dy_2) = \sum_{k=1}^n |\sigma_k(y_1) \sigma_k(y_2)| dy_1 \, dy_2$, and therefore $|K|(dy_1, dy_2) \lesssim dy_1 dy_2$, which as in the previous example yields $K \in \mathcal{K}^0$.

iv. The fractional kernel $K(dy_1, dy_1) = |y_1 - y_2|^{-\alpha} dy_1 dy_2$ with $\alpha \in (0, d)$ is in $\mathcal{K}^{\alpha/2}$ (Exercise 2.1).

For our existence and uniqueness proof we need the following simple bound:

**Lemma 1.24.** *For all $\alpha \in [0, 1)$ there exists a constant $C = C(\alpha)$ such that for all $\lambda > 0$:*

$$\int_0^t e^{2\lambda s} (t - s)^{-\alpha} ds \leqslant C e^{2\lambda t} \lambda^{\alpha - 1}, \qquad t \geqslant 0.$$

**Proof.** We have

$$\begin{aligned}\int_0^t e^{2\lambda s} (t - s)^{-\alpha} ds &= e^{2\lambda t} \int_0^t e^{-2\lambda s} s^{-\alpha} ds\\ &= e^{2\lambda t} \lambda^\alpha \int_0^t e^{-2\lambda s} (\lambda s)^{-\alpha} ds\\ &= e^{2\lambda t} \lambda^{\alpha - 1} \int_0^{\lambda t} e^{-2s} s^{-\alpha} ds,\end{aligned}$$

so that we can take $C = \int_0^\infty e^{-2s} s^{-\alpha} ds < \infty$ (here we need $\alpha < 1$, otherwise the singularity in $s = 0$ is non-integrable). □

**Theorem 1.25.** *Let $\alpha \in [0, 1)$ and let $W$ be a Gaussian martingale measure with covariance measure $K \in \mathcal{K}^\alpha$. Let $u_0$ be $\mathcal{F}_0 \otimes \mathcal{B}(\mathbb{R}^d)$-measurable and $L^2$-bounded. Let $f$ and $g$ be Lipschitz continuous. Then there exists a unique predictable process $u$ which is locally $L^2$-bounded and which is a mild solution to (1.1).*

**Proof.** For simplicity we assume that $C(T)$ from the definition of $\mathcal{K}^\alpha$ is uniformly bounded in $T$, so that for all $t>0$:

$$\sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d\times\mathbb{R}^d} p(t,x-y_1)p(t,x-y_2)|K|(dy_1,dy_2)\leqslant Ct^{-\alpha}.$$

For $\lambda\in\mathbb{R}$ we define the Banach space

$$\mathbb{M}_\lambda=\left\{u \text{ predictable s.t. } \|u\|_{\mathbb{M}_\lambda}=\sup_{t\geqslant 0,x\in\mathbb{R}^d} e^{-\lambda t}\mathbb{E}[|u(t,x)|^2]^{1/2}<\infty\right\}.$$

We will prove the theorem by a Picard iteration in $\mathbb{M}_\lambda$ for suitable $\lambda>0$: Define

$$\begin{aligned}(\Phi(u))(t,x):=&\int_{\mathbb{R}^d} p(t,x-y)u_0(y)dy+\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)f(u(s,y))dyds\\&+\int_{[0,t]\times\mathbb{R}^d} p(t-s,x-y)g(u(s,y))W(ds,dy).\end{aligned}$$

We first have to show that $\Phi$ maps $\mathbb{M}_\lambda$ to $\mathbb{M}_\lambda$. The first thing to show would be that $\Phi(u)$ is predictable, which can be done by an approximation argument and which we skip. Next, we have to show that $\Phi(u)$ has the right growth for its $L^2$-norm. Let $L\geqslant 0$ be a Lipschitz constant for both $f$ and $g$. Since $p(t,x-y)dy$ is a probability measure, Jensen's inequality applied to the first term and the Itô isometry applied to the third term yield

$$\begin{aligned}&\mathbb{E}[|(\Phi(u))(t,x)|^2]\\&\lesssim\mathbb{E}\left[\left|\int_{\mathbb{R}^d} p(t,x-y)u_0(y)dy\right|^2\right]+\mathbb{E}\left[\left|\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)f(u(s,y))dyds\right|^2\right]\\&\quad+\mathbb{E}\left[\left|\int_{[0,t]\times\mathbb{R}^d} p(t-s,x-y)g(u(s,y))W(ds,dy)\right|^2\right]\\&\leqslant\sup_{y\in\mathbb{R}^d}\mathbb{E}[|u_0(y)|^2]+\mathbb{E}\left[\left|\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)(|f(0)|+L|u(s,y)|)dyds\right|^2\right]\\&\quad+\mathbb{E}\left[\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d} p(t-s,x-y_1)p(t-s,x-y_2)g(u(s,y_1))g(u(s,y_2))K(dy_1,dy_2)ds\right].\end{aligned}$$

We treat the last two terms on the right hand side separately. For the middle term we apply Minkowski's inequality to pull the $L^2(\mathbb{P})$ norm inside the integral:

$$\begin{aligned}&\mathbb{E}\left[\left|\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)(|f(0)|+L|u(s,y)|)dyds\right|^2\right]\\&\leqslant\left(\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)(|f(0)|+L\|u(s,y)\|_{L^2(\mathbb{P})})dyds\right)^2\\&\leqslant\left(\int_0^t\int_{\mathbb{R}^d} p(t-s,x-y)(|f(0)|+L\|u\|_{\mathbb{M}_\lambda}e^{\lambda s})dyds\right)^2\\&\leqslant(t|f(0)|+L\|u\|_{\mathbb{M}_\lambda}C\lambda^{-1}e^{\lambda t})^2\lesssim e^{2\lambda t},\end{aligned}$$

if $\lambda>0$; here we used Lemma 1.24. For the stochastic integral we use that $K\in\mathcal{K}^\alpha$ and obtain

$$\begin{aligned}&\mathbb{E}\left[\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d} p(t-s,x-y_1)p(t-s,x-y_2)g(u(s,y_1))g(u(s,y_2))K(dy_1,dy_2)ds\right]\\&\lesssim\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d} p(t-s,x-y_1)p(t-s,x-y_2)|K|(dy_1,dy_2)(|g(0)|^2+L^2e^{2\lambda s}\|u\|_{\mathbb{M}_\lambda}^2)ds\\&\lesssim\int_0^t(t-s)^{-\alpha}(|g(0)|^2+L^2e^{2\lambda s}\|u\|_{\mathbb{M}_\lambda}^2)ds\lesssim|g(0)|^2t^{1-\alpha}+L^2\|u\|_{\mathbb{M}_\lambda}^2\lambda^{\alpha-1}e^{2\lambda t}\lesssim e^{2\lambda t},\end{aligned}$$

where we again applied Lemma 1.24 and we used that $\lambda > 0$. So for all $\lambda > 0$ we have $\Phi$: $\mathbb{M}_\lambda \to \mathbb{M}_\lambda$. Next we will see that $\Phi$ is a contraction for large enough $\lambda$, and therefore the claim follows from the Banach fixed point theorem.

Indeed, the same argument as above shows that

$$\begin{aligned}
&\mathbb{E}[|\Phi(u)(t,x) - \Phi(v)(t,x)|^2] \\
&\lesssim \mathbb{E}\left[\left|\int_0^t \int_{\mathbb{R}^d} p(t-s, x-y)(L|u(s,y) - v(s,y)|)dyds\right|^2\right] \\
&\quad + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} p(t-s, x-y_1)p(t-s,x-y_2)\mathbb{E}[Z(s,y_1,y_2)]K(dy_1,dy_2)ds,
\end{aligned}$$

where

$$\mathbb{E}[Z(s,y_1,y_2)] = \mathbb{E}[(g(u(s,y_1)) - g(v(s,y_1)))(g(u(s,y_2)) - g(v(s,y_2)))] \leqslant e^{2\lambda s} L^2 \|u-v\|_{\mathbb{M}_\lambda}^2.$$

As above, this yields

$$\mathbb{E}[|\Phi(u)(t,x) - \Phi(v)(t,x)|^2] \lesssim L^2\|u-v\|_{\mathbb{M}_\lambda}^2 \lambda^{-2} e^{2\lambda t} + L^2 \|u-v\|_{\mathbb{M}_\lambda}^2 \lambda^{\alpha-1} e^{2\lambda t}.$$

If $\lambda > 0$ is large enough (depending on $\alpha$ and $L$) we can bound the right hand side by $ce^{2\lambda t}\|u-v\|_{\mathbb{M}_\lambda}^2$ for some $c < 1$, and thus

$$\|\Phi(u) - \Phi(v)\|_{\mathbb{M}_\lambda} < \sqrt{c}\|u-v\|_{\mathbb{M}_\lambda},$$

i.e. $\Phi$ is a contraction.

Strictly speaking this only proves the existence and uniqueness of solutions in $\mathbb{M}_\lambda$, but there might be a locally $L^2$-bounded solution which grows too fast as $t \to \infty$ and is therefore not in $\mathbb{M}_\lambda$. But to rule out that possibility we could simply restrict to a compact time interval $[0,T]$ and use that any locally $L^2$-bounded predictable process restricted to $[0,T]$ is in $\mathbb{M}_\lambda([0,T])$ (with its obvious definition). Then we would get uniqueness on $[0,T]$ and since $T > 0$ is arbitrary we get uniqueness on $\mathbb{R}_+$. □

Question: Why did we need $\alpha \in [0,1)$? What goes wrong for $\alpha \geqslant 1$?

**Example 1.26.** This result does not cover any of the examples which we discussed in the introduction: For the KPZ equation

$$\partial_t h = \frac{1}{2}\partial_{xx} h - (\partial_x h)^2 + \xi.$$

we have the term $-(\partial_x h)^2$ which does not fit into Theorem 1.25. For the Zakai equation

$$\partial_t u = \frac{1}{2}\partial_{xx} u - \partial_x(bu) + u\xi$$

we have the term $-\partial_x(bu)$ which does not fit into Theorem 1.25. For the parabolic Anderson model

$$\partial_t u = \frac{1}{2}\Delta u + u\xi,$$

the space white noise $\xi$ does not depend on time, so it does not correspond to a Gaussian martingale measure.

i. However, things are not as bad as they look. For the KPZ equation we can use a trick: For simplicity we normalize the coefficients of the equation (set them all equal to $+1$), and we assume that

$$\partial_t h = \partial_{xx} h + (\partial_x h)^2 + \xi. \tag{1.4}$$

Let us pretend that $h$ and $\xi$ are smooth functions of space and time. Then

$$u = e^h$$

solves

$$\partial_t u = e^h \partial_t h = u(\partial_{xx} h + (\partial_x h)^2 + \xi),$$

and

$$u\partial_{xx} h = \partial_{xx} u - u(\partial_x h)^2,$$

leading to

$$\partial_t u = \partial_{xx} u + u\xi.$$

Now recall that $\xi$ is a space-time white noise, so we can interpret the equation for $u$ using Theorem 1.25 as

$$\partial_t u = \Delta u + u\partial_t W, \qquad u: \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}.$$

So $f = 0$ and $g(u) = u$ are Lipschitz continuous functions, and we are indeed in the setting of Theorem 1.25 as long as

$$\sup_{x\in\mathbb{R}} \mathbb{E}[u_0(x)^2] = \sup_{x\in\mathbb{R}} \mathbb{E}[e^{2h_0(x)}] < \infty.$$

(And in fact it is possible to extend Theorem 1.25 to allow for much more general initial conditions, see Theorem 2.4 in [48]). One can then show that almost surely $u(t,x) > 0$ for all $t, x$, and therefore we could define $h := \log u$.

The map $h \mapsto u = e^h$ is called the *Cole-Hopf transformation* and it was applied to the KPZ equation already by Bertini and Giacomin in [7]. This approach is not completely satisfactory, because through the transformation we avoided talking about what we mean by a solution to (1.4). The intrinsic solution theory to KPZ is much younger and it goes back to Hairer [27]. We will learn later in the course about Hairer's approach, for now let us just mention that while the Cole-Hopf transform does not define an equation for $h$ but only the solution, it turns out that it indeed gives the "physically meaningful" solution.

ii. The Zakai equation can be covered by slightly generalizing Theorem 1.25 and allowing for an additional term:

$$\partial_t u(t,x) = \Delta u(t,x) + f(x, u(t,x)) + \nabla \cdot h(x, u(t,x)) + g(x, u(t,x))\partial_t W,$$

see Exercise Sheet 3.

iii. But for the PAM there is nothing we can do, it is completely outside the realm of Theorem 1.25.

(Difficult) question: Consider the $\Phi_1^4$ equation:

$$\partial_t u = \Delta u - u^3 + \partial_t W,$$

where $u: \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and $\partial_t W$ is a space-time white noise. Can you solve this equation with Theorem 1.25? If not, do you have any ideas how we could try to "locally" solve the equation?

## 1.4 Regularity of the solutions

Let us study the space-time regularity of the solution. To simplify the computations we restrict to $g = 1$ and $f = 0$. But the result remains valid for general Lipschitz continuous $f$ and $g$, as long as $u_0$ is $L^q$-bounded for all $q \geqslant 1$ (see [51], Corollary 3.4). Note that in our special case the solution is explicit:

$$u(t,x) = (p(t) * u_0)(x) + \int_{[0,t]\times\mathbb{R}^d} p(t-s, x-y) W(ds, dy).$$

Since $p\in C^{\infty}((0,\infty),\mathbb{R}^d)$ (with a singularity at $t=0$) we have

$$\partial_t^{\nu}\partial_x^{\mu}(p(t)*u_0)(x)=(\partial_t^{\nu}\partial_x^{\mu}p(t)*u_0)(x),$$

and therefore $(p(t)*u_0)\in C^{\infty}((0,\infty),\mathbb{R}^d)$ (with a singularity at $t=0$ unless $u_0\in C^{\infty}(\mathbb{R}^d)$). Therefore, the regularity is dictated by the stochastic term

$$Z(t,x):=\int_{[0,t]\times\mathbb{R}^d}p(t-s,x-y)W(ds,dy),\qquad (t,x)\in\mathbb{R}_+\times\mathbb{R}^d,$$

which we also call the *stochastic convolution*.

The compute the regularity of $Z$ we need the following simple observation, which is often useful, not just in this section.

**Lemma 1.27. (Interpolation)** *Let $0\leqslant a\leqslant\min\{b,c\}$. Then we have for all $\lambda\in[0,1]$:*

$$a=a^{\lambda}a^{1-\lambda}\leqslant b^{\lambda}c^{1-\lambda}.$$

**Lemma 1.28.** *Let $K\in\mathcal{K}^{\alpha}$ for $\alpha\in[0,1)$. Then $Z$ is a Gaussian process and we have for all $q\geqslant 1$, $\beta\in(0,1-\alpha)$, and $T>0$:*

$$\mathbb{E}[|Z(t+s,x+y)-Z(t,x)|^q]^{1/q}\lesssim s^{\beta/2}+|y|^{\beta},\qquad (t,x),(s,y)\in[0,T]\times\mathbb{R}^d.$$

**Proof.** We only consider the special case $K(dz_1,dz_2)=\delta_{z_1}(dz_2)dz_1$ in $d=1$, for which $\alpha=1-\alpha=\frac{1}{2}$. The general case follows from similar arguments, but the computations become even more involved; see Appendix A.

$Z$ is the stochastic integral of $W$ against the deterministic function $p$, and therefore it is Gaussian (if we replace $p$ by a deterministic bounded elementary function then this follows from the Gaussianity of $W$, and it extends to $p$ by a limiting argument). Since $Z$ is centered Gaussian, the $q$-th moment is bounded by the $q/2$-th power of the second moment and we get

$$\begin{aligned}\mathbb{E}[|Z(t+s,x+y)-Z(t,x)|^q]^{1/q}\lesssim&\,\mathbb{E}[|Z(t+s,x+y)-Z(t+s,x)|^2]^{1/2}\\&+\mathbb{E}[|Z(t+s,x)-Z(t,x)|^2]^{1/2}.\end{aligned}$$

In the nonlinear case with nontrivial $f,g$ we would use the Burkholder-Davis-Gundy inequality instead of Gaussianity. The first term on the right hand side is bounded by

$$\begin{aligned}&\mathbb{E}[|Z(t+s,x+y)-Z(t+s,x)|^2]\\&=\int_0^{t+s}\int_{\mathbb{R}}(p(t+s-r,x+y-z)-p(t+s-r,x-z))^2dzdr\\&=\int_0^{t+s}\int_{\mathbb{R}}(p(r,z+y)-p(r,z))^2dzdr.\end{aligned}\tag{1.5}$$

By a Taylor expansion together with Exercise 2.3 we have for $c>1$:

$$\begin{aligned}|p(r,z+y)-p(r,z)|=&\left|\int_0^1\nabla p(r,z+\lambda y)\cdot y d\lambda\right|\\ \lesssim&\, r^{-1/2}|y|\int_0^1 p(cr,z+\lambda y)d\lambda.\end{aligned}$$

On the other hand we also have the trivial bound

$$|p(r,z+y)-p(r,z)|\leqslant p(r,z+y)+p(r,z),$$

and therefore by interpolation with $\beta \in [0,1]$:

$$
\begin{aligned}
&|p(r,z+y)-p(r,z)|\\
&\lesssim (r^{-1/2}|y|)^{\beta}\Big(\int_0^1 p(cr,z+\lambda y)d\lambda\Big)^{\beta}(p(r,z+y)+p(r,z))^{1-\beta}.
\end{aligned}
$$

We plug this back into (1.5) and obtain

$$
\begin{aligned}
&\mathbb{E}[|Z(t+s,x+y)-Z(t+s,x)|^2]\\
&\lesssim \int_0^{t+s}\int_{\mathbb{R}}\Big(r^{-1/2}|y|\int_0^1 p(cr,z+\lambda y)d\lambda\Big)^{2\beta}(p(r,z+y)+p(r,z))^{2-2\beta}dzdr\\
&\lesssim \int_0^{t+s} r^{-\beta}|y|^{2\beta}\Big(\int_{\mathbb{R}}\Big(\int_0^1 p(cr,z+\lambda y)d\lambda\Big)^2 dz\Big)^{\beta}\Big(\int_{\mathbb{R}}(p(r,z+y)+p(r,z))^2 dz\Big)^{1-\beta}dr,
\end{aligned}
$$

where we applied Hölder's inequality with $m=\frac{1}{\beta}$ and $m'=\frac{1}{1-\beta}$. Now recall that

$$
\int_{\mathbb{R}} p(t,z)^2 dz \lesssim t^{-\frac{1}{2}},
$$

and therefore

$$
\mathbb{E}[|Z(t+s,x+y)-Z(t+s,x)|^2] \lesssim \int_0^{t+s} r^{-\beta}|y|^{2\beta}r^{-\frac{1}{2}}dr \lesssim (2T)^{1-\beta-\frac{1}{2}}|y|^{2\beta} \lesssim |y|^{2\beta},
$$

if $\beta+\frac{1}{2}<1$, i.e. $\beta<\frac{1}{2}$. Similarly, we get for the time difference:

$$
\begin{aligned}
&\mathbb{E}[|Z(t+s,x)-Z(t,x)|^2]\\
&= \int_0^s\int_{\mathbb{R}} p(r,z)^2 dzdr + \int_0^t\int_{\mathbb{R}}(p(r+s,z)-p(r,z))^2 dzdr\\
&\lesssim \int_0^s r^{-\frac{1}{2}}dr + \int_0^t\int_{\mathbb{R}}\Big(\frac{s}{r}\int_0^1 p(c(r+\lambda s),z)d\lambda\Big)^{2\beta}(p(r+s,z)+p(r,z))^{2-2\beta}dzdr\\
&\lesssim s^{\frac{1}{2}} + \int_0^t\Big(\frac{s}{r}\Big)^{\beta}\Big(\int_{\mathbb{R}}\Big(\int_0^1 p(c(r+\lambda s),z)d\lambda\Big)^2 dz\Big)^{\beta}\\
&\qquad\qquad \times\Big(\int_{\mathbb{R}}(p(r+s,z)+p(r,z))^2 dz\Big)^{1-\beta}dr\\
&\lesssim s^{\frac{1}{2}} + \int_0^t s^{\beta}r^{-\beta-\frac{1}{2}}dr \lesssim T^{\frac{1}{2}-\beta}s^{\beta} + T^{1-\beta-\frac{1}{2}}s^{\beta} \lesssim s^{\beta}.
\end{aligned}
$$

This concludes the proof. □

Question: Why did we get less space regularity than time regularity?

**Remark 1.29.** Note the difference between the time regularity and the space regularity: $Z$ is "twice as regular" in the space variable ($\beta$-regularity as a function in $L^p$) as in the time-variable ($\frac{\beta}{2}$-regularity). If we trace back where this difference comes from, we see that it is due to the fact that by Exercise 2.3 for all $c>1$:

$$
|\partial_{x_i}p(t,x)| \lesssim t^{-\frac{1}{2}}p(ct,x), \qquad |\partial_t p(t,x)| \lesssim t^{-1}p(ct,x).
$$

So taking a time derivative gives us a singularity which is twice as bad as the singularity we get by taking a space derivative. Intuitively, this makes sense: convolving against $p$ in space-time essentially means that we invert the operator $\partial_t-\Delta$, taking $(\partial_t-\Delta)^{-1}$. And this should gain us two space derivatives, but only one time derivative. So time derivatives should "count twice as much" as space derivatives. Alternatively, we also see from the identity $\partial_t p=\Delta p$ that one time derivative "counts two space derivatives".

This also motivates the definition of the "parabolic distance" in Remark 1.32 below.

To obtain almost sure sample path properties from the moment bound in Lemma 1.28, we need a multi-parameter version of the Kolmogorov continuity criterion:

**Theorem 1.30.** *Let $(X_t)_{t\in\mathbb{R}^n}$ be a stochastic process indexed by $\mathbb{R}^n$ and assume that for all $m>0$ there exists a constant $C(m)>0$ such that*

$$\mathbb{E}[|X_t - X_s|^p]^{1/p} \leqslant C(m)|t-s|^\alpha, \qquad |s|,|t|\leqslant m,$$

*where $\alpha p > n$. Then $X$ has a continuous modification $\tilde{X}$ such that for all $m>0$ and all $\gamma\in\left(0,\alpha-\frac{n}{p}\right)$:*

$$\sup_{|s|,|t|\leqslant m} \frac{|\tilde{X}_t(\omega)-\tilde{X}_s(\omega)|}{|t-s|^\gamma} \leqslant \tilde{C}_m(\omega), \qquad \omega\in\Omega.$$

**Proof.** See [51], Theorem 1.2. □

The theorem remains true if $t=(t_1,\ldots,t_n)\in\mathbb{R}_+\times\mathbb{R}^{n-1}$.

**Corollary 1.31.** *Let $K\in\mathcal{K}^\alpha$. Then the solution $u$ to*

$$u(t,x) = (p(t)*u_0)(x) + \int_{[0,t]\times\mathbb{R}^d} p(t-s,x-y)W(ds,dy)$$

*restricted to $(0,\infty)\times\mathbb{R}^d$ has a continuous modification which is almost surely in*

$$C^{\gamma/2}_{\mathrm{loc}}((0,\infty)\times\mathbb{R}^d) := \{f\in C((0,\infty)\times\mathbb{R}^d) \colon f\chi\in C^{\gamma/2}((0,\infty)\times\mathbb{R}^d)\,\forall\chi\in C^\infty_c((0,\infty)\times\mathbb{R}^d)\},$$

*for all $\gamma<1-\alpha$.*

**Proof.** As discussed above we have $p(t)*u_0\in C^\infty((0,\infty)\times\mathbb{R}^d)$, so we only have to deal with $Z$. Lemma 1.28 yields for $s,t\in[0,m]$ and $|x|,|y|\leqslant m$ and $p\geqslant 1$:

$$\mathbb{E}[|Z(t,x)-Z(s,y)|^p]^{1/p} \lesssim |t-s|^{\beta/2} + |x-y|^\beta \lesssim |t-s|^{\beta/2} + |x-y|^{\beta/2} \lesssim |(t,x)-(s,y)|^{\beta/2},$$

for any $\beta<1-\alpha$. For $p\beta/2>d+1$ we get a locally $\frac{\beta}{2}-\frac{d+1}{p}-\varepsilon$ Hölder continuous modification, and since $p\geqslant 1$ is arbitrary the claim follows. □

**Remark 1.32.** In the corollary we "threw away" some space regularity of the $L^p$ norm of $Z$, by estimating

$$\mathbb{E}[|Z(t,x)-Z(t,y)|^p]^{1/p} \lesssim |x-y|^\beta \lesssim |x-y|^{\beta/2}$$

for $x,y$ in a bounded set. We could introduce the following *parabolic metric*:

$$d((t,x),(s,y)) := |t-s|^{1/2} + |x-y|.$$

With this metric we would get

$$\mathbb{E}[|Z(t,x)-Z(s,y)|^p]^{1/p} \lesssim d((t,x),(s,y))^\beta,$$

and then a slight refinement of Kolmogorov's continuity criterion would yield almost surely

$$|u(t,x)-u(s,y)| \lesssim d((t,x),(s,y))^\gamma,$$

locally uniformly in $(t,x),(s,y)$ and for all $\gamma<1-\alpha$.

Question: How would you define the equivalent of the parabolic distance if we would consider the SPDE

$$\partial_t u = \Delta^2 u + \partial_t W?$$

**Remark 1.33.** In the specific example of the space-time white noise in $d=1$ we get local $\frac{1}{2}-\varepsilon$ Hölder continuity in space and local $\frac{1}{4}-\varepsilon$ Hölder continuity in time. Therefore, for fixed $x\in\mathbb{R}$ the process

$$(u(t,x))_{t\geqslant 0}$$

is more irregular than a typical semimartingale path (which would be locally $\frac{1}{2}-\varepsilon$ Hölder continuous like the Brownian motion). And in fact this process is *not* a semimartingale, and in particular we cannot apply Itô's formula to $f(u(x,t))$. This is a problem for applications, but there are some alternative change of variable formulas which may be useful in different contexts. See Bellingeri [6] for various such results.

While $(u(t,x))_t$ for fixed $x$ is not a semimartingale, $(\int_{\mathbb{R}} u(t,x)\varphi(x)dx)_{t\geqslant 0}$ is a semimartingale for any $\varphi\in C_c^\infty(\mathbb{R})$ because $u$ solves the weak formulation (1.2) of the SPDE.

Question: How can it be that $Z$ is a stochastic integral but not a local martingale?

## 1.5 A glimpse into the variational approach

**Example 1.34.** Let us consider again the filtering problem which lead us to the Zakai equation, but now we allow the observation noise to be correlated with the noise in the signal dynamics:

$$dX_t = b(X_t)dt+\sigma_1 dB_t+\sigma_2 dW_t, \qquad X_0\sim u_0(x)dx,$$
$$Y_t = \int_0^t f(X_s)ds+W_t,$$

where $\sigma_1^2+\sigma_2^2=1$ and $\sigma_1^2>0$ (so $\sigma_2^2<1$) and where we recall that $B$ and $W$ are independent Brownian motions. In that case we still have

$$\mathbb{E}[\varphi(X_t)|\mathcal{F}_t^Y] = \frac{\int_{\mathbb{R}}\varphi(x)u(t,x)dx}{\int_{\mathbb{R}}u(t,x)dx},$$

but now $u$ solves a more complicated version of the Zakai equation:

$$\partial_t u = \frac{1}{2}\partial_{xx}u-\partial_x(bu)+(fu-\sigma_2\partial_x u)\partial_t Y, \qquad u(0)=u_0,$$

which under an equivalent probability measure becomes

$$\partial_t u = \frac{1}{2}\partial_{xx}u-\partial_x(bu)+(fu-\sigma_2\partial_x u)\partial_t V,$$

for a one-dimensional Brownian motion $V$. Let us set $b=f=0$, because we already know how to deal with those terms. Then we remain with

$$\partial_t u = \frac{1}{2}\partial_{xx}u+\sigma\partial_x u\partial_t W,$$

where for simplicity of notation we now write $-\sigma_2\to\sigma$ and $V\to W$, and where we recall that $\sigma^2<1$ which will be important later. If we try to solve this equation with the Walsh approach, we immediately run into a serious problem: For $u\in\mathbb{M}_\lambda$, where $\mathbb{M}_\lambda$ was defined in the proof of Theorem 1.25, the natural definition of $\Phi(u)$ is

$$\Phi(u)(t,x) = p(t)*u_0(x)+\int_0^t\left(\int_{\mathbb{R}}p(t-s,x-y)\partial_x u(s,y)dy\right)dW_s$$
$$= p(t)*u_0(x)+\int_0^t\left(\int_{\mathbb{R}}\partial_x p(t-s,x-y)u(s,y)dy\right)dW_s,$$

where we used that $W$ does not depend on the space variable. But then Itô's isometry gives

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_0^t\Big(\int_{\mathbb{R}}\partial_x p(t-s,x-y)u(s,y)dy\Big)dW_s\Big)^2\Big]\\&=\int_0^t\mathbb{E}\Big[\Big(\int_{\mathbb{R}}\partial_x p(t-s,x-y)u(s,y)dy\Big)^2\Big]ds\\&\lesssim\int_0^t\Big(\int_{\mathbb{R}}\|\partial_x p(t-s,x-y)u(s,y)\|_{L^2(\mathbb{P})}dy\Big)^2ds\\&\lesssim\int_0^t\Big(\int_{\mathbb{R}}|\partial_x p(t-s,x-y)|dy\Big)^2\|u\|_{\mathbb{M}_\lambda}^2e^{2\lambda s}ds.\end{aligned}$$

By Exercise 2.3 we have

$$\int_{\mathbb{R}}|\partial_x p(t-s,x-y)|dy\lesssim(t-s)^{-\frac{1}{2}}\int_{\mathbb{R}}p(c(t-s),x-y)dy=(t-s)^{-\frac{1}{2}},$$

and this estimate is sharp up to multiplication by a constant (you can check that by explicitly writing out $\partial_x p$). So we get

$$\int_0^t\Big(\int_{\mathbb{R}}|\partial_x p(t-s,x-y)|dy\Big)^2\|u\|_{\mathbb{M}_\lambda}^2e^{2\lambda s}ds\lesssim\int_0^t(t-s)^{-1}\|u\|_{\mathbb{M}_\lambda}^2e^{2\lambda s}ds=\infty.$$

And unfortunately, all of our estimates were quite sharp. The only possibly non-sharp estimate was when we pulled the $L^2(\mathbb{P})$ norm inside the time integral. Instead of doing that we could write a double integral. Then we could get a smaller upper bound if we were able to say that $\mathbb{E}[u(s,y_1)u(s,y_2)]$ becomes small when $|y_1-y_2|\gg 1$. But it seems unclear how to obtain such information about $u$, so we are stuck.

If on the other hand we knew a priori that $\mathbb{E}[|\partial_y u(t,y)|^2]$ is bounded, then we would not have to move the derivative $\partial_x$ onto $p$ and could instead write the integral as $\int_0^t(\int_{\mathbb{R}}p(t-s,x-y)\partial_y u(s,y)dy)dW_s$. This suggests that an approach which directly analyzes the regularity of $u$ and not just the "amplitude" $\mathbb{E}[|u(t,x)|^2]$ might be more suitable for this equation.

We take Example 1.34 as motivation to have a brief look into another method for solving SPDEs, the so called *variational approach*, which is exposed in the monograph [32] and a very accessible introduction is given in the lecture notes [44]. We limit our discussion to our example of the Zakai equation for correlated noise. To make it even simpler, we consider the equation on a bounded domain with periodic boundary conditions:

**Definition 1.35.** *The* $d$-dimensional torus *is* $\mathbb{T}^d=(\mathbb{R}/\mathbb{Z})^d=\mathbb{R}^d/\mathbb{Z}^d$*. We can identify any function* $f:\mathbb{T}^d\to\mathbb{R}$ *with a* 1-periodic *function* $\bar{f}:\mathbb{R}^d\to\mathbb{R}$*, i.e. an* $f$ *for which*

$$\bar{f}(x+k)=f(x),\qquad k\in\mathbb{Z}^d.$$

*Conversely, any* 1*-periodic function* $\bar{f}:\mathbb{R}^d\to\mathbb{R}$ *is canonically associated with a function* $f:\mathbb{T}^d\to\mathbb{R}$ *by setting*

$$f(x):=\bar{f}([x]),$$

*where* $[x]\in[-\frac{1}{2},\frac{1}{2})^d\subset\mathbb{R}^d$ *is the unique representative of* $x$ *that is in* $[-\frac{1}{2},\frac{1}{2})^d$*. In the following we will not distinguish between* $f$ *and* $\bar{f}$*. In that way we define for* $k\in\mathbb{N}_0$*:*

$$C^k(\mathbb{T}^d)=\{f\in C^k(\mathbb{R}^d):f\text{ is 1-periodic}\}.$$

*For integration purposes it is useful to identify* $\mathbb{T}^d$ *with* $[-\frac{1}{2},\frac{1}{2})^d$*. We thus write*

$$L^p(\mathbb{T}^d)=\left\{f\in L^p([-\frac{1}{2},\frac{1}{2})^d,dx),\ \textit{extended to a 1-periodic function on } \mathbb{R}^d\right\}.$$

*Then one can show that the* Fourier monomials

$$e_k(x):=e^{2\pi ik\cdot x},\qquad k\in\mathbb{Z},$$

*form an orthonormal basis of* $L^2(\mathbb{T}^d;\mathbb{C})$ *(i.e. with values in* $\mathbb{C}$*), which is equipped with the inner product*

$$\langle f,g\rangle_{L^2}:=\int_{\mathbb{T}^d}f(x)\overline{g(x)}dx,$$

*where* $\overline{g(x)}$ *is the complex conjugate of* $g$*.*

Question: Show that for $f,g\in C^1(\mathbb{T}^d)$: $\int_{\mathbb{T}^d}\partial_{x_i}f(x)g(x)dx=-\int_{\mathbb{T}^d}f(x)\partial_{x_i}g(x)dx$ (integration by parts without boundary terms).

**Example 1.36.** Let us consider again equation

$$\partial_t u=\frac{1}{2}\partial_{xx}u+\sigma\partial_x u\partial_t W,\qquad u(0)=u_0,$$

where $W$ is a one-dimensional Brownian motion, and for now we do not make any restriction on $\sigma$. To simplify the discussion, we consider the equation on $\mathbb{T}$, i.e. $u:\mathbb{R}_+\times\mathbb{T}\to\mathbb{R}$. Assume that $u_0$ only has finitely many Fourier coefficients, i.e. there exists $N\in\mathbb{N}_0$ with

$$u_0(x)=\sum_{|k|\leqslant N}e_k(x)\hat{u}_0(k),\qquad \hat{u}_0(k)=\langle u_0,e_k\rangle_{L^2}=\int_{\mathbb{T}}u_0(x)e^{-2\pi ik\cdot x}dx.$$

Since

$$\partial_x e_k=2\pi ike_k,\qquad \partial_{xx}e_k=-(2\pi k)^2e_k,$$

and since $W$ does not depend on the space variable, we formally get

$$u(t,x)=\sum_{|k|\leqslant N}e_k\hat{u}(t,k),\qquad \hat{u}(t,k)=\langle u(t,\cdot),e_k\rangle_{L^2},$$

where integration by parts yields (there are no boundary terms due to the periodicity of $u$):

$$\begin{aligned}\partial_t\hat{u}(t,k)&=\int_{\mathbb{T}}u(t,x)\overline{\partial_{xx}e_k(x)}dx-\sigma\int_{\mathbb{T}}u(t,x)\overline{\partial_x e_k(x)}\partial_t W\\&=-(2\pi k)^2\hat{u}(t,k)+\sigma 2\pi ik\hat{u}(t,k)\partial_t W.\end{aligned}$$

This is a decoupled family of linear ordinary stochastic differential equations, which we can solve explicitly, and if $\mathbb{E}[\|u_0\|_{L^2}^2]<\infty$, the solution satisfies

$$\max_{|k|\leqslant N}\mathbb{E}\left[\sup_{t\leqslant T}|\hat{u}(t,k)|^2\right]<\infty.$$

Since the equations for the Fourier coefficients are decoupled it is principle not even necessary to assume that $u_0$ only has finitely many non-zero Fourier coefficients. But then we should make sure that our candidate solution

$$u(t,x)=\sum_{k\in\mathbb{Z}}e_k(x)\hat{u}(t,k)$$

still lives in a reasonable function space.

This is the starting point of the variational approach to SPDEs: Assume again that $\hat{u}_0(k)=0$ for $|k|>N$. We just saw that there exists a unique solution, such that $u(t,\cdot)$ is in $C^\infty(\mathbb{T})$ for all $t\geqslant 0$ because it is a finite sum of Fourier monomials. Now we want to derive *energy estimates* for $u$ which do not depend on $N$. For that purpose we define a "Gelfand triple" consisting of three Hilbert spaces

$$H^{-1}(\mathbb{T})\subset L^2(\mathbb{T})\subset H^1(\mathbb{T}),$$

where

$$H^1(\mathbb{T})=\{f\in L^2(\mathbb{T}): f \text{ is weakly differentiable and } \partial_x f\in L^2\},$$

with norm

$$\|f\|_{H^1}^2:=\|f\|_{L^2}^2+\|\partial_x f\|_{L^2}^2,$$

and where $H^{-1}$ is the dual space of $H^1$. Recall that a function $g\in L^1(\mathbb{T})$ is called a weak derivative of $f$ and we write $\partial_x f:=g$ if for all $\varphi\in C^\infty(\mathbb{T})$: $\langle g,\varphi\rangle_{L^2}=-\langle f,\partial_x\varphi\rangle_{L^2}$. If $f\in C^1(\mathbb{T})$, then the weak derivative is just the usual derivative, and the weak derivative is always unique as an element of $L^1(\mathbb{T})$.

We apply Itô's formula to the $L^2$ norm (the "energy"), and we write $[X]$ for the quadratic variation, to avoid confusion with $\langle\cdot\rangle_{L^2}$:

$$\begin{aligned}
d\|u\|_{L^2(\mathbb{T})}^2 &= \int_{\mathbb{T}} d|u(t,x)|^2 dx = \int_{\mathbb{T}} 2u(t,x)du(t,x)dx+\int_{\mathbb{T}} d[u(t,x)]dx\\
&= \int_{\mathbb{T}} u(t,x)\partial_{xx}u(t,x)dx\,dt+2\sigma\int_{\mathbb{T}} u(t,x)\partial_x u(t,x)dx\,dW_t+\sigma^2\int_{\mathbb{T}}|\partial_x u(t,x)|^2 dx\,dt\\
&=-(1-\sigma^2)\int_{\mathbb{T}}|\partial_x u(t,x)|dx\,dt+2\sigma\int_{\mathbb{T}} u(t,x)\partial_x u(t,x)dx\,dW_t\\
&=-(1-\sigma^2)\|u\|_{H^1}^2 dt+(1-\sigma^2)\|u\|_{L^2}^2 dt+2\sigma\int_{\mathbb{T}} u(t,x)\partial_x u(t,x)dx\,dW_t,
\end{aligned}$$

where we applied integration by parts to move one of the derivatives from the Laplacian (and again we used that there are no boundary terms because $u$ is periodic). Note that

$$\int_{\mathbb{T}} u(t,x)\partial_x u(t,x)dx=\frac{1}{2}\int_{\mathbb{T}}\partial_x u(t,x)^2 dx=0$$

by the periodic boundary conditions, and therefore the stochastic integral vanishes. If the stochastic integral would have a less peculiar form, we could introduce the stopping times $\tau_n:=\inf\{t\geqslant 0: \|u(t)\|_{H^1}\geqslant n\}$ and derive bounds for the stopped process $u^{\tau_n}$, before sending $n\to\infty$ to derive bounds for $u$. The point is that after stopping, the stochastic integral is a true martingale, so its expectation vanishes. We do not carry out the stopping argument but directly use that the stochastic integral vanishes and therefore

$$\mathbb{E}[\|u(t)\|_{L^2}^2]+(1-\sigma^2)\int_0^t\mathbb{E}[\|u(s)\|_{H^1}^2]ds=(1-\sigma^2)\int_0^t\|u(s)\|_{L^2}^2 ds+\mathbb{E}[\|u_0\|_{L^2}^2].$$

If $\sigma^2\leqslant 1$, the two terms on the left hand side are positive, and therefore Gronwall's inequality gives us the a priori "energy estimate"

$$\sup_{t\leqslant T}\mathbb{E}[\|u(t)\|_{L^2}^2]+(1-\sigma^2)\int_0^T\mathbb{E}[\|u(s)\|_{H^1}^2]ds\lesssim_T\mathbb{E}[\|u_0\|_{L^2}^2].$$

If $u_0$ does not have finitely many Fourier coefficients, we define

$$u_0^N:=\sum_{|k|\leqslant N} e_k\hat{u}_0(k),$$

so that

$$\|u_0^N\|_{L^2}^2 = \sum_{|k|\leqslant N} |\hat{u}_0(k)|^2 \leqslant \sum_{k\in\mathbb{Z}} |\hat{u}_0(k)|^2 = \|u_0\|_{L^2}^2,$$

and therefore for the solution $u_N$ with initial condition $u_0^N$:

$$\sup_{t\leqslant T} \mathbb{E}[\|u_N(t)\|_{L^2}^2] + (1-\sigma^2)\int_0^T \mathbb{E}[\|u_N(s)\|_{H^1}^2] ds \lesssim_T \mathbb{E}[\|u_0\|_{L^2}^2].$$

If $(1-\sigma^2)>0$, this is sufficient to pass to the limit and to construct a unique weak solution $u\in C([0,T], L^2(\Omega; L^2(\mathbb{T}))\cap L^2([0,T]; L^2(\Omega; H^1(\mathbb{T})))$ for any $T>0$, where $L^p(X;Y)$ are the $L^p$ functions from $X\to Y$. This is the main idea of the variational approach and the strategy works in much more general situations, see [32, 44].

Question: Formally derive similar energy estimates for $u\colon \mathbb{R}_+\times\mathbb{T}^d\to\mathbb{R}$,

$$\partial_t u = \Delta u + f(u) + g(u)\partial_t W, \qquad u(0)=u_0,$$

where $W$ is a one-dimensional Brownian motion and $f$ and $g$ satisfy the *weak coercivity condition*

$$2f(u)u + g^2(u) \lesssim 1 + u^2.$$

Note that this condition seems useful for treating our previous example $f(u)=-u^3$.

**Example 1.37.** What goes wrong in Example 1.36 if $\sigma^2=1$ or even $\sigma^2>1$? For $\sigma^2=1$ we can find the solution explicitly: If $u_0\in C^2(\mathbb{T})$, then

$$u(t,x) := u_0(x+\sigma W_t)$$

solves by Itô's formula

$$du(t,x) = \frac{\sigma^2}{2}\partial_{xx}u(t,x)dt + \sigma\partial_x u(t,x)dW_t, \qquad u(0)=u_0,$$

which is our equation (since $\sigma^2=1$). Note that $u(t,\cdot)$ is exactly as regular as $u_0$, which is very different from what we observed for the case $\sigma^2<1$, where we had an $H^1$-estimate for $u$ depending only on the $L^2$-norm of $u_0$. This was due to regularization coming from the Laplacian, which we now lost. This becomes more clear if we write our equation in Stratonovich form:

$$\begin{aligned} du(t,x) &= \frac{1}{2}\partial_{xx}u(t,x)dt + \sigma\partial_x u(t,x)\circ dW_t - \frac{\sigma}{2}d[\partial_x u(\cdot,x), W]_t \\ &= \left(\frac{1}{2}-\frac{\sigma^2}{2}\right)\partial_{xx}u(t,x)dt + \sigma\partial_x u(t,x)\circ dW_t. \end{aligned}$$

So we see that for $\sigma^2=1$ the Laplacian disappears (and this is why we no longer see any regularization), and for $\sigma^2>1$ the factor in front of the Laplacian becomes negative, which should ring all alarm bells. Indeed, this would be like solving the heat equation backwards in time, which in general is a severely ill-posed problem: The forward heat equation maps any bounded measurable initial condition to $C^\infty$ at any positive time, so going backwards in time would mean that the solution has to become more and more irregular and it might stop being a function or even a generalized function.

Question: Define $v(t,x) := u(t, x-\sigma W_t)$ and show that at least formally

$$dv(t,x) = \left(\frac{1}{2}-\frac{\sigma^2}{2}\right)\partial_{xx}v(t,x)dt.$$

So for $\sigma^2>1$ $v$ would indeed be a solution to the backward heat equation.

## 1.6 Restrictions of our solution theories

**Example 1.38. (KPZ equation)** Consider again the KPZ equation

$$\partial_t h = \partial_{xx} h + \lambda(\partial_x h)^2 + \partial_t W,$$

where $W$ is a space-time white noise. Let us try to solve it now without applying the Cole-Hopf formula. In the Walsh approach there is no built in regularity analysis of $h$, but maybe we could try a different method that is better suited for regularity analysis, such as the variational approach?

Unfortunately not: If $\lambda = 0$ then the equation is well posed, and by the regularity analysis of Remark 1.33 we have $h(t,\cdot) \in C_{\mathrm{loc}}^{1/2-\varepsilon}$ for all $\varepsilon > 0$, and this is sharp, the solution is not better than that. For $\lambda \neq 0$ we do not expect to gain any regularity from the nonlinearity, so $\partial_x h$ would be the derivative of a non-differentiable function. We could try to make sense of it as a generalized function (we will discuss this more precisely later in the lecture), but then we would not be allowed to take the square. So we are stuck.

**Example 1.39. (Distributional regularity of SPDEs in $d > 1$)** Consider the stochatic heat equation with additive noise

$$\partial_t Z = \Delta Z + \partial_t W, \qquad Z(0) = 0,$$

where $\partial_t W$ is a space-time white noise. As we discussed, the solution should be given by

$$Z(t,x) = \int_{[0,t]\times\mathbb{R}^d} p(t-s, x-y) W(ds, dy),$$

but this is not defined if $d > 1$ because $W$ has a covariance measure in $\mathcal{K}^{-d/2}$ and $-d/2 \leqslant -1$ for $d \geqslant 2$. However, we could try to make sense of $Z$ as a generalized function: We formally define for $\varphi \in C_c^\infty$

$$Z(t)(\varphi) := \int_{\mathbb{R}^d} Z(t,x)\varphi(x) dx := \int_{[0,t]\times\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \varphi(x) p(t-s, x-y) dx \right) W(ds, dy).$$

We verify that the right hand side is well defined: Since $p(t-s,\cdot)$ is a probability density, Jensen's inequality gives

$$\begin{aligned}\int_0^t \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \varphi(x) p(t-s,x-y) dx \right|^2 dy ds &\leqslant \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\varphi(x)|^2 p(t-s,x-y) dx dy ds \\ &= t \|\varphi\|_{L^2}^2.\end{aligned}$$

On Sheet 4 you will use this point of view to show that almost surely $Z(t) \in C_{\mathrm{loc}}^{1-\frac{d}{2}-\varepsilon}$ for all $t > 0$ and all $\varepsilon > 0$, where $C_{\mathrm{loc}}^{-\alpha}$ is a space of *generalized functions* or *distributions*.

**Example 1.40. ($\Phi_d^4$ equation)** Consider now the $\Phi_d^4$ equation

$$\partial_t \phi = \Delta\phi - \phi^3 + \partial_t W, \qquad \phi(0) = \phi_0$$

where $\phi: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and where $\partial_t W$ is a space-time white noise. This equation is for example of interest in quantum field theory, and it arises as scaling limit of certain Ising type particle systems. As for the KPZ equation we do not expect that $\phi$ has better regularity than the solution $Z$ to the linear equation, so we expect $\phi(t) \in C_{\mathrm{loc}}^{1-d/2-\varepsilon}$ for all $t > 0$. If $d \geqslant 2$ this regularity is negative, and therefore the interpretation of the nonlinearity $\phi^3$ is dubious.

**Example 1.41. (PAM, once again)** The parabolic Anderson model

$$\partial_t u = \Delta u + u\xi,$$

where $\xi$ is a space white noise, is problematic for similar reasons: We will see that almost surely $\xi \in C_{\text{loc}}^{-d/2-\varepsilon}$, and therefore the meaning of the product $u\xi$ is unclear: In general, we can only multiply distributions with $C^\infty$ functions, and the solution $u$ will not be in $C^\infty$.

In all these examples we have problems because there is a lack of regularity. This is somewhat similar to the problem that we had when constructing the Itô integral: Since the Brownian motion is not of finite variation, it was not immediately clear how to make sense of $\int_0^t H_s dB_s$. But then we were able to use the flow of information generated by the underlying filtration together with the martingale structure of $B$ to make sense of the integral for adapted $H$. But in our SPDE examples the irregularity is always in the space variable, and there does not seem to be any useful flow of information in space. So martingale arguments seem quite useless here. Instead, we will find regularity based pathwise arguments to make sense of our equations. For that purpose we have to go well beyond what we could usually do in analysis and find new arguments and techniques to deal with very irregular equations (because we want to study very irregular noise). To learn the philosophy of the pathwise approach, it is easier to first study it on stochastic ordinary equations, where it is just Lyons's rough path theory [33].

# 2 A very short introduction to rough paths

Here we give a very brief introduction to the main ideas and techniques of Terry Lyons's rough path theory [33]. We use Gubinelli's approach [21], which is beautifully exposed in the monograph [18], see also [34] for nice lecture notes on rough paths.

## 2.1 Young's integral and Young equations

Brownian motion is the most important example among a class of Gaussian processes called *fractional Brownian motions*. A continuous and centered Gaussian process $(B_t)_{t \geqslant 0}$ with $B_0 = 0$ is called a *fractional Brownian motion with Hurst parameter* $H \in (0,1)$ if it has the covariance

$$\mathbb{E}[B_s B_t] = \Gamma(s,t) := \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}).$$

It is not entirely trivial to see that $\Gamma$ is indeed a covariance function (i.e. positive definite), but one can show that

$$\Gamma(s,t) = \int_{\mathbb{R}} \Phi(s,r)\Phi(t,r)dr,$$

for

$$\Phi(s,r) = \frac{1}{\gamma(H+1/2)}\big((s-r)_+^{H-1/2} - (-r)_+^{H-1/2}\big),$$

and from that representation we easily obtain that $\Gamma$ is indeed positive definite.

We would like to solve stochastic differential equations

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \qquad X_0 = x. \tag{2.1}$$

If $B$ is a Brownian motion ($H = 1/2$), then of course we can do this by Itô calculus. But if $H \neq 1/2$, then $B$ is no semimartingale and therefore we cannot use Itô calculus and the meaning of the integral

$$\int_0^t \sigma(X_s)dB_s$$

is unclear (one can also show that $B$ is almost surely not of finite variation, for any $H\in(0,1)$). However, there is a way to define such integrals in a pathwise sense if $B$ is not of finite variation, at least if $\sigma(X)$ and $B$ have "compatible regularities".

**Lemma 2.1.** *If $B$ is a fractional Brownian motion of Hurst index $H$, then almost surely $(B_t)_{t\in[0,T]}\in C^{\alpha}([0,T])$ for all $\alpha\in(0,H)$ and all $T>0$, where*

$$C^{\alpha}([0,T])=C^{\alpha}([0,T],\mathbb{R})=\left\{f\in C([0,T]):\|f\|_{\alpha}=\sup_{0\leqslant s<t\leqslant T}\frac{|f(t)-f(s)|}{|t-s|^{\alpha}}<\infty\right\}$$

*is the space of $\alpha$-Hölder continuous functions.*

**Proof.** We have

$$\mathbb{E}[|B_t-B_s|^2]=\Gamma(t,t)+\Gamma(s,s)-2\Gamma(s,t)=t^{2H}+s^{2H}-(s^{2H}+t^{2H}-|t-s|^{2H})=|t-s|^{2H}.$$

Since $B$ is Gaussian we get

$$\mathbb{E}[|B_t-B_s|^p]^{1/p}\simeq\mathbb{E}[|B_t-B_s|^2]^{1/2}=|t-s|^H$$

for all $p>0$. For $p>1/H$ we thus obtain that $B$ has an $\alpha$-Hölder continuous modification for any $\alpha\in(0,H-1/p)$. Since $B$ itself is continuous, it is indistinguishable from this modification, and thus $B$ is a.s. $\alpha$-Hölder continuous for any $\alpha\in(0,H-1/p)$. Since $p>0$ is arbitrary, the claim follows. $\square$

For $H>1/2$ will make use of this regularity to solve SDEs like (2.1) with the help of *Young integration*. For $H\leqslant 1/2$ this no longer works and in that case we will use rough path integration (at least if $H>1/3$). Both rough path integration and Young integration rely on the following fundamental result:

**Theorem 2.2. (Sewing lemma)** *Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|$, let $T>0$ and let $\Xi:[0,T]^2\to\mathcal{X}$ be continuous and such that $\Xi_{t,t}=0$ for all $t\in[0,T]$. We assume that there exist $C,\varepsilon>0$ such that for all $0\leqslant s<u<t\leqslant T$:*

$$\|\delta\Xi_{s,u,t}\|:=\|\Xi_{s,t}-\Xi_{s,u}-\Xi_{u,t}\|\leqslant C|t-s|^{1+\varepsilon}.$$

*Then there exists a unique function $\mathcal{I}\Xi:[0,T]\to\mathcal{X}$ with $\mathcal{I}\Xi_0=0$ and such that*

$$\|\mathcal{I}\Xi_{s,t}-\Xi_{s,t}\|\lesssim C|t-s|^{1+\varepsilon},\qquad 0\leqslant s<t\leqslant T,\tag{2.2}$$

*where we use the notation*

$$y_{s,t}:=y_t-y_s.$$

*Moreover, $\mathcal{I}\Xi$ is continuous and*

$$\mathcal{I}\Xi_t=\lim_{n\to\infty}\sum_{k=0}^{K_n-1}\Xi_{t_k^n\wedge t,t_{k+1}^n\wedge t},\qquad t\in[0,T],\tag{2.3}$$

*where the convergence is uniform in $t$ and $\{0=t_0^n<t_1^n<\cdots<t_{K^n}^n=T\}$ is any sequence of partitions with mesh size going to $0$, i.e. $\max_k|t_{k+1}^n-t_k^n|\to 0$.*

**Proof.**

1. Construction of $\mathcal{I}\Xi$: We define with $t_k^n=k2^{-n}T$

$$\mathcal{I}_t^n=\sum_{k=0}^{2^n-1}\Xi_{t_k^n\wedge t,t_{k+1}^n\wedge t}.$$

Since $t_k^n = t_{2k}^{n+1} < t_{2k+1}^{n+1} < t_{2k+2}^{n+1} = t_{k+1}^n$, we have

$$\begin{aligned}
&\|\mathcal{I}_t^n - \mathcal{I}_t^{n+1}\|\\
&= \left\| \sum_{k=0}^{2^n-1} \left( \Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} - \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} \right) \right\|\\
&\leqslant \sum_{k=0}^{2^n-1} \left\| \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} - \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{2n} \wedge t} \right\|\\
&= \sum_{k=0}^{2^n-1} \left\| \delta \Xi_{t_{2k}^{n+1} \wedge t, t_{2k+1}^{n+1} \wedge t, t_{2k+2}^{n+1} \wedge t} \right\|\\
&\leqslant \sum_{k=0}^{2^n-1} C(T 2^{-n})^{1+\varepsilon} \lesssim C T^{1+\varepsilon} 2^{-n\varepsilon},
\end{aligned}$$

and the right hand side does not depend on $t$ and it is summable in $n$. Therefore, $(\mathcal{I}^n)_n$ is a Cauchy sequence in $C([0,T], \mathcal{X})$ (continuity of $\mathcal{I}^n$ follows from continuity of $\Xi$ and since $\Xi_{s,s} = 0$), and thus it converges to a limit $\mathcal{I}\Xi \in C([0,T], \mathcal{X})$. This essentially concludes the main part of the proof. Everything which follows now is only needed for proving that $\mathcal{I}\Xi$ satisfies (2.2).

2. We can slightly improve the estimate above by noting that for $k$ with $t_k^n \geqslant t$ we have

$$\Xi_{t_k^n \wedge t, t_{k+1}^n \wedge t} - \Xi_{t_k^n \wedge t, t_{2k+1}^{n+1} \wedge t} - \Xi_{t_{2k+1}^{n+1} \wedge t, t_{k+1}^n \wedge t} = 0,$$

and therefore we only have to sum $\lceil 2^n t / T \rceil$ terms and this gives the bound

$$\|\mathcal{I}_t^n - \mathcal{I}_t^{n+1}\| \leqslant \left\lceil \frac{2^n t}{T} \right\rceil C(T 2^{-n})^{1+\varepsilon} \lesssim C t T^{\varepsilon} 2^{-n\varepsilon}$$

whenever $\frac{2^n t}{T} > 1/2$. Similarly we get for $s < t$ with $\frac{2^n |t-s|}{T} > 1/2$ ($\Leftrightarrow 2^n |t-s| > T/2$)

$$\|\mathcal{I}_{s,t}^n - \mathcal{I}_{s,t}^{n+1}\| \lesssim C|t-s| T^{\varepsilon} 2^{-n\varepsilon}. \tag{2.4}$$

3. $\mathcal{I}\Xi$ satisfies (2.2): Let $0 \leqslant s < t \leqslant T$ and let $n \in \mathbb{N}$ be maximal such that $2^{-n} T > |t-s|$ (so in particular $2^{-n} T < 2|t-s| \Leftrightarrow 2^n |t-s| > T/2$). Then

$$\|\mathcal{I}\Xi_{s,t} - \Xi_{s,t}\| \leqslant \|\mathcal{I}\Xi_{s,t} - \mathcal{I}_{s,t}^n\| + \|\mathcal{I}_{s,t}^n - \Xi_{s,t}\|.$$

We treat the two terms on the right hand side separately. For the first one we apply the bound (2.4) and obtain

$$\|\mathcal{I}\Xi_{s,t} - \mathcal{I}_{s,t}^n\| \leqslant \sum_{k=n}^{\infty} \|\mathcal{I}_{s,t}^{k+1} - \mathcal{I}_{s,t}^k\| \lesssim \sum_{k=n}^{\infty} C|t-s| T^{\varepsilon} 2^{-n\varepsilon} \lesssim C|t-s|^{1+\varepsilon},$$

where we applied the simple lemma which follows after this proof. To bound the remaining term we note that if $k$ is such that $s \in (t_k^n, t_{k+1}^n]$, then $t \in (t_k^n, t_{k+2}^n)$ and thus

$$\begin{aligned}
\|\mathcal{I}_{s,t}^n - \Xi_{s,t}\| &= \|\Xi_{t_k^n, t_{k+1}^n \wedge t} + \Xi_{t_{k+1}^n \wedge t, t} - \Xi_{t_k^n, s} - \Xi_{s,t}\|\\
&\leqslant \|\Xi_{t_k^n, t} - \Xi_{t_k^n, t_{k+1}^n \wedge t} - \Xi_{t_{k+1}^n \wedge t, t}\| + \|\Xi_{t_k^n, t} - \Xi_{t_k^n, s} - \Xi_{s,t}\|\\
&= \|\delta \Xi_{t_k^n, t_{k+1}^n \wedge t, t}\| + \|\delta \Xi_{t_k^n, s, t}\|\\
&\lesssim C|t - s + 2^{-n} T|^{1+\varepsilon} \lesssim C|t-s|^{1+\varepsilon}.
\end{aligned}$$

Therefore, $\mathcal{I}\Xi$ satisfies (2.2).

4. Finally we show that (2.2) uniquely characterizes $\mathcal{I}\Xi$ as the limit in (2.3). Indeed, we have

$$\begin{aligned}\left\|\mathcal{I}\Xi_t - \sum_{k=0}^{K_n-1} \Xi_{t_k^n\wedge t, t_{k+1}^n\wedge t}\right\| &\leqslant \sum_{k=0}^{K_n-1} \left\|\mathcal{I}\Xi_{t_k^n\wedge t, t_{k+1}^n\wedge t} - \Xi_{t_k^n\wedge t, t_{k+1}^n\wedge t}\right\| \\ &\lesssim C \sum_{k=0}^{K_n-1} |t_{k+1}^n\wedge t - t_k^n \wedge t|^{1+\varepsilon} \\ &\leqslant C \max_k |t_{k+1}^n - t_k^n|^{\varepsilon} \sum_{k=0}^{K_n-1} |t_{k+1}^n - t_k^n| \\ &= C \max_k |t_{k+1}^n - t_k^n|^{\varepsilon} T,\end{aligned}$$

and by assumption the right hand side converges to zero as $n \to \infty$. This concludes the proof.

□

**Lemma 2.3.** *Let $\alpha > 0$. Then for all $n \in \mathbb{N}$*

$$\sum_{k=0}^{n} 2^{k\alpha} = \frac{2^{(n+1)\alpha} - 1}{2^{\alpha} - 1} \simeq 2^{n\alpha}, \qquad \sum_{k=n}^{\infty} 2^{-k\alpha} = 2^{-n\alpha} \sum_{k=0}^{\infty} 2^{-k\alpha} = \frac{2^{-n\alpha}}{1 - 2^{-\alpha}} \simeq 2^{-n\alpha}.$$

Question: Show that $\delta\Xi_{s,u,t} \equiv 0$ for all $s < u < t$ if and only if there exists a function $f\colon [0,T] \to \mathcal{X}$ such that $\Xi_{s,t} = f_t - f_s$.

**Remark 2.4.** If $\Xi$ was given by the increments of a function, i.e. $\Xi_{s,t} = x_t - x_s$ for some $x\colon [0,T] \to \mathcal{X}$, then we would have $\Xi_{s,t} = \Xi_{s,u} + \Xi_{u,t}$, so we could call $\Xi$ an *additive function*. In that case $\delta\Xi_{s,u,t} = x_{s,t} - x_{s,u} - x_{u,t} = 0$ and $\mathcal{I}_t^n = x_t - x_0$ for all $n$ and thus $\mathcal{I}\Xi_t = x_t - x_0$.

So $\|\delta\Xi_{s,u,t}\|$ measures "how far $\Xi$ is from being an additive function". If $\|\delta\Xi_{s,u,t}\| \lesssim |t-s|^{1+\varepsilon}$, then in general $\Xi$ is not an additive function, but the sewing lemma shows that there exists a unique (up to addition of constants) additive function $\mathcal{I}\Xi$ such that $\|\mathcal{I}\Xi_{s,t} - \Xi_{s,t}\| \lesssim |t-s|^{1+\varepsilon}$. This additive function is obtained by "sewing together" $\Xi$, since $\mathcal{I}\Xi_t = \lim_{n\to\infty} \sum_{k=0}^{K_n-1} \Xi_{t_k^n\wedge t, t_{k+1}^n\wedge t}$.

**Exercise 2.1.** Show that if $x\colon [0,T] \to \mathbb{R}$ is $\alpha$-Hölder continuous for $\alpha > 1$, then $x$ is constant: $x_t = x_0$ for all $t \in [0,T]$.

**Corollary 2.5. (Young integral)** *Let $T > 0$ and let $\alpha, \beta \in (0,1]$ be such that $\alpha + \beta > 1$. Let $x \in C^{\alpha}([0,T])$ and $y \in C^{\beta}([0,T])$. Then the Young integral*

$$\int_0^t x_s dy_s = \lim_{n\to\infty} \sum_{k=0}^{2^n-1} x_{t_k^n} y_{t_k^n\wedge t, t_{k+1}^n\wedge t}, \qquad t_k^n = k 2^{-n} T,$$

*exists and it is the unique function such that for all $[s,t] \subset [0,T]$:*

$$\left| \int_s^t x_r dy_r - x_s y_{s,t} \right| \lesssim \|x\|_{\alpha} \|y\|_{\beta} |t-s|^{\alpha+\beta}. \tag{2.5}$$

*Moreover,*

$$\left\| \int_0^{\cdot} x_s dy_s \right\|_{\beta} \lesssim (1 + T^{\alpha})(|x_0| + \|x\|_{\alpha}) \|y\|_{\beta}. \tag{2.6}$$

**Proof.** We define

$$\Xi_{s,t} := x_s(y_t - y_s),$$

which is continuous in $(s,t)$, satisfies $\Xi_{t,t} = x_t(y_t - y_t) = 0$, and

$$\begin{aligned}
|\delta\Xi_{s,u,t}| &= |x_s(y_t - y_s) - x_s(y_u - y_s) - x_u(y_t - y_u)| \\
&= |(x_s - x_u)(y_t - y_u)| \\
&\leqslant \|x\|_\alpha \|y\|_\beta |u-s|^\alpha |t-u|^\beta \\
&\leqslant \|x\|_\alpha \|y\|_\beta |t-s|^{\alpha+\beta},
\end{aligned}$$

and since $\alpha + \beta > 1$ we can apply the sewing lemma. To prove (2.6) we use (2.5) and the $\beta$-Hölder continuity of $y$:

$$\begin{aligned}
\left|\int_s^t x_r d y_r\right| &\leqslant \left|\int_s^t x_r d y_r - x_s y_{s,t}\right| + |x_s y_{s,t}| \\
&\lesssim \|x\|_\alpha \|y\|_\beta |t-s|^{\alpha+\beta} + |x_{0,s} y_{s,t}| + |x_0 y_{s,t}| \\
&\leqslant T^\alpha \|x\|_\alpha \|y\|_\beta |t-s|^\beta + \|x\|_\alpha T^\alpha \|y\|_\beta |t-s|^\beta + |x_0| \|y\|_\beta |t-s|^\beta \\
&\lesssim (1 + T^\alpha)(|x_0| + \|x\|_\alpha) \|y\|_\beta |t-s|^\beta.
\end{aligned}$$

□

Question: Let $y_t = B_t(\omega)$ for a typical sample path of a Brownian motion. Which $x$ can we integrate using the Young integral? Is $x_t = f(B_t(\omega))$ for a smooth function $f$ ok? Or $x_t = X_t(\omega)$, where $X$ solves the Itô SDE $d X_t = b(X_t) d t + \sigma(X_t) d B_t$?

**Remark 2.6.** If $y$ is of finite variation (e.g. Lipschitz continuous), then by definition the Young integral agrees with the Lebesgue-Stieltjes integral $\int_0^\cdot x_s d y_s$ (since both are limits of the same sums).

**Remark 2.7.** Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and let $L(\mathcal{X}, \mathcal{Y})$ be the space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$. Then for $x \in C^\alpha([0,T], L(\mathcal{X}, \mathcal{Y}))$ and $y \in C^\beta([0,T], \mathcal{X})$ the Young integral $\int_0^\cdot x_s d y_s \in C^\beta([0,T], \mathcal{Y})$ can be constructed in the same way and it satisfies the same estimates.

If we want use Young integration to solve the equation

$$d x_t = b(x_t) d t + \sigma(x_t) d y_t$$

for $y \in C^\alpha([0,T])$ (for example $y = B(\omega)$ for a fractional Brownian motion with Hurst index $H > \alpha$), then we should assume that $\sigma(x) \in C^\beta([0,T])$ with $\beta > 1 - \alpha$. We will see that for nice functions $\sigma$ the path $\sigma(x)$ has the same regularity as $x$, and therefore we need $x \in C^\beta([0,T])$. Then $\int_0^\cdot \sigma(x_s) d y_s \in C^\alpha([0,T])$, and since $\int_0^\cdot b(x_s) d s$ is Lipschitz, also $x \in C^\alpha([0,T])$ and this means that we need to take $\beta = \alpha$ and then we require $\alpha > 1 - \alpha$ or equivalently $\alpha > 1/2$. This is what we call the *Young regime*.

Essentially we have everything we need to solve linear equations. But to handle non-linear equations we need to understand how nonlinearities interact with Hölder regularity. To state the result we need the following function space and norm:

$$C_b^\gamma(\mathbb{R}^d, \mathbb{R}^n) := \{f \in C^{\lfloor\gamma\rfloor}(\mathbb{R}^d, \mathbb{R}^n) : \|f\|_{C_b^\gamma} < \infty\},$$

where

$$\|f\|_{C_b^\gamma} := \sum_{|\mu| \leqslant \lfloor\gamma\rfloor} \|\partial^\mu f\|_\infty + \max_{|\mu| = \lfloor\gamma\rfloor} \sup_{x \neq y} \frac{|\partial^\mu f(x) - \partial^\mu f(y)|}{|x-y|^{\gamma - \lfloor\gamma\rfloor}}.$$

In words, $C_b^{\gamma}(\mathbb{R}^d, \mathbb{R}^n)$ is the space of functions which are $\lfloor\gamma\rfloor$ times continuously differentiable with bounded partial derivatives, and the partial derivatives of order $\lfloor\gamma\rfloor$ are $\gamma - \lfloor\gamma\rfloor$ Hölder continuous. For $\gamma = 1$ this means that $f$ is Lipschitz continuous. For $\gamma = \lfloor\gamma\rfloor$ this means that the partial derivatives of order $\lfloor\gamma\rfloor$ are continuous. We also write

$$\|x\|_{C_b^{\alpha}} := \|x\|_{\alpha} + \|x\|_{\infty}$$

for $\alpha \in (0,1]$ and $x \in C^{\alpha}([0,T])$.

**Lemma 2.8.** *Let $\alpha \in (0,1]$ and let $x, \tilde{x} \in C^{\alpha}([0,T], \mathbb{R}^d)$. Let $\sigma \in C_b^1(\mathbb{R}^d, \mathbb{R}^n)$. Then $\sigma(x) \in C^{\alpha}([0,T], \mathbb{R}^n)$ and*

$$\|\sigma(x)\|_{\alpha} \leqslant \|\sigma\|_{C_b^1} \|x\|_{\alpha}.$$

*For $\sigma \in C_b^2(\mathbb{R}^d, \mathbb{R}^n)$ we have*

$$\|\sigma(x) - \sigma(\tilde{x})\|_{\alpha} \lesssim \|\sigma\|_{C_b^2} (1 + \|x - \tilde{x}\|_{\infty}) \|x - \tilde{x}\|_{\alpha}.$$

**Proof.** The first bound is easy:

$$|\sigma(x)_{s,t}| \leqslant \|\sigma\|_{C_b^1} |x_{s,t}| \leqslant \|\sigma\|_{C_b^1} \|x\|_{\alpha} |t-s|^{\alpha}.$$

The bound for the difference is a bit more technical. Taylor's formula gives

$$\begin{aligned}
|\sigma(x)_{s,t} - \sigma(\tilde{x})_{s,t}| &= \left| \int_0^1 \nabla\sigma(\tilde{x}_t + \lambda(x_t - \tilde{x}_t)) \cdot (x_t - \tilde{x}_t) d\lambda - \int_0^1 \nabla\sigma(\tilde{x}_s + \lambda(x_s - \tilde{x}_s)) \cdot (x_s - \tilde{x}_s) d\lambda \right| \\
&\leqslant \left| \int_0^1 [\nabla\sigma(\tilde{x}_t + \lambda(x_t - \tilde{x}_t)) - \nabla\sigma(\tilde{x}_s + \lambda(x_s - \tilde{x}_s))] \cdot (x_t - \tilde{x}_t) d\lambda \right| \\
&\quad + \left| \int_0^1 \nabla\sigma(\tilde{x}_s + \lambda(x_s - \tilde{x}_s)) \cdot (x_{s,t} - \tilde{x}_{s,t}) d\lambda \right| \\
&\leqslant \|\sigma\|_{C_b^2} |\tilde{x}_t + \lambda(x_t - \tilde{x}_t) - (\tilde{x}_s + \lambda(x_s - \tilde{x}_s))| \|x - \tilde{x}\|_{\infty} \\
&\quad + 2\|\sigma\|_{C_b^1} \|x - \tilde{x}\|_{\alpha} |t-s|^{\alpha} \\
&\lesssim \|\sigma\|_{C_b^2} \|x - \tilde{x}\|_{\alpha} \|x - \tilde{x}\|_{\infty} |t-s|^{\alpha} + \|\sigma\|_{C_b^1} \|x - \tilde{x}\|_{\alpha} |t-s|^{\alpha} \\
&\lesssim \|\sigma\|_{C_b^2} (1 + \|x - \tilde{x}\|_{\infty}) \|x - \tilde{x}\|_{\alpha} |t-s|^{\alpha}.
\end{aligned}$$

□

Question: Show that for $\sigma \in C_b^1$ and $x \in C^{\alpha}$ we have

$$|\sigma(x)_{s,t} - \sigma'(x_s) x_{s,t}| \leqslant \|\sigma\|_{C_b^2} \|x\|_{\alpha}^2 |t-s|^{2\alpha}.$$

Deduce that for $\alpha > \frac{1}{2}$ we have

$$\sigma(x_t) = \sigma(x_0) + \int_0^t \sigma'(x_s) dx_s.$$

**Theorem 2.9.** **(Young equation)** *Let $\alpha \in (\frac{1}{2}, 1]$, let $y \in C^{\alpha}([0,T], \mathbb{R}^n)$ and let $b \in C_b^1(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in C_b^2(\mathbb{R}^d, \mathbb{R}^{d \times n})$. Then for all $x_0 \in \mathbb{R}^d$ there exists a unique solution $x \in C^{\alpha}([0,T], \mathbb{R}^d)$ to the Young integral equation*

$$\begin{aligned}
x_t &= x_0 + \int_0^t b(x_s) ds + \int_0^t \sigma(x_s) dy_s, \\
&:= x_0 + \int_0^t b(x_s) ds + \sum_{j=1}^n \int_0^t \sigma_{\cdot,j}(x_s) dy_s^j, \qquad t \in [0,T].
\end{aligned}$$

*Moreover, $x$ depends locally Lipschitz continuously on $(x_0, y)$: If $\tilde{x}$ solves the same equation for $\tilde{x}_0, \tilde{y}$, then there exists $K > 0$ depending only on $b, \sigma, T$ and $|x_0|, |\tilde{x}_0|$ and $\|y\|_\alpha, \|\tilde{y}\|_\alpha$, such that for all $\alpha' \in (\frac{1}{2}, \alpha)$:*

$$\|x - \tilde{x}\|_{\alpha'} \leqslant K(|x_0 - \tilde{x}_0| + \|y - \tilde{y}\|_\alpha).$$

**Proof.**

1. We use a Picard iteration on a small time interval. Let $\alpha' \in (\frac{1}{2}, \alpha)$. For $\tau \in (0, 1 \wedge T]$ we consider

$$B_\tau := \{x \in C([0, \tau], \mathbb{R}^d) : x(0) = x_0, \|x\|_{\alpha'} \leqslant 1\},$$

where we write $\|x\|_{\alpha'}$ and $\|y\|_\alpha$ for the norms restricted to $[0, \tau]$. For $x \in B_\tau$ we define

$$\Phi(x)_t := x_0 + \int_0^t b(x_s) ds + \int_0^t \sigma(x_s) dy_s, \qquad t \in [0, \tau].$$

2. We first show that $\Phi$ leaves $B_\tau$ invariant if $\tau > 0$ is sufficiently small. We have $\Phi(x)_0 = x_0$ and by the bound (2.6) for the Young integral

$$\begin{aligned}\|\Phi(x)\|_{\alpha'} &\overset{(2.6)}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty + (1 + \tau^\alpha)(|\sigma(x_0)| + \|\sigma(x)\|_{\alpha'})\|y\|_{\alpha'} \\ &\overset{\text{Lem. 2.8}}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty + (1 + \tau^\alpha)(\|\sigma\|_\infty + \|\sigma\|_{C_b^1}\|x\|_{\alpha'})\|y\|_{\alpha'} \\ &\overset{(2.7)}{\lesssim} \tau^{1-\alpha'}\|b\|_\infty + \tau^{\alpha-\alpha'}\|\sigma\|_{C_b^1}\|y\|_\alpha,\end{aligned}$$

where we used that

$$\|y\|_{\alpha'} = \sup_{0 \leqslant s < t \leqslant \tau} \frac{|y_{s,t}|}{|t - s|^{\alpha'}} \leqslant \sup_{0 \leqslant s < t \leqslant \tau} \frac{|y_{s,t}|}{|t - s|^\alpha} \sup_{0 \leqslant s < t \leqslant \tau} \frac{|t - s|^\alpha}{|t - s|^{\alpha'}} = \|y\|_\alpha \tau^{\alpha-\alpha'}, \qquad (2.7)$$

and also that $\tau \leqslant 1$ so any positive power of $\tau$ can be bounded by 1. So if $\tau \in (0, T \wedge 1]$ is small enough (depending only on $b, \sigma, y$ but not on $x_0$), then $\Phi$ leaves $B_\tau$ invariant.

3. Next, we show that $\Phi$ is a contraction on the complete metric space $(B_\tau, \|\cdot\|_\alpha)$ (possibly after further decreasing the value of $\tau$). Note that $\|\cdot\|_\alpha$ is only a seminorm because $\|c\|_\alpha = 0$ for any constant function $c$. But since in $B_\tau$ we fix the initial value $x_0$, $\|\cdot\|_\alpha$ becomes a norm. Using the completeness of $\mathbb{R}^d$ it is not difficult to show that $(B_\tau, \|\cdot\|_\alpha)$ is indeed complete.

   To see the contraction property of $\Phi$, note that $\alpha + \alpha' > 1$ by assumption (since $\alpha, \alpha' > \frac{1}{2}$), and therefore the bound for the Young integral below is justified:

$$\begin{aligned}\|\Phi(x) - \Phi(\tilde{x})\|_{\alpha'} &\leqslant \left\| \int_0^\cdot (b(x_s) - b(\tilde{x}_s)) ds \right\|_{\alpha'} + \left\| \int_0^\cdot (\sigma(x_s) - \sigma(\tilde{x}_s)) dy_s \right\|_{\alpha'} \\ &\leqslant \tau^{1-\alpha'}\|b(x) - b(\tilde{x})\|_\infty + \tau^{\alpha-\alpha'} \left\| \int_0^\cdot (\sigma(x_s) - \sigma(\tilde{x}_s)) dy_s \right\|_\alpha \\ &\overset{(2.6)}{\lesssim} \tau^{1-\alpha'}\|b(x) - b(\tilde{x})\|_\infty + \tau^{\alpha-\alpha'}(1 + \tau^\alpha)\|\sigma(x) - \sigma(\tilde{x})\|_{\alpha'}\|y\|_\alpha \\ &\overset{\text{Lem. 2.8}}{\lesssim} \tau^{1-\alpha'}\|b\|_{C_b^1}\|x - \tilde{x}\|_\infty + \tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}(1 + \|x - \tilde{x}\|_\infty) \\ &\qquad\qquad \times \|x - \tilde{x}\|_{\alpha'}\|y\|_\alpha \\ &\leqslant \tau^{1-\alpha'}\|b\|_{C_b^1}\tau^{\alpha'}\|x - \tilde{x}\|_{\alpha'} + \tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}\|x - \tilde{x}\|_{\alpha'}\|y\|_\alpha \\ &\lesssim (\tau\|b\|_{C_b^1} + \tau^{\alpha-\alpha'}\|\sigma\|_{C_b^2}\|y\|_{\alpha'})\|x - \tilde{x}\|_{\alpha'},\end{aligned}$$

where we used that since $x$ and $\tilde{x}$ both start in $x_0$:

$$\|x-\tilde{x}\|_\infty \leqslant \tau^{\alpha'}\|x-\tilde{x}\|_{\alpha'} \lesssim 1.$$

So if $\tau$ is sufficiently small, then $\Phi$ is indeed a contraction and there exists a unique fixed point.

4. The length $\tau$ of the interval on which $\Phi$ is a contraction does not depend on $x_0$, so now we can interate the construction on $[\tau, 2\tau]$, then on $[2\tau, 3\tau]$, etc., until we reach $[0,T]$. This concludes the proof of existence and uniqueness of solutions. However, so far we only showed that the solution $x$ is in $C^{\alpha'}([0,T],\mathbb{R}^d)$. To show that $x$ is even $\alpha$-Hölder continuous we use the fact that $x$ solves the equation and that the right hand side of the equation is in $C^\alpha$ for any $x\in C^{\alpha'}$.

5. The continuous dependence of $x$ on $(x_0, y)$ will be shown on Sheet 5.

□

Question: $C^\alpha([0,T],\mathbb{R}^d)$ is compactly embedded in $C^{\alpha'}([0,T],\mathbb{R}^d)$ for all $\alpha'<\alpha$. Use this to show existence of solutions to Young equations under the assumption that $\sigma\in C_b^1$ (instead of $\sigma\in C_b^2$).

**Corollary 2.10.** *Let $(B_t)_{t\in[0,T]}$ be an $n$-dimensional fractional Brownian motion of Hurst parameters $H>\frac{1}{2}$ (i.e. the components $(B^1,\ldots,B^n)$ are i.i.d. and each component is a fractional Brownian motion). Let $b\in C_b^1(\mathbb{R}^d,\mathbb{R}^d)$ and $\sigma\in C_b^2(\mathbb{R}^d,\mathbb{R}^{d\times n})$ and let $x_0\in\mathbb{R}^d$. Let $\alpha\in(\frac{1}{2},H)$ Then there exists a unique (up to indistinguishability) process $X$ such that almost surely $(X_t)_{t\in[0,T]}\in C^\alpha([0,T],\mathbb{R}^d)$ and*

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

*where the integral against $B$ is a Young integral which is well defined almost surely.*

*If $(B^m)_{m\in\mathbb{N}}\subset C^1([0,T],\mathbb{R}^d)$ is a sequence of paths such that almost surely $\|B-B^m\|_{C_b^\alpha}\to 0$, then $X=\lim_{m\to\infty}X^m$, where*

$$\partial_t X_t^m = b(X_t^m) + \sigma(X_t^m)\partial_t B_t^m, \qquad X_0^m = X_0.$$

**Remark 2.11.** In this result we first freeze the realization $B(\omega)$ of the noise, and then we do deterministic analysis with this given path. This is very different from Itô stochastic differential equations, which do not make sense for a fixed $\omega$ and for which the solution is only defined "modulo null sets". Also, if here we take the canonical probability space

$$\Omega = C^\alpha([0,T],\mathbb{R}^n),$$

for $\alpha\in(\frac{1}{2},H)$, then the map $\omega\mapsto X(\omega)$ is continuous. While for Itô SDEs it is only measurable.

Our theorem excludes exactly the most interesting case $H=\frac{1}{2}$, which corresponds to the Brownian motion. On Exercise Sheet 5 we will see that the conditions on the Young integral are sharp. So to treat the Brownian case with a similar philosophy and to cover the case $H<\frac{1}{2}$ we need to do more. The solution is to "enrich" the path $y$ by equipping it with more information. In that way we can construct a continuous pathwise integral which applies to paths of regularity $<\frac{1}{2}$.

## 2.2 Idea and definition of a rough path

To be able to treat the Brownian motion, we would like to extend the theory of Young equations to driving signals $y \in C^\alpha([0,T],\mathbb{R}^n)$ with $\alpha \leqslant \frac{1}{2}$. Unfortunately this is not possible.

**Example 2.12.** Consider the case $d = n = 2$ and

$$x_t^1 = \int_0^t dy_s^1, \qquad x_t^2 = \int_0^t x_s^1 dy_s^2, \tag{2.8}$$

with initial condition $y(0) = 0$. This equation has the explicit solution $x^1 = y^1$ and $x^2 = \int_0^\cdot y^1 dy^2$. For $m \in \mathbb{N}$ we set

$$y_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}.$$

From Sheet 5 we know that $(y^m)$ converges to 0 uniformly and in $C^\alpha$ for all $\alpha < 1/2$, and that $x_t^{2,m} \to \frac{t}{2}$ as $m \to \infty$. But of course, the solution to (2.8) with $y \equiv 0$ is equal to $(0,0)$ and not $(0,t/2)$, and therefore $x$ does not depend continuously on $y$ in $C^\alpha$-norm if $\alpha < 1/2$.

The problem is that the fast oscillations of $y^m$ interact with the nonlinearity in our integral equation and this interaction creates nontrivial effects, even though the amplitude of $y^m$ is very small.

Note that all paths involved in this example are smooth ($C^\infty$), so the problem is not the lack of regularity but the topologies in which the $(y^m)$ converge. In the following we will introduce rough path topologies which help us to overcome this lack of continuity.

We could imagine two possible approaches for doing so. The naive one would be to try to find a better suited function space, which contains Brownian paths and paths of solutions to SDEs and in which the integral $\int_0^\cdot x_s dy_s$ becomes a continuous functional (i.e. not to work with Hölder norms). However, this is impossible! A counterexample by Lyons shows that there cannot exist a Banach space $\mathcal{X}$ of real-valued functions on $[0,1]$ such that $\mathcal{X}$ contains almost all sample paths of the Brownian motion and such that there exists a continuous functional $\mathcal{X}^2 \ni (x,y) \mapsto I(x,y)$ with $I(x,y) = \int_0^1 x_s \partial_s y_s$ whenever $y \in C^1$; see Section 1.5.1 of [34].

The second approach is to accept this lack of continuity, and to *enhance* the path $y$ to make the map $y \mapsto x$ continuous. To understand this philosophy, let us consider the following trivial example:

**Example 2.13.** The map $f: \mathbb{R} \to \mathbb{R}$,

$$f(x) = \begin{cases} -1, & x < 0, \\ +1, & x \geqslant 0, \end{cases}$$

is obviously discontinuous in 0. The problem is that we can approach 0 from the left or from the right, and $\mathbb{R}$ is not rich enough to encode the information "from where we are coming". To obtain a continuous map, we could enhance the input space to encode the information whether we approach 0 from the left or from the right. More precisely, we consider

$$\mathcal{X} := ((-\infty,0) \times \{-\}) \cup (\{0\} \times \{-,+\}) \cup ((0,\infty) \times \{+\}) \subset \mathbb{R} \times \{-,+\}$$

instead, where $\mathbb{R} \times \{-,+\}$ is equipped with the product topology (and $\{-,+\}$ is equipped with the discrete topology). Note that $\mathbb{R} \subset \mathcal{X}$, i.e. there exists a (non-canonical) injection $\Phi$ from $\mathbb{R}$ to $\mathcal{X}$ by setting $\Phi(x) = (x,+)$ if $x \geqslant 0$ and $\Phi(x) = (x,-)$ if $x < 0$. We define

$$g(x,+) = +1, \qquad g(x,-) = -1.$$

Then $g$ is continuous, and we have

$$f(x) = g(\Phi(x)).$$

In other words, we have decomposed $f$ as the concatenation of the continuous map $g$ with the discontinuous and non-canonical map $\Phi$.

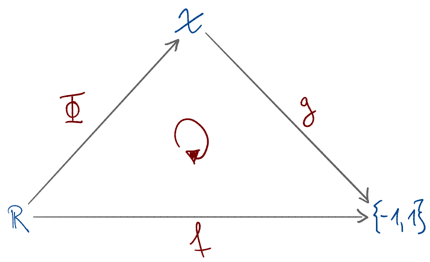


**Figure 2.1.** Commuting diagram.

**Example 2.14.** To understand how we should enhance our path space let $y \in C^1([0,T], \mathbb{R}^d)$ and let us try to construct $\int_0^1 f(y_s) dy_s$ in a way that depends continuously on $y$ in $C^\alpha$ topology for $\alpha < 1/2$, where $f: \mathbb{R}^d \to L(\mathbb{R}^d, \mathbb{R})$ is a smooth bounded function with bounded derivatives. Since $y$ and $f(y)$ are both Lipschitz continuous, we have along a sequence of partitions with mesh size going to zero:

$$\int_0^1 f(y_s) dy_s = \lim_{n \to \infty} \sum_k f(y_{t_k^n}) y_{t_k^n, t_{k+1}^n}.$$

But we would like to control this using only the $C^\alpha$-norm of $y$, and for that purpose we want to apply the sewing lemma. With $\Xi_{s,t} = f(y(s)) y_{s,t}$ we get

$$|\delta \Xi_{s,u,t}| = |-f(y)_{s,u} y_{u,t}| \leqslant \|f\|_{C_b^1} \|y\|_\alpha^2 |t-s|^{2\alpha}.$$

Since $\alpha < 1/2$, this is not good enough to apply the sewing lemma, and therefore we are stuck. But we can try to improve the approximation:

$$\begin{aligned}
\int_0^1 f(y_s) dy_s &= \sum_{i=1}^d \int_0^1 f_i(y_s) dy_s^i \\
&= \sum_k \left( \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} f_i(y_{t_k^n}) dy_s^i + \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} (f_i(y_s) - f_i(y_{t_k^n})) dy_s^i \right) \\
&= \sum_k \left[ \sum_{i=1}^d f_i(y_{t_k^n}) y^i_{t_k^n, t_{k+1}^n} + \sum_{j=1}^d \partial_j f_i(y_{t_k^n}) \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} (y_s^j - y_{t_k^n}^j) dy_s^i \right] \\
&\quad + \sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} \left( f_i(y_s) - f_i(y_{t_k^n}) - \sum_{j=1}^d \partial_j f_i(y_{t_k^n}) (y_s^j - y_{t_k^n}^j) \right) dy_s^i.
\end{aligned}$$

For the last term on the right hand side we expect to get

$$\sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} O(|t_{k+1}^n - t_k^n|^{2\alpha}) dy_s^i = \sum_k O(|t_{k+1}^n - t_k^n|^{3\alpha}) \leqslant \max_k |t_{k+1}^n - t_k^n|^{3\alpha - 1},$$

which converges to 0 if $\alpha > \frac{1}{3}$ (we have not shown fully rigorously that this term is really of order $O(|t_{k+1}^n - t_k^n|^{3\alpha})$, but it can be justified).

On the other hand, for the second term on the right hand side we expect to have

$$\sum_k \sum_{i=1}^d \int_{t_k^n}^{t_{k+1}^n} \sum_{j=1}^d \partial_j f_i(y_{t_k^n})(y_s^j - y_{t_k^n}^j) dy_s^i = \sum_k O(|t_{k+1}^n - t_k^n|^{2\alpha}),$$

and if $\alpha \leqslant \frac{1}{2}$, then this is not negligible (unlike for the Young case $\alpha > \frac{1}{2}$). This suggests to consider a different $\Xi$ in the sewing lemma, namely

$$\Xi_{s,t} := f(y_s) y_{s,t} + \mathrm{D} f(y_s) \mathbb{Y}_{s,t},$$

where $\mathrm{D}f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_d f \end{pmatrix}$ is the derivative of $f$ and

$$\mathbb{Y}_{s,t} := \int_s^t (y_r - y_s) \otimes dy_r = \left( \int_s^t (y_r^i - y_s^i) dy_r^j \right)_{i,j=1,\ldots,d} \in \mathbb{R}^{d \otimes d}.$$

and

$$\mathrm{D} f(y) \mathbb{Y} = \sum_{i=1}^d \partial_i f(y) \mathbb{Y}^{i,\cdot} = \sum_{i,j=1}^d \partial_i f_j(y) \mathbb{Y}^{i,j}, \qquad \mathbb{Y} \in \mathbb{R}^{d \otimes d}.$$

Then

$$\delta \Xi_{s,u,t} = -f(y)_{s,u} y_{u,t} + \mathrm{D} f(y_s)(\mathbb{Y}_{s,t} - \mathbb{Y}_{s,u}) - \mathrm{D} f(y_u) \mathbb{Y}_{u,t}, \tag{2.9}$$

and

$$\begin{aligned}
\mathbb{Y}_{s,t} - \mathbb{Y}_{s,u} &= \int_s^t (y_r - y_s) \otimes \mathrm{d} y_s - \int_s^u (y_r - y_s) \otimes \mathrm{d} y_r \\
&= \int_u^t (y_r - y_u) \otimes \mathrm{d} y_r + y_u \otimes y_{u,t} - y_s \otimes y_{s,t} + y_s \otimes y_{s,u} \\
&= \mathbb{Y}_{u,t} + y_{s,u} \otimes y_{u,t}.
\end{aligned}$$

Therefore, we obtain in (2.9)

$$\begin{aligned}
|\delta \Xi_{s,u,t}| &\leqslant |-f(y)_{s,u} y_{u,t} + \mathrm{D} f(y_s) y_{s,u} \otimes y_{u,t}| + |-\mathrm{D} f(y)_{s,u} \mathbb{Y}_{u,t}| \\
&\leqslant \|f\|_{C_b^2} \|y\|_\alpha^3 |t-s|^{3\alpha} + \|\sigma\|_{C_b^2} \|y\|_\alpha \|\mathbb{Y}\|_{2\alpha} |t-s|^{3\alpha} = O(|t-s|^{3\alpha}).
\end{aligned} \tag{2.10}$$

So if $\alpha > \frac{1}{3}$, we can apply the sewing lemma to bound the integral in terms of $f$, $y$, and $\mathbb{Y}$. In other words, the knowledge of the simple functional $\mathbb{Y}$ allows us to construct the integral $\int f(y) dy$ for all $f \in C_b^2(\mathbb{R}^d, L(\mathbb{R}^d, \mathbb{R}))$ – as a continuous functional of $(y, \mathbb{Y})$. As we will see soon, it also allows us to solve differential equations driven by $(y, \mathbb{Y})$ and that the solution depends continuously on the signal.

Question (difficult): What could we do if $\alpha \in (\frac{1}{4}, \frac{1}{3}]$?

**Notation 2.15.** *Since* $\mathbb{y}$ *is more complicated to type than* $\mathbb{Y}$*, we will always write* $\mathbb{Y}$ *for the iterated integrals. For consistency we also write* $Y$ *for the path, instead of* $y$*.*

Let us write $\Delta_T = \{(s,t) \in [0,T]^2 : s \leqslant t\}$ and

$$C_2^{2\alpha}(\Delta_T, \mathbb{R}^{d \otimes d}) := \{f : \Delta_T \to \mathbb{R}^{d \otimes d} : \|f\|_{2\alpha} < \infty\},$$

where

$$\|f\|_{2\alpha} := \sup_{0 \leqslant s < t \leqslant T} \frac{|f_{s,t}|}{|t-s|^{2\alpha}}.$$

**Definition 2.16. (Rough path)** *Let* $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ *and* $d \in \mathbb{N}$. *A* $d$*-dimensional* $\alpha$-rough path *is a pair* $(Y, \mathbb{Y}) =: \boldsymbol{Y}$ *with* $Y \in C^\alpha([0,T], \mathbb{R}^d)$ *and* $\mathbb{Y} \in C_2^{2\alpha}(\Delta_T, \mathbb{R}^{d \otimes d})$, *such that* Chen's relation

$$\delta \mathbb{Y}_{s,u,t} = Y_{s,u} \otimes Y_{u,t} \tag{2.11}$$

*holds for all* $0 \leqslant s \leqslant u \leqslant t \leqslant T$. *We define*

$$|||\boldsymbol{Y}|||_\alpha := \|Y\|_\alpha + \sqrt{\|\mathbb{Y}\|_{2\alpha}}.$$

*We say that a sequence of* $\alpha$*-rough paths* $(\boldsymbol{Y}^m)$ *converges to* $\boldsymbol{Y}$ *in* $\alpha$*-rough path topology if*

$$\lim_{m \to \infty} |||\boldsymbol{Y}^m - \boldsymbol{Y}|||_\alpha := \lim_{m \to \infty} \left( \|Y^m - Y\|_\alpha + \sqrt{\|\mathbb{Y}^m - \mathbb{Y}\|_{2\alpha}} \right) = 0.$$

**Remark 2.17.**

i. We think of $\mathbb{Y}$ as postulating "iterated integrals" of $Y$,

$$\mathbb{Y}_{s,t} = \int_s^t \int_s^{r_1} dY_{r_2} \otimes dY_{r_1} = \int_s^t Y_r \otimes dY_r - Y_s \otimes Y_{s,t}.$$

Since $Y \in C^\alpha$ for $\alpha \leqslant \frac{1}{2}$, the right hand side is not well defined in general, so the left hand side should be read as its definition.

ii. The space of rough paths is not a linear space, because Chen's relation (2.11) is not preserved under linear operations. Intuitively, knowing $\int_s^t Y_r \otimes dY_r$ and $\int_s^t \tilde{Y_r} \otimes d\tilde{Y_r}$ does not mean that we know $\int_s^t (Y_r + \tilde{Y_r}) \otimes d(Y_r + \tilde{Y_r})$.

iii. Also, $|||\boldsymbol{Y}|||_\alpha$ is of course not a norm. The reason for considering $\sqrt{\|\mathbb{Y}\|_{2\alpha}}$ rather than $\|\mathbb{Y}\|_{2\alpha}$ is that the natural dilation on rough path space is $(Y, \mathbb{Y}) \mapsto (\lambda Y, \lambda^2 \mathbb{Y})$. Indeed,

$$\int_s^t (\lambda Y)_{r_1} \otimes d(\lambda Y)_{r_1} - (\lambda Y)_s \otimes (\lambda Y)_{s,t} = \lambda^2 \mathbb{Y}_{s,t}.$$

So by taking $\sqrt{\|\mathbb{Y}\|_{2\alpha}}$, we make $|||\cdot|||_\alpha$ homogeneous under dilations.

**Lemma 2.18.** *Let* $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ *and* $d \in \mathbb{N}$. *An alternative definition of a* $d$*-dimensional* $\alpha$*-rough path is as follows: It is a pair* $(Y, I)$ *with* $Y \in C^\alpha([0,T], \mathbb{R}^d)$, $I \in C^\alpha([0,T], \mathbb{R}^{d \otimes d})$, *such that*

$$\sup_{(s,t) \in \Delta_T} \frac{|I_{s,t} - Y_s \otimes Y_{s,t}|}{|t-s|^{2\alpha}} < \infty. \tag{2.12}$$

*The link with Definition 2.16 is*

$$I_{s,t} = \mathbb{Y}_{s,t} + Y_s \otimes Y_{s,t}.$$

**Proof.** If $(Y, I)$ are as claimed, then the function $\mathbb{Y}_{s,t} := I_{s,t} - Y_s \otimes Y_{s,t}$ satisfies Chen's relation:

$$\begin{aligned} \delta \mathbb{Y}_{s,u,t} =& \delta_{s,u,t} I - Y_s \otimes Y_{s,t} + Y_s \otimes Y_{s,u} + Y_u \otimes Y_{u,t} \\ =& 0 + Y_{s,u} \otimes Y_{u,t}, \end{aligned}$$

since $I$ is an additive function. By assumption, $\mathbb{Y} \in C_2^{2\alpha}$.

Conversely, if $(Y, \mathbb{Y})$ is an $\alpha$-rough path, then we define

$$I_t := \mathbb{Y}_{0,t} + Y_0 \otimes Y_{0,t}.$$

Then

$$\begin{aligned} I_{s,t} &= \mathbb{Y}_{0,t} + Y_0 \otimes Y_{0,t} - (\mathbb{Y}_{0,s} + Y_0 \otimes Y_{0,s}) \\ &\overset{\text{Chen}}{=} \mathbb{Y}_{s,t} + Y_{0,s} \otimes Y_{s,t} + Y_0 \otimes Y_{0,t} - Y_0 \otimes Y_{0,s} \\ &= \mathbb{Y}_{s,t} + Y_s \otimes Y_{s,t}, \end{aligned}$$

and therefore $|I_{s,t} - Y_s \otimes Y_{s,t}| \lesssim |t-s|^{2\alpha}$. □

Question: Show that if $B$ is a $d$-dimensional Brownian motion, then $\mathbb{B}_{s,t} := \int_s^t B_{s,r} \otimes dB_r$ (Itô integral) and $\tilde{\mathbb{B}}_{s,t} := \int_s^t B_{s,r} \otimes \circ dB_r$ (Stratonovich integral) both satisfy Chen's relation.

**Example 2.19.** Let $\alpha \in (\frac{1}{3}, \frac{1}{2})$.

i. Let $\beta > \frac{1}{2}$ and $Y \in C^\beta([0,T], \mathbb{R}^d)$. Then we could define

$$I_t := \int_0^t Y_s \otimes dY_s,$$

and by the estimate (2.5) for the Young integral we have

$$|I_{s,t} - Y_s \otimes Y_{s,t}| \lesssim \|Y\|_\beta^2 |t-s|^{2\beta} \leqslant T^{2(\beta-\alpha)} \|Y\|_\beta^2 |t-s|^{2\alpha}.$$

Therefore, $(Y, I)$ is an $\alpha$-rough path in the sense of Lemma 2.18. However, while $I = \int_0^\cdot Y_s \otimes dY_s$ is a canonical choice, it is by far not the only option: Indeed, for any $Z \in C^{2\alpha}([0,T], \mathbb{R}^{d\otimes d})$ we could also define

$$I_t := \int_0^t Y_s \otimes dY_s + Z_t.$$

Indeed, since $Z \in C^{2\alpha}$ this $I$ obviously still satisfies the estimate (2.12).

ii. More concretely, let us take $d = 2$ and $Y \equiv 0$. Then a possible choice for $I$ would be $I_t \equiv 0$, but we could also take

$$I_t = \begin{pmatrix} 0 & \frac{t}{2} \\ -\frac{t}{2} & 0 \end{pmatrix}.$$

If we consider

$$Y_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}, \qquad I_t^m = \int_0^t Y_s^m \otimes dY_s^m,$$

then $Y^m \to Y$ in $C^\alpha$ and by Example 2.12 $I_t^m \to I_t$ for all $t \in [0,T]$. In fact one can strengthen this result and show that $(\boldsymbol{Y}^m)$ converges to $\boldsymbol{Y}$ in $\alpha$-rough path topology. So by keeping track of $I_t$, we remember that we approximated $Y \equiv 0$ by the oscillatory paths $(Y^m)$. This is reminiscent of Example 2.13, where by enhancing the state 0 to $(0,-)$ and $(0,+)$ we could keep track whether we had approached 0 from the left or from the right, respectively.

**Lemma 2.20.** *Let $f \in C_b^2(\mathbb{R}^d, L(\mathbb{R}^d, \mathbb{R}))$ and let $\boldsymbol{Y}$ be a $d$-dimensional $\alpha$-rough path for $\alpha \in (\frac{1}{3}, \frac{1}{2}]$. Then for all $t \in [0,T]$ the integral*

$$\int_0^t f(Y_s) d\boldsymbol{Y}_s := \mathcal{I}\Xi_t, \qquad \textit{for } \Xi_{s,t} := f(Y_s) Y_{s,t} + \mathrm{D}f(Y_s) \mathbb{Y}_{s,t},$$

*is well defined, and it is the unique function such that*

$$\left|\int_s^t f(Y_r) d\boldsymbol{Y}_r - f(Y_s) Y_{s,t} - \mathrm{D} f(Y_s) \mathbb{Y}_{s,t}\right| \lesssim \|f\|_{C_b^2} (\|Y\|_\alpha^3 + \|Y\|_\alpha \|\mathbb{Y}\|_{2\alpha}) |t-s|^{3\alpha}$$

*for all* $(s,t) \in \Delta_T$. *If* $(\boldsymbol{Y}^m)_{m\in\mathbb{N}}$ *is a sequence of* $\alpha$*-rough paths converging to* $\boldsymbol{Y}$ *in rough path topology, then*

$$\int_0^t f(Y_s^m) d\boldsymbol{Y}_s^m \longrightarrow \int_0^t f(Y_s) d\boldsymbol{Y}_s.$$

**Proof.** For the first part of the statement it suffices to combine the sewing lemma, Theorem 2.2, with the estimate (2.10). To obtain (2.10) we did not use that $Y$ was a smooth path, but only that its iterated integrals satisfy Chen's relation.

For the continuity statement we note that $\int_0^t f(Y_s^m) d\boldsymbol{Y}_s^m - \int_0^t f(Y_s) d\boldsymbol{Y}_s = \mathcal{I}\Xi_t^m$, where

$$\Xi_{s,t}^m = (f(Y_s^m) Y_{s,t}^m + \mathrm{D} f(Y_s^m) \mathbb{Y}_{s,t}^m - f(Y_s) Y_{s,t} - \mathrm{D} f(Y_s) \mathbb{Y}_{s,t}),$$

and therefore as in (2.10)

$$\begin{aligned} |\delta \Xi_{s,u,t}^m| \leqslant & |-f(Y^m)_{s,u} Y_{u,t}^m + f(Y)_{s,u} Y_{u,t} + \mathrm{D} f(Y_s^m) Y_{s,u}^m \otimes Y_{u,t}^m - \mathrm{D} f(Y_s) Y_{s,u} \otimes Y_{u,t}| \\ & + |-\mathrm{D} f(Y^m)_{s,u} \mathbb{Y}_{u,t}^m + \mathrm{D} f(Y)_{s,u} \mathbb{Y}_{u,t}|. \end{aligned}$$

By using a Taylor expansion and rebracketing like $ab - cd = a(b-d) + (a-c)d$ we obtain that

$$\begin{aligned} |\delta \Xi_{s,u,t}^m| \lesssim & \|f\|_{C_b^2} (\|Y\|_\alpha^2 + \|Y^m\|_\alpha^2) \|Y - Y^m\|_\alpha |t-s|^{3\alpha} \\ & + (\|\mathbb{Y}^m\|_{2\alpha} + \|\mathbb{Y}\|_{2\alpha}) \|f\|_{C_b^1} \|Y - Y^m\|_{2\alpha} |t-s|^{3\alpha} \\ & + \|f\|_{C_b^1} (\|Y\|_\alpha + \|Y^m\|_\alpha) \|\mathbb{Y} - \mathbb{Y}^m\|_{2\alpha} |t-s|^{3\alpha}, \end{aligned}$$

and the right hand side converges to 0 as $m \to \infty$. □

**Example 2.21.** As an application, we obtain that in the setting of Example 2.19

$$\lim_{m\to\infty} \int_0^t f\left(\frac{1}{m}\cos(m^2 s)\right) m\cos(m^2 s) ds = \lim_{m\to\infty} \int_0^t f(Y_s^{m,1}) dY_s^{m,2} = \lim_{m\to\infty} \int_0^t F(Y_s^m) dY_s^m,$$

where $F(y) = (\begin{smallmatrix} 0 & f(y^1) \end{smallmatrix})$, i.e. $F(y) y = f(y^1) y^2$. Now observe that with $t_k^n = \frac{kt}{n}$:

$$\begin{aligned} \int_0^t F(Y_s^m) dY_s^m &= \lim_{n\to\infty} \left( \sum_{k=0}^{n-1} F(Y_{t_k^n}^m) Y_{t_k^n, t_{k+1}^n}^m + O\left(\frac{1}{n}\right) \right) \\ &\overset{(*)}{=} \lim_{n\to\infty} \sum_{k=0}^{n-1} \left( F(Y_{t_k^n}^m) Y_{t_k^n, t_{k+1}^n}^m + F'(Y_{t_k^n}^m) \mathbb{Y}_{t_k^n, t_{k+1}^n}^m \right) \\ &= \int_0^t F(Y_s^m) d\boldsymbol{Y}_s^m, \end{aligned}$$

where in $(*)$ we used that $m$ is fixed and thus $Y^m$ is smooth and $\mathbb{Y}_{t_k^n, t_{k+1}^n}^m = O\left(\frac{1}{n}\right)$. Since $\boldsymbol{Y}^m \to \boldsymbol{Y}$ in rough path topology, where $\boldsymbol{Y} \leftrightarrow (0, I)$ with $I_t = \begin{pmatrix} 0 & \frac{t}{2} \\ -\frac{t}{2} & 0 \end{pmatrix}$, we get

$$\lim_{m\to\infty} \int_0^t f\left(\frac{1}{m}\cos(m^2 s)\right) m\cos(m^2 s) ds = \int_0^t F(Y_s) d\boldsymbol{Y}_s.$$

Let us compute the integral on the right hand side: $\int_0^t F(Y_s)d\boldsymbol{Y}_s = \mathcal{I}\Xi_t$, where

$$\Xi_{s,t} = f(Y_s^1)Y_{s,t}^2 + f'(Y_s^1)\mathbb{Y}_{s,t}^{1,2} = 0 = 0 + f'(0)\frac{t-s}{2}.$$

Therefore,

$$\int_0^t F(Y_s)d\boldsymbol{Y}_s = \mathcal{I}\Xi_t = f'(0)\frac{t}{2}.$$

Of course, it is possible to compute this limit directly. But with our technology it comes out nearly for free.

Question: Let $Z^m \to Z$ in $C^{\frac{1}{2}+\varepsilon}$. Show that $\int_0^t f(Z_s^m) m \cos(m^2 s) ds \to 0$.

## 2.3 Controlled paths

Throughout this section we fix $\alpha \in (\frac{1}{3}, \frac{1}{2}]$ and $T > 0$. In the previous section we defined rough paths and we showed that for any $\alpha$-rough path $\boldsymbol{Y}$ we can construct the integral $\int_0^\cdot f(Y_s)d\boldsymbol{Y}_s$ as a continuous map in $\alpha$-rough path topology. But ultimately our goal is to solve integral equations

$$dX_t = b(X_t)dt + \sigma(X_t)d\boldsymbol{Y}_t,$$

and the integral $\int_0^\cdot \sigma(X_s)d\boldsymbol{Y}_s$ is of a different form than $\int_0^\cdot f(Y_s)d\boldsymbol{Y}_s$, because the integrand is not just a function of $Y_s$.

A potential solution would be to enhance our rough path $\boldsymbol{Y}$ so that it also "contains $X$". This strategy works and Terry Lyons originally used it in [33], when he developed the general theory of rough paths. Here we follow instead the later approach of Gubinelli [21], who extends the integral $\int_0^\cdot X_s d\boldsymbol{Y}_s$ to more general integrands, while still keeping its continuity properties. The space of integrands should include functions $X_s = f(Y_s)$ for $f \in C_b^2$, and it should include $\sigma(X_s)$, where $X$ solves our integral equation. So in particular it has to contain functions of regularity $C^\alpha$. But as Example 2.12 shows, we cannot hope to have a continuous integral $\int_0^t X_s d\boldsymbol{Y}_s$ for generic $X \in C^\alpha$.

So we need to impose some structure on $X$, and this structure should be richer than just requiring sufficient regularity. To understand what we need, let us recall what we used to derive the estimate (2.10) which allowed us to construct $\int_0^\cdot f(Y_s)d\boldsymbol{Y}_s$ as a continuous map:

- $f(Y), \mathrm{D}f(Y) \in C^\alpha$;
- $|f(Y)_{s,u} - \mathrm{D}f(Y_s)Y_{s,u}| \lesssim |u-s|^{2\alpha}$;
- $(Y, \mathbb{Y})$ is an $\alpha$-rough path (in particular Chen's relation holds).

So whenever similar conditions hold, we could hope to apply the sewing lemma. Note that the second condition simply says that the increments of $f(Y)$ are well approximated by the increments of $Y$, times a "derivative" $\mathrm{D}f(Y_s)$. This motivates the following definition:

**Definition 2.22.** *Let $Y \in C^\alpha([0,T], \mathbb{R}^d)$. A path $X \in C^\alpha([0,T], \mathbb{R}^m)$ is* controlled by $Y$ *if there exists $X' \in C^\alpha([0,T], L(\mathbb{R}^d, \mathbb{R}^m))$, such that $R^X \in C_2^{2\alpha}([0,T], \mathbb{R}^m)$, where*

$$R_{s,t}^X := X_{s,t} - X_s' Y_{s,t}.$$

*In that case we write*

$$(X, X') \in \mathscr{D}_Y^{2\alpha}([0,T], \mathbb{R}^m)$$

*or simply $(X, X') \in \mathscr{D}_Y^{2\alpha}$, and we define*

$$\|X, X'\|_{X,2\alpha} := \|X'\|_\alpha + \|R^X\|_{2\alpha}.$$

*$\mathscr{D}_Y^{2\alpha}$ is a Banach space with respect to the norm $|X_0| + |X_0'| + \|X, X'\|_{X,2\alpha}$.*

Question: Find an example where $X$ is controlled by $Y$ but $X'$ is not unique, i.e. there exist $X' \neq \tilde{X}'$ such that that $(X, X'), (X, \tilde{X}') \in \mathscr{D}_Y^{2\alpha}$. (Hint: what if $Y$ is actually $C^{2\alpha}$ and not just $C^\alpha$)?

**Notation.** *In the following we will often have estimates up to $T$-dependent constants. To simplify the presentation we do not keep track of them explicitly and write $\lesssim_T$ instead. But later it will be important to have a locally uniform control of the $T$-dependence, so by convention $a \lesssim_T b$ means $a \leqslant C(T) b$ for an increasing function $C: \mathbb{R}_+ \to \mathbb{R}_+$. For example, $|t-s|^{3\alpha} \lesssim_T |t-s|^{2\alpha}$ for $s, t \in [0, T]$*

**Theorem 2.23.** *Let $\boldsymbol{Y}$ be a $d$-dimensional $\alpha$-rough path and let $(X, X') \in \mathscr{D}_Y^{2\alpha}([0, T], L(\mathbb{R}^d, \mathbb{R}^m))$. Then for all $t \in [0, T]$ the sewing integral*

$$\int_0^t X_s d\boldsymbol{Y}_s := \mathcal{I}\Xi_t, \qquad \Xi_{s,t} := X_s Y_{s,t} + X_s' \mathbb{Y}_{s,t}$$

*is well defined and satisfies*

$$\left| \int_s^t X_r d\boldsymbol{Y}_r - X_s Y_{s,t} - X_s' \mathbb{Y}_{s,t} \right| \lesssim (\|R^X\|_{2\alpha} \|Y\|_\alpha + \|X'\|_\alpha \|\mathbb{Y}\|_{2\alpha}) |t-s|^{3\alpha} \tag{2.13}$$

*for all $(s, t) \in \Delta_T$. Consequently, the map*

$$\mathscr{D}_Y^{2\alpha}([0, T], L(\mathbb{R}^d, \mathbb{R}^m)) \ni (X, X') \mapsto \left( \int_0^\cdot X_s d\boldsymbol{Y}_s, X \right) \in \mathscr{D}_Y^{2\alpha}([0, T], \mathbb{R}^m)$$

*is a continuous linear operator and satisfies*

$$\left\| \int_0^\cdot X_s d\boldsymbol{Y}_s, X \right\|_{X,2\alpha} \lesssim_T (\|R^X\|_{2\alpha} \|Y\|_\alpha + \|X'\|_\alpha \|\mathbb{Y}\|_{2\alpha}) + \|X'\|_\infty \|\mathbb{Y}\|_{2\alpha} + \|X\|_\alpha. \tag{2.14}$$

**Proof.** Estimate (2.13) follows easily from the sewing lemma and we leave its proof as an exercise (see also the derivation of (2.10)). Given (2.13), we get

$$\left| \int_s^t X_s d\boldsymbol{Y}_s - X_s Y_{s,t} \right| \lesssim (\|R^X\|_{2\alpha} \|Y\|_\alpha + \|X'\|_\alpha \|\mathbb{Y}\|_{2\alpha}) |t-s|^{3\alpha} + \|X'\|_\infty \|\mathbb{Y}\|_{2\alpha} |t-s|^{2\alpha},$$

and now we simply estimate $|t-s|^{3\alpha} \lesssim_T |t-s|^{2\alpha}$. The estimate for the derivative is trivial: $\|X\|_\alpha \leqslant \|X\|_\alpha$. □

Question: Let $\boldsymbol{Y} = (Y, \mathbb{Y})$ be an $\alpha$-rough path and let $Z \in C^{2\alpha}([0, T], \mathbb{R}^{d \otimes d})$ and $\tilde{\boldsymbol{Y}} = (Y, \tilde{\mathbb{Y}})$ with $\tilde{\mathbb{Y}}_{s,t} = \mathbb{Y}_{s,t} + Z_{s,t}$. Let $(X, X') \in \mathscr{D}_Y^{2\alpha}$ and compute

$$\int_0^t X_s d\tilde{\boldsymbol{Y}}_s - \int_0^t X_s d\boldsymbol{Y}_s.$$

**Remark 2.24.** In the setting of Theorem 2.23 let $\tilde{\boldsymbol{Y}} = (\tilde{Y}, \tilde{\mathbb{Y}})$ be another $d$-dimensional $\alpha$-rough path, and let $(\tilde{X}, \tilde{X}') \in \mathcal{D}_{\tilde{Y}}^{2\alpha}([0, T], L(\mathbb{R}^d, \mathbb{R}^m))$. Define

$$\rho_\alpha(\boldsymbol{Y}, \tilde{\boldsymbol{Y}}) = \|Y - \tilde{Y}\|_\alpha + \|\mathbb{Y} - \tilde{\mathbb{Y}}\|_{2\alpha},$$
$$d_{Y,\tilde{Y},2\alpha}(X, X', \tilde{X}, \tilde{X}') = \|X' - \tilde{X}'\|_\alpha + \|R^X - R^{\tilde{X}}\|_{2\alpha},$$

and $M = \max\{\|Y\|_\alpha, \|\mathbb{Y}\|_{2\alpha}, |X_0'|, \|X, X'\|_{Y,2\alpha}, \|\tilde{Y}\|_\alpha, \|\tilde{\mathbb{Y}}\|_{2\alpha}, |\tilde{X}_0'|, \|\tilde{X}, \tilde{X}'\|_{Y,2\alpha}\}$. Set

$$(Z, Z') = \left(\int_0^\cdot X_s d\boldsymbol{Y}_s, X\right), \qquad (\tilde{Z}, \tilde{Z}') = \left(\int_0^\cdot \tilde{X}_s \tilde{\boldsymbol{Y}}_s, \tilde{X}\right).$$

You show in Exercise 7.1 that

$$d_{Y,\tilde{Y},2\alpha}(Z, Z', \tilde{Z}, \tilde{Z}') \lesssim_T M(\rho_\alpha(\boldsymbol{Y}, \tilde{\boldsymbol{Y}}) + |X_0' - \tilde{X}_0'| + d_{Y,\tilde{Y},2\alpha}(X, X', \tilde{X}, \tilde{X}')).$$

Question: Let $(X, X') \in \mathscr{D}_Y^{2\alpha}([0,T], \mathbb{R}^m)$ and let $A \in \mathbb{R}^{k\times m}$. Show that $AX$ is a controlled path. What is its derivative?

We have shown that controlled paths are stable under integration against $\boldsymbol{Y}$. When solving an equation of the type

$$dX_t = b(X_t)dt + \sigma(X_t)d\boldsymbol{Y}_t,$$

we not only need to integrate against $\boldsymbol{Y}$, but we also need to apply a nonlinear map $\sigma$ to a controlled path. The next theorem shows that controlled paths are stable under the application of nonlinear maps.

**Theorem 2.25.** *Let $Y \in C^\alpha([0,T], \mathbb{R}^d)$ and let $(X, X') \in \mathscr{D}_Y^{2\alpha}([0,T], \mathbb{R}^m)$. Let $f \in C_b^2(\mathbb{R}^m, \mathbb{R}^n)$. Then*

$$(f(X), \mathrm{D}f(X)X') \in \mathscr{D}_Y^{2\alpha}([0,T], \mathbb{R}^n),$$

*and*

$$\|f(X), \mathrm{D}f(X)X'\|_{Y,2\alpha} \lesssim_T (1+M)\|f\|_{C_b^2}(1+\|Y\|_\alpha)^2(|X_0'| + \|X, X'\|_{Y,2\alpha}), \tag{2.15}$$

*with $M = |X_0'| + \|X, X'\|_{Y,2\alpha}$. If $(\tilde{X}, \tilde{X}') \in \mathscr{D}_Y^{2\alpha}([0,T], \mathbb{R}^m)$ is another controlled path with $|\tilde{X}_0'| + \|\tilde{X}, \tilde{X}'\|_{X,2\alpha} \leqslant M$ and if $f \in C_b^3$, then*

$$\begin{aligned}\|(f(X), \mathrm{D}f(X)X') - (f(\tilde{X}), \mathrm{D}f(\tilde{X})\tilde{X}')\|_{Y,2\alpha}& \\ \lesssim_{T,M} \|f\|_{C_b^3}(1+\|Y\|_\alpha)^2(|X_0 - \tilde{X}_0| + |X_0' - \tilde{X}_0'| + \|(X, X') - (\tilde{X}, \tilde{X}')\|_{Y,2\alpha}).&\end{aligned} \tag{2.16}$$

**Proof.** We show (2.15). First, we control the derivative $f(X)' = \mathrm{D}f(X)X'$:

$$\begin{aligned}|(\mathrm{D}f(X)X')_{s,t}| &\leqslant |\mathrm{D}f(X)_{s,t}X_t'| + |\mathrm{D}f(X_s)X_{s,t}'| \\ &\leqslant \|f\|_{C_b^2}\|X\|_\alpha|t-s|^\alpha\|X'\|_\infty + \|f\|_{C_b^1}\|X'\|_\alpha|t-s|^\alpha \\ &\leqslant \|f\|_{C_b^2}\|X\|_\alpha|t-s|^\alpha(|X_0'| + T^\alpha\|X'\|_\alpha) + \|f\|_{C_b^1}\|X'\|_\alpha|t-s|^\alpha \\ &\lesssim_T \|f\|_{C_b^2}(1+\|X\|_\alpha)(|X_0'| + \|X'\|_\alpha)|t-s|^\alpha.\end{aligned}$$

To bound $\|X\|_\alpha$ note that

$$\begin{aligned}|X_{s,t}| &\leqslant |X_{s,t} - X_{s,t}'Y_{s,t}| + |X_s'Y_{s,t}| \\ &\lesssim_T \|R^X\|_{2\alpha}|t-s|^\alpha + \|X'\|_\infty\|Y\|_\alpha|t-s|^\alpha \\ &\lesssim_T (|X_0'| + \|X, X'\|_{Y,2\alpha})(1+\|Y\|_\alpha)|t-s|^\alpha \\ &\lesssim_T M(1+\|Y\|_\alpha)|t-s|^\alpha.\end{aligned}$$

Next, we show that $f(X)$ is controlled with derivative $\mathrm{D}f(X_s)X_s'$:

$$\begin{aligned}|f(X)_{s,t} - \mathrm{D}f(X_s)X_s'Y_{s,t}| &\leqslant |f(X)_{s,t} - \mathrm{D}f(X_s)X_{s,t}| + |\mathrm{D}f(X_s)X_{s,t} - \mathrm{D}f(X_s)X_s'Y_{s,t}| \\ &\leqslant (\|f\|_{C_b^2}\|X\|_\alpha^2|t-s|^{2\alpha} + \|f\|_{C_b^1}\|R^X\|_{2\alpha})|t-s|^{2\alpha} \\ &\lesssim \|f\|_{C_b^2}(1+M)(1+\|Y\|_\alpha)^2(|X_0'| + \|X, X'\|_{Y,2\alpha})|t-s|^{2\alpha}.\end{aligned}$$

The derivation of (2.16) is more involved. Conceptually it is similar to the proof of the second estimate in Lemma 2.8, but technically it is more complex. See Lemma 7.3 in Friz-Hairer [18]. □

**Remark 2.26.** In the setting of Theorem 2.25 let $\tilde{\boldsymbol{Y}}=(\tilde{Y},\tilde{\mathbb{Y}})$ be another $d$-dimensional $\alpha$-rough path, and let $(\tilde{X},\tilde{X}')\in\mathscr{D}_{\tilde{Y}}^{2\alpha}([0,T],L(\mathbb{R}^m))$. Let $\rho_\alpha(\boldsymbol{Y},\tilde{\boldsymbol{Y}})$, $d_{Y,\tilde{Y},2\alpha}(X,X',\tilde{X},\tilde{X}')$, and $M$ be as in Remark 2.24. Set

$$(Z,Z')=(f(X),\mathrm{D}f(X)X'), \qquad (\tilde{Z},\tilde{Z}')=(f(\tilde{X}),\mathrm{D}F(\tilde{X})\tilde{X}').$$

Then Theorem 7.5 of Friz-Hairer [18] shows that

$$d_{Y,\tilde{Y},2\alpha}(Z,Z',\tilde{Z},\tilde{Z}')\leqslant_{T,M}(\rho_\alpha(\boldsymbol{Y},\tilde{\boldsymbol{Y}})+|X_0-\tilde{X}_0|+|X_0'-\tilde{X}_0'|+d_{Y,\tilde{Y},2\alpha}(X,X',\tilde{X},\tilde{X}')).$$

We now have all the ingredients that we need in order to solve rough differential equations of the type $dX_t=b(X_t)dt+\sigma(X_t)d\boldsymbol{Y}_t$, $X_0=x$, where $\boldsymbol{Y}=(Y,\mathbb{Y})$ is an $\alpha$-rough path and we look for solutions $(X,X')\in\mathscr{D}_Y^{2\alpha}$. For simplicity of notation we will take $b=0$ from now on, but it is not difficult to adapt the arguments to include a drift $b$. By definition, $(X,X')$ solves the equation if

$$X'=\sigma(X), \qquad X_t=x+\int_0^t\sigma(X_s)d\boldsymbol{Y}_s, \qquad t\in[0,T].$$

From Theorem 2.23 we know that $\int_0^\cdot\sigma(X_s)d\boldsymbol{Y}_s$ is the unique function which satisfies

$$\left|\int_s^t\sigma(X_s)d\boldsymbol{Y}_s-\sigma(X_s)Y_{s,t}-\mathrm{D}\sigma(X_s)\sigma(X_s)\mathbb{Y}_{s,t}\right|\lesssim|t-s|^{3\alpha}$$

for all $(s,t)\in\Delta_T$, where we used that $(\sigma(X))'=\mathrm{D}\sigma(X)X'=\mathrm{D}\sigma(X)\sigma(X)$. In other words, we have the following simple observation, which often is useful:

**Lemma 2.27.** *Let $\boldsymbol{Y}=(Y,\mathbb{Y})$ be a $d$-dimensional $\alpha$-rough path, let $\sigma\in C_b^2(\mathbb{R}^m,L(\mathbb{R}^d,\mathbb{R}^m))$, and let $x\in\mathbb{R}^m$. Let $X\colon[0,T]\to\mathbb{R}^m$. Then $(X,\sigma(X))\in\mathscr{D}_Y^{2\alpha}$ and*

$$X_t=x+\int_0^t\sigma(X_s)d\boldsymbol{Y}_s, \qquad t\in[0,T],$$

*if and only if $X_0=x$ and for all $(s,t)\in\Delta_T$*

$$|X_{s,t}-\sigma(X_s)Y_{s,t}-\mathrm{D}\sigma(X_s)\sigma(X_s)\mathbb{Y}_{s,t}|\lesssim|t-s|^{3\alpha}.$$

Question: How does this lemma look like in the Young case? Can you find a formulation for classical ODEs which is equivalent to the formulation as a differential equation?

**Theorem 2.28.** *Let $\boldsymbol{Y}=(Y,\mathbb{Y})$ be a $d$-dimensional $\alpha$-rough path, let $\sigma\in C_b^3(\mathbb{R}^m,L(\mathbb{R}^d,\mathbb{R}^m))$, and let $x\in\mathbb{R}^m$. Then there exists a unique solution $(X,X')\in\mathscr{D}_Y^{2\alpha}([0,T],\mathbb{R}^m)$ to the equation*

$$X_t=x+\int_0^t\sigma(X_s)d\boldsymbol{Y}_s, \qquad X_t'=\sigma(X_t), \qquad t\in[0,T].$$

**Proof.** Now that we know that the maps $(X,X')\mapsto(\sigma(X),\mathrm{D}\sigma(X)X')$ and $(\sigma(X),\mathrm{D}\sigma(X)X')\mapsto(\int_0^\cdot\sigma(X_s)d\boldsymbol{Y}_s,\sigma(X))$ are bounded and continuous, the proof is conceptually very similar to the one in the Young case (Theorem 2.9), although of course more technical. See Theorem 8.4 of Friz-Hairer [18]. □

**Remark 2.29.** One of the key results of rough path theory is the continuity of the Itô-Lyons map: In the setting of Theorem 2.28, write

$$X = \Phi(x, \boldsymbol{Y}).$$

It follows from Remarks 2.24 and 2.26 that if $\tilde{\boldsymbol{Y}}$ is another $\alpha$-rough path and if $\tilde{x} \in \mathbb{R}^m$, then

$$\|\Phi(x, \boldsymbol{Y}) - \Phi(\tilde{x}, \tilde{\boldsymbol{Y}})\|_{\alpha'} \lesssim_{T,M} (|x - \tilde{x}| + \rho_\alpha(\boldsymbol{Y}, \tilde{\boldsymbol{Y}})),$$

where $M = \max\{|x|, |\tilde{x}|, |||\boldsymbol{Y}|||_\alpha, |||\tilde{\boldsymbol{Y}}|||_\alpha\}$. See Theorem 8.5 of Friz-Hairer [18].

Question: Let as in Example 2.19

$$Y_t^m = \begin{pmatrix} \frac{1}{m}\cos(m^2 t) \\ \frac{1}{m}\sin(m^2 t) \end{pmatrix}, \qquad I_t^m = \int_0^t Y_s^m \otimes dY_s^m.$$

Let $\sigma \in C_b^3(\mathbb{R}^2, L(\mathbb{R}^d, \mathbb{R}^2))$. Let

$$X_t^m = x + \int_0^t \sigma(X_s^m) \partial_s Y_s^m ds, \qquad t \in [0, T].$$

Derive the equation that $X = \lim_{m\to\infty} X^m$ solves.

## 2.4 Stochastic processes as rough paths

We have seen that if $Y \in C^\alpha([0, T], \mathbb{R}^d)$, then there is no unique choice of a second order process $\mathbb{Y}$ which turns $(Y, \mathbb{Y})$ into an $\alpha$-rough path. Indeed it is not even obvious whether such a $\mathbb{Y}$ exists at all (by the *Lyons-Victoir extension theorem* it does, but we will not prove this). However, if $Y$ is a stochastic process, then there often is a canonical choice for $\mathbb{Y}$.

### 2.4.1 Brownian motion

Let us start with the easiest example, where $B$ is a $d$-dimensional standard Brownian motion on $[0, T]$. In that case we have almost surely $B \in C^\alpha([0, T], \mathbb{R}^d)$ whenever $\alpha < 1/2$, and we define

$$\mathbb{B}_{s,t}^{\text{Ito}} = \int_s^t B_r \otimes dB_r - B_s \otimes B_{s,t},$$

where the stochastic integral $dB_r$ is understood in the Itô sense; in particular, $\mathbb{B}$ is continuous. By construction we also have Chen's relation. It thus only remains to show that almost surely $|\mathbb{B}_{s,t}^{\text{Ito}}| \lesssim |t - s|^{2\alpha}$. This is a consequence of the following result.

**Theorem 2.30. (Kolmogorov's continuity criterion for rough paths)** *Let $(X, \mathbb{X})$ be a stochastic process which almost surely satisfies Chen's relation. Assume that there exist $p \geqslant 2$, $\beta > \frac{1}{p}$, $C > 0$, such that for all $(s, t) \in \Delta_T$*

$$\mathbb{E}[|X_{s,t}|^p]^{1/p} \leqslant C|t - s|^\beta, \qquad \mathbb{E}[|\mathbb{X}_{s,t}|^{p/2}]^{2/p} \leqslant C|t - s|^{2\beta}.$$

*Then there exists a modification $\boldsymbol{X} = (\tilde{X}, \tilde{\mathbb{X}})$ of $(X, \mathbb{X})$ satisfying Chen's relation, and such that*

$$\mathbb{E}[|||\tilde{\boldsymbol{X}}|||_\alpha^p]^{1/p} \lesssim C$$

*for all $\alpha \in (0, \beta - \frac{1}{p})$.*

**Proof.** See Friz-Hairer [18], Theorem 3.1. □

To apply this to $(B, \mathbb{B}^{\mathrm{Ito}})$ we only need to bound $\mathbb{E}[|\mathbb{B}^{\mathrm{Ito}}_{s,t}|^{p/2}]$ for sufficiently large $p$. We apply the Burkholder-Davis-Gundy inequality twice to obtain

$$\mathbb{E}[|\mathbb{B}^{\mathrm{Ito}}_{s,t}|^{p/2}] \simeq \mathbb{E}\left[\left(\int_s^t |B_r - B_s|^2 ds\right)^{p/4}\right] \leqslant \mathbb{E}\left[\sup_{r \in [s,t]} |B_r - B_s|^{p/2}\right] |t-s|^{p/4}$$
$$\simeq |t-s|^{p/4} \times |t-s|^{p/4} = |t-s|^{p/2}.$$

Moreover, $\mathbb{E}[|B_{s,t}|^p] \simeq |t-s|^{p/2}$, by another application of the Burkholder-Davis-Gundy inequality or alternatively because $B$ is Gaussian. Taking $p > 6$ we obtain that $\boldsymbol{B}^{\mathrm{Ito}} = (B, \mathbb{B}^{\mathrm{Ito}})$ is almost surely an $\alpha$-rough path for any $\alpha \in (\frac{1}{3}, \frac{1}{2} - \frac{1}{p})$, which is a non-empty interval because $p > 6$. Note that we do not have to take a modification of $(B, \mathbb{B}^{\mathrm{Ito}})$, because the pair is already continuous.

We also define the Stratonovich iterated integrals of $\mathbb{B}$ by

$$\mathbb{B}^{\mathrm{Strat}}_{s,t} := \int_s^t (B_r - B_s) \otimes \circ dB_r = \mathbb{B}^{\mathrm{Ito}}_{s,t} + \frac{1}{2} \mathbb{I}_d (t-s),$$

where $\mathbb{I}_d$ is the identity matrix on $\mathbb{R}^d$ and $\circ dB_r$ denotes the Stratonovich integral. We set $\boldsymbol{B}^{\mathrm{Strat}} = (B, \mathbb{B}^{\mathrm{Strat}})$.

**Theorem 2.31.** *If $(X, X')$ is an adapted process such that almost surely $(X, X') \in \mathscr{D}^{2\alpha}_B$, then almost surely*

$$\int_0^{\cdot} X_s d\boldsymbol{B}^{\mathrm{Ito}}_s = \int_0^{\cdot} X_s dB_s, \qquad \int_0^{\cdot} X_s d\boldsymbol{B}^{\mathrm{Strat}}_s = \int_0^{\cdot} X_s \circ dB_s,$$

*where the left hand sides denote the rough path integrals of $X$ with respect to $\boldsymbol{B}^{\mathrm{Ito}}$ and $\boldsymbol{B}^{\mathrm{Strat}}$ respectively, and the right hand sides denote the Itô and the Stratonovich integral (for the Stratonovich integral we should also assume that $X$ is a semimartingale).*

**Proof.** By stopping in

$$\tau_n = \inf\{t \geqslant 0 : \|X, X'\|_{\mathscr{D}^{2\alpha}_B([0,t])} \geqslant n\},$$

we may assume that $\mathbb{E}[\|X, X'\|^2_{B,2\alpha}] < \infty$. For the Itô integral, we estimate

$$\begin{aligned}
\mathbb{E}\left[\left|\int_s^t X_r dB_r - X_s B_{s,t} - X'_s \mathbb{B}_{s,t}\right|^2\right]^{1/2} &= \mathbb{E}\left[\left|\int_s^t (X_r - X_s - X'_s B_{s,r}) dB_r\right|^2\right]^{1/2} \\
&= \mathbb{E}\left[\int_s^t |X_r - X_s - X'_s B_{s,r}|^2 dr\right]^{1/2} \\
&\leqslant \mathbb{E}[\|X, X'\|^2_{B,2\alpha}]^{1/2} \left(\int_s^t |r-s|^{4\alpha} dr\right)^{1/2} \\
&\simeq \mathbb{E}[\|X, X'\|^2_{B,2\alpha}]^{1/2} |t-s|^{\frac{1}{2} + 2\alpha}.
\end{aligned}$$

Since $\frac{1}{2} + 2\alpha > 1$, this gives for fixed $t \in [0, T]$ and with $t^n_k = kt/n$:

$$\mathbb{E}\left[\left|\int_0^t X_s dB_s - \sum_{k=0}^{n-1} (X_{t^n_k} B_{t^n_k, t^n_{k+1}} + X'_{t^n_k} \mathbb{B}_{t^n_k, t^n_{k+1}})\right|^2\right]^{1/2} \lesssim \mathbb{E}[\|X, X'\|^2_{B,2\alpha}]^{1/2} n \left|\frac{t}{n}\right|^{\frac{1}{2} + 2\alpha} \to 0.$$

Therefore, $\mathcal{I}_n = \sum_{k=0}^{n-1} (X_{t^n_k} B_{t^n_k, t^n_{k+1}} + X'_{t^n_k} \mathbb{B}_{t^n_k, t^n_{k+1}})$ converges in $L^2$ to $\int_0^t X_s dB_s$. But it also converges almost surely to $\int_0^t X_s d\boldsymbol{B}^{\mathrm{Ito}}_s$, and therefore almost surely $\int_0^t X_s dB_s = \int_0^t X_s d\boldsymbol{B}^{\mathrm{Ito}}_s$. A priori the null set depends on $t$, but both processes are continuous, and therefore they are indistinguishable.

The Stratonovich integral can be rewritten in terms of the Itô integral, and this can be used to establish the claim also follows in that case. See Section 5 of Friz-Hairer [18] for details. □

Question: Let $X_t = x + \int_0^t \sigma(X_s) d\boldsymbol{B}_s^{\mathrm{Ito}}$ and determine which "Stratonovich equation" $X$ solves (i.e. derive an equation for $X$ which involves only an integral against $\boldsymbol{B}^{\mathrm{Strat}}$).

### 2.4.2 Fractional Brownian motion

What follows is not part of the videos and not relevant for the exam.

Recall that $B$ is a $d$-dimensional fractional Brownian motion with Hurst parameter $H \in (0,1)$ if $B_0 = 0$ and if the components of $(B^1, \ldots, B^d)$ are independent continuous centered Gaussian processes with covariance

$$\mathbb{E}[B_s^i B_t^i] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}), \qquad i \in \{1, \ldots, d\}, s, t \in [0, T].$$

Our aim is to construct the interated integrals $\mathbb{B}_{s,t} = (\int_s^t B_{s,r}^i d B_r^j)_{i,j}$. First observe that for $i = j$ we could simply set

$$\int_s^t B_{s,r}^i d B_r^i := \frac{1}{2}(B_{s,t}^i)^2.$$

This is the only possible choice under which we have the integration by parts rule from classical calculus, and it satisfies Chen's relation on the diagonal::

$$\frac{1}{2}(B_{s,t}^i)^2 - \frac{1}{2}(B_{s,u}^i)^2 - \frac{1}{2}(B_{u,t}^i)^2 = \frac{1}{2}(B_{u,t}^i B_{s,u}^i + B_{s,u}^i B_{u,t}^i) = (B_{s,u} \otimes B_{u,t})^{i,i}.$$

Moreover,

$$\left|\frac{1}{2}(B_{s,t}^i)^2\right| \leqslant \frac{|t-s|^{2\alpha}}{2} \|B\|_\alpha^2,$$

so we get the right regularity.

**Remark 2.32.** For any sequence of partitions $(t_k)$ of $[0,t]$ we have

$$\frac{1}{2}(B_t^i)^2 = \frac{1}{2}\sum_k [(B_{t_{k+1}}^i)^2 - (B_{t_k}^i)^2] = \sum_k \frac{B_{t_k}^i + B_{t_{k+1}}^i}{2} B_{t_k, t_{k+1}}^i,$$

so our definition of $\int_s^t B_{s,r}^i d B_r^i$ corresponds to Riemann sums taking the average of the left point and the right point of the integrand. In the Itô case we can also take left-point Riemann sums (replacing $\frac{B_{t_k}^i + B_{t_{k+1}}^i}{2}$ by $B_{t_k}^i$). We could try to do the same for $H < \frac{1}{2}$. The difference between the two Riemann sums is $\frac{1}{2}\sum_k (B_{t_k,t_{k+1}}^i)^2$, which should converge to the quadratic variation. But for $H < \frac{1}{2}$ the quadratic variation does not exist, because

$$\mathbb{E}\left[\sum_k (B_{t_k,t_{k+1}}^i)^2\right] = \sum_k |t_{k+1} - t_k|^{2H} \to \infty$$

if $H < \frac{1}{2}$.

The off-diagonal terms are more complicated.

**Lemma 2.33.** *Let $(B_t)_{t\in[0,1]}$ and $(W_t)_{t\in[0,1]}$ be indepdendent one-dimensional fractional Brownian motions with Hurst parameter $H \in (0, \frac{1}{2})$. Define for $n \in \mathbb{N}$*

$$I_n(B, dW)(t) := \sum_{k=0}^{k_t^n - 1} B_{\tau_k^n} W_{\tau_k^n, \tau_{k+1}^n}, \qquad t \in [0,1]$$

*where* $\tau_k^n := k2^{-n}$ *and* $k_t^n = \max\{k : \tau_k^n \leqslant t\}$. *Then for all* $t \in [0,1]$ *and* $p \in (1,\infty)$ *we have*

$$\mathbb{E}[|I_{n+1}(B,dW)(t) - I_n(B,dW)(t)|^2]^{1/2} \lesssim 2^{-n(2H-1/2)}\sqrt{t}.$$

*So for* $H > \frac{1}{4}$ *the sequence* $(I_n(B,dW)(t))_n$ *converges in* $L^2(\Omega)$ *to a limit* $I(B,dW)_t$.

**Proof.** Let us write $I_n(t) = I_n(B,dW)(t)$. Then

$$I_{n+1}(t) - I_n(t) = -\sum_{k=0}^{k_t^n - 1} B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}} W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}},$$

and therefore by independence of $B$ and $W$

$$\begin{aligned}
&\mathbb{E}[|I_{n+1}(t) - I_n(t)|^2] \\
&= \sum_{k,\ell=0}^{k_t^n - 1} \mathbb{E}[B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}} B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}]\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}] \\
&\leqslant \sum_{k=0}^{k_t^n - 1} (2^{-n})^{2H}(2^{-n})^{2H} \\
&+ 2\sum_{k=0}^{k_t^n - 1}\sum_{\ell=0}^{k-1} \big(\mathbb{E}[|B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}}|^2]\mathbb{E}[|B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}|^2]\big)^{1/2}|\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}]|.
\end{aligned}$$

The first term on the right hand side is clearly bounded by $\lesssim 2^n t 2^{-n4H} = (2^{-n(2H-1/2)}\sqrt{t})^2$. To bound the second term we need the following estimate, which we leave as an exercise:

$$|\mathbb{E}[W_{s,s+h} W_{t,t+h}]| \lesssim (t-s)^{2H-2} h^2, \qquad 0 \leqslant s < t, \qquad 0 < h \leqslant t - s.$$

This leads to

$$\begin{aligned}
&\sum_{k=0}^{k_t^n - 1}\sum_{\ell=0}^{k-1} \big(\mathbb{E}[|B_{\tau_{2k}^{n+1},\tau_{2k+1}^{n+1}}|^2]\mathbb{E}[|B_{\tau_{2\ell}^{n+1},\tau_{2\ell+1}^{n+1}}|^2]\big)^{1/2}|\mathbb{E}[W_{\tau_{2k+1}^{n+1},\tau_{2k+2}^{n+1}} W_{\tau_{2\ell+1}^{n+1},\tau_{2\ell+2}^{n+1}}]| \\
&\lesssim 2^{-n2H}\sum_{k=0}^{k_t^n - 1}\sum_{\ell=0}^{k-1} |\tau_{2k+1}^{n+1} - \tau_{2\ell+1}^{n+1}|^{2H-2}|2^{-n}|^2 \lesssim 2^{-n2H}\sum_{k=1}^{k_t^n - 1}\sum_{\ell=0}^{k-1} |(k-\ell)2^{-n}|^{2H-2}|2^{-n}|^2 \\
&= 2^{-n4H}\sum_{k=1}^{k_t^n - 1}\sum_{\ell=0}^{k-1} |k-\ell|^{2H-2} \lesssim 2^{-n4H}\int_1^{k_t^n - 1}\int_0^{x-1} |x-y|^{2H-2} dy dx.
\end{aligned}$$

The integral on the right hand side is bounded by

$$\int_1^{k_t^n - 1}\int_0^{x-1} |x-y|^{2H-2} dy dx = \int_1^{k_t^n - 1} \frac{1 - |x|^{2H-1}}{1-2H} dx \lesssim |k_t^n| \leqslant 2^n t,$$

and this completes the proof. □

**Remark 2.34.** The threshold $H > \frac{1}{4}$ does not appear because our estimates are inadequate. For $H \leqslant \frac{1}{4}$ the sequence $(I_n(B,dW))$ does not converge (see [15]) and there is no known canonical definition of $\int_0^\cdot B_s dW_s$. See [50, 43] for two non-canonical constructions based on renormalization arguments (roughly speaking based on subtracting random diverging counterterms from the diverging sequence $I_n(B,dW)$).

So far we only control $I(B,dW)$ in $L^2(\Omega)$. To show that it has sufficient regularity we need the following deep result on moments of polynomials of Gaussian random variables. Maybe later in the course we will see a proof in a special case.

**Theorem 2.35. (Gaussian hypercontractivity)** *Let $I$ be an index set and let $(X_i)_{i\in I}$ be a centered Gaussian process. Let $P:\mathbb{R}^m\to\mathbb{R}$ be a polynomial of degree $n$. Then for all $0<p<\infty$ there exists a constant $C_{n,p}>0$ (which is independent of $m$) such that*

$$C_{n,p}^{-1}\mathbb{E}[|P(X_{i_1},\ldots,X_{i_m})|^p]^{1/p}\leqslant\mathbb{E}[|P(X_{i_1},\ldots,X_{i_m})|^2]^{1/2}\leqslant C_{n,p}\mathbb{E}[|P(X_{i_1},\ldots,X_{i_m})|^p]^{1/p}.$$

**Proof.** See Janson [31], Theorem 3.50. □

**Theorem 2.36.** *Let $(B_t)_{t\in[0,1]}$ be a $d$-dimensional fractional Brownian motion with Hurst index $H\in(\frac{1}{3},\frac{1}{2})$. For $i\in\{1,\ldots,d\}$ we set $\mathbb{B}^{ii}_{s,t}:=\frac{1}{2}(B^i_{s,t})^2$ and for $i\neq j$ let*

$$\mathbb{B}^{ij}_{s,t}:=I(B^i,dB^j)_{s,t}-B^i_sB^j_{s,t}.$$

*Then*

$$\mathbb{E}[|\mathbb{B}_{s,t}|^p]^{1/p}\lesssim|t-s|^{2H}$$

*for all $p\in(0,\infty)$, and in particular we can apply Theorem 2.30 to obtain a modification $\tilde{\mathbb{B}}$ of $\mathbb{B}$ such that $(B,\tilde{\mathbb{B}})$ is an $\alpha$-rough path for all $\alpha\in(1/3,H)$.*

**Proof.** Let $i\neq j$ and $0\leqslant s<t\leqslant 1$. Define

$$J_n(t):=I_n(B^i,dB^j)(t)+B^i_{\tau^n_{k^n_t}}B^j_{\tau^n_{k^n_t},\tau^n_{k^n_t}+t},$$

where $I_n$ is as in Lemma 2.33. Using similar arguments as in the proof of Lemma 2.33, one can show that

$$\mathbb{E}[|(J_n)_{s,t}-I(B^i,\mathrm{d}B^j)_{s,t}|^2]^{1/2}\lesssim 2^{-n(2H-1/2)}\sqrt{t-s}.$$

The extra term in (2.36) makes the calculation longer but not more difficult; similarly estimating the difference of the time increments is more technical but not more difficult than estimating the difference at a fixed time.

Pick now $n_0$ with $2^{-n_0-1}\leqslant|t-s|<2^{-n_0}$. Using the same decomposition that appeared in the proof of Theorem 2.2, we get

$$\begin{aligned}\mathbb{E}[|I(B^i,dB^j)_{s,t}-B^i_sB^j_{s,t}|^p]^{1/p}&\leqslant\mathbb{E}[|I(B^i,dB^j)_{s,t}-(J_{n_0})_{s,t}|^p]^{1/p}+\mathbb{E}[|(J_{n_0})_{s,t}-B^i_sB^j_{s,t}|^p]^{1/p}\\&\lesssim|t-s|^{2H},\end{aligned}$$

where the last step used Gaussian hypercontractivity. Now the claim follows from Kolmogorov's continuity criterion for rough paths. □

# 3 Besov spaces and paraproducts

To develop a similar "rough path type" approach for SPDEs we need some tools from analysis, for example tempered distributions, notions of regularity for them, and results which control the product of a distribution and a function in the case of compatible regularities. An excellent reference for further reading is the monograph by Bahouri, Chemin and Danchin [3], where much of the material here is taken from.

## 3.1 Tempered distributions

**Definition 3.1.**

i. *The* Schwartz functions *are*

$$\mathscr{S} = \{\varphi \in C^\infty(\mathbb{R}^d, \mathbb{C}) : \|\varphi\|_{k,\mathscr{S}} < \infty \forall k \in \mathbb{N}_0\},$$

*where*

$$\|\varphi\|_{k,\mathscr{S}} = \sup_{|\mu| \leqslant k} \|(1 + |\cdot|^k) \partial^\mu \varphi\|_\infty.$$

ii. *The* (Schwartz) distributions *or* (tempered) distributions *are the linear maps* $u$: $\mathscr{S} \to \mathbb{C}$ *which satisfy*

$$|u(\varphi)| \leqslant C \|\varphi\|_{k,\mathscr{S}}$$

*for some* $C > 0$ *and* $k \in \mathbb{N}_0$. *In that case we write* $u \in \mathscr{S}'$.

iii. *A sequence* $(u_n) \subset \mathscr{S}'$ converges *to* $u \in \mathscr{S}'$ *if* $u_n(\varphi) \to u(\varphi)$ *for all* $\varphi \in \mathscr{S}$. *One can show that then there exist* $k \in \mathbb{N}_0$ *and* $C > 0$ *with* $|u_n(\varphi)| \leqslant C \|\varphi\|_{k,\mathscr{S}}$ *for all* $n \in \mathbb{N}$*; see [49], Theorem V.7.*

**Example 3.2.** Clearly $L^p = L^p(\mathbb{R}^d) \subset \mathscr{S}'$ for all $p \in [1, \infty]$ if we identify $u \in L^p$ with the map $\varphi \mapsto \int_{\mathbb{R}^d} u(x) \varphi(x) dx$. Also, the space of finite signed measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is contained in $\mathscr{S}'$. Another example of a tempered distribution is $\varphi \mapsto \partial^\mu \varphi(x)$ for $\mu \in \mathbb{N}_0^d$ and $x \in \mathbb{R}^d$. A continuous function $u$ is in $\mathscr{S}'$ if and only if it has at most polynomial growth at infinity.

- Question: Find yet another example of a tempered distribution, which is not of the same type as the ones above.

Using duality, we can define many linear maps on tempered distributions. If $A : \mathscr{S} \to \mathscr{S}$ is such that there exists a linear map $A^T : \mathscr{S} \to \mathscr{S}$ which satisfies for all $\varphi, \psi \in \mathscr{S}$

$$\int_{\mathbb{R}^d} (A\varphi)(x) \psi(x) dx = \int_{\mathbb{R}^d} \varphi(x) (A^T \psi)(x) dx,$$

and such that for all $m \in \mathbb{N}_0$ there exist $k_m \in \mathbb{N}_0$, $C_m > 0$ with $\|A^T \varphi\|_{m,\mathscr{S}} \leqslant C_m \|\varphi\|_{k_m,\mathscr{S}}$. Then we define for $u \in \mathscr{S}'$ the tempered distribution $Au$ via

$$(Au)(\varphi) := u(A^T \varphi).$$

**Example 3.3.**

i. We can *differentiate* distributions: For $\mu \in \mathbb{N}_0^d$ and $A = \partial^\mu$ we have $A^T = (-1)^{|\mu|} \partial^\mu$.

ii. We can *multiply* distributions with (very) nice functions: For $f \in C^\infty$ with all partial derivatives of at most polynomial growth and $A\varphi = f\varphi$ we have $A^T = A$.

iii. We can take the *Fourier transform* of distributions: For the Fourier transform

$$A\varphi(z) = \mathscr{F}\varphi(z) := \hat{\varphi}(z) := \int_{\mathbb{R}^d} \varphi(x) e^{-2\pi i x \cdot z} dx$$

we have $A^T = A$. See Theorem 1.19 of [3].

iv. We can take the *inverse Fourier transform* of distributions: For the inverse Fourier transform

$$A\varphi(z) = \mathscr{F}^{-1}\varphi(z) = \int_{\mathbb{R}^d} \varphi(x) e^{2\pi i x \cdot z} dx$$

we have $A^T = A$.

v. We can *convolve* distributions with nice functions: For $\chi \in \mathscr{S}$ and the convolution

$$A\varphi = \chi * \varphi = \int_{\mathbb{R}^d} \chi(\cdot - y)\varphi(y)\mathrm{d}y$$

we have $A^T\varphi = (\chi(-\cdot)) * \varphi$. In this case one can show that $\chi * u \in C^\infty \cap \mathscr{S}'$ for all $u \in \mathscr{S}'$.

vi. We can *rescale* distributions: For $\lambda > 0$ and $\alpha, \beta \in \mathbb{R}$ and

$$A\varphi(x) = \lambda^\alpha \varphi(\lambda^\beta x)$$

we have $A^T\psi(x) = \lambda^{\alpha - \beta d}\psi(\lambda^{-\beta}x)$. More generally, if $L \in \mathbb{R}^{d\times d}$ is an invertible matrix $\mathbb{R}^d$ and $A\varphi(x) = \varphi(Lx)$, then $A^T\psi(x) = |\det L|^{-1}\psi(L^{-1}x)$.

The main reason for considering test functions in $\mathscr{S}$ rather than the simpler space $C_c^\infty \subset \mathscr{S}$ is that in this way we can define the Fourier transform by duality. Since $C_c^\infty$ is not closed under Fourier transformation (if $\varphi \in C_c^\infty$, then in general $\mathscr{F}\varphi \notin C_c^\infty$), this would not be possible if we allowed only test functions in $C_c^\infty$.

- Question: Check that if $\pi$ is a finite measure on $\mathbb{R}^d$, then

$$\mathscr{F}\pi(z) = \int_{\mathbb{R}^d} e^{-2\pi i x\cdot z}\pi(dx),$$

in the sense that $(\mathscr{F}\pi)(\varphi) = \int_{\mathbb{R}^d} \mathscr{F}\pi(z)\varphi(z)dz$.

**Example 3.4.** Let $u \in \mathscr{S}'$ and $\varphi, \psi \in \mathscr{S}$. The following relations will be used all the time:

- $\mathscr{F}^{-1}\mathscr{F}u = \mathscr{F}\mathscr{F}^{-1}u = u$ for all $u \in \mathscr{S}'$ (see Theorem 1.19 of [3]).
- Parseval's identity:

$$\langle \varphi, \psi \rangle_{L^2(\mathbb{R}^d, \mathbb{C})} = \int_{\mathbb{R}^d} \varphi(x)\psi(x)^* dx = \int_{\mathbb{R}^d} \hat{\varphi}(x)\hat{\psi}(x)^* dx = \langle \hat{\varphi}, \hat{\psi} \rangle_{L^2(\mathbb{R}^d, \mathbb{C})}$$

  where $(\cdot)^*$ denotes complex conjugation. See Theorem 1.25 of [3] which proves the identity for $\varphi = \psi$; apply polarization to deduce it for general $\varphi, \psi$. By extension we have $u(\varphi^*) = \hat{u}(\hat{\varphi}^*)$.
- $\widehat{\partial^\mu u} = (2\pi i x)^\mu \hat{u}$ (Proposition 1.24 of [3]), where $(2\pi i x)^\mu = (2\pi i)^{|\mu|} x_1^{\mu_1} \times \ldots \times x_d^{\mu_d}$.
- $\widehat{u\varphi} = \hat{u} * \hat{\varphi}$ (Proposition 1.24 of [3]).
- $\widehat{u * \varphi} = \hat{u}\hat{\varphi}$ (Proposition 1.24 of [3]).
- $\mathrm{supp}(\varphi * \psi) \subset \overline{\mathrm{supp}(\varphi) + \mathrm{supp}(\psi)} := \overline{\{x + y \colon x \in \mathrm{supp}(\varphi), y \in \mathrm{supp}(\psi)\}}$ (easy to check).

**Remark 3.5.** The cited results in [3] sometimes have additional factors of $2\pi$. This is because in [3] the Fourier transform is defined as $\mathscr{F}\varphi(z) = \int_{\mathbb{R}^d} \varphi(x)e^{-ix\cdot z}dx$. Including the additional $2\pi$ in the exponential has the advantage that Parseval's identity and many other results hold without factors. The only point where we see the $2\pi$ appear is in $\widehat{\partial^\mu u} = (-2\pi i x)^\mu u$, which would be $\widehat{\partial^\mu u} = (-ix)^\mu u$ with the definition in [3].

**Definition 3.6.** *Sometimes we will talk about the* support *of $u \in \mathscr{S}'$. If $U \subset \mathbb{R}^d$ is open, then $u$* vanishes on $U$ *if $u(\varphi) = 0$ for all $\varphi \in \mathscr{S}$ with* $\mathrm{supp}(\varphi) \subset U$. *The union of all open sets on which $u$ vanishes is still an open set. The* support $\mathrm{supp}(u)$ *of $u$ is the complement of this open set (and therefore always closed).*

We will also constantly use the following fundamental inequalities:

**Lemma 3.7.** **(Hölder's inequality)** *Let $p, q, r \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then*

$$\|uv\|_{L^r} \leqslant \|u\|_{L^p} \|v\|_{L^q}.$$

**Proof.** See [3], Proposition 1.1. □

**Lemma 3.8.** **(Young's inequality for convolutions)** *Let $p, q, r \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Then*

$$\|u * v\|_{L^r} \leqslant \|u\|_{L^p} \|v\|_{L^q}.$$

**Proof.** See [3], Lemma 1.4. □

We will refer to this as "Young's inequality", omitting "for convolutions". If we need Young's inequality for products we will distinguish this by explicitly mentioning "for products".

- Question: Check that the support of the Dirac delta $\delta(\varphi) = \varphi(0)$ is $\{0\}$.

  Define $U = \mathbb{R}^d \setminus \{0\}$. If $\varphi$ is such that $\operatorname{supp}(\varphi) \subset U$, then $\varphi(0) = \delta(\varphi) = 0$. On the other hand the only larger open set $U' \supset U$ would be $U' = \mathbb{R}^d$, and it is not true that $\varphi(0) = 0$ for all $\varphi$ with $\operatorname{supp}(\varphi) \subset \mathbb{R}^d$. Therefore, $U$ is the largest open set on which $\delta$ vanishes and thus $\operatorname{supp}(\varphi) = \{0\}$.

## 3.2 Besov spaces

The main difficulty we encountered when trying to solve singular SPDEs like the KPZ equation was that we had to multiply distributions. For $u \in \mathscr{S}'$ and $\varphi \in C^\infty$ with partial derivatives of polynomial growth we can define the product $u\varphi$ by duality. But if $v$ is a non-smooth function or even a distribution, then the duality approach breaks down and we need other arguments to define $uv$. If say $u, v \in L^2$, then $uv \in L^1$ is of course also well defined, so we might hope to find another approach that makes sense of $uv$ for all $u, v \in \mathscr{S}'$. But this is not possible:

**Example 3.9.** **(Schwartz)** In $d = 1$ we can turn $\frac{1}{x}$ into a tempered distribution via the so called *principal value*. The details of that construction are not important for us, but with it we formally obtain for the Dirac delta $\delta$ (i.e. $\delta(\varphi) = \varphi(0)$)

$$0 = (\delta \times x) = (\delta \times x) \times \frac{1}{x} \neq \delta \times \left(x \times \frac{1}{x}\right) = \delta \times 1 = \delta.$$

This example shows that a general extension of the product $uv$ to distributions or non-smooth functions $v$ is not possible. But we will see that if we restrict both $u$ and $v$ to suitable subspaces of $\mathscr{S}'$, then $uv$ is canonically defined as a continuous extension of the product of Schwartz functions. Of course, the example $u, v \in L^2$ from above works, but we are interested in situations where at least one of $u, v$ is a bona fide distribution and not a function. The simplest solution is to require $u$ and $v$ to have compatible regularity. For that purpose we need to introduce regularities on distribution spaces.

To measure the regularity of distributions we first note that if $u \in \mathscr{S}'$ with $\operatorname{supp}(\hat{u}) \subset K$, where $K \subset \mathbb{R}^d$ is a compact set, then there exists $\varphi \in C_c^\infty$ with $\varphi|_K \equiv 1$ and therefore

$$u = \mathscr{F}^{-1}(\hat{u}) = \mathscr{F}^{-1}(\varphi \hat{u}) = (\mathscr{F}^{-1} \varphi) * u.$$

Since $\mathscr{F}^{-1}\varphi \in \mathscr{S}$ we deduce that $u \in C^{\infty}$. Moreover, if $|z| \simeq \lambda$ for all $z \in \operatorname{supp}(\hat{u})$, then essentially we can picture $u$ as a sine-function with period $(2\pi\lambda)^{-1}$ (imagine $d=1$ for simplicity). So if $\lambda$ is small, then $u$ is smooth and oscillating slowly but if $\lambda \gg 1$, then $u$ is very wild. This suggests that smooth functions have some decay in their Fourier transform. It turns out that measuring the size of single Fourier coefficients does not provide enough information and instead it is more useful to group the different frequency ranges into blocks. More precisely, we would like to decompose

$$u = \mathscr{F}^{-1}(\hat{u}) = \mathscr{F}^{-1}(\mathbb{1}_{[0,1)}(|\cdot|)\hat{u}) + \sum_{j \geqslant 0}^{\infty} \mathscr{F}^{-1}(\mathbb{1}_{[2^j, 2^{j+1})}(|\cdot|)\hat{u}) = \Delta_{-1}u + \sum_{j=0}^{\infty} \Delta_j u.$$

Then $\Delta_j u$ is the projection of $u$ onto its frequencies of order $\simeq 2^j$. Since frequencies of order $2^j$ correspond to spatial scales of order $2^{-j}$, the sum $\sum_{i \leqslant j} \Delta_i u$ provides a description of $u$ up to the spatial scale $2^{-j}$. For a smooth function $u$ this should already give a very accurate picture of $u$, and therefore we expect $\Delta_j u$ to rapidly decay as $j \to \infty$. Measuring the strength of that decay will provide us with a notion of regularity.

But there are two problems with the above formal decomposition: First of all it is not even well defined, because we are multiplying $\hat{u} \in \mathscr{S}'$ with non-smooth indicator functions. And even in situations where we can make sense of this product, it still turns out that the operation $u \mapsto \Delta_j u$ is quite badly behaved. For example, we would like to estimate $\|\Delta_j u\|_{L^p} \leqslant \|\mathscr{F}^{-1}(\mathbb{1}_{[2^j, 2^{j+1})}(|\cdot|))\|_{L^1} \|u\|_{L^p}$ via Young's inequality, but the $L^1$ norm on the right hand side is infinite because while $\mathscr{F}^{-1}(\mathbb{1}_{[2^j, 2^{j+1})}(|\cdot|)) \in C^{\infty}$, it is not in $L^1$.

**Definition 3.10.** *A* (smooth, dyadic) partition of unity *consists of two positive and radial functions $\rho_{-1}, \rho_0 \in C_c^{\infty}(\mathbb{R}^d, \mathbb{R})$ such that the support of $\rho_{-1}$ is contained in a ball $B \subset \mathbb{R}^d$, the support of $\rho_0$ is contained in an annulus $A = \{x \in \mathbb{R}^d : 0 < a \leqslant |x| \leqslant b\}$, and with*

$$\rho_j := \rho_0(2^{-j}\cdot)$$

*the following conditions are satisfied:*

*i.* $\sum_{j=-1}^{\infty} \rho_j(x) = 1$ *for all* $x \in \mathbb{R}^d$*;*

*ii.* $\operatorname{supp}(\rho_i) \cap \operatorname{supp}(\rho_j) = \emptyset$ *if* $|i-j| > 1$.

Such a partition of unity exists, see Proposition 2.10 of [3]. For the rest of these lecture notes we fix one partition of unity $(\rho_j)_{j \geqslant -1}$.

**Definition 3.11.** *For $u \in \mathscr{S}'$ and $j \geqslant -1$ we define the* Littlewood-Paley blocks *of $u$ as*

$$\Delta_j u = \mathscr{F}^{-1}(\rho_j \hat{u}).$$

**Notation.** *We write $K_j = \mathscr{F}^{-1}\rho_j$, so that $\Delta_j u = K_j * u$. We also use the notation*

$$\Delta_{\leqslant j} u = \sum_{i \leqslant j} \Delta_i u, \qquad \Delta_{<j} u = \sum_{i<j} \Delta_i u, \qquad \Delta_{\geqslant j} u = \sum_{i \geqslant j} \Delta_i u, \qquad \Delta_{>j} u = \sum_{i>j} \Delta_i u,$$

$$K_{\leqslant j} = \sum_{i \leqslant j} K_i, \qquad K_{<j} = \sum_{i<j} K_i.$$

*The kernels $K_j, K_{<j}, K_{\leqslant j}$ are all bounded in $L^1$, uniformly in $j$. Moreover, for $j \geqslant 0$ we have the scaling relation $K_j = 2^{jd} K_0(2^j \cdot)$.*

Note that no nice kernel exists for $\Delta_{>j}$, because we would need to take $K_{>j} = \delta - K_{\leqslant j}$, where $\delta$ is the Dirac delta and $K_{\leqslant j}$ is a smooth kernel.

- Question: Check that $\|K_j\|_{L^1} = \|K_0\|_{L^1}$ for $j \geqslant 0$.

  We have

$$K_j(x) = \mathscr{F}^{-1}\rho_j(x) = \int_{\mathbb{R}^d} \rho_0(2^{-j}z) e^{2\pi i x \cdot z} dz = 2^{jd} \int_{\mathbb{R}^d} \rho_0(z) e^{2\pi i x \cdot 2^j z} dz = 2^{jd} K_0(2^j x),$$

  and therefore

$$\int_{\mathbb{R}^d} |K_j(x)| dx = \int_{\mathbb{R}^d} |2^{jd} K_0(2^j x)| dx = \int_{\mathbb{R}^d} |K_0(x)| dx.$$

We have

$$\Delta_{\leqslant j} u(\varphi) = \sum_{i \leqslant j} \mathscr{F}^{-1}(\rho_i \mathscr{F} u)(\varphi) = \sum_{i \leqslant j} (\rho_i \mathscr{F} u)(\mathscr{F}^{-1}\varphi) = u\left(\sum_{i \leqslant j} \mathscr{F}(\rho_i \mathscr{F}^{-1}\varphi)\right),$$

and therefore

$$u = \sum_{j \geqslant -1} \Delta_j u = \lim_{j \to \infty} \Delta_{\leqslant j} u$$

for all $u \in \mathscr{S}'$.

If $u \in L^p$, then by Young's inequality we have uniformly in $j$:

$$\|\Delta_j u\|_{L^p} \leqslant \|K_j\|_{L^1} \|u\|_{L^p} \lesssim \|u\|_{L^p}.$$

As discussed above, we want to describe the regularity of $u \in \mathscr{S}'$ by the decay (or growth) of $\Delta_j u$. For that purpose we first have to decide how to measure the size of $\Delta_j u$. A canonical choice is to consider the $L^p$-norm for $p \in [1, \infty]$. Recall that $\mathscr{F}\Delta_j u$ is compactly supported, and therefore $\Delta_j u \in C^\infty \cap \mathscr{S}'$. In particular,

$$\|\Delta_j u\|_{L^p} := \begin{cases} (\int_{\mathbb{R}^d} |\Delta_j u(x)|^p dx)^{1/p}, & p < \infty, \\ \operatorname{ess\,sup}_{x \in \mathbb{R}^d} |\Delta_j u(x)|, & p = \infty, \end{cases}$$

is well defined (but possibly infinite). Actually $\|\Delta_j u\|_{L^\infty} = \sup_{x \in \mathbb{R}^d} |\Delta_j u(x)|$ because $\Delta_j u$ is continuous and therefore the essential supremum is the supremum.

**Definition 3.12.** *For $\alpha \in \mathbb{R}$ and $p, q \in [1, \infty]$ the* Besov space $B^\alpha_{p,q}$ *is defined as*

$$B^\alpha_{p,q} = \left\{ u \in \mathscr{S}' : \|u\|_{B^\alpha_{p,q}} = \|(2^{j\alpha} \|\Delta_j u\|_{L^p})_{j \geqslant -1}\|_{\ell^q_j} < \infty \right\}.$$

*$B^\alpha_{p,q}$ is a Banach space for all $\alpha, p, q$.*

So $p$ describes the integrability and $\alpha$ describes the regularity (i.e. the decay of the blocks). The index $q$ provides some fine-tuning and it is not very important: Indeed we have for $q_1 \leqslant q_2$ and $\alpha \in \mathbb{R}$ the inclusions

$$B^\alpha_{p,q_1} \subset B^\alpha_{p,q_2} \subset B^{\alpha'}_{p,q_1}$$

whenever $\alpha' < \alpha$ (Exercise: if this is not clear to you, check that $\|\cdot\|_{\ell^{q_2}} \leqslant \|\cdot\|_{\ell^{q_1}}$ for $q_1 \leqslant q_2$; see also Sheet 2).

The Gaussian noises that we will consider have the same regularity index $\alpha$ in any Besov space $B^\alpha_{p,q}$, which, as we will see, is a consequence of the comparability of moments for Gaussian random variables. Therefore, we mainly work in the easiest setting $p = q = \infty$, for which we introduce a special notation:

$$\mathscr{C}^\alpha = B^\alpha_{\infty,\infty}, \qquad \|\cdot\|_\alpha = \|\cdot\|_{B^\alpha_{\infty,\infty}}.$$

**Exercise 3.1.** Let $\delta$ denote the Dirac delta, $\delta(\varphi) = \varphi(0)$. Show that $\delta \in B^{-d(1-1/p)}_{p,\infty}$ for all $p \in [1, \infty]$. So when dealing with equations involving the Dirac delta (say as initial condition), it may be advantageous to work in Besov spaces with finite integrability index.

**Exercise 3.2.** Show that $\|u\|_\alpha \lesssim \|u\|_\beta$ for $\alpha \leqslant \beta$, that $\|u\|_{L^\infty} \lesssim \|u\|_\alpha$ for $\alpha > 0$, that $\|u\|_\alpha \lesssim \|u\|_{L^\infty}$ for $\alpha \leqslant 0$, and that $\|\Delta_{\leqslant j} u\|_{L^\infty} \lesssim 2^{-j\alpha}\|u\|_\alpha$ for $\alpha < 0$ and $\|\Delta_{>j} u\|_{L^\infty} \lesssim 2^{-j\alpha}\|u\|_\alpha$ for $\alpha > 0$. We will often use these inequalities without explicitly mentioning it.

If $\alpha \in \mathbb{R}_+ \setminus \mathbb{N}_0$, then one can show that $\mathscr{C}^\alpha$ is the space of bounded $\lfloor \alpha \rfloor$ times continuously differentiable functions, with bounded partial derivatives, and whose partial derivatives of order $\lfloor \alpha \rfloor$ are $(\alpha - \lfloor \alpha \rfloor)$-Hölder continuous, with norm equivalent to

$$\|u\|_\alpha \simeq \|u\|_{C_b^\alpha} = \sum_{|\mu| \leqslant \lfloor \alpha \rfloor} \|\partial^\mu u\|_\infty + \sum_{|\mu| = \lfloor \alpha \rfloor} \sup_{x \neq y} \frac{|\partial^\mu u(x) - \partial^\mu u(y)|}{|x - y|^{\alpha - \lfloor \alpha \rfloor}}.$$

But for $k \in \mathbb{N}_0$ the space $\mathscr{C}^k$ is strictly larger than $C_b^k$. In fact $\mathscr{C}^0$ is even larger than $L^\infty$. We will see the equivalence for $\alpha \in (0, 1)$ as an exercise below, but before that we need the following Bernstein inequality, which is very useful when dealing with functions with compactly supported Fourier transform.

**Lemma 3.13. (Bernstein type inequality)** *Let $B \subset \mathbb{R}^d$ be a ball, let $k \in \mathbb{N}_0$, and let $1 \leqslant p \leqslant q \leqslant \infty$. There exists a constant $C > 0$ which depends only on $k, B, p, q$, such that for all $\lambda > 0$ and for all $u \in L^p$ with $\operatorname{supp}(\mathscr{F} u) \subseteq \lambda B$ we have*

$$\max_{\mu \in \mathbb{N}^d : |\mu| = k} \|\partial^\mu u\|_{L^q} \leqslant C \lambda^{k + d(\frac{1}{p} - \frac{1}{q})} \|u\|_{L^p}.$$

**Proof.** Let $\psi \in C_c^\infty$ with $\psi \equiv 1$ on $B$ and write $\psi_\lambda(x) = \psi(\lambda^{-1} x)$. Young's inequality gives

$$\|\partial^\mu u\|_{L^q} = \|\partial^\mu \mathscr{F}^{-1}(\psi_\lambda \hat{u})\|_{L^q} = \|(\partial^\mu \mathscr{F}^{-1}(\psi_\lambda)) * u\|_{L^q} \leqslant \|\partial^\mu \mathscr{F}^{-1}(\psi_\lambda)\|_{L^r} \|u\|_{L^p},$$

where $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r}$. Now it suffices to note that (if $r < \infty$)

$$\begin{aligned}
\|\partial^\mu \mathscr{F}^{-1}(\psi_\lambda)\|_{L^r} &= \left( \int_{\mathbb{R}^d} |\partial^\mu (\lambda^d \mathscr{F}^{-1} \psi(\lambda x))|^r dx \right)^{1/r} \\
&= \left( \lambda^{(|\mu| + d) r} \int_{\mathbb{R}^d} |(\partial^\mu \mathscr{F}^{-1} \psi)(\lambda x)|^r dx \right)^{1/r} \\
&= \left( \lambda^{(|\mu| + d) r - d} \int_{\mathbb{R}^d} |\partial^\mu \mathscr{F}^{-1} \psi(x)|^r dx \right)^{1/r} \\
&= \lambda^{|\mu| + d(1 - \frac{1}{r})} \|\partial^\mu \mathscr{F}^{-1} \psi\|_{L^r}.
\end{aligned}$$

The claim follows by plugging in the equality $1 - \frac{1}{r} = \frac{1}{p} - \frac{1}{q}$. The case $r = \infty$ is similar and slightly easier. □

- Complete the proof by treating the case $r = \infty$.

**Corollary 3.14.** *For $\alpha \in \mathbb{R}$, $\mu \in \mathbb{N}_0^d$, and $u \in \mathscr{C}^\alpha$, we have*

$$\|\partial^\mu u\|_{\alpha - |\mu|} \lesssim \|u\|_\alpha.$$

**Proof.** We have by the Bernstein type inequality:

$$\begin{aligned}
\|\Delta_j (\partial^\mu u)\|_{L^\infty} &= \|K_j * (\partial^\mu u)\|_{L^\infty} = \|\partial^\mu (K_j * u)\|_{L^\infty} = \|\partial^\mu \Delta_j u\|_{L^\infty} \\
&\lesssim 2^{j|\mu|} \|\Delta_j u\|_{L^\infty} \lesssim 2^{-j(\alpha - |\mu|)} \|u\|_\alpha.
\end{aligned}$$

□

We can also use the Bernstein type inequality to see the claimed equivalence of $\mathscr{C}^\alpha$ and $C_b^\alpha$, at least in the case $\alpha \in (0,1)$ (the case $\alpha > 1$ is similar but more technical). The next exercise is extremely instructive, because it is based on many arguments we will often encounter later (convergent and divergent geometric series, smuggling in constant terms into integrals against $K_j$, treating small and large scales separately).

**Exercise 3.3.** Let $\alpha \in (0,1)$. Then $\mathscr{C}^\alpha = C_b^\alpha$ is the space of bounded $\alpha$-Hölder continuous functions, and

$$\|u\|_\alpha \simeq \|u\|_{C_b^\alpha}.$$

Another simple application of the Bernstein type inequality is the Besov embedding theorem:

**Theorem 3.15. (Besov embedding)** *Let $1 \leqslant p_1 \leqslant p_2 \leqslant \infty$ and $1 \leqslant q_1 \leqslant q_2 \leqslant \infty$, and let $\alpha \in \mathbb{R}$. Then for all $u \in \mathscr{S}'$*

$$\|u\|_{B_{p_2,q_2}^{\alpha - d\left(\frac{1}{p_1} - \frac{1}{p_2}\right)}} \lesssim \|u\|_{B_{p_1,q_1}^\alpha}.$$

- Question: Give the proof yourself. It is only 2 lines.

The next lemma, a characterization of Besov regularity for functions that can be decomposed into pieces which are localized in Fourier space, will also be immensely useful. Recall that an *annulus* is a set $A = \{x \in \mathbb{R}^d : a \leqslant |x| \leqslant b\}$ for some $0 < a < b$, and a *ball* is a set $B = \{x \in \mathbb{R}^d : |x| \leqslant b\}$.

**Lemma 3.16.**

i. *Let $A \subset \mathbb{R}^d$ be an annulus, let $\alpha \in \mathbb{R}$, and let $(u_j)_{j \geqslant 0}$ be a sequence of smooth functions with $\operatorname{supp}(\mathscr{F} u_j) \subset 2^j A$ and such that $\|u_j\|_{L^\infty} \lesssim 2^{-j\alpha}$ for all $j$. Then*

$$u = \sum_{j \geqslant 0} u_j \in \mathscr{C}^\alpha \qquad \text{and} \qquad \|u\|_\alpha \lesssim \sup_{j \geqslant 0} \{2^{j\alpha} \|u_j\|_{L^\infty}\}.$$

ii. *Let $B \subset \mathbb{R}^d$ be a ball, let $\alpha > 0$, and let $(u_j)_{j \geqslant -1}$ be a sequence of smooth functions with $\operatorname{supp}(\mathscr{F} u_j) \subset 2^j B$ and such that $\|u_j\|_{L^\infty} \lesssim 2^{-j\alpha}$ for all $j$. Then*

$$u = \sum_{j \geqslant -1} u_j \in \mathscr{C}^\alpha \qquad \text{and} \qquad \|u\|_\alpha \lesssim \sup_{j \geqslant -1} \{2^{j\alpha} \|u_j\|_{L^\infty}\}.$$

**Proof.**

i. If $\mathscr{F} u_j$ is supported in $2^j A$, then $\Delta_i u_j \neq 0$ only if $2^i \simeq 2^j$ and therefore

$$\begin{aligned}\|\Delta_i u\|_{L^\infty} &\leqslant \sum_{j: 2^j \simeq 2^i} \|\Delta_i u_j\|_{L^\infty} \lesssim \sup_{k \geqslant -1} \{2^{k\alpha} \|u_k\|_{L^\infty}\} \sum_{j: 2^j \simeq 2^i} 2^{-j\alpha} \\ &\simeq \sup_{k \geqslant -1} \{2^{k\alpha} \|u_k\|_{L^\infty}\} 2^{-i\alpha}.\end{aligned}$$

ii. If $\mathscr{F} u_j$ is supported in $2^j B$, then $\Delta_i u_j \neq 0$ only if $2^i \lesssim 2^j$. Therefore,

$$\begin{aligned}\|\Delta_i u\|_{L^\infty} &\leqslant \sum_{j: 2^j \gtrsim 2^i} \|\Delta_i u_j\|_{L^\infty} \leqslant \sup_{k \geqslant -1} \{2^{k\alpha} \|u_k\|_{L^\infty}\} \sum_{j: 2^j \gtrsim 2^i} 2^{-j\alpha} \\ &\simeq \sup_{k \geqslant -1} \{2^{k\alpha} \|u_k\|_{L^\infty}\} 2^{-i\alpha},\end{aligned}$$

where in the last step we used that $\alpha > 0$. □

A similar result also holds for Besov spaces $B^\alpha_{p,q}$ with general $p,q\in[1,\infty]$. As a first application, one can use this lemma to show that while the norm $\|\cdot\|_{B^\alpha_{p,q}}$ depends on the specific partition of unity used to define it, the space $B^\alpha_{p,q}$ does not and every other partition of unity induces an equivalent norm.

- Use the previous lemma to prove that if $(\tilde\rho_j)_{j\geqslant -1}$ is another dyadic partition of unity, and $\tilde\Delta_j u=\mathscr{F}^{-1}(\tilde\rho_j\mathscr{F}u)$ and $\|u\|_{\tilde B^\alpha_{p,q}}=\|(2^{j\alpha}\|\tilde\Delta_j u\|_{L^p})_{j\geqslant -1}\|_{\ell^q_j}$, then $\|u\|_{\tilde B^\alpha_{p,q}}\simeq\|u\|_{B^\alpha_{p,q}}$.

## 3.3 The paraproduct and the resonant product

Now that we know how to measure the regularity of distributions, let us come back to the problem of multiplying distributions. We will follow Bony [8] who introduced *paraproducts* which provide a useful tool to decompose the multiplication into simpler problems. The usefulness of the paraproduct comes from the following simple observation:

**Lemma 3.17.** *There exists an annulus $A$ such that for all $j\geqslant 1$ and all $i\leqslant j-2$*

$$\operatorname{supp}(\mathscr{F}(\Delta_i u\Delta_j v))\subset 2^j A,\qquad u,v\in\mathscr{S}'.$$

*Moreover, there exists a ball $B$ such that for all $i,j\geqslant -1$ with $|i-j|\leqslant 1$*

$$\operatorname{supp}(\mathscr{F}(\Delta_i u\Delta_j v))\subset 2^j B.$$

**Proof.** This is quite simple: If $i\geqslant 0$, then

$$\begin{aligned}\operatorname{supp}(\mathscr{F}(\Delta_i u\Delta_j v)) &= \operatorname{supp}(\mathscr{F}\Delta_i u*\mathscr{F}\Delta_j u)\subset\overline{\operatorname{supp}(\mathscr{F}\Delta_i u)+\operatorname{supp}(\mathscr{F}\Delta_j u)}\\ &\subset 2^i\tilde A+2^j\tilde A=2^j(2^{i-j}\tilde A+\tilde A)\end{aligned}$$

for an annulus $\tilde A$. By our assumptions on the dyadic partition of unity we can choose $\tilde A$ such that $2^{-k}\tilde A\cap\tilde A=\emptyset$ for all $k\geqslant 2$ and therefore $2^{i-j}\tilde A+\tilde A\subset A$ for a new annulus $A$ and all $i\leqslant j-2$. The argument for $i=-1$ is similar.

If on the other hand $|i-j|\leqslant 1$, then all we can say is that $\operatorname{supp}(\mathscr{F}(\Delta_i u\Delta_j v))\subset 2^j B$ for a ball $B$. □

For the following heuristic discussion we assume that $u$ and $\mathscr{F}u$ are functions. Recall that if $u$ is such that $\mathscr{F}u(z)\neq 0$ only for $|z|\simeq\lambda$, then $u$ is essentially a superposition of $e^{2\pi i z\cdot x}$ for $|z|\simeq\lambda$. So on spatial scales $r\ll\lambda^{-1}$, $u$ is nearly constant:

$$|u(x+r)-u(x)|\lesssim\lambda r\ll 1.$$

On spatial scales $r\gg\lambda^{-1}$, $u$ looks highly oscillatory, and its average over a region of size $r$ nearly vanishes: Let's consider $d=1$ for simplicity, then

$$\left|\frac{1}{r}\int_x^{x+r}e^{2\pi i\lambda y}dy\right|=\frac{1}{r}\left|\frac{e^{2\pi i\lambda(x+r)}-e^{2\pi i\lambda x}}{2\pi i\lambda}\right|\leqslant\frac{2}{r\lambda}\ll 1.$$

We say that $u$ "lives on the spatial scale $2^{-j}$".

Therefore, Lemma 3.17 shows that multiplying $\Delta_j v$, a function that lives on the spatial scale $2^{-j}$, with $\Delta_{\leqslant j-2}u$, we obtain a new function $\Delta_{\leqslant j-2}u\Delta_j v$ which still lives on the spatial scale $2^{-j}$. The multiplication does not create any effects on larger or smaller scales. If on the other hand $|i-j|\leqslant 1$, then $\Delta_i u$ and $\Delta_j v$ live on the spatial scale $2^{-j}$, but multiplying the two together can create effects on the scale 1, i.e. small scale contributions work together to create an effect on large scales. Therefore, the large scale contributions of $\sum_{|i-j|\leqslant 1}\Delta_i u\Delta_j v$ might diverge, and we interpret this as a *resonance phenomenon*.

**Example 3.18.** Below we see a slowly oscillating function $u$ (red curve) and a fast sine curve $v$ (blue curve). The product $uv$ is shown under the two curves. We see that the local fluctuations of $uv$ are due to $v$, and that $uv$ is essentially oscillating with the same speed as $v$.

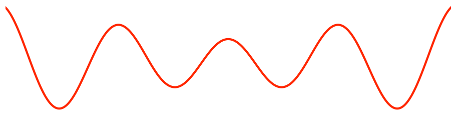

**Figure 3.1.** $u$ oscillates slowly.

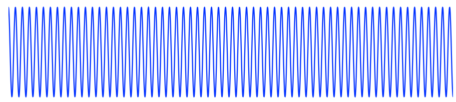

**Figure 3.2.** $v$ is a fast sine curve.

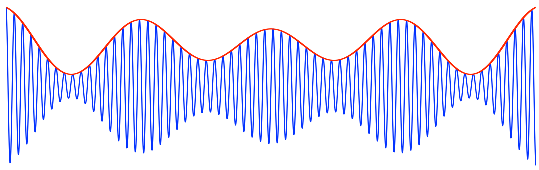

**Figure 3.3.** $uv$ still lives on the same scale as $v$.

Formally we can decompose the product $uv$ of two distributions as

$$uv = \sum_{i,j \geqslant -1} \Delta_i u \, \Delta_j v = u \varolessthan v + u \varogreaterthan v + u \odot v.$$

Here $u \varolessthan v$ is the part of the double sum with $i \leqslant j-2$, $u \varogreaterthan v$ is the part with $i \geqslant j+2$, and $u \odot v$ is the "diagonal" part, where $|i-j| \leqslant 1$. More precisely, we define

$$u \varolessthan v = v \varogreaterthan u = \sum_{j \geqslant -1} \Delta_{\leqslant j-2} u \, \Delta_j v \qquad \text{and} \qquad u \odot v = \sum_{i,j:|i-j| \leqslant 1} \Delta_i u \, \Delta_j v.$$

We call $u \varolessthan v$ and $u \varogreaterthan v$ *paraproducts*, and $u \odot v$ the *resonant product*.

Bony's crucial observation is that $u \varolessthan v$ (and thus $u \varogreaterthan v$) is always a well-defined distribution. Heuristically, $u \varolessthan v$ behaves at large frequencies (i.e. small spatial scales) like $v$ and thus retains the same regularity, and $u$ provides only a frequency modulation of $v$. This can also be seen in Example 3.18 above, where the product $uv$ is actually equal to the paraproduct $u \varolessthan v$ because $u$ has no rapidly oscillating components. The only difficulty in constructing $uv$ for arbitrary distributions lies in handling the diagonal term $u \odot v$. The following key estimates provide the analytically precise formulation of the preceding heuristic discussion:

**Theorem 3.19. (Paraproduct estimates)** *For all $\beta \in \mathbb{R}$ and $u, v \in \mathscr{S}'$ we have*

$$\|u \varolessthan v\|_\beta \lesssim \|u\|_{L^\infty} \|v\|_\beta, \tag{3.1}$$

*and for $\alpha < 0$ furthermore*

$$\|u \varolessthan v\|_{\alpha+\beta} \lesssim \|u\|_\alpha \|v\|_\beta. \tag{3.2}$$

*If $\alpha + \beta > 0$ we have*

$$\|u \odot v\|_{\alpha+\beta} \lesssim \|u\|_\alpha \|v\|_\beta. \tag{3.3}$$

**Proof.** By Lemma 3.17 there exists an annulus $A$ such that $\operatorname{supp}(\mathscr{F}(\Delta_{\leqslant j-2} u \Delta_j v)) \subset 2^j A$, and for $u \in L^\infty$ we have

$$\|\Delta_{\leqslant j-2} u \Delta_j v\|_{L^\infty} \leqslant \|\Delta_{\leqslant j-2} u\|_{L^\infty} \|\Delta_j v\|_{L^\infty} \lesssim \|u\|_{L^\infty} 2^{-j\beta} \|v\|_\beta.$$

Inequality (3.1) now follows from Lemma 3.16. The proof of (3.2) and (3.3) works in the same way, except that for estimating $u \odot v$ we need $\alpha + \beta > 0$ because now the terms of the series are supported in balls and not in annuli. □

The ill–posedness of $u \odot v$ for $\alpha + \beta \leqslant 0$ can be interpreted as a resonance effect since $u \odot v$ contains exactly those part of the double series where $u$ and $v$ are in the same frequency range. As discussed above, the paraproduct $u \otimes v$ can be interpreted as frequency modulation of $v$.

In combination with the estimates from Exercise 3.2 above we deduce the following simple corollary:

**Corollary 3.20.** *Let $\alpha + \beta > 0$. Then the product $(u, v) \mapsto uv$ of Schwartz functions can be extended to a bounded bilinear operator from $\mathscr{C}^\alpha \times \mathscr{C}^\beta$ to $\mathscr{C}^{\alpha \wedge \beta}$. While $u \otimes v$, $u \otimes v$, and $u \odot v$ depend on our specific dyadic partition of unity, the product $uv$ does not.*

The condition $\alpha + \beta > 0$ is sharp:

Question: Find a counterexample to show that for $\alpha + \beta < 0$ there exists no continuous extension of the product to $\mathscr{C}^\alpha \times \mathscr{C}^\beta$. *Hint: Think of $u(x) = n^{-\tilde{\alpha}} e^{inx}$ for $\tilde{\alpha} > \alpha$.*

**Example 3.21.** Let $\alpha, \beta \in \mathbb{R}$ and consider the functions $u_n(x) = n^{-\tilde{\alpha}} e^{inx}$ on $\mathbb{R}$, and $v_n(x) = n^{-\tilde{\beta}} e^{-inx}$. It is easy to see that $\|u_n\|_\alpha \to 0$ and $\|v_n\|_\alpha \to 0$ for all $\alpha < \tilde{\alpha}$ and $\beta < \tilde{\beta}$. Nonetheless

$$u_n v_n \equiv n^{-(\tilde{\alpha} + \tilde{\beta})}$$

diverges to $\infty$ if $\tilde{\alpha} + \tilde{\beta} < 0$, and stays constant for $\tilde{\alpha} + \tilde{\beta} = 0$. To obtain a counterxample for $\alpha + \beta = 0$ we could consider a superposition $u = \sum_n n^{-\alpha} e^{2\pi i C 2^n x}$ for $C > 0$ such that $\Delta_j u = n^{-1/2} e^{2\pi i C 2^j x}$, and similarly for $v$.

**Example 3.22.** Let $d = 1$ and $u \in \mathscr{C}^\alpha$ and $v \in \mathscr{C}^\beta$ with $\alpha + \beta > 1$. Then $\partial_t v \in \mathscr{C}^{\beta - 1}$ by Corollary 3.14, and since $\alpha + \beta - 1 > 0$ the product $u \partial_t v$ is well defined. We could also construct this product by differentiating the Young integral:

$$u \partial_t v := \partial_t \int_0^t u_s d y_s,$$

where the derivative is taken in the sense of distributions. Conversely, integration is a linear map and thus we can make sense of $\int_0^\cdot (u \partial_t v) dt$ and this gives an alternative construction of the Young integral.

**Example 3.23.** Let $(B_t)_{t \in [0,T]}$ be a Brownian motion and extend $B|_{(-\infty,0]} \equiv 0$ and $B|_{[T,\infty)} \equiv B_T$. Then $B \in \mathscr{C}^\alpha$ for all $\alpha < 1/2$ almost surely, and as we saw above this implies $\partial_t B \in \mathscr{C}^{\alpha - 1}$. Therefore, the sum of the regularities is $2\alpha - 1 < 0$ and the product $B \partial_t B$ is ill-defined. This manifests itself in the probabilistic phenomenon that there are different reasonable interpretations for the integral $\int_0^t B_s d B_s = \int_0^t (B_s \partial_s B_s) ds$, for example Itô, Stratonovich, or backward Itô, roughly speaking because different approximations lead to different limits.

In $d = 1$ there is actually a certain stiffness, because if $(B^n)$ is a sequence of smooth paths that converge to $B$ in $\mathscr{C}^\alpha$, then we always have

$$B^n \partial_t B^n = \frac{1}{2} \partial_t ((B^n)^2) \to \frac{1}{2} \partial_t (B^2),$$

i.e. there appears to be a canonical interpretation for the product $B\partial_t B$, which corresponds to the Stratonovich integral $\int_0^t B_s \circ dB_s$. However, nobody is forcing us to approximate $B$ with smooth functions, and if we take piecewise constant approximations and compute $\int_0^t B_s^n dB_s^n$ as a discrete sum, then the limit is the Itô integral $\int_0^t B_s dB_s$.

- Show that for $f \in \mathscr{C}^\varepsilon$ and the Dirac delta $\delta$ the product $\delta f$ is well defined and a continuous function of $f$, despite the fact that $\delta \in \mathscr{C}^{-d}$. Do you have any intuition why?

**Remark 3.24.** So far we only considered tempered distributions on $\mathbb{R}^d$, but the same works also on the torus $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$. In that case we simply have $\mathscr{S}(\mathbb{T}^d) = C^\infty(\mathbb{T}^d)$, and $\mathscr{S}'(\mathbb{T}^d)$ consists of all linear maps $u: C^\infty(\mathbb{T}^d) \to \mathbb{C}$ such that $|u(\varphi)| \leqslant C \sum_{|\mu| \leqslant k} \|\partial^\mu \varphi\|_\infty$ for some $C, k > 0$.

The Fourier transform $\mathscr{F}_{\mathbb{T}^d} u(k) = \hat{u}(k) := u(e^{-2\pi i k \cdot (\cdot)})$ is defined pointwise for all $k \in \mathbb{Z}^d$ and all $u \in \mathscr{S}'(\mathbb{T}^d)$, with inverse Fourier transform $\mathscr{F}_{\mathbb{T}^d}^{-1} \eta(x) = \sum_{k \in \mathbb{Z}^d} e^{2\pi i k \cdot x} \eta(k)$. A sequence $(\eta(k))_{k \in \mathbb{Z}^d}$ is the Fourier transform of some $u \in \mathscr{S}'(\mathbb{T}^d)$ if and only if $|\eta(k)|$ grows at most polynomially in $k$.

The Littlewood-Paley blocks are then defined in exactly the same way, and all the results of this section continue to hold. The proofs are mostly the same, only that we have to find a replacement for some scaling arguments: For example it is not true that $(\mathscr{F}_{\mathbb{T}^d}^{-1} \rho_0(2^{-j} \cdot)) = 2^{jd} (\mathscr{F}_{\mathbb{T}^d}^{-1} \rho_0)(2^j \cdot)$.

Usually we can apply the Poisson summation formula to overcome this difficulty: For $\varphi \in \mathscr{S}(\mathbb{R}^d)$ we have

$$\mathscr{F}_{\mathbb{T}^d}^{-1} \varphi(x) := \sum_{k \in \mathbb{Z}^d} e^{2\pi i k \cdot x} \varphi(k) = \sum_{k \in \mathbb{Z}^d} \mathscr{F}^{-1} \varphi(x + k). \tag{3.4}$$

See e.g. [25] for details. We will not work out the details here, and if necessary we simply apply the results of this section on $\mathbb{T}^d$ without commenting on it.

# 4 First examples of pathwise SPDEs

Here we discuss how to use our new tools to solve some interesting SPDEs. We work on the torus $\mathbb{T}^d$, and as mentioned before, all results from Section 3 remain valid on $\mathbb{T}^d$.

**Convention.** From now on we slightly re-define the space $\mathscr{C}^\alpha$: We define it as the closure of $C^\infty(\mathbb{T}^d)$ with respect to $\|\cdot\|_\alpha$. This is a strict subspace of $B^\alpha_{\infty,\infty}$, so all results from Section 3 remain valid. Some aspects of this new space are nicer, for example it is separable while $B^\alpha_{\infty,\infty}$ is not (generally $L^\infty$-based function spaces tend not to be separable). An alternative characterization of our new $\mathscr{C}^\alpha$ is

$$\mathscr{C}^\alpha = \left\{ u \in \mathscr{S}' : \lim_{j \to \infty} 2^{j\alpha} \|\Delta_j u\|_{L^\infty} = 0 \right\}.$$

So to show that $u \in \mathscr{S}'$ is in the new $\mathscr{C}^\alpha$ it suffices to show that $\|u\|_{\alpha'} < \infty$ for some $\alpha' > \alpha$.

## 4.1 Regularity of the linearized $\Phi^4_d$ equation

Equipped with these tools, we now aim to solve the $\Phi^4_d$ equation $\phi: \mathbb{R}_+ \times \mathbb{T}^d \to \mathbb{R}$,

$$\partial_t \phi = \Delta \phi - \phi^3 + \phi + \xi, \tag{4.1}$$

where $\xi$ is a space-time white noise on $\mathbb{R}_+ \times \mathbb{T}^d$ (we now use the notation $\xi$ rather than $\partial_t W$ because Itô integration will not play a big role any more and rather we understand $\xi$ as a space-time distribution).

Originally we were interested in the equation on $\mathbb{R}^d$ instead of $\mathbb{T}^d$, but as will understand later, in infinite volume there are considerable technical problems additionally to the regularity problems that already appear in finite volume. After we understand how to deal with the regularity problems, we could return to the infinite volume problem. But we will not do this here and instead refer to [38].

Let $(P_t)_{t\geqslant 0}$ be the semigroup generated by $\Delta$, i.e.

$$P_t u = p(t,\cdot) * u$$

for the heat kernel

$$p(t,x) = \sum_{k\in\mathbb{Z}^d} e^{2\pi i k\cdot x} e^{-|2\pi k|^2 t} = \sum_{k\in\mathbb{Z}^d} (4\pi t)^{-d/2} \exp\left(-\frac{|x+k|^2}{4t}\right),$$

where the second equality follows from the Poisson summation formula (3.4). Recall that the definition of a mild solution to (4.1) is

$$\begin{aligned}\phi(t) =& P_t\phi_0 + \int_0^t P_{t-s}(-\phi(s)^3 + \phi(s) + \xi(s))ds\\ =& P_t\phi_0 + \int_0^t P_{t-s}(-\phi(s)^3 + \phi(s))ds + \int_0^t P_{t-s}\xi(s)ds.\end{aligned}$$

We do not expect any substantial cancellations between the different terms on the right hand side, so $\phi$ should have at best the same regularity as

$$Z(t) = \int_0^t P_{t-s}\xi(s)ds.$$

So to see in which function space we can hope to solve the equation for $\phi$, let us compute the regularity of $Z$. For that purpose we use the following type of *Kolmogorov continuity criterion*, where we write for a Banach space $X$ and $T>0$, $\gamma\in[0,1]$:

$$\|u\|_{C_T^\gamma X} := \sup_{t\in[0,T]} \|u(t)\|_X + \sup_{0\leqslant s<t\leqslant T} \frac{\|u(t)-u(s)\|_X}{|t-s|^\gamma}.$$

Note that for $\gamma=0$ this is equivalent to the supremum norm, and in that case we also write $C_T X := C_T^0 X$.

**Proposition 4.1.** *Let $(u(t))_{t\in[0,T]}$ be a stochastic process with values in $\mathscr{S}'(\mathbb{T}^d)$ and assume that for all $j\geqslant -1$, for all $0\leqslant s<t\leqslant T$, and for all $x\in\mathbb{T}^d$*

$$\mathbb{E}[|\Delta_j u(0,x)|^p]^{1/p} + \frac{\mathbb{E}[|\Delta_j u(t,x) - \Delta_j u(s,x)|^p]^{1/p}}{|t-s|^\gamma} \leqslant K 2^{-j\alpha}, \tag{4.2}$$

*where $\gamma>1/p$. Then (there exists a modification such that) for all $\gamma'\in(0,\gamma-\frac{1}{p})$ and all $\alpha'<\alpha$*

$$\mathbb{E}\Big[\|u\|^p_{C_T^{\gamma'}\mathscr{C}^{\alpha'-d/p}}\Big]^{1/p} \lesssim \mathbb{E}\Big[\|u\|^p_{C_T^{\gamma'}B^{\alpha'}_{p,p}}\Big]^{1/p} \lesssim K. \tag{4.3}$$

**Proof.** The first inequality in (4.3) is simply the Besov embedding theorem. To see the second inequality, note that

$$\begin{aligned}\mathbb{E}\Big[\|u(t)-u(s)\|^p_{B^{\alpha'}_{p,p}}\Big] &= \sum_{j\geqslant -1} 2^{j\alpha' p}\int_{\mathbb{T}^d} \mathbb{E}[|\Delta_j u(t,x) - \Delta_j u(s,x)|^p]dx\\ &\leqslant \sum_{j\geqslant -1} 2^{j\alpha' p}\int_{\mathbb{T}^d} K^p|t-s|^{\gamma p}2^{-j\alpha p}dx \lesssim K^p|t-s|^{\gamma p},\end{aligned}$$

where we used that $\alpha' < \alpha$ and that $\mathbb{T}^d$ has finite volume. Similarly $\mathbb{E}[\|u(0)\|_{B^{\alpha'}_{p,p}}^p] \lesssim K^p$, and since $\gamma - 1/p > 0$ we can apply Kolmogorov's continuity criterion (for Banach space valued processes) and obtain for $\gamma' < \gamma - 1/p$

$$\mathbb{E}\Big[\|u\|_{C_T^{\gamma'} B^{\alpha'}_{p,p}}^p\Big]^{1/p} \lesssim K.$$

□

**Remark 4.2.** The proof crucially used that $\int_{\mathbb{T}^d} 1 dx < \infty$. On $\mathbb{R}^d$ a uniform bound as in (4.2) would only show that $u$ has trajectories in a weighted Besov space, and this would lead to the technical difficulties mentioned above.

As an application, we get the regularity of $Z$:

**Proposition 4.3.** *We have for all $\gamma < 1/2$ and all $\alpha < 1 - d/2$ and all $p \geqslant 1$:*

$$\mathbb{E}\big[\|Z\|_{C_T^0 \mathscr{C}^\alpha}^p + \|Z\|_{C_T^\gamma \mathscr{C}^{\alpha-1}}^p\big] < \infty.$$

**Proof.** By Lemma 4.1 it suffices to show for all $\lambda \in [0,1]$, for all $p \in [1,\infty)$, and for all $x \in \mathbb{T}^d$ and $0 \leqslant s < t \leqslant T$

$$\mathbb{E}[|\Delta_j Z(t,x) - \Delta_j Z(s,x)|^p]^{1/p} \lesssim 2^{j(d/2-1+\lambda)} |t-s|^{\lambda/2}. \tag{4.4}$$

Indeed, if we apply this inequality with $\lambda = 1$ we get $\mathbb{E}\big[\|Z\|_{C_T^\gamma \mathscr{C}^{\alpha-1}}^p\big] < \infty$, and if we apply it with $\lambda \simeq 0$ (but positive) and $p$ large enough so that $p\lambda > 1$ we get $\mathbb{E}[\|Z\|_{C_T^\varepsilon \mathscr{C}^\alpha}^p] < \infty$ for some $\varepsilon > 0$ and thus $\mathbb{E}\big[\|Z\|_{C_T^0 \mathscr{C}^\alpha}^p\big] < \infty$.

So let us derive (4.4). Since $\Delta_j Z(t) - \Delta_j Z(s)$ is a Gaussian random variable (it is a linear functional of the Gaussian process $\xi$), we have with $c_p = \mathbb{E}[|X|^p]$ for a $\mathcal{N}(0,1)$ variable $X$, and using the orthogonality of the integrals $\int_0^s$ and $\int_s^t$,

$$\begin{aligned}
&\mathbb{E}[|\Delta_j Z(t,x) - \Delta_j Z(s,x)|^p]^{2/p} = c_p^{2/p} \mathbb{E}[|\Delta_j Z(t,x) - \Delta_j Z(s,x)|^2]\\
&\simeq \mathbb{E}\left[\left|\int_s^t \int_{\mathbb{T}^d} (K_j * p(t-r))(x-z) \xi(r,z) \mathrm{d}z \mathrm{d}r\right|^2\right]\\
&\quad + \mathbb{E}\left[\left|\int_0^s \int_{\mathbb{T}^d} (K_j * (p(t-r) - p(s-r)))(x-z) \xi(r,z) \mathrm{d}z \mathrm{d}r\right|^2\right]\\
&= \int_s^t \int_{\mathbb{T}^d} |(K_j * p(t-r))(x-z)|^2 \mathrm{d}z \mathrm{d}r + \int_0^s \int_{\mathbb{T}^d} |(K_j * (p(t-r) - p(s-r)))(x-z)|^2 \mathrm{d}z \mathrm{d}r\\
&= \int_s^t \sum_{k \in \mathbb{Z}^d} |\rho_j(k)|^2 e^{-2|2\pi k|^2 (t-r)} \mathrm{d}r + \int_0^s \sum_{k \in \mathbb{Z}^d} |\rho_j(k)|^2 e^{-2|2\pi k|^2 (s-r)} (e^{-|2\pi k|^2 (t-s)} - 1)^2 \mathrm{d}r\\
&\lesssim \sum_{k \in \mathbb{Z}^d} |\rho_j(k)|^2 [\min\{|t-s|, |k|^{-2}\} + \min\{|k|^{-2}, |k|^{-2} |k|^2 |t-s|\}]\\
&\lesssim 2^{jd} \min\{|t-s|, 2^{-2j}\},
\end{aligned}$$

because $\rho_j(k) \neq 0$ for $O(2^{jd})$ values of $k$. Now (4.4) follows by interpolation. □

These computations are essentially sharp and a slight modification of the proof shows that $\mathbb{E}[\|Z(t)\|_{B^{1-d/2}_{p,p}}^p] = \infty$ for all $p \in [1,\infty)$ and all $t > 0$. Therefore, $Z$ is function valued if and only if $d = 1$, and the regularity gets worse with increasing dimension. This is a first indication why also the solution theory gets more and more complicated with increasing dimension.

**Remark 4.4.** Let us try to get an intuitive understanding why this is the regularity of $Z$: By Exercise 9.1 the white noise in dimension $d$ has regularity $-\frac{d}{2}-\varepsilon$. The space-time white noise has a $d+1$ dimensional index, so we might be tempted to guess that it has regularity $-\frac{d+1}{2}-\varepsilon$. But time and space are not equal for our problem, recall the parabolic distance from Remark 1.32! In fact, one time derivative counts two space derivatives. So from this "parabolic regularity" point of view, we would guess that $\xi$ has regularity $-\frac{d+2}{2}-\varepsilon=-\frac{d}{2}-1-\varepsilon$. This can be made rigorous, but we will not need it. In any case, then convolving $\xi$ against the heat semigroup should gain two derivatives, so we would guess that $Z$ has regularity $1-\frac{d}{2}-\varepsilon$. Which is exactly what we proved.

Question: Where did we use the Gaussianity of $Z$?

## 4.2 Schauder estimates

Here we study the regularizing effect of the heat semigroup $(P_t)_{t\geqslant 0}$ generated by $\Delta$. Recall that $P_t u = p(t,\cdot) * u$ for the heat kernel

$$p(t,x)=\sum_{k\in\mathbb{Z}^d} e^{2\pi i k\cdot x} e^{-|2\pi k|^2 t}=\sum_{k\in\mathbb{Z}^d}(4\pi t)^{-d/2}\exp\left(-\frac{|x+k|^2}{4t}\right),$$

We start with the following fundamental bound:

**Lemma 4.5.** *We have for all $\beta\geqslant 0$ and $\alpha\in\mathbb{R}$:*

$$\|P_t u\|_{\alpha+\beta}\lesssim(1\vee t^{-\frac{\beta}{2}})\|u\|_\alpha \tag{4.5}$$

*and for $\gamma\in(0,2]$*

$$\|P_t u-u\|_{\alpha-\gamma}\lesssim(1\wedge t^{\frac{\gamma}{2}})\|u\|_\alpha, \qquad \|P_t u-u\|_{L^\infty}\lesssim(1\wedge t^{\frac{\gamma}{2}})\|u\|_\gamma. \tag{4.6}$$

**Proof.** Estimate (4.5) is shown in Exercise 9.2. To see (4.6) we may assume that $t\leqslant 1$ because the estimate for $t>1$ is trivial. We use that $\partial_t P_t u=\Delta u$, and therefore

$$\begin{aligned}\|P_t u-u\|_{\alpha-\gamma} &= \left\|\int_0^t \partial_s P_s u\, ds\right\|_{\alpha-\gamma} \leqslant \int_0^t \|\Delta P_s u\|_{\alpha-\gamma} ds\\ &\overset{(4.5)}{\lesssim} \int_0^t s^{-\frac{(\alpha-\gamma)-(\alpha-2)}{2}}\|u\|_\alpha ds=\int_0^t s^{\frac{\gamma}{2}-1}\|u\|_\alpha ds\\ &\lesssim t^{\frac{\gamma}{2}}\|u\|_\alpha.\end{aligned}$$

The second estimate in (4.6) is roughly speaking the first estimate for $\alpha=\gamma$. Except that this would only give a bound in $\mathscr{C}^0$, which is worse than the claimed $L^\infty$ bound. To get the $L^\infty$ bound we thus distinguish the small and large blocks: We have the two estimates

$$\|\Delta_j(P_t u-u)\|_{L^\infty}\lesssim 2^{j(2-\gamma)}\|P_t u-u\|_{\gamma-2}\lesssim 2^{j(2-\gamma)}t\|u\|_\gamma$$

and

$$\|\Delta_j(P_t u-u)\|_{L^\infty}\leqslant\|P_t\Delta_j u\|_{L^\infty}+\|\Delta_j u\|_{L^\infty}\lesssim\|\Delta_j u\|_{L^\infty}\lesssim 2^{-j\gamma}\|u\|_\gamma.$$

Let $2^{-j_0}\simeq t^{1/2}$. We apply the first estimate for $j\leqslant j_0$ and the second one for $j>j_0$. Then

$$\|P_t u-u\|_{L^\infty}\lesssim\sum_{j\leqslant j_0}2^{j(2-\gamma)}t\|u\|_\gamma+\sum_{j>j_0}2^{-j\gamma}\|u\|_\gamma\lesssim t^{\frac{\gamma}{2}-1}t\|u\|_\gamma+t^{\frac{\gamma}{2}}\|u\|_\gamma\simeq t^{\frac{\gamma}{2}}\|u\|_\gamma.$$

□

**Theorem 4.6. (Schauder estimates)** *Let $\alpha \in \mathbb{R}$ and let $(P_t)_{t \geqslant 0}$ be the semigroup generated by $\Delta$. Given $f \in C_T \mathscr{C}^\alpha$ and $\varphi \in \mathscr{C}^{\alpha+2}$, let $u$ be the mild solution to*

$$(\partial_t - \Delta) u = f, \qquad u(0) = \varphi.$$

*Then we have for all $\delta \in [0, 2]$*

$$\|u\|_{C_T \mathscr{C}^{\alpha+\delta}} \lesssim (T \vee T^{1-\frac{\delta}{2}}) \|f\|_{C_T \mathscr{C}^\alpha} + \|\varphi\|_{\alpha+2}.$$

**Proof.** By definition $u(t) = P_t \varphi + \int_0^t P_{t-s} f(s) ds$. We apply (4.5) with $\beta \to 0$, $\alpha \to \alpha + 2$, and obtain

$$\|(t \mapsto P_t \varphi)\|_{C_T \mathscr{C}^{\alpha+2}} \lesssim \|\varphi\|_{\alpha+2}.$$

The continuity of $t \mapsto P_t \varphi$ in $t = 0$ is a bit subtle and here we need the new definition of $\mathscr{C}^\alpha$, more precisely that $2^{j\alpha} \|\Delta_j \varphi\|_{L^\infty} \to 0$ for $j \to \infty$. We leave that problem as an exercise.

For the space-time convolution we have for $\delta \in [0, 2)$

$$\left\| \int_0^t P_{t-s} f(s) ds \right\|_{\alpha+\delta} \leqslant \int_0^t \|P_{t-s} f(s)\|_{\alpha+\delta} ds \lesssim \int_0^t (1 \vee |t-s|^{-\frac{\delta}{2}}) \|f\|_{C_T \mathscr{C}^{\alpha-2}} ds$$
$$\lesssim (t \vee t^{1-\frac{\delta}{2}}) \|f\|_{C_T \mathscr{C}^{\alpha-2}}.$$

For $\delta = 2$ there is a problem because $|t-s|^{-1}$ barely fails to be integrable. In that case we have to be slightly more careful and use two different estimates, one for $s$ close to 0, and one for $s$ close to $t$, see Lemma A.9 in [22] for details. □

## 4.3 The $\Phi_1^4$ equation

We now assume that $d = 1$ and we fix $\alpha \in (0, 1/2)$ and consider a general initial condition $\phi_0 \in \mathscr{C}^\alpha$ (not necessarily $\phi_0 = 0$). We also replace the drift $-\phi^3 + \phi$ by $-\phi^3$ to simplify the notation, because the linear term does not introduce any additional difficulties. Then the mild formulation of our equation is

$$\phi(t) = P_t \phi_0 + \int_0^t P_{t-s}(-\phi(s)^3) \mathrm{d}s + Z(t) = \int_0^t P_{t-s}(-\phi(s)^3) \mathrm{d}s + \tilde{Z}(t)$$

for $\tilde{Z}(t) = Z(t) + P_t \phi_0$. Theorem 4.6 applied to $P_t \phi_0$ shows that $\tilde{Z} \in C_T \mathscr{C}^\alpha$ for all $T > 0$.

**Theorem 4.7.** *Let $d = 1$ and $\alpha \in (0, 1/2)$ and $\phi_0 \in \mathscr{C}^\alpha$. There exists a random maximal existence time $T^* \in (0, \infty]$ and a unique $\phi \in C_{T^*} \mathscr{C}^\alpha := \bigcup_{T < T^*} C_T \mathscr{C}^\alpha$ such that*

$$\phi(t) = P_t \phi_0 + \int_0^t P_{t-s}(-\phi(s)^3) ds + Z(t), \qquad t \in [0, T^*).$$

*If $T^* < \infty$, then $\lim_{t \uparrow T^*} \|\phi(t)\|_\alpha = \infty$.*

**Proof.** We set up a Picard iteration by defining

$$\Psi : C_T \mathscr{C}^\alpha \to C_T \mathscr{C}^\alpha, \qquad \Psi(\phi)(t) = \tilde{Z}(t) + \int_0^t P_{t-s}(-\phi(s)^3) \mathrm{d}s$$

We have by the Schauder estimates with $\delta = 0$ and by the paraproduct estimates (Corollary 3.20, here we need $\alpha > 0$) for $t \in [0, T]$ and some $K > 0$ whose value may change in each line:

$$\|\Psi(\phi)\|_{C_T \mathscr{C}^\alpha} \overset{\text{Schauder}}{\leqslant} \|\tilde{Z}(t)\|_{C_T \mathscr{C}^\alpha} + KT \|\phi^3\|_{C_T \mathscr{C}^\alpha}$$
$$\overset{\text{Cor. 3.20}}{\leqslant} \|\tilde{Z}(t)\|_{C_T \mathscr{C}^\alpha} + KT \|\phi^3\|_{C_T \mathscr{C}^\alpha},$$

which shows that $\Psi$ maps to $C_T\mathscr{C}^\alpha$. Set

$$M = \sup_{t\leqslant 1} \|\tilde{Z}(t)\|_\alpha.$$

If $T\in(0,1]$ is small enough, depending on $M$, then $\Psi$ leaves the ball $B_{C_T\mathscr{C}^\alpha}(0,2M)$ in $C_T\mathscr{C}^\alpha$ with center 0 and radius $2M$ invariant. Moreover, since

$$\|\phi^3-\psi^3\|_\alpha \lesssim (\|\phi\|_\alpha^2+\|\psi\|_\alpha^2)\|\phi-\psi\|_\alpha,$$

we get that for a possibly even smaller $T>0$ the map $\Psi$ is a contraction on $B_{C_T\mathscr{C}^\alpha}(0,2M)$. By the Banach fixed point theorem we thus find a unique solution to (4.3) on the interval $[0,T]$, where $T$ depends on $\tilde{Z}$ through $M$ and thus may be random. We can iterate the construction, but since the initial condition is part of $\tilde{Z}$, the time interval in the second iteration step might be strictly smaller. Ultimately we get the existence of a $T^*\in(0,\infty]$ and a unique solution $\phi\in C_T\mathscr{C}^\alpha$ for all $T<T^*$, such that in the case $T^*<\infty$ we have $\lim_{t\uparrow T^*}\|\phi(t)\|_\alpha=\infty$. In other words, $[0,T^*)$ is the maximal existence interval, and the solution blows up at $T^*$, or it exists for all times and $T^*=\infty$. □

Question: What goes wrong if $d\geqslant 2$?

**Remark 4.8.**

i. By slightly refining the analysis we could show that the solution $\phi$ depends continuously on $(\phi_0,Z)\in\mathscr{C}^\alpha\times C_T\mathscr{C}^\alpha$.

ii. Actually we would not expect the solution to blow up, and indeed it does not. But to see this we would have to use the sign of the nonlinearity: The above analysis works also for the equation with $+\phi^3$ instead of $-\phi^3$, and in that case we expect it to blow up in finite time.

**Remark 4.9.**

i. Our arguments break down in $d>1$, because then we no longer have $\alpha>0$ and therefore we cannot estimate $\|\phi^3\|_{C_T\mathscr{C}^\alpha}\lesssim\|\phi\|^3_{C_T\mathscr{C}^\alpha}$ (and in fact we do not have an estimate for $\phi^3$ at all).

ii. When we applied the Schauder estimates in the proof, we did not use the regularizing effect of $(P_t)$ at all and we chose $\delta=0$ in Theorem 4.6. If we would instead take $\delta=2$, then we would obtain that

$$\phi-\tilde{Z}\in C_T\mathscr{C}^{\alpha+2}.$$

This observation is the starting point for our solution of the $\Phi^4_2$ equation.

Question: Which equation should $v=\phi-\tilde{Z}$ solve?

## 4.4 The $\Phi^4_2$ equation

In $d=2$ we have $Z\in C_T\mathscr{C}^{0-}:=\bigcap_{\varepsilon>0}C_T\mathscr{C}^{-\varepsilon}$, and therefore even $Z(s)^3$ is ill-defined, let alone $\phi(s)^3$. We ignore this problem for now and decompose $\phi=Z+v$, a strategy which goes back Da Prato and Debussche [16]. Then $v$ should solve

$$(\partial_t-\Delta)v=-\phi^3=-(Z^3+3Z^2v+3Zv^2+v^3), \qquad v(0)=\phi_0-Z(0)=\phi_0,$$

where we used that $Z(0)=0$. If we ignore that the products $Z^3$ and $Z^2$ are ill defined and if we simply apply the paraproduct estimates to them anyways, then we get $Z^2,Z^3\in C_T\mathscr{C}^{0-}$.

Question: Convince yourself that if we ignore the constraint $\alpha+\beta>0$ for $f\odot g$, then the product $fg=f\varolessthan g+f\varogreaterthan g+f\odot g$ of $f\in\mathscr{C}^\alpha$ and $g\in\mathscr{C}^\beta$ should have regularity $fg\in\mathscr{C}^{\alpha\wedge\beta\wedge(\alpha+\beta)}$.

So let us continue our computations under the assumption that $Z^2, Z^3 \in C_T\mathscr{C}^{0-}$ are given – this may look like a bold assumption, but $Z$ is a a Gaussian process, and in probability theory we deal with ill-defined nonlinear operations of Gaussian processes all the time: Recall the integral $\int B \otimes dB$ for a Brownian motion. The reason why we can do more in the stochastic setting is that the analytic paraproduct theory gives only worst case estimates, while in the stochastic case there are cancellations coming from independence properties of the noise, so we are not in a worst case situation.

If $Z^2, Z^3 \in C_T\mathscr{C}^{0-}$ are given, the right hand side of the equation for $v$ is well defined as long as we can estimate $v \in C_T\mathscr{C}^{0+} = \bigcup_{\varepsilon>0} C_T\mathscr{C}^{\varepsilon}$, and in that case this right hand side is in $C_T\mathscr{C}^{0-}$. But this can easily be achieved with our Schauder estimates.

**Theorem 4.10.** *Let $d=2$ and $\alpha \in (1,2)$ and $\phi_0 \in \mathscr{C}^{\alpha}$. Assume that $Z, Z^2, Z^3 \in C(\mathbb{R}_+, \mathscr{C}^{\alpha-2})$ are given. Then there exists a time $T^* \in (0,\infty]$ depending on $Z, Z^2, Z^3$ and $\phi_0$ and a unique $v \in C_{T^*}\mathscr{C}^{\alpha} := \bigcup_{T<T^*} C_T\mathscr{C}^{\alpha}$ such that*

$$v(t) = P_t\phi_0 + \int_0^t P_{t-s}(-(Z^3 + 3Z^2 v + 3Zv^2 + v^3)(s))\mathrm{d}s, \qquad t \in [0, T^*).$$

*Moreover, $v$ depends continuously on $(\phi_0, Z, Z^2, Z^3)$.*

**Proof.** Exercise 10.1. □

This works as long as $Z^2, Z^3$ are given. To attempt constructing theses products we use a Fourier truncation and set

$$Z_{\varepsilon}(t) := \mathscr{F}_{\mathbb{T}^d}^{-1}(\varphi_{\varepsilon}(\cdot)\mathscr{F}_{\mathbb{T}^d} Z(t,\cdot)),$$

where $\varphi_{\varepsilon}(k) = \varphi(\varepsilon k)$ for a compactly supported bounded and even function $\varphi$ which is continuous around 0 and which satisfies $\varphi(0) = 1$ (think of $\varphi \in C_c^{\infty}$ with $\varphi(0) = 1$ or $\varphi = \mathbb{1}_{[-1,1]}$). If we can show that $Z_{\varepsilon}^k$ converges in $C_T\mathscr{C}^{0-}$ to a limit as $\varepsilon \to 0$ and that the limit does not on the specific truncation function $\varphi$, then we can just define $Z^k$ as that limit.

Unfortunately, already $Z_{\varepsilon}^2$ diverges:

**Lemma 4.11.** *We have*

$$\mathbb{E}[\Delta_j(Z_{\varepsilon}^2)(t,x)] = \delta_{j,-1} C_{\varepsilon}(t),$$

*where $\delta$ is the Kronecker delta and for $t > 0$*

$$C_{\varepsilon}(t) = t + \sum_{k \neq 0} \frac{|\varphi(\varepsilon k)|^2}{2|2\pi k|^2}(1 - e^{-2|2\pi k|^2 t}) = \begin{cases} O(1), & d=1, \\ O(|\log(\varepsilon)|), & d=2, \\ O(\varepsilon^{2-d}), & d \geqslant 3. \end{cases}$$

**Proof.** We have with $\mathscr{F}_{\mathbb{T}^d} p^{\varepsilon}(t,k) = \varphi_{\varepsilon}(k)\mathscr{F}_{\mathbb{T}^d} p(t,k) = \varphi_{\varepsilon}(k) e^{-|2\pi k|^2 t}$:

$$\begin{aligned}
\mathbb{E}[\Delta_j(Z_{\varepsilon}^2)(t,x)] &= \int_{\mathbb{T}^d} dy K_j(x-y)\mathbb{E}[(Z_{\varepsilon}^2)(t,y)] \\
&= \int_{\mathbb{T}^d} dy K_j(x-y) \int_0^t ds \int_{\mathbb{T}^d} dz p^{\varepsilon}(t-s, y-z)^2 \\
&= \int_{\mathbb{T}^d} dy K_j(x-y) \int_0^t ds \sum_k |\varphi_{\varepsilon}(k) e^{-|2\pi k|^2 t}|^2 \\
&= \left(\int_{\mathbb{T}^d} dy K_j(y)\right) \times \left(t + \sum_{k \neq 0} \frac{|\varphi(\varepsilon k)|^2}{2|2\pi k|^2}(1 - e^{-2|2\pi k|^2 t})\right),
\end{aligned}$$

from where the claim follows. □

One can show that if a sequence of fixed degree polynomials of Gaussians converges in probability, then they also converge in $L^p$ for any $p>0$. So for $d>1$, $(Z_\varepsilon^2)_\varepsilon$ does not converge since the expectations $(\mathbb{E}[\Delta_{-1}Z_\varepsilon^2])_\varepsilon$ diverge. We will cure this problem by *renormalizing* $Z_\varepsilon^2$. For that purpose we need a bit of Gaussian analysis

## 4.5 Elements of Gaussian analysis

Unlike for $k=1$ and/or $d=1$, the construction of $Z^k$ for $k>1$ and $d>1$ is nontrivial. Here we introduce the main tools needed for the construction (renormalization, hypercontractivity), and sketch the estimates. The presentation is inspired by [37, 40].

**Definition 4.12.** *For $x\in\mathbb{R}$ and $t\geqslant 0$ we define the* Hermite polynomials *recursively via*

$$H_0(x,t)=1, \qquad H_n(x,t)=x H_{n-1}(x,t)-t\partial_x H_{n-1}(x,t).$$

The first few Hermite polynomials are

$$H_0(x,t)=1, \qquad H_1(x,t)=x, \qquad H_2(x,t)=x^2-t,$$
$$H_3(x,t)=x^3-3tx, \qquad H_4(x,t)=x^4-6tx^2+3t^2.$$

Recall that $Z(t)=\int_0^t P_{t-s}\xi(s)ds$, so

$$Z_\varepsilon(t)=\int_0^t P_{t-s}^\varepsilon\xi(s)ds, \qquad P_t^\varepsilon f=\mathscr{F}_{\mathbb{T}^d}^{-1}(\varphi_\varepsilon e^{-|2\pi\cdot|^2t}\mathscr{F}_{\mathbb{T}^d}f).$$

For $n=2$ we get with $C_\varepsilon(t)=\mathbb{E}[Z_\varepsilon(t,x)^2]$:

$$H_2(Z_\varepsilon(t,x),C_\varepsilon(t))=Z_\varepsilon(t,x)^2-C_\varepsilon(t)=Z_\varepsilon(t,x)^2-\mathbb{E}[Z_\varepsilon(t,x)^2],$$

and this suggests that $H_2(Z_\varepsilon(t,x),C_\varepsilon(t))$ might be better behaved than $Z_\varepsilon(t,x)^2$. It turns out that also the higher order Hermite polynomials $H_n(Z_\varepsilon(t,x),C_\varepsilon(t))$ are better behaved than $Z_\varepsilon(t,x)^n$. Intuitively, this can be explained by the fact that the $H_n(Z_\varepsilon(t,x),C_\varepsilon(t))$ are orthogonal projections:

**Exercise 4.1.** Show that for $t>0$

$$H_n(x,t)=(-t)^n e^{x^2/(2t)}\partial_x^n e^{-x^2/(2t)}. \tag{4.7}$$

Conclude that the family $(H_n(\cdot,t))_{n\in\mathbb{N}_0}$ is orthogonal with respect to the centered Gaussian measure with variance $t$. Show also that

$$\partial_x H_n=nH_{n-1}, \qquad \partial_t H_n=-\frac{n(n-1)}{2}H_{n-2}=-\frac{1}{2}\partial_x^2 H_n, \tag{4.8}$$

i.e. that each Hermite polynomial solves the backward heat equation $\left(\partial_t+\frac{1}{2}\partial_x^2\right)H_n=0$.

Question: Compute $\mathbb{E}[H_n(X,1)^2]$ for $X\sim\mathcal{N}(0,1)$.

**Remark 4.13.** One can also show that if $(X,Y)$ are jointly centered Gaussian, then

$$\mathbb{E}[H_m(X,\mathbb{E}[X^2])H_n(Y,\mathbb{E}[Y^2])]=\delta_{m,n}n!\mathbb{E}[XY]^n,$$

see Lemma 1.1.1 in [42].

With the help of Hermite polynomials we can do efficient computations. However, to control

$$\Delta_j(H_n(Z_\varepsilon(t,\cdot),C_\varepsilon(t)))(x)=\int K_j(x-y)H_n(Z_\varepsilon(t,y),C_\varepsilon(t))\mathrm{d}y$$

we also need to understand sums of Hermite polynomials. This leads to the so called *Wiener-Ito chaos*.

**Definition 4.14.** *Let $E := \mathbb{R}_+ \times \mathbb{T}^d$. For $n \in \mathbb{N}$ we write*

$$L_s^2(E^n) := L^2(E^n)/\sim,$$

*where the equivalence relation $f \sim g$ holds if for some permutation $\sigma$ of $\{1, \ldots, n\}$:*

$$f(z_1, \ldots, z_n) = g(z_{\sigma(1)}, \ldots, z_{\sigma(n)})$$

*as elements of $L^2(E^n)$. We interpret $L_s^2(E^n)$ as the* symmetric *functions in $L^2(E^n)$, which are such that $f(z_1, \ldots, z_n) = f(z_{\sigma(1)}, \ldots, z_{\sigma(n)})$ for any permutation $\sigma$. Indeed, every equivalence class has exactly one symmetric representative, and for a general $f \in L^2(E^n)$ we can always write its symmetric representative as*

$$\tilde{f}(z_1, \ldots, z_n) := \frac{1}{n!} \sum_{\sigma \in \Sigma_n} f(z_{\sigma(1)}, \ldots, z_{\sigma(n)}),$$

*where $\Sigma_n$ is the symmetric group on $\{1, \ldots, n\}$. For $f \in L_s^2(E^n)$ we write*

$$\|f\|_{L_s^2(E^n)} := \|\tilde{f}\|_{L^2(E^n)}.$$

Note that for $f \in L^2(E^n)$ with symmetric representative $\tilde{f}$ we have by the triangle inequality for the $L^2(E^n)$-norm:

$$\|\tilde{f}\|_{L^2(E^n)} \leqslant \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \|f(z_{\sigma(1)}, \ldots, z_{\sigma(n)})\|_{L^2(E^n)} = \frac{1}{n!} n! \|f\|_{L^2(E^n)} = \|f\|_{L^2(E^n)}.$$

**Definition 4.15.** *For $f \in L^2(E^n)$ with symmetric representative $\tilde{f}$ we define the $n$-th* Wiener Itô integral *as*

$$W_n(f) := W_n(\tilde{f}) := n! \int_{[0,\infty) \times \mathbb{T}^d} \int_{[0,t_n] \times \mathbb{T}^d} \cdots \int_{[0,t_2] \times \mathbb{T}^d} \tilde{f}(z_1, \ldots, z_n) \xi(dt_1, dx_1) \cdots \xi(dt_n, dz_n).$$

**Lemma 4.16.** *The Wiener-Itô integral is a (multiple of a) linear isometry:*

$$\|W_n(f)\|_{L^2(\Omega)}^2 = \mathbb{E}[W_n(f)^2] = n! \|f\|_{L_s^2(E^n)}^2 := n! \|\tilde{f}\|_{L^2(E^n)}^2.$$

**Proof.** This follows by repeated application of Itô's isometry. ☐

**Definition 4.17.** *We write $\mathcal{H}_n \subset L^2(\Omega)$ for the image of $W_n$, and we call $\mathcal{H}_n$ the $n$-th* Wiener-Itô chaos*. One can show that for $\sigma(\xi) := \sigma(\xi(\varphi) : \varphi \in L^2(E))$ we have the* chaos decomposition

$$L^2(\Omega, \sigma(\xi)) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n.$$

Our next goal is to relate Wiener-Itô integrals and Hermite polynomials. For that purpose we need the following simple observation, which shows that Hermite polynomials are intimately connected to martingales.

**Lemma 4.18.** *Let $M$ be a continuous local martingale with $M_0 = 0$. Then*

$$H_n(M_t, \langle M \rangle_t) = n \int_0^t H_{n-1}(M_s, \langle M \rangle_s) dM_s.$$

**Proof.** We apply Ito's formula to $H_n(M_t, \langle M\rangle_t)$: Since $H_n(0,0)=0$ and $\left(\partial_t+\frac{1}{2}\partial_x^2\right)H_n\equiv 0$ by (4.8), we get

$$\begin{aligned}H_n(M_t,\langle M\rangle_t)=&\int_0^t \partial_x H_n(M_s,\langle M\rangle_s)dM_s+\int_0^t\left(\partial_t+\frac{1}{2}\partial_x^2\right)H_n(M_s,\langle M\rangle_s)d\langle M\rangle_s\\ =&n\int_0^t H_{n-1}(M_s,\langle M\rangle_s)dM_s,\end{aligned}$$

where the last step follows from the first identity in (4.8). □

- ○ Question: Define $\mathcal{E}(M)_t:=\sum_{n=0}^\infty \frac{1}{n!}H_n(M_t,\langle M\rangle_t)$. (Formally) show that $\mathcal{E}(M)_t=e^{M_t-\frac{1}{2}\langle M\rangle_t}$.

To apply this result, we need to find a suitable continuous martingale. But for $\varphi\in L^2(E)$ the process $M_t^\varphi=\xi(\varphi\mathbb{1}_{[0,t]})$ is an integral against a Gaussian martingale measure and therefore a continuous (Gaussian) martingale in the filtration

$$\mathcal{F}_t=\sigma\{\xi(\varphi\mathbb{1}_{[0,s]}):\varphi\in L^2(E),s\leqslant t\},$$

with deterministic quadratic variation

$$\langle M^\varphi\rangle_t=\int_0^t\int_{\mathbb{T}^d}\varphi(r,x)^2dxdr,\qquad \langle M^\varphi,M^\psi\rangle_t=\int_0^t\int_{\mathbb{T}^d}\varphi(r,x)\psi(r,x)dxdr.$$

**Corollary 4.19.** *For $\varphi\in L^2(E)$ we have with $\varphi^{\otimes n}(z_1,\ldots,z_n):=\varphi(z_1)\cdots\varphi(z_n)$:*

$$W_n(\varphi^{\otimes n})=H_n(\xi(\varphi),\|\varphi\|_{L^2(E)}^2).$$

**Proof.** Consider the continuous martingale $M_t^\varphi=\xi(\mathbb{1}_{[0,t]}\varphi)$. Then

$$\begin{aligned}H_n(\xi(\varphi),\|\varphi\|_{L^2(E)}^2)=&H_n(M_\infty^\varphi,\langle M^\varphi\rangle_\infty)\\ \overset{\text{Lem. 4.18}}{=}&n!\int_0^\infty\int_0^{t_n}\cdots\int_0^{t_2}dM_{t_1}^\varphi dM_{t_2}^\varphi\cdots dM_{t_n}^\varphi.\end{aligned}$$

We have for all adapted and progressively measurable, square-integrable processes $F$:

$$\int_0^\infty F(t)dM_t^\varphi=\int_{[0,\infty)\times\mathbb{T}^d}\varphi(t,x)F(t)\xi(dt,dx).$$

Indeed, for step functions $F$ this is obvious and for general $F$ it follows by an approximation argument. Therefore,

$$n!\int_0^\infty\int_0^{t_n}\cdots\int_0^{t_2}dM_{t_1}^\varphi dM_{t_2}^\varphi\cdots dM_{t_n}^\varphi=W_n(\varphi^{\otimes n}),$$

and the proof is complete. □

Consequently, for all $\varphi\in L^2(E)$ the process

$$W_n(\varphi^{\otimes n}\mathbb{1}_{[0,t]}^{\otimes n})=n\int_0^t W_{n-1}\big(\varphi^{\otimes(n-1)}\mathbb{1}_{[0,s]}^{\otimes(n-1)}\big)dM_s^\varphi,\qquad t\geqslant 0,$$

is a continuous martingale, and the quadratic covariation of two such martingales is

$$\begin{aligned}&\langle W_n(\varphi^{\otimes n}\mathbb{1}_{[0,\cdot]}^{\otimes n}),W_n(\psi^{\otimes n}\mathbb{1}_{[0,\cdot]}^{\otimes n})\rangle_t\\ =&n^2\int_E W_{n-1}\big(\varphi^{\otimes n}(z,\cdot)\mathbb{1}_{[0,s]}^{\otimes(n-1)}\big)W_{n-1}\big(\psi^{\otimes n}(z,\cdot)\mathbb{1}_{[0,s]}^{\otimes(n-1)}\big)\mathbb{1}_{[0,t]}(s)dz,\end{aligned}$$

where we write $z=(s,x)\in E$. Linear combinations of such $\varphi^{\otimes n}$ are dense in $L^2_s(E^n)$, so by approximation we deduce that for all $\varphi\in L^2_s(E^n)$ the process $W_n(\varphi\mathbb{1}^{\otimes n}_{[0,t]})$, $t\geqslant 0$, is a martingale with quadratic variation

$$\langle W_n(\varphi\mathbb{1}^{\otimes n}_{[0,\cdot]})\rangle_t=n^2\int_E W_{n-1}\big(\varphi(z,\cdot)\mathbb{1}^{\otimes(n-1)}_{[0,s]}\big)^2\mathbb{1}_{[0,t]}(s)dz.$$

**Remark 4.20.** To see that $L^2(\Omega,\sigma(\xi))$ has the chaos decomposition property, it suffices to apply Corollary 4.19 and to note that the monomial $x^n$ can be written as a linear combination of $H_k(x,t)$ with $k\leqslant n$. Therefore, any random variable which is orthogonal to $\bigoplus_{n=0}^\infty\mathcal{H}_n$ is orthogonal to all polynomials $\xi(\varphi)^n$ for all $\varphi\in L^2(E)$ and all $n\in\mathbb{N}_0$. See Theorem 1.1.1 of [42] for details.

To apply the Besov-Kolmogorov type result from Proposition 4.1 we need to control high moments. This can be done with the help of the following theorem:

**Theorem 4.21. (Gaussian hypercontractivity)** *For all $p\in(0,\infty)$ there exists a constant $C_p>0$ such that for all $n\in\mathbb{N}_0$ and all $\varphi\in L^2_s(E^n)$:*

$$\mathbb{E}[|W_n(\varphi)|^p]\leqslant C_p^n(n!)^{p/2}\|\varphi\|^p_{L^2(E^n)}=C_p^n\mathbb{E}[|W_n(\varphi)|^2]^{p/2}.$$

**Proof. (of Theorem 4.21)** For $p<2$ there is nothing to show, so let $p\geqslant 2$. By the Burkholder-Davis-Gundy inequality, together with the Minkowski inequality $\|\int_E(\ldots)\mathrm{d}z\|_{L^{p/2}(\Omega)}\leqslant\int_E\|\ldots\|_{L^{p/2}(\Omega)}dz$, we have

$$\begin{aligned}
\mathbb{E}[|W_n(\varphi)|^p]&\leqslant\mathbb{E}\Big[\sup_{t\geqslant 0}|W_n(\varphi\mathbb{1}^{\otimes n}_{[0,t]})|^p\Big]\leqslant C_p\mathbb{E}\Big[\Big(n^2\int_E W_{n-1}\big(\varphi(z,\cdot)\mathbb{1}^{\otimes(n-1)}_{[0,s]}\big)^2dz\Big)^{p/2}\Big]\\
&\leqslant C_p\Big(n^2\int_E\mathbb{E}\big[\big|W_{n-1}\big(\varphi(z_1,\cdot)\mathbb{1}^{\otimes(n-1)}_{[0,s_1]}\big)\big|^p\big]^{2/p}dz_1\Big)^{p/2}\\
&\leqslant C_p^2\Big(n^2(n-1)^2\int_E\int_E\mathbb{E}\big[\big|W_{n-2}\big(\varphi(z_1,z_2,\cdot)\mathbb{1}^{\otimes(n-2)}_{[0,s_2]}\big)\big|^p\big]^{2/p}\mathbb{1}_{s_2\leqslant s_1}dz_2dz_1\Big)^{p/2}\\
&\leqslant\cdots\leqslant C_p^n\Big((n!)^2\int_E\ldots\int_E|\varphi(z_1,\ldots,z_n)|^2\mathbb{1}_{s_n\leqslant\cdots\leqslant s_1}dz_n\ldots dz_1\Big)^{p/2}\\
&=C_p^n\Big(n!\int_E\ldots\int_E|\varphi(z_1,\ldots,z_n)|^2dz_n\ldots dz_1\Big)^{p/2},
\end{aligned}$$

where in the last step we used that $\varphi$ is symmetric in its $n$ arguments. In that case the right hand side equals $C_p^n(n!)^{p/2}\|\varphi\|^p_{L^2(E^n)}$, and this completes the proof. □

Can you do similar things if there is no designated time variable, e.g. for space white noise on $\mathbb{T}^d$?

**Remark 4.22.** It seems to have been important that there was a designated "time variable" in the computations above. But in fact we can use similar arguments whenever $\xi$ is a white noise on $L^2(E)$ and $E=I\times E'$ for some interval $I\subset\mathbb{R}$, for example if $E=\mathbb{T}^d$ we can interpret this as $E\simeq[0,1]\times\mathbb{T}^{d-1}$ and this gives us an "artificial" time variable to work with.

Similar martingale arguments also allow to prove hypercontractivity results in discrete settings, see for example Lemma 5.1 in [36].

**Theorem 4.23.** *Let $d = 2$ and define*

$$Z_\varepsilon^{:n:}(t,x) := H_n(Z_\varepsilon(t,x), C_\varepsilon(t)), \qquad t \geqslant 0, x \in \mathbb{T}^2.$$

*Let $\alpha < 2$, $\gamma < 1/2$, and $T > 0$. There exists $Z^{:n:} \in C_T \mathscr{C}^{\alpha-2} \cap C_T^\gamma \mathscr{C}^{\alpha-3}$ such that for all $p \in [1,\infty)$*

$$\lim_{\varepsilon \to 0} \mathbb{E}\big[ \|Z_\varepsilon^{:n:} - Z^{:n:}\|^p_{C_T^0 \mathscr{C}^{\alpha-2}} + \|Z_\varepsilon^{:n:} - Z^{:n:}\|^p_{C_T^\gamma \mathscr{C}^{\alpha-3}}\big] = 0. \tag{4.9}$$

**Proof.** (Voluntary, not contained in the videos)

1. We have

$$Z_\varepsilon(t,x) = \int_0^t \int_{\mathbb{T}^d} p^\varepsilon(t-s, x-y)\xi(s,y)\mathrm{d}y\mathrm{d}s = \xi(\mathbb{1}_{[0,t]} p^\varepsilon(t-\cdot, x-\cdot))$$

and in Lemma 4.11 we saw that $C_\varepsilon(t) = \|\mathbb{1}_{[0,t]} p^\varepsilon(t-\cdot, x-\cdot)\|^2_{L^2}$, so that

$$Z_\varepsilon^{:n:}(t,x) = W_n(\mathbb{1}_{[0,t]}^{\otimes n} p^\varepsilon(t-\cdot, x-\cdot)^{\otimes n}),$$

and by construction $W_n$ is a continuous linear map, which means that we can exchange it with integrals and we obtain

$$\begin{aligned}
\Delta_j Z_\varepsilon^{:n:}(t,x) &= \int_{\mathbb{T}^d} K_j(x-y) W_n(\mathbb{1}_{[0,t]}^{\otimes n} p^\varepsilon(t-\cdot, y-\cdot)^{\otimes n})\mathrm{d}y\\
&= W_n\bigg(\int_{\mathbb{T}^d} K_j(x-y)\mathbb{1}_{[0,t]}^{\otimes n} p^\varepsilon(t-\cdot, y-\cdot)^{\otimes n}\mathrm{d}y\bigg).
\end{aligned}$$

So by Theorem 4.21 we have

$$\begin{aligned}
\mathbb{E}[|\Delta_j Z_\varepsilon^{:n:}(t,x)|^p]^{2/p} &\lesssim \int_{E^n} \left| \int_{\mathbb{T}^d} K_j(x-y) \prod_{j=1}^n [\mathbb{1}_{s_j \leqslant t} p^\varepsilon(t-s_j, y-x_j)]\mathrm{d}y \right|^2 \mathrm{d}z_n \ldots \mathrm{d}z_1\\
&= \int_{E^n} \int_{\mathbb{T}^d} K_j(x-y_1) K_j(x-y_2)\\
&\quad \times \prod_{j=1}^n [\mathbb{1}_{s_j \leqslant t} p^\varepsilon(t-s_j, y_1-x_j) p^\varepsilon(t-s_j, y_2-x_j)]\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_n\ldots\mathrm{d}z_1\\
&= \int_{[0,t]^n} \int_{\mathbb{T}^d} K_j(x-y_1) K_j(x-y_2) \prod_{j=1}^n \tilde{p}^\varepsilon(2(t-s_j), y_2-y_1)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}s_n\ldots\mathrm{d}s_1\\
&= \int_{\mathbb{T}^d} K_j(x-y_1)K_j(x-y_2)\bigg(\int_0^t \tilde{p}^\varepsilon(2s, y_2-y_1)\mathrm{d}s\bigg)^n \mathrm{d}y_1\mathrm{d}y_2,
\end{aligned}$$

where $\mathscr{F}_{\mathbb{T}^d}\tilde{p}^\varepsilon(t,k) = \varphi_\varepsilon(k)^2 \mathscr{F}_{\mathbb{T}^d} p(t,k)$ and we used the semigroup property $p(r,\cdot) * p(s,\cdot) = p(r+s,\cdot)$. Now we apply Parseval's identity and obtain

$$\begin{aligned}
&\mathbb{E}[|\Delta_j Z_\varepsilon^{:n:}(t,x)|^p]^{2/p}\\
&\lesssim \sum_{k_1,\ldots,k_n} \rho_j(k_1+\cdots+k_n)^2 \prod_{j=1}^n \bigg(\int_0^t |\mathscr{F}_{\mathbb{T}^d}\tilde{p}^\varepsilon(2s,k_j)|^2 \mathrm{d}s\bigg)\\
&= n! \sum_{k_1,\ldots,k_n} \rho_j(k_1+\cdots+k_n)^2 \varphi_\varepsilon(k_1)^2 \ldots \varphi_\varepsilon(k_n)^2 \prod_{j=1}^n \bigg(\mathbb{1}_{k_j=0}\, t + \mathbb{1}_{k_j \neq 0} \frac{1-e^{-4|2\pi k_j|^2 t}}{4|2\pi k_j|^2}\bigg)\\
&\lesssim_{n,t} \sum_{k_1,\ldots,k_n} \rho_j(k_1+\cdots+k_n)^2 \prod_{j=1}^n \frac{1}{1+|k_j|^2}.
\end{aligned}$$

2. Next, we use Lemma 4.24 below and get for all $\delta > 0$ with $n\delta < 2$ and with $\langle k\rangle = \sqrt{1+|k|^2}$

$$
\begin{aligned}
&\sum_{k_1,\ldots,k_n} \rho_j(k_1+\cdots+k_n)^2 \prod_{j=1}^{n} \langle k_j\rangle^{-2} \\
&\lesssim \sum_{\ell_{n-1}} \sum_{k_1,\ldots,k_{n-2}} \rho_j(k_1+\cdots+k_{n-2}+\ell_{n-1})^2 \prod_{j=1}^{n-2} \langle k_j\rangle^{-2} \sum_{k_{n-1}} \langle \ell_{n-1}-k_{n-1}\rangle^{-2+\delta} \langle k_{n-1}\rangle^{-2+\delta} \\
&\lesssim \sum_{\ell_{n-1}} \sum_{k_1,\ldots,k_{n-2}} \rho_j(k_1+\cdots+k_{n-2}+\ell_{n-1})^2 \prod_{j=1}^{n-2} \langle k_j\rangle^{-2} \langle \ell_{n-1}\rangle^{-2+2\delta} \\
&\lesssim \cdots \lesssim \sum_{\ell_1} \rho_j(\ell_1)^2 \langle \ell_1\rangle^{-2+n\delta} \lesssim 2^{2j} 2^{j(-2+n\delta)} = 2^{jn\delta}.
\end{aligned}
$$

3. Thus, we have shown that $\mathbb{E}[|\Delta_j Z_\varepsilon^{:n:}(t,x)|^p]^{1/p} \lesssim 2^{jn\delta/2}$ for all $\delta > 0$ and $p > 0$. Combining the above analysis with similar arguments as in the proof of Lemma 4.3, we also get

$$\mathbb{E}[|\Delta_j Z_\varepsilon^{:n:}(t,x) - \Delta_j Z_\varepsilon^{:n:}(s,x)|^p]^{1/p} \lesssim 2^{j(n\delta/2+\lambda)} |t-s|^{\lambda/2}$$

for all $\lambda \in [0,1]$, and therefore it follows from Lemma 4.1 that we can bound

$$\mathbb{E}\big[\|Z_\varepsilon^{:n:}\|^p_{C_T^0 \mathscr{C}^{\alpha-2}} + \|Z_\varepsilon^{:n:}\|^p_{C_T^\gamma \mathscr{C}^{\alpha-3}}\big]$$

uniformly in $\varepsilon$. Since moreover $\varphi_{\varepsilon'}(k)^2 - \varphi_\varepsilon(k)^2 \to 0$ as $\varepsilon, \varepsilon' \to 0$ for all $k$, and the difference is uniformly bounded, we see from the dominated convergence theorem that $(Z_\varepsilon^{:n:})_\varepsilon$ is a Cauchy sequence in $L^p(C_T^0 \mathscr{C}^{\alpha-2} \cap C_T^\gamma \mathscr{C}^{\alpha-3})$. Therefore it converges to the limit

$$Z^{:n:}(t,x) = W_n(\mathbb{1}_{[0,t]}^{\otimes n} p(t-\cdot, x-\cdot)^{\otimes n}),$$

which is of course not defined pointwise in $x$ but only as a distribution. □

**Lemma 4.24.** *Let $\alpha, \beta < d$ be such that $\alpha + \beta > d$ and write $\langle k\rangle = \sqrt{1+|k|^2}$. Then*

$$\sum_{k' \in \mathbb{Z}^d} \langle k-k'\rangle^{-\alpha} \langle k'\rangle^{-\beta} \lesssim \langle k\rangle^{d-\alpha-\beta}.$$

**Proof.** See [40], Lemma 4.1. □

**Remark 4.25.** Most of the proof is independent of the dimension. However, in step 2. we crucially used that $d=2$. Convince yourself that in $d=3$ the same construction works for $Z^{:2:}$ but not for $Z^{:3:}$, and in $d=4$ it does not even work for $Z^{:2:}$.

For more details on Wiener-Ito chaos and related concepts see [31, 42].

## 4.6 Revisiting the $\Phi_2^4$ equation

In Section 4.4 we saw that if we were able to construct $Z^2$, $Z^3$ in $d=2$, then we would be able to solve the $\Phi_2^4$ equation by setting $\phi = Z + v$, where

$$(\partial_t - \Delta) v = -(Z^3 + 3Z^2 v + 3Z v^2 + v^3), \qquad v(0) = \phi_0.$$

However, we just showed that we can only construct the renormalized products

$$Z^{:n:}(t,x) = \lim_{\varepsilon \to 0} Z_\varepsilon^{:n:}(t,x),$$

where the convergence is in $C_T\mathscr{C}^{\alpha-2}\cap C_T^\gamma\mathscr{C}^{\alpha-3}$ whenever $\alpha<2$, $\gamma<1/2$, and $T>0$. So it is natural to replace the equation for $v$ with the "renormalized" equation

$$(\partial_t-\Delta)v=-(Z^{:3:}+3Z^{:2:}v+3Zv^2+v^3),\qquad v(0)=\phi_0,$$

which can be solved by the same arguments as in Section 4.4.

Of course, this raises the question whether we still have a useful interpretation for the new equation. For that purpose we go to the approximations $Z_\varepsilon^{:n:}$. Since the solution $v$ depends continuously on the data, we have $v=\lim_{\varepsilon\to 0}v_\varepsilon$, where

$$\begin{aligned}(\partial_t-\Delta)v_\varepsilon &= -(Z_\varepsilon^{:3:}+3Z_\varepsilon^{:2:}v_\varepsilon+3Z_\varepsilon v_\varepsilon^2+v_\varepsilon^3)\\ &= -\sum_{k=0}^{3}\binom{3}{k}H_k(Z_\varepsilon,C_\varepsilon)v_\varepsilon^{3-k}\\ &\overset{\text{Ex. 11.2}}{=}-H_3(Z_\varepsilon+v_\varepsilon,C_\varepsilon)\\ &= -(\phi_\varepsilon^3-3C_\varepsilon\phi_\varepsilon),\end{aligned}$$

where $\phi_\varepsilon=Z_\varepsilon+v_\varepsilon$, and we recall that Exercise 11.3 shows that

$$H_n(x+y,t)=\sum_{k=0}^{n}\binom{n}{k}H_k(x,t)y^{n-k}.$$

Therefore, $\phi=Z+v$ is given as the limit of $\phi_\varepsilon$ solving the renormalized equation

$$(\partial_t-\Delta)\phi_\varepsilon=-(\phi_\varepsilon^3-3C_\varepsilon\phi_\varepsilon)+\xi_\varepsilon,$$

where $\xi_\varepsilon=\mathscr{F}_{\mathbb{T}^d}^{-1}(\varphi_\varepsilon\mathscr{F}_{\mathbb{T}^d}\xi)$.

Combining Exercise 11.2 with a similar analysis as above, we can construct for all $n$ (locally in time) the limit $\phi$ of $\phi_\varepsilon$ solving

$$(\partial_t-\Delta)\phi_\varepsilon=-\phi_\varepsilon^{:n:}+\xi_\varepsilon:=-H_n(\phi_\varepsilon,C_\varepsilon)+\xi_\varepsilon,$$

which we would call the $\Phi_2^{n+1}$ model.

**Remark 4.26. (Voluntary, not contained in the videos)** We have some freedom how to choose the renormalization, because of course also the equation with renormalization $C_\varepsilon+\lambda$ for $\lambda\in\mathbb{R}$ converges to a limit. Our renormalization depends on time:

$$(\partial_t-\Delta)\phi_\varepsilon(t,x)=-(\phi_\varepsilon^3(t,x)-3C_\varepsilon(t)\phi_\varepsilon(t,x))+\xi_\varepsilon(t,x),$$

where

$$C_\varepsilon(t)=t+\sum_{k\neq 0}\frac{|\varphi(\varepsilon k)|^2}{2|2\pi k|^2}(1-e^{-2|2\pi k|^2t}).$$

It would be more natural to use a time-independent renormalization because the original equation was time-homogeneous in law and thus we would like also the renormalized equation to be time-homogeneous in law. For that purpose, note that with $c_\varepsilon=\sum_{k\neq 0}\frac{|\varphi(\varepsilon k)|^2}{2|2\pi k|^2}$ we have for all $t>0$

$$\lim_{\varepsilon\to 0}(C_\varepsilon(t)-c_\varepsilon)=t,$$

i.e. the difference of $C_\varepsilon(t)$ and the constant $c_\varepsilon$ converges to a finite limit. This suggests that also the solution $\tilde{\phi}_\varepsilon$ to

$$(\partial_t-\Delta)\tilde{\phi}_\varepsilon=-(\tilde{\phi}_\varepsilon^3-3c_\varepsilon\tilde{\phi}_\varepsilon)+\xi_\varepsilon$$

should converge, and indeed it is possible to show that by allowing for a "singularity" near 0.

## 4.7 Local subcriticality

We will see in the next section that for the $\Phi^4_3$ equation the trick of decomposing $\phi = Z + v$ is no longer sufficient. But then we will learn new tools ("paracontrolled distributions") to deal with the three-dimensional case. These new tools break down in dimension 4, but now the problem is more serious and at the moment there is no known theory for the $\Phi^4_4$ equation, which would be the physically relevant model (see the discussion in the appendix).

Here I want to briefly sketch why dimension 4 is different from lower dimensions. Regularity structures and paracontrolled distributions take a new point of view on regularity, and from that new point of view the solution $\phi$ to the $\Phi^4_d$ equation actually has relatively good regularity. For example, our solution to the $\Phi^4_2$ equation can be decomposed as

$$\phi = Z + C_T \mathscr{C}^{2-},$$

meaning that $\phi - Z \in C_T \mathscr{C}^{2-}$. So $\phi$ can be interpreted as a perturbation of the Gaussian $Z$, and on small scales this perturbation is nearly constant (because it has such good regularity). If we set the problem up in such a way that the initial condition for the perturbation is 0 (which we could do by changing the initial condition for $Z$), then the perturbation is negligible on small scales.

This can also be seen from a scaling argument: Consider the space-time rescaling operation $\mathcal{S}_\lambda u(t, x) = u(\lambda^2 t, \lambda x)$ and set $\phi_\lambda = \lambda^{d/2-1} \mathcal{S}_\lambda \phi$. Then

$$\begin{aligned} \partial_t \phi_\lambda =& \lambda^{2+d/2-1} \mathcal{S}_\lambda (\Delta \phi - \phi^3 + \xi) \\ =& (\Delta \phi_\lambda - \lambda^{4-d} \phi_\lambda^3 + \lambda^{1+d/2} \mathcal{S}_\lambda \xi). \end{aligned}$$

**Exercise. (Exercise 11.3)** Let $\xi$ be a space-time white noise and set $\xi_\lambda := \lambda^{1+d/2} \mathcal{S}_\lambda \xi$, interpreted rigorously as $\xi_\lambda(f) = \lambda^{1+d/2} \xi(\lambda^{-2-d} \mathcal{S}_{\lambda^{-1}} f)$. Show that $\xi_\lambda$ is a new space-time white noise.

By the exercise we have

$$\partial_t \phi_\lambda = \Delta \phi_\lambda - \lambda^{4-d} \phi_\lambda^3 + \xi_\lambda, \tag{4.10}$$

and since we are interested in small scales we want to take $\lambda \ll 1$. So we see that the equation should behave very differently depending on whether the dimension is $d < 4$, $d = 4$, or $d > 5$. For $d < 4$ the nonlinearity formally vanishes as $\lambda \to 0$ (in the language of Hairer [28] the equation is *locally subcritical*). For $d = 4$ the equation is formally scaling invariant (*critical*). For $d > 5$ the nonlinearity dominates the small scales (*supercritical*). We say that 4 is the *critical dimension*.

**Remark 4.27.** Another way of checking whether an equation is (sub-/super-)critical is based on regularity, and it is often easier to apply: We formally treat the space-time white noise as an element of $\mathcal{C}^{-\frac{d}{2}-1}$ and we are allowed to multiply distributions by applying the formal regularity estimates for the product which we get by ignoring the constraint $\alpha + \beta > 0$ for the resonant product; we also gain two derivatives by Schauder estimates. Then we guess that the solution to the $\Phi^4_d$ equation should have the regularity $\mathcal{C}^{1-\frac{d}{2}}$ (two derivatives more than the noise). Therefore, in $d \geqslant 2$ we guess $\phi^3 \in \mathcal{C}^{3-\frac{3}{2}d}$. Now we have to compare this regularity with the regularity of the noise:

$$(\partial_t - \Delta) \phi = - \underbrace{\phi^3}_{3-\frac{3}{2}d} + \underbrace{\xi}_{-\frac{d}{2}-1},$$

and

$$3 - \frac{3}{2} d > -\frac{d}{2} - 1 \qquad \Leftrightarrow \qquad 4 > d.$$

For $d=4$ the nonlinear term and the noise have the same regularity, so we say that the equation is critical. For $d>4$ the nonlinear term has worse regularity than the noise (and in particular our ansatz $\phi\in\mathcal{C}^{1-\frac{d}{2}}$ was not justified) and thus the equation is supercritical.

**Remark 4.28.** In the locally subcritical case we formally obtain for $\lambda\to 0$ the equation

$$\partial_t\phi=\Delta\phi+\xi,$$

whose solution we denote by $Z$. There is a small subtlety here because actually the white noise $\xi_\lambda$ should depend on $\lambda$, but let us ignore this. By this discussion we expect that for $d<4$ and on small scales $\phi$ should resemble $Z$, and as a first step we can then make sense of $Z^3$ (with suitable renormalization). Since the difficulty we have in making sense of $\phi^3$ comes from its irregularity, and irregularity is a small scale property and on small scales $\phi$ is essentially $Z$, we can thus hope to make sense of $\phi^3$. Regularity structures and paracontrolled distributions provide different tools for implementing this intution. Regularity structures are based on controlling increments, while paracontrolled distributions are based on Fourier descriptions of regularity.

Paracontrolled distributions are less general than regularity structures. Regularity structures provide a general theory to treat subcritical equations, and by now there exist deep black box type results that give a (local-in-time) renormalization and existence-uniqueness theory for large classes of subcritical equations [28, 9, 13, 10].

The following intuitive description might or might not be useful: Paracontrolled distributions are at the moment restricted to equations where the scaling exponent $\alpha$ in the factor $\lambda^\alpha$ that we pick up in front of the nonlinearity by a scaling argument as above is bounded from below by some $\alpha\geqslant\alpha_0>0$ (with $\alpha_0$ depending on the specific equation), while in regularity structures it suffices if $\alpha>0$. On the other hand, paracontrolled distributions are based on classical tools and function spaces from PDE theory, which might make them easier to learn and easier to implement in some applications.

Yet another formal explanation is that we may consider regularity structures as generalized Taylor expansions, and then the restriction of paracontrolled distributions is that we can only deal with first order expansions while regularity structures allow expansions of arbitrary order; see however [4, 35] for some progress towards generalizations of paracontrolled distributions.

Question: In which dimension is $(\partial_t-\Delta)u=u\xi$ (sub-/super-)critical?

# 5 The $\Phi^4_3$ equation and paracontrolled distributions

## 5.1 Tree notation

Let us now consider the case $d=3$. Then $Z\in C_T\mathscr{C}^{-1/2-}$ by Lemma 4.3, so if we assume that as for $\Phi^4_2$ we can construct the renormalized powers $Z^{:n:}$ with their canonical regularities, then we would get $Z^{:2:}\in C_T\mathscr{C}^{-1-}$ and $Z^{:3:}\in C_T\mathscr{C}^{-3/2-}$. Therefore, we expect that by Schauder estimates the solution $v$ to

$$(\partial_t-\Delta)v=-(Z^{:3:}+3Z^{:2:}v+3Zv^2+v^3)$$

has 2 degrees of regularity more than $Z^{:3:}$, i.e. $v\in C_T\mathscr{C}^{1/2-}$. But now we have a problem, because the product $Z^{:2:}v$ is ill-defined for $Z^{:2:}\in C_T\mathscr{C}^{-1-}$ and $v\in C_T\mathscr{C}^{1/2-}$ since the sum of the regularities is well below 0. We try to overcome this problem by cancelling again the most irregular term on the right hand side. Let us write

$$Y=\int_0^t P_{t-s}(-Z^{:3:}(s))ds,$$

so that $(\partial_t - \Delta)Y = -Z^{:3:}$, and thus we expect $Y \in C_T\mathscr{C}^{1/2-}$. We set $v^{(2)} = \phi - Z - Y$. Then $v = Y + v^{(2)}$, and therefore

$$(\partial_t - \Delta)v^{(2)} = -(3Z^{:2:}(Y + v^{(2)}) + 3Z(Y + v^{(2)})^2 + (Y + v^{(2)})^3).$$

The right hand side contains the product $Z^{:2:}(Y + v^{(2)})$ of $Z^{:2:} \in C_T\mathscr{C}^{-1-}$ and $Y + v^{(2)} \in C_T\mathscr{C}^{1/2-}$. However, $Y$ is explicit and thus $Z^{:2:}Y$ might be constructed using stochastic arguments. The problematic part is $Z^{:2:}v^{(2)}$, which will at best have the regularity of $Z^{:2:} \in C_T\mathscr{C}^{-1-}$. So by Schauder estimates we expect $v^{(2)} \in C_T\mathscr{C}^{1-}$, and this means that the product $Z^{:2:}v^{(2)}$ is ill-defined.

Unfortunately, this problem is more substantial than before, and we cannot solve it by expanding $v^{(2)}$ further, say by setting $v^{(2)} = U + v^{(3)}$ for some $U$: In that case we would have the product $Z^{:2:}v^{(3)}$ in the equation for $v^{(3)}$, and by the same arguments as above we then expect $v^{(3)} \in C_T\mathscr{C}^{1-}$ so that the product $Z^{:2:}v^{(3)}$ is ill-defined.

The solution is to use a more sophisticated expansion of $v^{(2)}$, based on paraproducts. But before that, let us introduce more efficient notation: Already for the considerations above we needed $Z, Z^{:n:}, Y$, and soon we would need even more letters. To make the variable names more transparent and compact, we use trees to denote the polynomials of $\xi$ that appear in the expansion of $\phi$. We use a dot $\bullet$ to represent an instance of $\xi$, and a line $\backslash$ to denote the convolution with the heat kernel. If two trees are multiplied with each other, we simply join them. So for example

$$Z = \RSI, \qquad Z^2 = \RSV, \qquad Z^3 = \RSW, \qquad Y = -\RSY, \qquad YZ = -\RSVY.$$

Sometimes we will also write $\RSY^2$ etc., if this seems simpler than writing out the full tree. For now we do not worry about the fact that many of these trees can at best be defined by using stochastic arguments, and that even this needs a proof and probably some renormalization. Instead, we work under the assumption that each tree in our expansion is given with its canonical regularity.

**Remark 5.1.** At this point a very attentive reader might recall from Remark 4.25 that our construction of $Z^{:3:}$ fails in $d = 3$. This will not be a problem because we only encounter $\RSY$ in the expansions below, and while $\RSW$ is not an element of any $C_T\mathscr{C}^\alpha$ space (continuous function of time with values in a space of distributions), it is a space-time distribution and we can construct its convolution against the heat kernel.

Question: Write out $\RSVY$.

## 5.2 The main commutator estimate

Let us write

$$\mathscr{L} = \partial_t - \Delta.$$

With our new notation, the expansion for $u := v^{(2)} = \phi - \RSI + \RSY$ can be rewritten as

$$\begin{aligned}\mathscr{L}u &= -\Big(3\RSV\Big(-\RSY + u\Big) + 3\RSI\Big(-\RSY + u\Big)^2 + \Big(-\RSY + u\Big)^3\Big)\\ &= -3\Big(u - \RSY\Big)\RSV + \tau_0 + \tau_1 u + \tau_2 u^2 - u^3,\end{aligned}$$

where we wrote

$$\tau_0 := -3\RSI \cdot \RSY^2 + \RSY^3, \qquad \tau_1 := 6\RSVY - 3\RSY^2, \qquad \tau_2 := -3\RSI + 3\RSY.$$

Let us use the underbrace $\underbrace{X}_{\alpha}$ to express that we expect $X \in C_T \mathscr{C}^{\alpha}$. Then

$$\mathscr{L}u = -3(\underbrace{u}_{1-} - \underbrace{\Psi}_{\frac{1}{2}-})\underbrace{\vee}_{-1-} + \underbrace{\tau_0}_{-\frac{1}{2}-} + \underbrace{\tau_1}_{-\frac{1}{2}-}\underbrace{u}_{1-} + \underbrace{\tau_2}_{-\frac{1}{2}-}\underbrace{u^2}_{1-} - \underbrace{u^3}_{1-},$$

which, as discussed before, means that $\left(u - \Psi\right)\vee$ is ill defined. Note however that if $\tau_2$, $\tau_1, \tau_0$ are given with their canonical regularities, then this is the only ill defined product.

The idea is now to decompose $\left(u - \Psi\right)\vee$ further with the help of paraproducts: Ignoring the fact that the resonant product is ill defined, we would expect to have the regularities

$$\left(u - \Psi\right)\vee = \underbrace{\left(u - \Psi\right)\varolessthan\vee}_{-1-} + \underbrace{\left(u - \Psi\right)\varogreaterthan\vee}_{-\frac{1}{2}-} + \underbrace{\left(u - \Psi\right)\odot\vee}_{-\frac{1}{2}-},$$

and therefore

$$\mathscr{L}u = -3\left(u - \Psi\right)\varolessthan\vee + C_T\mathscr{C}^{-1/2-},$$

by which we mean that $\mathscr{L}u + 3\left(u - \Psi\right)\varolessthan\vee \in C_T\mathscr{C}^{-1/2-}$. In other words, we expect that $\mathscr{L}u$ is given by a paraproduct plus a more regular remainder. We will see below that the convolution with the heat kernel in a certain sense commutes with the paraproduct (modulo a more regular remainder term), and this leads to the *paracontrolled ansatz*

$$\phi = \uparrow - \Psi + u, \qquad u = u' \varolessthan \Upsilon + u^{\sharp}, \qquad u' = -3\left(u - \Psi\right) \in C_T\mathscr{C}^{1/2-},$$
$$u \in C_T\mathscr{C}^{1-}, \qquad u^{\sharp} \in C_T\mathscr{C}^{3/2-}.$$

This is somewhat similar to the notion of a controlled path in rough path theory, recall Definition 2.22: We consider $u$ with a very specific structure which is adapted to the problem at hand.

Why should this specific decomposition be useful? Well, now we can expand the ill-defined resonant product further as

$$u \odot \vee = \underbrace{(u' \varolessthan \Upsilon)}_{1-} \odot \underbrace{\vee}_{-1-} + \underbrace{u^{\sharp}}_{\frac{3}{2}-} \odot \underbrace{\vee}_{-1-},$$

and the second term on the right hand side is well defined. Therefore, it remains to understand the resonant product $(u' \varolessthan \Upsilon) \odot \vee$.

Recall that we saw in Example 3.18 that the paraproduct $u' \varolessthan \Upsilon$ is a "frequency modulation" of $\Upsilon$ and it looks like $\Upsilon$ on small scales. But the difficulty we have with defining $(u' \varolessthan \Upsilon) \odot \vee$ comes from the fact that small scale contributions of $u' \varolessthan \Upsilon$ and $\vee$ might create diverging resonances in the product. So if we understand how the small scale contributions of $\Upsilon$ interact with those of $\vee$ and that no diverging resonances arise in the product, then we might also hope that $u' \varolessthan \Upsilon$ has no diverging resonances with $\vee$. This can be made precise with the help of the following commutator estimate, the main technical result in paracontrolled distributions. Here we write $\mathscr{C}^{\infty} = \bigcap_{\alpha > 0} \mathscr{C}^{\alpha}$.

**Lemma 5.2. (Commutator estimate, Lemma 2.4 of [22])** *Let $\alpha \in (0, 1)$ and let $\beta$, $\gamma \in \mathbb{R}$ be such that $\alpha + \beta + \gamma > 0$ and $\beta + \gamma < 0$. Then the trilinear operator $C: \mathscr{C}^{\infty} \to \mathscr{C}^{\infty}$,*

$$C(f, g, h) = ((f \varolessthan g) \odot h) - f(g \odot h),$$

*satisfies*

$$\|C(f,g,h)\|_{\alpha+\beta+\gamma} \lesssim \|f\|_\alpha \|g\|_\beta \|h\|_\gamma. \tag{5.1}$$

*Therefore, it has a unique extension to a bounded trilinear operator from $\mathscr{C}^\alpha \times \mathscr{C}^\beta \times \mathscr{C}^\gamma$ to $\mathscr{C}^{\alpha+\beta+\gamma}$.*

**Remark 5.3.**

i. It would be aesthetically more pleasing to have the same estimate for

$$(f \otimes g) \odot h - f \otimes (g \odot h), \tag{5.2}$$

but unfortunately this is not true.

ii. The constraint $\beta + \gamma < 0$ is not very important: If $\beta + \gamma > 0$, then we can easily control $C$ using the estimates for paraproduct and resonant product. If $\beta + \gamma = 0$, we have to "give up a bit of regularity" and slightly lower $\beta$ or $\gamma$.

Question: Using Lemma 5.2, convince yourself heuristically that the operator in (5.2) does not satisfy the desired estimate.

To prove Lemma 5.2 we need the following auxiliary result:

**Lemma 5.4.** *Let $\alpha \in (0,1)$ and $\beta \in \mathbb{R}$ and let $f \in \mathscr{C}^\alpha$ and $g \in \mathscr{C}^\beta$. Then for all $j \geqslant -1$*

$$\|\Delta_j(f \otimes g) - f \Delta_j g\|_{L^\infty} \lesssim 2^{-j(\alpha+\beta)} \|f\|_\alpha \|g\|_\beta.$$

**Proof.** We give the proof on $\mathbb{R}^d$. On $\mathbb{T}^d$ it would be slightly more complicated as there is no exact scaling relation $K_j(x) = 2^{jd} K_0(2^j x)$, but we could use Poisson summation to overcome this problem.

Recall that $f \otimes g = \sum_i \Delta_{<i-1} f \Delta_i g$ and that $\mathscr{F}(\Delta_{<i-1} f \Delta_i g)$ is supported on an annulus $2^i A$. Therefore,

$$\begin{aligned}
\Delta_j(f \otimes g) - f \Delta_j g &= \sum_{i:2^i \simeq 2^j} (\Delta_j(\Delta_{<i-1} f \Delta_i g) - f \Delta_j \Delta_i g) \\
&= \sum_{i:2^i \simeq 2^j} (\Delta_j(\Delta_{<i-1} f \Delta_i g) - \Delta_{<i-1} f \Delta_j \Delta_i g) - \sum_{i:2^i \simeq 2^j} \Delta_{\geqslant i-1} f \Delta_j \Delta_i g.
\end{aligned}$$

Since $\alpha > 0$, the second term on the right hand side is easily estimated by

$$\left\| \sum_{i:2^i \simeq 2^j} \Delta_{\geqslant i-1} f \Delta_j \Delta_i g \right\|_{L^\infty} \lesssim \sum_{i:2^i \simeq 2^j} 2^{-i\alpha} \|f\|_\alpha 2^{-i\beta} \|g\|_\beta \simeq 2^{-j(\alpha+\beta)} \|f\|_\alpha \|g\|_\beta.$$

For the remaining term we write the action of $\Delta_j$ as a convolution:

$$\begin{aligned}
&|(\Delta_j(\Delta_{<i-1} f \Delta_i g) - \Delta_{<i-1} f \Delta_j \Delta_i g)(x)| \\
&= \left| \int K_j(x-y)(\Delta_{<i-1} f(y) - \Delta_{<i-1} f(x)) \Delta_i g(y) \mathrm{d}y \right| \\
&\overset{\text{Taylor}}{\lesssim} \int |K_j(x-y)| \max_{|\mu|=1} \|\partial^\mu \Delta_{<i-1} f\|_{L^\infty} |x-y| \|\Delta_i g\|_{L^\infty} \mathrm{d}y \\
&\simeq 2^{i(1-\alpha-\beta)} \max_{|\mu|=1} \|\partial^\mu f\|_{\alpha-1} \|g\|_\beta \int |K_j(x-y)| |x-y| \mathrm{d}y \\
&\lesssim 2^{i(1-\alpha-\beta)} \|f\|_\alpha \|g\|_\beta \int |K_j(y)| |y| \mathrm{d}y,
\end{aligned}$$

where we used that $1-\alpha>0$ and thus $\sum_{j<i-1} 2^{j(1-\alpha)} \simeq 2^{i(1-\alpha)}$. If $j=-1$, this estimate is sufficient. For $j \geqslant 0$ we have

$$\int |K_j(y)||y|\mathrm{d}y = \int |2^{jd} K_0(2^j y)| 2^{-j} |2^j y| \mathrm{d}y = 2^{-j} \int |K_0(y)| \times |y| \mathrm{d}y \simeq 2^{-j},$$

from where our claim follows. □

**Proof. (of Lemma 5.2)**

- Decomposition $C = C_1 + C_2$:

  Since $f, g, h \in \mathscr{S}$, all the series are absolutely summable and we can exchange the order of summation at will. By the spectral support properties of the paraproduct, there exists $N \in \mathbb{N}$ such that $\Delta_i(\Delta_k f \circledcirc g) = 0$ for $k > N$. This allows us to decompose

$$\begin{aligned} C(f,g,h) &= \sum_{|i-j|\leqslant 1} [\Delta_i(f \circledcirc g) - f \Delta_i g] \Delta_j h \\ &= \sum_{|i-j|\leqslant 1} [\Delta_i(\Delta_{\lesssim i} f \circledcirc g) - \Delta_{\lesssim i} f \Delta_i g] \Delta_j h + \sum_{|i-j|\leqslant 1} \Delta_{\gtrsim i} f \Delta_i g \Delta_j h \\ &=: C_1(f,g,h) + C_2(f,g,h), \end{aligned}$$

  where $C_1$ and $C_2$ are defined by the equality, and where

$$\Delta_{\lesssim i} f := \Delta_{\leqslant i+N} f, \qquad \Delta_{\gtrsim i} f := f - \Delta_{\lesssim i} f.$$

- Bound for $C_1$:

  For fixed $i$, the term

$$A_i := \sum_j \mathbb{1}_{|i-j|\leqslant 1} [\Delta_i(\Delta_{\lesssim i} f \circledcirc g) - \Delta_{\lesssim i} f \Delta_i g] \Delta_j h$$

  has spectral support in a ball $2^i B$. Therefore, the regularity of $C_1 = \sum_i A_i$ can be controlled using Lemma 3.16 if we derive a suitable $L^\infty$-bound for $A_i$. Since $\alpha \in (0,1)$, Lemma 5.4 yields

$$\begin{aligned} &\left\| \sum_j \mathbb{1}_{|i-j|\leqslant 1} [\Delta_i(\Delta_{\lesssim i} f \circledcirc g) - \Delta_{\lesssim i} f \Delta_i g] \Delta_j h \right\|_{L^\infty} \\ &\lesssim \sum_j \mathbb{1}_{|i-j|\leqslant 1} 2^{-i(\alpha+\beta)} \|\Delta_{\lesssim i} f\|_\alpha \|g\|_\beta 2^{-j\gamma} \|h\|_\gamma \\ &\lesssim 2^{-i(\alpha+\beta+\gamma)} \|f\|_\alpha \|g\|_\beta \|h\|_\gamma, \end{aligned}$$

  so since $\alpha+\beta+\gamma>0$ the claimed regularity for $C_1(f,g,h)$ indeed follows from Lemma 3.16.

- Bound for $C_2$:

  We write

$$C_2(f,g,h) = \sum_{|i-j|\leqslant 1} \Delta_{\gtrsim i} f \Delta_i g \Delta_j h = \sum_k \left( \sum_{|i-j|\leqslant 1} \mathbb{1}_{i \lesssim k} \Delta_k f \Delta_i g \Delta_j h \right),$$

  and the term inside the brackets has spectral supprt in a ball $2^k B$. Moreover, since $\beta+\gamma<0$,

$$\begin{aligned} \left\| \sum_{|i-j|\leqslant 1} \mathbb{1}_{i \lesssim k} \Delta_k f \Delta_i g \Delta_j h \right\|_{L^\infty} &\lesssim 2^{-k\alpha} \|f\|_\alpha \sum_{i \lesssim k} 2^{-i(\beta+\gamma)} \|g\|_\beta \|h\|_\gamma \\ &\lesssim 2^{-k(\alpha+\beta+\gamma)} \|f\|_\alpha \|g\|_\beta \|h\|_\gamma, \end{aligned}$$

and now we use once more that $\alpha+\beta+\gamma>0$ and apply Lemma 3.16 to conclude the proof of the estimate for $C(f,g,h)$.

- Existence/uniqueness of the extension:

  Recall that we redefined $\mathscr{C}^\delta$ as the closure of $\mathscr{C}^\infty$ with respect to the $\|\cdot\|_\delta$ norm. So $\mathscr{C}^\infty$ is dense and the existence and uniqueness of the continuous extension follows from the fact that any uniformly continuous map on a dense subset can be uniquely extended to a continuous map on the full space. $\square$

So if $\phi$ satisfies the paracontrolled ansatz, our commutator estimate suggests to define

$$\left(u'\circledcirc\mathsf{Y}\right)\odot\vee := C(\underbrace{u'}_{\frac{1}{2}-},\underbrace{\mathsf{Y}}_{1-},\underbrace{\vee}_{-1-})+\underbrace{u'}_{\frac{1}{2}-}\underbrace{\left(\mathsf{Y}\odot\vee\right)}_{0-}.$$

The sum of the regularities of the arguments of $C$ is strictly positive, and therefore $C(u', \mathsf{Y},\vee)$ is well defined and in $C_T\mathscr{C}^{1/2-}$ by Lemma 5.2. For the remaining term the resonant product $\mathsf{Y}\odot\vee$ is still not defined, but if we make it part of the data, call it $\overset{\vee}{\mathsf{Y}}$, and assume it has its canonical regularity $\overset{\vee}{\mathsf{Y}}\in C_T\mathscr{C}^{0-}$, then $u'\overset{\vee}{\mathsf{Y}}$ is well defined and in $C_T\mathscr{C}^{0-}$.

In the following we make this rigorous by defining a suitable Banach space of paracontrolled distributions in which we can set up a fixed point iteration to solve the $\Phi^4_3$ equation. To simplify the presentation we consider a linearized equation which, from the regularity analysis point of view, contains nearly all the difficulties of the $\Phi^4_3$ equation.

Question: Why do we need $\alpha<1$ in the commutator estimate?

## 5.3 A linearized $\Phi^4_3$ equation

Consider now the linearized equation

$$(\partial_t-\Delta)v=v\eta.$$

Since in this equation we only have to deal with two noise terms, we abandon the tree notation and write $\eta$ instead of $\vee$. But as discussed above, for dealing with the full $\Phi^4_3$ equation it is convenient to use tree notation.

In the previous discussion we considered function spaces of the type $C_T\mathscr{C}^\alpha$, thus we only quantified the space regularity. It turns out that now we also need to make more precise assumptions on the time regularity. For that purpose we define the following function spaces:

**Definition 5.5.** *For $\alpha\in(0,2]$ and $T>0$ we set*

$$\mathscr{L}^\alpha_T := C^{\alpha/2}_T L^\infty\cap C_T\mathscr{C}^\alpha,$$

*equipped with the norm $\|u\|_{\mathscr{L}^\alpha_T}=\|u\|_{C^{\alpha/2}_T}+\|u\|_{C_T\mathscr{C}^\alpha}$. We also write*

$$\mathscr{L}^\alpha=C^{\alpha/2}_{\mathrm{loc}}(\mathbb{R}_+,L^\infty)\cap C(\mathbb{R}_+,\mathscr{C}^\alpha),$$

*with the obvious definition of the space of locally Hölder continuous functions $C^{\alpha/2}_{\mathrm{loc}}(\mathbb{R}_+, L^\infty)$.*

We fix $\alpha \in (\frac{2}{3}, 1)$, and we assume that $\eta \in C(\mathbb{R}_+, \mathscr{C}^{\alpha-2})$. We define

$$\mathcal{I}\eta(t) = \int_0^t P_{t-s}\eta(s)ds, \qquad t \geqslant 0,$$

which, as we will see below, satisfies $\|\mathcal{I}\eta\|_{\mathscr{L}_T^\alpha} \lesssim (1 \vee T)\|\eta\|_{C_T\mathscr{C}^{\alpha-2}}$ for all $T > 0$. We also assume that $\mathcal{I}\eta \odot \eta \in C(\mathbb{R}_+, \mathscr{C}^{2\alpha-2})$ is given with its natural regularity (note that $2\alpha - 2 < 0$). We write

$$\mathcal{T}^\alpha = C(\mathbb{R}_+, \mathscr{C}^{\alpha-2}) \times C(\mathbb{R}_+, \mathscr{C}^{2\alpha-2})$$

and

$$\mathbb{Z} = (\eta, \mathcal{I}\eta \odot \eta)$$

for a generic element of $\mathcal{T}^\alpha$, as well as for $T > 0$

$$\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha} := \|\eta\|_{C_T\mathscr{C}^{\alpha-2}} + \|\mathcal{I}\eta \odot \eta\|_{C_T\mathscr{C}^{2\alpha-2}}.$$

We also need a modified version of the paraproduct:

**Definition 5.6.** *We define the modified paraproduct as*

$$u \prec\!\!\prec v(t) = \int_0^t P_{t-s}(u(s) \otimes \mathscr{L}v(s))ds + P_t(u(0) \otimes v(0)),$$

*whenever this is well defined.*

**Remark 5.7.** This definition is due to Bailleul, Bernicot and Frey [5], although they did not include the term $P_t(u(0) \otimes v(0))$, which we will need to compare $u \prec\!\!\prec v$ and $u \otimes v$. The beauty of their definition is that we trivially have $\mathscr{L}(u \prec\!\!\prec v) = u \otimes \mathscr{L}v$. This will be very convenient for deriving *paracontrolled Schauder estimates*.

You should compare the following definition with Definition 2.22 of a controlled path. What are the similarities and what are the differences?

**Definition 5.8.** *Let $\beta \in (\frac{2-\alpha}{2}, \alpha]$ and $T > 0$. A function $v \in \mathscr{L}_T^\beta$ is* paracontrolled by *$\mathbb{Z} \in \mathcal{T}^\alpha$ if there exists $v' \in \mathscr{L}_T^\beta$, such that $v^\sharp \in \mathscr{L}_T^{2\beta}$, where*

$$v^\sharp := v - v' \prec\!\!\prec \mathcal{I}\eta.$$

*In that case we write*

$$v \in \mathscr{D}_T^\beta(\mathbb{Z}) \qquad \text{or} \qquad (v, v') \in \mathscr{D}_T^\beta(\mathbb{Z}),$$

*and we define*

$$\|v\|_{\mathscr{D}_T^\beta} := \|v\|_{\mathscr{L}_T^\beta} + \|v'\|_{\mathscr{L}_T^\beta} + \|v^\sharp\|_{\mathscr{L}_T^{2\beta}},$$

*$\mathscr{D}_T^\beta(\mathbb{Z})$ is a Banach space. If there is no ambiguity about $v'$, we also write $v \in \mathscr{D}_T^\beta(\mathbb{Z})$, or $v \in \mathscr{D}_T^\beta$ if there is no ambiguity about $\mathbb{Z}$.*

*If $\tilde{\mathbb{Z}} \in \mathcal{T}^\alpha$ and if $\tilde{v} \in \mathscr{D}_T^\beta(\tilde{\mathbb{Z}})$, then we also define*

$$\|v - \tilde{v}\|_{\mathscr{D}_T^\beta} = \|v - \tilde{v}\|_{\mathscr{L}_T^\beta} + \|v' - \tilde{v}'\|_{\mathscr{L}_T^\beta} + \|v^\sharp - \tilde{v}^\sharp\|_{\mathscr{L}_T^{2\beta}}.$$

Note that $\|v - \tilde{v}\|_{\mathscr{D}_T^\beta}$ is only formal notation, this is not a norm since $v$ and $\tilde{v}$ do not even live in the same space.

Question: Find an example where $v$ is paracontrolled by $\mathbb{Z}$ but $v'$ is not unique, i.e. there exist $v' \neq \tilde{v}'$ such that that $(v, v'), (v, \tilde{v}') \in \mathscr{D}_T^\beta$. You may use that for $\gamma \in (0, 2]$

$$\left\| \left( t \mapsto \int_0^t P_{t-s} f(s) ds \right) \right\|_{\mathscr{L}_T^\gamma} \lesssim (T \vee 1) \| f \|_{C_T \mathscr{C}^{\gamma-2}}.$$

*Hint: what if $\eta$ is actually in $C_T \mathscr{C}^{2\beta-2}$ and not just in $\mathscr{L}_T^\alpha$?*

Our next aim is to make sense of the product $v\eta$ for $v \in \mathscr{D}_T^\beta$. We use the paracontrolled structure and set

$$v\eta := v \otimes \eta + v \ominus \eta + \underbrace{v^\sharp}_{2\beta} \odot \underbrace{\eta}_{\alpha-2} + (v' \prec\!\!\prec \mathcal{I}\eta) \odot \eta,$$

and since $2\beta + \alpha - 2 > 0$ by assumption on $\beta$ ($\beta > \frac{2-\alpha}{2}$), the resonant product $v^\sharp \odot \eta$ is well defined. If instead of the modified paraproduct $v' \prec\!\!\prec \mathcal{I}\eta$ we had $v' \otimes \mathcal{I}\eta$, then we could set

$$(v' \otimes \mathcal{I}\eta) \odot \eta = C(\underbrace{v'}_{\beta}, \underbrace{\mathcal{I}\eta}_{\alpha}, \underbrace{\eta}_{\alpha-2}) + \underbrace{v'}_{\beta} \underbrace{(\mathcal{I}\eta \odot \eta)}_{2\alpha-2},$$

and using that $\beta + 2\alpha - 2 \geqslant 2\beta + \alpha - 2 > 0$, the right hand side would be well defined. Therefore, we should compare the modified paraproduct with the "usual" paraproduct and hopefully we can show that the difference is more regular. For that purpose we need the following auxiliary estimate. Recall that

$$\varphi(\mathrm{D}) u = \mathscr{F}^{-1}(\varphi \mathscr{F} u) = \mathscr{F}^{-1} \varphi * u.$$

**Lemma 5.9. (Improved version of [46], Lemma 5.3.20)** *Let $\alpha < 1$, $\beta \in \mathbb{R}$, and $\delta > -1$. Then we have uniformly in $\varepsilon > 0$ for $u \in \mathscr{C}^\alpha$ and $v \in \mathscr{C}^\beta$:*

$$\begin{aligned} &\| \varphi(\varepsilon \mathrm{D})(u \otimes v) - u \otimes (\varphi(\varepsilon \mathrm{D}) v) \|_{\alpha+\beta+\delta} \\ &\lesssim \varepsilon^{-\delta} \left( \| x \mathscr{F}^{-1} \varphi \|_{B_{1,\infty}^{\delta+1}} + \| \mathscr{F}^{-1} \varphi \|_{B_{1,\infty}^\delta} \right) \| u \|_\alpha \| v \|_\beta. \end{aligned}$$

**Proof. (Voluntary, not part of the lecture videos)** Let us write $\hat{\varphi}_\varepsilon = \mathcal{F}^{-1}(\varphi(\varepsilon \cdot))$. The proof is similar to that of Lemma 5.4, and we argue again on $\mathbb{R}^d$ because here we can use scaling.

1. By Lemma 3.16 it suffices to control for all $j$ the $L^\infty$ norm of

$$\begin{aligned} &\hat{\varphi}_\varepsilon * (\Delta_{<j-1} u \Delta_j v)(x) - \Delta_{<j-1} u \otimes \hat{\varphi}_\varepsilon * v(x) \\ &= \int \hat{\varphi}_\varepsilon(x-y)(\Delta_{<j-1} u(y) - \Delta_{<j-1} u(x)) \Delta_j v(y) \mathrm{d}y. \end{aligned}$$

2. For $\varepsilon 2^j \geqslant 1$ we use that $\Delta_{<j-1} u \Delta_j v$ and $\Delta_j$ are both spectrally supported in an annulus $2^j A$, and thus we can find $\psi \in C_c^\infty$ with $\psi|_{A=1}$ and replace $\hat{\varphi}_\varepsilon$ by $\hat{\varphi}_\varepsilon * \hat{\psi}_{2^{-j}}$, where

$$\hat{\psi}_{2^{-j}} = \mathscr{F}^{-1}(\psi(2^{-j} \cdot)).$$

   As in Lemma 5.4 we get, using that $\alpha < 1$,

$$\begin{aligned} &\left| \int \hat{\varphi}_\varepsilon * \hat{\psi}_{2^{-j}}(x-y)(\Delta_{<j-1} u(y) - \Delta_{<j-1} u(x)) \Delta_j v(y) dy \right| \\ &\lesssim \int |\hat{\varphi}_\varepsilon * \hat{\psi}_{2^{-j}}(y)| \times |y| dy 2^{j(1-\alpha-\beta)} \| u \|_\alpha \| v \|_\beta, \end{aligned}$$

so it suffices to bound the integral on the right hand side. For that purpose note first that

$$
\begin{aligned}
y\hat{\varphi}_\varepsilon * \hat{\psi}_{2^{-j}}(y) =& y\mathscr{F}^{-1}(\varphi(\varepsilon\cdot)\psi(2^{-j}\cdot))(y)\\
=& \varepsilon\mathscr{F}^{-1}((\nabla\varphi)(\varepsilon\cdot)\psi(2^{-j}\cdot))(y) + 2^{-j}\mathscr{F}^{-1}(\varphi(\varepsilon\cdot)(\nabla\psi)(2^{-j}\cdot))(y)\\
=& \varepsilon^{1-d}\mathscr{F}^{-1}((\nabla\varphi)\psi(\varepsilon^{-1}2^{-j}\cdot))(\varepsilon^{-1}y)\\
&+ 2^{-j}\varepsilon^{-d}\mathscr{F}^{-1}(\varphi(\nabla\psi)(\varepsilon^{-1}2^{-j}\cdot))(\varepsilon^{-1}y),
\end{aligned}
$$

and therefore

$$
\int |y\hat{\varphi}_\varepsilon * \hat{\psi}_{2^{-j}}(y)|\mathrm{d}y = \varepsilon\|\mathscr{F}^{-1}(\nabla\varphi\psi(\varepsilon^{-1}2^{-j}\cdot))\|_{L^1} + 2^{-j}\|\mathscr{F}^{-1}(\varphi(\nabla\psi)(\varepsilon^{-1}2^{-j}\cdot))\|_{L^1}.
$$

The first term on the right hand side can be estimated by

$$
\begin{aligned}
\|\mathscr{F}^{-1}(\nabla\varphi\psi(\varepsilon^{-1}2^{-j}\cdot))\|_{L^1} \leqslant& \sum_{i:2^i\sim\varepsilon 2^j} \|\Delta_i\mathscr{F}^{-1}(\nabla\varphi) * \mathscr{F}^{-1}\psi(\varepsilon^{-1}2^{-j}\cdot)\|_{L^1}\\
\leqslant& \sum_{i:2^i\sim\varepsilon 2^j} \|\Delta_i\mathscr{F}^{-1}(\nabla\varphi)\|_{L^1}\|\mathscr{F}^{-1}\psi\|_{L^1}\\
\lesssim& \sum_{i:2^i\sim\varepsilon 2^j} \|\Delta_i(x\mathscr{F}^{-1}\varphi)\|_{L^1}\\
\simeq& (\varepsilon 2^j)^{-\delta-1}\|x\mathscr{F}^{-1}\varphi\|_{B^{\delta+1}_{1,\infty}},
\end{aligned}
$$

and by the same argument

$$
\|\mathscr{F}^{-1}(\varphi(\nabla\psi)(\varepsilon^{-1}2^{-j}\cdot))\|_{L^1} \lesssim (\varepsilon 2^j)^{-\delta}\|\mathscr{F}^{-1}\varphi\|_{B^{\delta}_{1,\infty}},
$$

so overall

$$
\begin{aligned}
&\|\hat{\varphi}_\varepsilon * (\Delta_{<j-1}u\Delta_j v) - \Delta_{<j-1}u \circledcirc \varphi_\varepsilon * v\|_{L^\infty}\\
\lesssim& \Big(\varepsilon(\varepsilon 2^j)^{-\delta-1}\|x\mathscr{F}^{-1}\varphi\|_{B^{\delta+1}_{1,\infty}} + 2^{-j}(\varepsilon 2^j)^{-\delta}\|\mathscr{F}^{-1}\varphi\|_{B^{\delta}_{1,\infty}}\Big)2^{j(1-\alpha-\beta)}\|u\|_\alpha\|v\|_\beta\\
\lesssim& \varepsilon^{-\delta}2^{-j(\delta+\alpha+\beta)}\Big(\|x\mathscr{F}^{-1}\varphi\|_{B^{\delta+1}_{1,\infty}} + \|\mathscr{F}^{-1}\varphi\|_{B^{\delta}_{1,\infty}}\Big)\|u\|_\alpha\|v\|_\beta,
\end{aligned}
$$

which is the desired bound if $\varepsilon 2^j \geqslant 1$.

3. For $\varepsilon 2^j < 1$ we do not make use of the function $\psi$. Then we obtain

$$
\begin{aligned}
&\left|\int \hat{\varphi}_\varepsilon(\Delta_{<j-1}u(y) - \Delta_{<j-1}u(x))\Delta_j v(y)dy\right|\\
\lesssim& \int |\hat{\varphi}_\varepsilon(y)| \times |y| dy 2^{j(1-\alpha-\beta)}\|u\|_\alpha\|v\|_\beta\\
=& \varepsilon\int |\mathscr{F}^{-1}(y)y| dy 2^{j(1-\alpha-\beta)}\|u\|_\alpha\|v\|_\beta\\
=& \varepsilon 2^j 2^{-j(\alpha+\beta)}\|x\mathscr{F}^{-1}\varphi\|_{L^1}\|u\|_\alpha\|v\|_\beta\\
\lesssim& (\varepsilon 2^j)^{-\delta}2^{-j(\alpha+\beta)}\|x\mathscr{F}^{-1}\varphi\|_{B^{\delta+1}_{1,\infty}}\|u\|_\alpha\|v\|_\beta,
\end{aligned}
$$

where in the last step we used that $\delta > -1$ and thus $\varepsilon 2^j \leqslant (\varepsilon 2^j)^{-\delta}$ and $\|x\mathscr{F}^{-1}\varphi\|_{L^1} \lesssim \|x\mathscr{F}^{-1}\varphi\|_{B^{\delta+1}_{1,\infty}}$. □

We will apply this with $\varphi(x) = e^{-|2\pi x|^2}$ and $\varepsilon = t^{1/2}$, so that $\varphi(\varepsilon\mathrm{D})u = P_t u$.

**Corollary 5.10.** *Let $\alpha \in (0,1)$ and $\beta \in \mathbb{R}$. We have for $u$ and $v$ such that the right hand side is finite:*

$$
\|u \prec\!\!\prec v - u \circledcirc v\|_{C_T\mathscr{C}^{\alpha+\beta}} \lesssim (1 \vee T)\|u\|_{\mathscr{L}^\alpha_T}(\|v(0)\|_\beta + \|\mathscr{L}v\|_{C_T\mathscr{C}^{\beta-2}}).
$$

**Proof.** We write

$$v(t)=\int_0^t P_{t-s}\mathscr{L}v(s)ds+P_tv(0)$$

and decompose

$$\begin{aligned}
&u \prec\!\!\prec v(t)-u\varolessthan v(t)\\
&=\int_0^t P_{t-s}(u(s)\varolessthan\mathscr{L}v(s))ds+P_t(u(0)\varolessthan v(0))\\
&\quad -u(t)\varolessthan\left(\int_0^t P_{t-s}\mathscr{L}v(s)ds+P_tv(0)\right)\\
&=\int_0^t P_{t-s}((u(s)-u(t))\varolessthan\mathscr{L}v(s))ds+\int_0^t[P_{t-s}(u(t)\varolessthan\mathscr{L}v(s))-u(t)\varolessthan P_{t-s}\mathscr{L}v(s)]ds\\
&\quad +P_t((u(0)-u(t))\varolessthan v(0))+[P_t(u(t)\varolessthan v(0))-u(t)\varolessthan P_tv(0)]\\
&=:A_1+A_2+A_3+A_4.
\end{aligned}$$

Now we simply estimate each $A_i$ separately and we apply the commutator estimate for paraproduct and $P_t$ (Lemma 5.9), the estimate $\|P_tf\|_{\gamma+\delta}\lesssim(1\vee t^{-\delta/2})\|f\|_\gamma$, and the time regularity of $u$ along the way: For example we get with $\alpha'<\alpha$

$$\begin{aligned}
\|A_1\|_{\alpha'+\beta}&\lesssim\int_0^t\|P_{t-s}((u(s)-u(t))\varolessthan\mathscr{L}v(s))\|_{\alpha'+\beta}\mathrm{d}s\\
&\overset{(4.5)}{\lesssim}\int_0^t(1\vee(t-s)^{-(2+\alpha')/2})\|(u(s)-u(t))\varolessthan\mathscr{L}v(s)\|_{\beta-2}\mathrm{d}s\\
&\lesssim\int_0^t(1\vee(t-s)^{-1-\alpha'/2})\|u(s)-u(t)\|_{L^\infty}\|\mathscr{L}v\|_{C_T\mathscr{C}^{\beta-2}}\mathrm{d}s\\
&\lesssim\int_0^t(1\vee(t-s)^{-1-\alpha'/2})(1\wedge|t-s|^{\alpha/2})\|u\|_{\mathscr{L}_T^\alpha}\|\mathscr{L}v\|_{C_T\mathscr{C}^{\beta-2}}\mathrm{d}s\\
&\lesssim(T\vee T^{(\alpha-\alpha')/2})\|u\|_{\mathscr{L}_T^\alpha}\|\mathscr{L}v\|_{C_T\mathscr{C}^{\beta-2}}.
\end{aligned}$$

As indicated in the proof of Theorem 4.6, we can refine the analysis to avoid the loss of regularity and get the same result for $\alpha'=\alpha$.

For $A_2$ we have

$$\begin{aligned}
\|A_2\|_{\alpha'+\beta}&\leqslant\int_0^t\|P_{t-s}(u(t)\varolessthan\mathscr{L}v(s))-u(t)\varolessthan P_{t-s}\mathscr{L}v(s)\|_{\alpha'+\beta}ds\\
&\overset{\text{Lem. }5.9}{\lesssim}\int_0^t(1\vee(t-s)^{-(2-(\alpha-\alpha'))/2})\|u(t)\|_\alpha\|\mathscr{L}v(s)\|_{\beta-2}ds\\
&\lesssim(T\vee T^{(\alpha-\alpha')/2})\|u\|_{\mathscr{L}_T^\alpha}\|\mathscr{L}v\|_{C_T\mathscr{C}^{\beta-2}},
\end{aligned}$$

and again we can improve the estimate to handle also $\alpha'=\alpha$. □

Question: Derive the bound for $A_3$.

**Definition 5.11.** *For $v\in\mathscr{D}_T^\beta(\mathbb{Z})$ we define*

$$\begin{aligned}
v\eta:=&v\varolessthan\eta+v\varogreaterthan\eta+v^\sharp\odot\eta+(v'\prec\!\!\prec\mathcal{I}\eta-v'\varolessthan\mathcal{I}\eta)\odot\eta\\
&+C(v',\mathcal{I}\eta,\eta)+v'(\mathcal{I}\eta\odot\eta).
\end{aligned}$$

**Lemma 5.12.** *For $\alpha\in(\frac{2}{3},1)$, $\beta\in(\frac{2-\alpha}{2},\alpha]$, $\mathbb{Z}\in\mathcal{T}^\alpha$ and $v\in\mathscr{D}_T^\beta(\mathbb{Z})$ we have*

$$\|v\eta\|_{C_T\mathscr{C}^{\alpha-2}}+\|v\eta-v\varolessthan\eta\|_{C_T\mathscr{C}^{\alpha+\beta-2}}\leqslant M_T\|v\|_{\mathscr{D}_T^\beta},$$

*where $M_T > 0$ is a polynomial function of $\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}$. Moreover, if $\tilde{v} \in \mathscr{D}_T^\beta \tilde{\mathbb{Z}}$, then*

$$\|v\eta - \tilde{v}\tilde{\eta}\|_{C_T \mathscr{C}^{\alpha+\beta-2}} \leqslant M_T \big( \|\mathbb{Z} - \tilde{\mathbb{Z}}\|_{\mathcal{T}_T^\alpha} + \|v - \tilde{v}\|_{\mathscr{D}_T^\beta} \big),$$

*where now $M_T > 0$ is a polynomial function of $\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}$, $\|\tilde{\mathbb{Z}}\|_{\mathcal{T}_T^\alpha}$, $\|v\|_{\mathscr{D}_T^\beta(\mathbb{Z})}$, $\|\tilde{v}\|_{\mathscr{D}_T^\beta(\tilde{\mathbb{Z}})}$, and only of $\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}$ if $\mathbb{Z} = \tilde{\mathbb{Z}}$.*

**Proof.** We simply have to estimate each term in the definition of $v\eta$:

$$\begin{aligned}
\|v \otimes \eta\|_{C_T \mathscr{C}^{\alpha-2}} &\overset{\beta>0}{\lesssim} \|v\|_{C_T \mathscr{C}^\beta} \|\eta\|_{C_T \mathscr{C}^{\alpha-2}}, \\
\|v \ominus \eta\|_{C_T \mathscr{C}^{\alpha+\beta-2}} &\overset{\alpha-2<0}{\lesssim} \|v\|_{C_T \mathscr{C}^\beta} \|\eta\|_{C_T \mathscr{C}^{\alpha-2}}, \\
\|v^\sharp \odot \eta\|_{C_T \mathscr{C}^{\alpha+2\beta-2}} &\overset{\alpha+2\beta-2>0}{\lesssim} \|v^\sharp\|_{C_T \mathscr{C}^{2\beta}} \|\eta\|_{C_T \mathscr{C}^{\alpha-2}}, \\
\|(v' \prec\!\!\prec \mathcal{I}\eta - v' \otimes \mathcal{I}\eta) \odot \eta\|_{C_T \mathscr{C}^{2\alpha+\beta-2}} &\overset{\beta<1}{\lesssim} \|v'\|_{\mathscr{L}_T^\beta} (\|\mathcal{I}\eta(0)\|_\alpha + \|\mathscr{L}\mathcal{I}\eta\|_{C_T \mathscr{C}^{\alpha-2}}) \|\eta\|_{C_T \mathscr{C}^{\alpha-2}} \\
&= \|v'\|_{\mathscr{L}_T^\beta} \|\eta\|_{C_T \mathscr{C}^{\alpha-2}}^2, \\
\|C(v', \mathcal{I}\eta, \eta)\|_{C_T \mathscr{C}^{2\alpha+\beta-2}} &\overset{\beta+2\alpha-2>0}{\lesssim} \|v'\|_{C_T \mathscr{C}^\beta} \|\mathcal{I}\eta\|_{C_T \mathscr{C}^\alpha} \|\eta\|_{C_T \mathscr{C}^{\alpha-2}}, \\
\|v'(\mathcal{I}\eta \odot \eta)\|_{C_T \mathscr{C}^{2\alpha-2}} &\overset{\beta+2\alpha-2>0}{\lesssim} \|v'\|_{C_T \mathscr{C}^\beta} \|\mathcal{I}\eta \odot \eta\|_{C_T \mathscr{C}^{2\alpha-2}},
\end{aligned}$$

The term with the lowest regularity is $v \otimes \eta \in C_T \mathscr{C}^{\alpha-2}$, and if we subtract it from the product the remaining terms all have at least regularity $\alpha + \beta - 2$. Therefore, the first claimed estimate follows.

The estimate for the difference of the two products follows from the same arguments by using the bi- or tri-linearity of all the operators. □

To recap, we defined a Banach space of paracontrolled functions $\mathscr{D}_T^\beta$ that can be decomposed into a paraproduct plus a more regular remainder term, and we showed that for such functions the right hand side of our equation is well defined and again given as a paraproduct plus a more regular remainder term. The last ingredient we need to set up a paracontrolled Picard iteration is a paracontrolled Schauder estimate. By the definition of the modified paraproduct this is quite easy, but as a preparatory step we need to extend our Schauder estimates from Theorem 4.6 to parabolic spaces:

**Theorem 5.13.** *Let $\gamma \in (0, 2]$. We have for all $\gamma' \in (0, \gamma]$:*

$$\begin{aligned}
\|(t \mapsto P_t \varphi)\|_{\mathscr{L}_T^\gamma} &\lesssim \|\varphi\|_\gamma, \\
\left\| \left( t \mapsto \int_0^t P_{t-s} f(s) ds \right) \right\|_{\mathscr{L}_T^{\gamma'}} &\lesssim T^{\frac{\gamma-\gamma'}{2}} (T \vee 1) \|f\|_{C_T \mathscr{C}^{\gamma-2}}.
\end{aligned}$$

**Proof. (Voluntary, not part of the lecture videos)** The space regularity is controlled in Theorem 4.6. Therefore, we only have to control the time regularity.

1. Recall from (4.6) that for $\gamma \in (0, 2]$

$$\|(P_t - \mathrm{id}) u\|_{L^\infty} \lesssim (1 \wedge t^{\frac{\gamma}{2}}) \|u\|_\gamma.$$

   This gives

$$\|P_t \varphi - P_s \varphi\|_{L^\infty} = \|(P_{t-s} - \mathrm{id}) P_s \varphi\|_{L^\infty} \lesssim (1 \wedge |t-s|^{\gamma/2}) \|\varphi\|_\gamma,$$

   and thus

$$\|(t \mapsto P_t \varphi)\|_{C_T^\gamma L^\infty} \lesssim \|\varphi\|_\gamma.$$

2. For the space-time convolution, it suffices to show the bound

$$\left\| \left( t \mapsto \int_0^t P_{t-s} f(s) ds \right) \right\|_{C_T^{\gamma/2} L^\infty} \lesssim (1 \vee T) \| f \|_{C_T \mathscr{C}^\alpha}. \tag{5.3}$$

Indeed, the estimate for the $C_T^{\gamma'/2} L^\infty$ norm then follows from $|t-s|^{(\gamma-\gamma')/2} \leqslant T^{(\gamma-\gamma')/2}$ for $s, t \in [0, T]$. The bound in (5.3) is shown as follows:

$$\begin{aligned}\left\| \int_0^t P_{t-r} f(r) dr - \int_0^s P_{s-r} f(r) dr \right\|_{L^\infty} \leqslant & \left\| \int_s^t P_{t-r} f(r) dr \right\|_{L^\infty} \\ & + \left\| (P_{t-s} - \mathrm{id}) \int_0^s P_{s-r} f(r) dr \right\|_{L^\infty},\end{aligned}$$

and the second term on the right hand side is controlled by (4.6) together with the Schauder estimates from Theorem 4.6:

$$\begin{aligned}\left\| (P_{t-s} - \mathrm{id}) \int_0^s P_{s-r} f(r) dr \right\|_{L^\infty} & \overset{(4.6)}{\lesssim} |t-s|^{\gamma/2} \left\| \int_0^s P_{s-r} f(r) dr \right\|_\gamma \\ & \overset{\text{Thm. } 4.6}{\lesssim} |t-s|^{\gamma/2} (T \vee 1) \| f \|_{C_T \mathscr{C}^{\gamma-2}}.\end{aligned}$$

To deal with the remaining term, we assume first that $|t-s| \leqslant 1$. In that case we decompose it into high and low frequencies and we obtain for $j_0 \geqslant -1$ such that $2^{-j_0} \simeq |t-s|^{1/2}$:

$$\begin{aligned}& \left\| \int_s^t P_{t-r} f(r) dr \right\|_{L^\infty} \\ & \leqslant \sum_{j \leqslant j_0} \int_s^t \| P_{t-r} \Delta_j f(r) \|_{L^\infty} dr + \left\| \Delta_{> j_0} \int_s^t P_{t-r} f(r) dr \right\|_{L^\infty} \\ & \lesssim \sum_{j \leqslant j_0} |t-s| 2^{-j(\gamma-2)} \| f \|_{C_T \mathscr{C}^{\gamma-2}} + 2^{-j_0 \gamma} \left\| \int_s^t P_{t-r} f(r) dr \right\|_\gamma \\ & \overset{\text{Thm. } 4.6}{\lesssim} (|t-s| 2^{-j_0(\gamma-2)} + 2^{-j_0 \gamma} (|t-s| \vee 1)) \| f \|_{C_T \mathscr{C}^{\gamma-2}} \\ & \simeq |t-s|^{\gamma/2} \| f \|_{C_T \mathscr{C}^\alpha}.\end{aligned}$$

If $|t-s| > 1$, we simply bound

$$\left\| \int_s^t P_{t-r} f(r) dr \right\|_{L^\infty} \lesssim (1 \vee |t-s|) \| f \|_{C_T \mathscr{C}^{\gamma-2}} \lesssim |t-s|^{\gamma/2} (T \vee 1) \| f \|_{C_T \mathscr{C}^{\gamma-2}}.$$

□

**Proposition 5.14.** **("Paracontrolled Schauder estimate")** *Let $\alpha \in (\frac{2}{3}, 1)$ and $\beta \in (\frac{2-\alpha}{2}, \alpha]$. Then we have for all $v, v'$ such that the right hand side is finite:*

$$\| v - v' \prec\!\!\prec \mathcal{I}\eta \|_{\mathscr{L}_T^{2\beta}} \lesssim \| v(0) \|_{2\beta} + T^{(\alpha-\beta)/2} (1 \vee T) \| \mathscr{L} v - v' \circledcirc \eta \|_{C_T \mathscr{C}^{\alpha+\beta-2}}.$$

**Proof.** We have by definition of the modified paraproduct

$$\mathscr{L}(v - v' \prec\!\!\prec \mathcal{I}\eta) = \mathscr{L} v - v' \circledcirc \eta, \qquad (v - v' \prec\!\!\prec \mathcal{I}\eta)(0) = v(0),$$

so the claim follows from the Schauder estimates in parabolic spaces, Theorem 5.13. □

Question: Let $(\mathcal{X}, d)$ be a complete metric space and let $\Psi: \mathcal{X} \to \mathcal{X}$ be such that for some $k \in \mathbb{N}$ the map $\Psi^k = \underbrace{\Psi \circ \cdots \circ \Psi}_{k \text{ times}}$ is a contraction. Show that $\Psi$ has a unique fixed point.

Finally, we can set up a Picard iteration in a space of paracontrolled distributions:

**Theorem 5.15.** *Let $T>0$, $\alpha\in(\frac{2}{3},1)$, $\beta\in(\frac{2-\alpha}{2},\alpha)$, $\mathbb{Z}\in\mathcal{T}^\alpha$, and $v_0\in\mathscr{C}^{2\beta}$. Then for all $T>0$ there exists a unique $v\in\mathscr{D}_T^\beta(\mathbb{Z})$ such that*

$$v(t)=P_t v_0+\int_0^t P_{t-s}(v\eta)(s)ds,\qquad t\in[0,T].$$

*Moreover, $v$ depends continuously on $(v_0,\mathbb{Z})$.*

**Proof.** Throughout the proof, $M_T$ denotes a changing constant which depends polynomially on $\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}$ and which is increasing in $T$, but which is independent of $v_0$. We show below that the map

$$\Psi\colon\mathscr{D}_T^\beta\to\mathscr{D}_T^\beta,\qquad \Psi(v,v',v^\sharp):=(\Psi v,(\Psi v)',(\Psi v)^\sharp):=(w,w',w^\sharp),$$

where

$$\begin{aligned}w(t)&:=P_t v_0+\int_0^t P_{t-s}(v(s)\eta(s))ds,\qquad t\in[0,T]\\ w'&:=v,\\ w^\sharp&:=w-w'\prec\!\!\prec\mathcal{I}\eta,\end{aligned}$$

is well defined and that for $T>0$ sufficiently small (depending only on $\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}$ but not on $v_0$), $\Psi^2=\Psi\circ\Psi$ is a contraction. Then we show that this implies that $\Psi$ has a unique fixed point. In that way we obtain a unique solution on a small time interval, and as in the proof of Theorem 4.10 we can then extend the solution to $[T,2T]$, $[2T,3T]$, and so on.

1. We have by the estimate for the paracontrolled product in Lemma 5.12

   $$\|\mathscr{L}w-v\circledcirc\eta\|_{C_T\mathscr{C}^{\alpha+\beta-2}}=\|v\eta-v\circledcirc\eta\|_{C_T\mathscr{C}^{\alpha+\beta-2}}\leqslant M_T\|v\|_{\mathscr{D}_T^\beta}.$$

   Thus, the paracontrolled Schauder estimates from Proposition 5.14 give

   $$\|w^\sharp\|_{\mathscr{L}_T^{2\beta}}\leqslant M_T\big(\|v_0\|_{2\beta}+T^{(\alpha-\beta)/2}\|v\|_{\mathscr{D}_T^\beta}\big).\tag{5.4}$$

   Note that $\|v_0\|_{2\beta}$ does not come with a small factor, but this term drops out if we compare $\Psi v-\Psi\tilde{v}$.

2. Again by Lemma 5.12

   $$\|\mathscr{L}w\|_{C_T\mathscr{C}^{\alpha-2}}=\|v\eta\|_{C_T\mathscr{C}^{\alpha-2}}\leqslant M_T\|v\|_{\mathscr{D}_T^\beta},$$

   so the Schauder estimates in parabolic spaces from Theorem 5.13 give

   $$\|w\|_{\mathscr{L}_T^\beta}\leqslant M_T\big(\|v_0\|_{2\beta}+T^{(\alpha-\beta)/2}\|v\|_{\mathscr{D}_T^\beta}\big).\tag{5.5}$$

3. The estimates (5.4) and (5.5) both come with a factor $T^{(\alpha-\beta)/2}$, so for small $T$, with the maximal size of $T$ depending depending on $M_T$ but not on $v_0$, we get

   $$\|\Psi v-\Psi\tilde{v}\|_{\mathscr{L}_T^\beta}+\|(\Psi v)^\sharp-(\Psi\tilde{v})^\sharp\|_{\mathscr{L}_T^{2\beta}}\leqslant\frac{1}{2}\|v-\tilde{v}\|_{\mathscr{D}_T^\beta}.$$

4. It remains to bound the derivative $w'=v$: We have

   $$\|w'\|_{\mathscr{L}_T^\beta}=\|v\|_{\mathscr{L}_T^\beta},\tag{5.6}$$

   so $w'$ has the right regularity. But unlike for $w$ and $w^\sharp$, we do not gain a small factor and thus $\Psi$ is not a contraction.

5. Therefore, we consider $\Psi^2 = \Psi \circ \Psi$ instead. We first apply (5.4) and then (5.4), (5.5), (5.6) to obtain

$$\begin{aligned}\|(\Psi^2 v)^\sharp - (\Psi^2 \tilde{v})^\sharp\|_{\mathscr{L}_T^{2\beta}} &\leqslant T^{(\alpha-\beta)/2} M_T \|(\Psi v, \Psi v', \Psi v^\sharp) - (\Psi \tilde{v}, \Psi \tilde{v}', \Psi \tilde{v}^\sharp)\|_{\mathscr{D}_T^\beta} \\ &\leqslant T^{(\alpha-\beta)/2} M_T \|v - \tilde{v}\|_{\mathscr{D}_T^\beta},\end{aligned}$$

and similarly we get from (5.5) and then (5.4), (5.5), (5.6):

$$\|(\Psi^2 v) - (\Psi^2 \tilde{v})\|_{\mathscr{L}_T^\beta} \leqslant T^{(\alpha-\beta)/2} M_T \|v - \tilde{v}\|_{\mathscr{D}_T^\beta}.$$

For the derivative, we now get by applying (5.6) and then (5.5):

$$\begin{aligned}\|(\Psi^2 v)' - (\Psi^2 \tilde{v})'\|_{\mathscr{L}_T^\beta} &= \|\Psi v - \Psi \tilde{v}\|_{\mathscr{L}_T^\beta} \\ &\leqslant T^{(\alpha-\beta)/2} M_T \|v - \tilde{v}\|_{\mathscr{D}_T^\beta},\end{aligned}$$

which proves the contraction property of $\Psi^2$ (for small $T > 0$).

6. By Banach's fixed point theorem, $\Psi^2$ has a unique fixed point, and it only remains to show that also $\Psi$ has a unique fixed point. Uniqueness is clear, because any fixed point for $\Psi$ is one for $\Psi^2$ as well. So it suffices to show that the fixed point $v$ of $\Psi^2$ is also a fixed point of $\Psi$. Let us write $w := \Psi v$. Then

$$\Psi w = \Psi^2 v = v,$$

and thus

$$\Psi^2 w = \Psi(\Psi w) = \Psi v = w.$$

Hence, $w$ is a fixed point of $\Psi^2$, and by uniqueness we have $w = v$, i.e. $\Psi v = v$. □

**Remark 5.16.** This concludes the lecture taught at FU Berlin in summer 2020. The following material is included for interested readers, but it is not relevant for the course.

After we solved the linearized equation, solving the full $\Phi_3^4$ equation is not much more difficult, although the notation gets more tedious, see the next subsection. Since the $\Phi_3^4$ equation is non-linear, we can only solve it on a small time interval. To construct it for all times, we need to make use of the negative sign of the nonlinearity, which played no role in our estimations so far. In that way we can obtain very strong estimates for the solution which are way better than estimates for linear equations, see for example [39].

Section 6 solves the two-dimensional parabolic Anderson model for Dirac delta initial conditions, presents the relation with the Anderson Hamiltonian, and studies its long time behavior on the torus.

Section 7 presents a link between paracontrolled distributions and Hairer's regularity structures.

Section 8 presents some results on nonlinear stochastic wave equations.

## 5.4 The full $\Phi_3^4$ equation

To study the full $\Phi_3^4$ equation we first need to define a space of extended data where all analytically ill-defined trees live. To simplify the renormalization, we replace the operator $\Delta$ in the equation by $A = \Delta - 1$. This has the advantage that now the semigroup $(e^{tA} = e^{-t} P_t)_{t \geqslant 0}$ is integrable over all of $\mathbb{R}_+$. Of course, we can recover the original equation by considering $(\partial_t - A)\phi = -\phi^3 + \phi + \xi$ instead, but as discussed before the linear term $+\phi$ on the right hand side poses no additional difficulty for our small scale solution theory and therefore we simply omit it. On the other hand this term might have a strong effect on the long time behavior of the solution, but at least for now we do not care about that.

In the following we write for a Banach space $X$

$$C_{\mathbb{R}}X = C(\mathbb{R}, X) \qquad \text{and} \qquad C_{\mathrm{pg}}(X) = \{u \in C_{\mathbb{R}}X : \exists k \in \mathbb{N}\, s.t.\ \|u(t)\|_X \lesssim 1 + |t|^k\},$$

equipped with the distance

$$d(u,v)_{C_{\mathbb{R}}X} = \sum_{n=1}^{\infty} 2^{-n}(\|u|_{[-n,n]} - v|_{[-n,n]}\|_{C([-n,n],X)} \wedge 1)$$

under which $C_{\mathbb{R}}X$ is complete (of course the subspace $C_{\mathrm{pg}}X$ of polynomially growing functions is not closed). By adapting the proof of Schauder 4.6 is not hard to see that for $u \in C_{\mathrm{pg}}\mathscr{C}^{\alpha}$ we have

$$t \mapsto \int_{-\infty}^{t} e^{(t-s)A} u(s) \mathrm{d}s \in C_{\mathrm{pg}}\mathscr{C}^{\alpha+2}.$$

**Definition 5.17.** *Let $\alpha \in (1/3, 1/2)$ and let*

$$\mathcal{T}^{\alpha} \subset C_{\mathbb{R}}\mathscr{C}^{\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-2} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha} \times C_{\mathbb{R}}\mathscr{C}^{\alpha} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-1},$$

*be the closure of the image of the map*

$$\begin{gathered}\Theta : C_{\mathrm{pg}}L^{\infty} \times \mathbb{R} \times \mathbb{R} \\ \to C_{\mathbb{R}}\mathscr{C}^{\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-2} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha} \times C_{\mathbb{R}}\mathscr{C}^{\alpha} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{\alpha-1} \times C_{\mathbb{R}}\mathscr{C}^{2\alpha-1}, \\ \Theta(Z, c_1, c_2) = \left(\uparrow, \vee, Y, \Psi, \Psi\!\uparrow, \Psi\!\vee, Y\!\vee\right),\end{gathered}$$

*where*

$$\begin{aligned}
\uparrow(t) &= Z(t), \\
\vee(t) &= (\uparrow(t)^2 - c_1), \\
Y(t) &= \int_{-\infty}^{t} e^{(t-s)A} \vee(s) \mathrm{d}s, \\
\Psi(t) &= \int_{-\infty}^{t} e^{(t-s)A} (\uparrow(s)^3 - 3c_1 \uparrow(s)) \mathrm{d}s, \\
\Psi\!\uparrow(t) &= \Psi(t) \odot \uparrow(t), \\
\Psi\!\vee(t) &= \Psi(t) \odot \vee(t) - 6c_2 \uparrow(t), \\
Y\!\vee(t) &= \left(\int_{-\infty}^{t} e^{(t-s)A} \vee(s) \mathrm{d}s\right) \odot \vee(t) - 2c_2.
\end{aligned}$$

*We write $\mathcal{T} = \bigcup_{\alpha \in (1/3,1/2)} \mathcal{T}^{\alpha}$. We also write $\|\mathbb{Z}\|_{\mathcal{T}_T^{\alpha}}$ $\|\mathbb{Z} - \tilde{\mathbb{Z}}\|_{\mathcal{T}_T^{\alpha}}$ for the canonical norm on the product space $\mathcal{T}^{\alpha}|_{[0,T]}$ of functions in $\mathcal{T}^{\alpha}$ restricted to the time interval $[0,T]$.*

**Remark 5.18.** We did not use the canonical regularities that we would guess from paraproduct and Schauder estimates. It would be more natural to take

$$\Psi \in C_{\mathbb{R}}\mathscr{C}^{3\alpha-1}, \qquad \Psi\!\uparrow \in C_{\mathbb{R}}\mathscr{C}^{4\alpha-2}, \qquad \Psi\!\vee \in C_{\mathbb{R}}\mathscr{C}^{5\alpha-3}, \qquad Y\!\vee \in C_{\mathbb{R}}\mathscr{C}^{4\alpha-2}.$$

But since we are interested in the case $\alpha = 1/2 - \varepsilon$ and to simplify the presentation we formally identify $2\alpha + k\alpha = 1 + k\alpha$ whenever $k > 0$.

Our aim is now to solve the $\Phi_3^4$ equation for given data $\mathbb{Z} \in \mathcal{T}$. For that purpose we define a Banach space of distributions, depending on $\tau$, in which we can use paracontrolled arguments to make sense of the equation and in which we can set up a Picard iteration:

**Definition 5.19.** *Let $\alpha \in (1/3, 1/2)$, $\beta \in (1/3, \alpha]$, $T > 0$, and $\mathbb{Z} \in \mathcal{T}^\alpha$. We say that*

$$(\phi, v', v^\sharp) \in C_T\mathscr{C}^{\alpha-1} \times C_T\mathscr{C}^{\beta} \times C_T\mathscr{C}^{\beta+1}$$

*is paracontrolled by $\mathbb{Z}$ and write $(\phi, v', v^\sharp) \in \mathcal{D}^\beta_T\mathbb{Z}$ if*

$$\phi = \RSI - \RSW + v, \qquad v = v' \prec\!\!\prec \RSV + v^\sharp.$$

*If there is no ambiguity about $v'$ and $v^\sharp$, we also write $v \in \mathcal{D}^\beta_T\mathbb{Z}$ or $\phi \in \mathcal{D}^\beta_T\mathbb{Z}$.*

**Definition 5.20.** *For $(\phi, v', v^\sharp) \in \mathcal{D}^\beta_T\mathbb{Z}$ we define the space-time distribution*

$$-\phi^{:3:} := -(\partial_t - A)\RSW - 3\Big(v - \RSW\Big)\RSVee + \tau_0 + \tau_1 v + \tau_2 v^2 - v^3$$

*where the coefficients $\tau_0, \tau_1, \tau_2 \in C_{\mathbb{R}}\mathscr{C}^{\alpha-1}$ are defined as*

$$\tau_0 := -3\RSI \times \RSW^2 + \RSW^3 := -3\Big(\RSI \varolessthan \RSW^2 + \RSI \varogreaterthan \RSW^2 + \RSI \odot \Big(\RSW \odot \RSW\Big) + 2C\Big(\RSW, \RSW, \RSI\Big) + \RSV\RSWVo\Big) + \RSW^3,$$

$$\tau_1 := 6\RSWVee - 3\RSW^2 := 6\RSWVo + 6\RSW \varolessthan \RSI + 6\RSW \varogreaterthan \RSI - 3\RSW^2,$$

$$\tau_2 := -3\RSI + 3\RSW,$$

*and where*

$$\begin{aligned}(v + \RSW)\RSVee := {} & (v + \RSW)\varolessthan\RSVee + (v + \RSW)\varogreaterthan\RSVee + v^\sharp \odot \RSVee + (v' \prec\!\!\prec \RSV - v' \varolessthan \RSV) \odot \RSVee \\ & + C(v', \RSV, \RSVee) + v'\RSVVo + \RSWVo.\end{aligned}$$

The following lemma gives a more intuitive representation of $\phi^{:3:}$ in the case of regular data $\mathbb{Z}$:

**Lemma 5.21.** *If $\mathbb{Z} = \Theta(Z, c_1, c_2)$ for $Z \in C_{\mathrm{pg}}L^\infty$, then*

$$-\phi^{:3:} = -\phi^3 + 3(c_1 + c_2)\phi.$$

**Proof.** If $c_1 = c_2 = 0$, then this immediately follows from the considerations at the beginning of this section, so we only have to keep track where $c_1$ and $c_2$ appear. This can be done in a lengthy but straightforward computation, noting that now all products are well defined and we can combine all the paraproducts and commutators etc. to form usual products. $\square$

While $(\partial_t - A)\RSW$ is not necessarily a function of time with values in a space of distributions, the renormalized cube $\phi^{:3:}$ is indeed a function of time once we subtract its most singular contribution:

**Lemma 5.22.** *Let $\alpha \in (1/3, 1/2)$, $\beta \in (1/3, \alpha]$, $T > 0$, and $\mathbb{Z}, \tilde{\mathbb{Z}} \in \mathcal{T}^\alpha$ as well as $(\phi, v', v^\sharp) \in \mathcal{D}^\beta_T\mathbb{Z}$ and $(\tilde{\phi}, \tilde{v}', \tilde{v}^\sharp) \in \mathcal{D}^\beta_T\tilde{\mathbb{Z}}$. Then*

$$\begin{aligned}& \Big\|\phi^{:3:} - (\partial_t - A)\RSW\Big\|_{C_T\mathscr{C}^{2\alpha-2}} + \Big\|\phi^{:3:} - (\partial_t - A)\RSW - 3\Big(v - \RSW\Big)\varolessthan\RSVee\Big\|_{C_T\mathscr{C}^{\alpha-1}} \\ & \leqslant P(\|\mathbb{Z}\|_{\mathcal{T}^\alpha_T})(1 + \|v'\|^3_{C_T\mathscr{C}^\beta} + \|v^\sharp\|^3_{C_T\mathscr{C}^{\beta+1}})\end{aligned}$$

*for a polynomial $P$, and also*

$$\begin{aligned}& \Big\|\phi^{:3:} - (\partial_t - A)\RSW - \Big(\tilde{\phi}^{:3:} - (\partial_t - A)\tilde{\RSW}\Big)\Big\|_{C_T\mathscr{C}^{2\alpha-2}} \\ & \quad + \Big\|\phi^{:3:} - (\partial_t - A)\RSW - 3\Big(v - \RSW\Big)\varolessthan\RSVee - \Big(\tilde{\phi}^{:3:} - (\partial_t - A)\tilde{\RSW} - 3\Big(\tilde{v} - \tilde{\RSW}\Big)\varolessthan\tilde{\RSVee}\Big)\Big\|_{C_T\mathscr{C}^{\alpha-1}} \\ & \leqslant M_T(\|\mathbb{Z} - \tilde{\mathbb{Z}}\|_{\mathcal{T}^\alpha_T} + \|v' - \tilde{v}'\|_{C_T\mathscr{C}^\beta} + \|v^\sharp - \tilde{v}^\sharp\|_{C_T\mathscr{C}^{\beta+1}}),\end{aligned}$$

*where for another polynomial $P$*

$$M_T = P(\|\mathbb{Z}\|_{\mathcal{T}_T^\alpha}, \|\tilde{\mathbb{Z}}\|_{\mathcal{T}_T^\alpha}, \|v'\|_{C_T\mathscr{C}^{\beta+1}}, \|v^\sharp\|_{C_T\mathscr{C}^\beta}, \|\tilde{v}'\|_{C_T\mathscr{C}^\beta}, \|\tilde{v}^\sharp\|_{C_T\mathscr{C}^{\beta+1}}).$$

**Proof.** This easily follows by combining the Definition 5.20 of $\phi^{:3:}$ with the paraproduct estimates from Theorem 3.19 and the commutator estimate from Lemma 5.2. □

Now consider the equation

$$(\partial_t - A)\phi = -\phi^{:3:} + \xi,$$

where we wrote

$$\xi := (\partial_t - A)\,\text{¡}\,,$$

which is a space-time distribution. We are looking for paracontrolled solutions, so we should decompose $\phi = \text{¡} - \Psi + v$, and the equation for $v$ is

$$(\partial_t - A)v = -\phi^{:3:} + (\partial_t - A)\Psi, \qquad v(0) = \phi_0 - \text{¡}\,(0) + \Psi(0),$$

which we can control with Lemma 5.22. From here we can set up a Picard iteration and solve the equation uniquely on a small time interval, just as in the case of the linear equation of Section 5.3. Unlike in the linear case, now the length of the time interval on which we obtain a contraction depends on the initial condition, and therefore we only obtain local existence up to an explosion time. To show that the explosion time is infinite, we have to use the sign of the nonlinearity $-\phi^{:3:}$ and apply more refined estimates, see [39].

# 6 Parabolic Anderson model & Anderson Hamiltonian

## 6.1 The parabolic Anderson model

Let now $\xi$ be a space white noise on $\mathbb{T}^2$, i.e. a centered Gaussian process with values in $\mathscr{S}'$ such that for all $\varphi, \psi \in C^\infty(\mathbb{T}^2)$ we have $\mathbb{E}[\xi(\varphi)\xi(\psi)] = \langle \varphi, \psi \rangle_{L^2(\mathbb{T}^2)}$. We want to study the parabolic Anderson model (PAM)

$$\mathscr{L}u = (\partial_t - \Delta)u = u\xi,$$

which is a continuum model for a branching population in a random potential: We consider independent diffusing particles on $\mathbb{T}^2$ that in the point $x$ branch with rate $\xi(x)^+ = \max\{\xi(x), 0\}$ and get killed with rate $\xi(x)^- = \max\{-\xi(x), 0\}$. Of course $\xi(x)$ does not make any sense because $\xi$ is only a distribution, and also the solution $u$ is not integer valued; but we can derive $u$ as a continuum limit of a discrete model behaving as described above [36].

**Exercise 6.1.** Show that if $\eta$ is a space white noise on $\mathbb{T}^d$, then

$$\mathbb{E}[\|\eta\|_{-d/2-\kappa}^p] < \infty$$

for all $\kappa, p > 0$.

*Hint: Compare with Lemma 4.3.*

Thus we have $\xi \in \mathscr{C}^{-1-\kappa}$ for all $\kappa > 0$, which is the same regularity that we had for the tree $\vee$ in the linearized $\Phi_3^4$ equation. In other words, the parabolic Anderson model is a simplified linearized $\Phi_3^4$ equation and we can solve it by using the same arguments as in Section 5.3. Since $\xi$ does not depend on time we can now work with "extended data" that does not depend on time either: Let $\alpha \in (2/3, 1)$ so that $\xi \in \mathscr{C}^{\alpha-2}$, and consider

$$X = (1 - \Delta)^{-1}\xi \in \mathscr{C}^\alpha.$$

Then

$$\mathscr{L}X = -\Delta X = (1-\Delta)X - X = \xi - X = \xi + \mathscr{C}^{\alpha}.$$

**Exercise 6.2.** Let $\rho: \mathbb{R}^2 \to \mathbb{R}$ be an even, compactly supported function which is continuous in 0 and which satisfies $\rho(0) = 1$. Define

$$\xi^{\varepsilon} := \mathscr{F}^{-1}(\rho(\varepsilon \cdot)\mathscr{F}\xi), \qquad X^{\varepsilon} := (1-\Delta)^{-1}\xi^{\varepsilon} = \mathscr{F}^{-1}(\rho(\varepsilon \cdot)\mathscr{F}X),$$

as well as

$$c^{\varepsilon} := \mathbb{E}[X^{\varepsilon}(0)\xi^{\varepsilon}(0)].$$

Show that for all $x \in \mathbb{T}^2$

$$c^{\varepsilon} = \mathbb{E}[X^{\varepsilon}(x)\xi^{\varepsilon}(x)] = \mathbb{E}[X^{\varepsilon} \odot \xi^{\varepsilon}(x)] = O(|\log \varepsilon|)$$

and that there exists $X \diamond \xi$ such that for all $p \in [1, \infty)$

$$\lim_{\varepsilon \to 0} \mathbb{E}[\|\xi^{\varepsilon} - \xi\|_{\alpha-2}^p + \|X^{\varepsilon} - X\|_{\alpha}^p + \|X \diamond \xi - (X^{\varepsilon} \odot \xi^{\varepsilon} - c^{\varepsilon})\|_{2\alpha-2}^p] = 0.$$

*Hint: Use the tools from Section 4.5.*

From here it is not difficult to slightly adapt the arguments from Section 5.3 to show that for all $\beta \in (2/3, \alpha]$ and for all $u_0$ with $u_0 - u_0 \circledcirc X \in \mathscr{C}^{\alpha+\beta}$ there exists a unique paracontrolled solution $u = u \prec\!\!\prec X + u^{\sharp}$ with $u \in \mathscr{L}_T^{\beta}$ and $u^{\sharp} \in \mathscr{L}^{\alpha+\beta}$ to

$$\mathscr{L}u = u\xi := u \circledcirc \xi + u \circledcirc \xi + u^{\sharp} \odot \xi + (u \prec\!\!\prec X - u \circledcirc X) \odot \xi + C(u, X, \xi) + u(X \diamond \xi)$$

with initial condition $u(0) = u_0$. Note that unlike before we do not have $\mathscr{L}(u \prec\!\!\prec X) = u \circledcirc \xi$, but instead

$$\mathscr{L}(u \prec\!\!\prec X) = u \circledcirc \mathscr{L}X = u \circledcirc \xi - u \circledcirc X,$$

but the term $u \circledcirc X$ has positive regularity and therefore $\int_0^t P_{t-s}(u \circledcirc X) \mathrm{d}s \in \mathscr{L}_T^{\alpha+\beta}$.

Moreover, since $X \diamond \xi = \lim_{\varepsilon} (X^{\varepsilon} \odot \xi^{\varepsilon} - c^{\varepsilon})$, we have $u = \lim_{\varepsilon} u^{\varepsilon}$, where

$$\mathscr{L}u^{\varepsilon} = u^{\varepsilon}(\xi^{\varepsilon} - c^{\varepsilon}).$$

(It is a good exercise to convince yourself of this! Where does the $c^{\varepsilon}$ enter the equation?)

Our aim is now to analyze this equation in a bit more detail. We will first extend the solution theory to much more general initial conditions, then we will present a strong maximum principle, then we will study the Anderson Hamiltonian, i.e. the infinitesimal generator of the solution semigroup, and finally we will combine all these tools to obtain a quite precise understanding of the long time behavior of the (periodic) parabolic Anderson model.

## 6.2 General Besov spaces and more general initial conditions

As before, the condition on the initial condition is quite unnatural. A canonical initial condition for our population would be the Dirac delta, which would model the start from a unit mass at 0, and then we could explore how this mass diffuses through the system. By Exercise 3.1 we have $\delta \in B_{p,\infty}^{-2(1-1/p)}$ for all $p \in [1, \infty)$ (since we redefined the space $\mathscr{C}^{\gamma}$ as the closure of the smooth functions we only have $\delta \in \mathscr{C}^{-2-\kappa}$ for $\kappa > 0$).

Our estimate $\|P_t \varphi\|_{\beta+\gamma} \lesssim (1 + t^{-\gamma/2})\|\varphi\|_{\beta}$ can be easily generalized to

$$\|P_t \varphi\|_{B_{p,q}^{\beta+\gamma}} \lesssim (1 + t^{-\gamma/2})\|\varphi\|_{B_{p,q}^{\beta}}.$$

To obtain an integrable singularity at $t=0$ we need $\gamma<2$, so since the initial condition for the paracontrolled remainder would be $u^\sharp(0)=\delta-\delta\varolessthan X\in\mathscr{C}^{-2-\kappa}$ we could at best obtain

$$u^\sharp(t)=\underbrace{P_t u^\sharp(0)}_{\mathscr{C}^{-2\kappa}}+\underbrace{\int_0^t P_{t-s}\mathscr{L}u^\sharp(s)\mathrm{d}s}_{??}\in\mathscr{C}^{-2\kappa}.$$

Of course, this is way too irregular to make sense of $u^\sharp(t)\odot\xi$. If on the other hand we could work in the space $B_{1,\infty}$, then we would have $u^\sharp(0)\in B^0_{1,\infty}$ and then $P_t u^\sharp(0)\in B^{2-\kappa}_{1,\infty}$, which gives us some hope to make sense of $u^\sharp(t)\odot\xi$, because $\alpha+2-\kappa>0$.

This is the purpose of the current subsection: We develop the solution theory for paracontrolled distributions in general Besov spaces, and we finally give details on how to include singularities at $t=0$ which allow us to take less regular initial conditions. We start by defining the following function spaces:

**Definition 6.1.** *Let $p\in[1,\infty]$ and $\alpha\in\mathbb{R}$, $\beta\in(0,2)$, $\gamma\in[0,1)$. Then we define*

$$\mathscr{C}^\alpha_p:=B^\alpha_{p,\infty},\qquad \mathscr{L}^\beta_{p,T}:=C^{\beta/2}_T L^p\cap C_T\mathscr{C}^\beta_p,$$

*both equipped with their canonical norm, as well as*

$$\mathscr{M}^\gamma_T\mathscr{C}^\alpha_p:=\{f\in C([0,T],\mathscr{S}'):t\mapsto t^\gamma\varphi(t)\in C_T\mathscr{C}^\alpha_p\},$$
$$\mathscr{L}^{\gamma,\beta}_{p,T}:=\{\varphi\in C([0,T],\mathscr{S}'):t\mapsto t^\gamma\varphi(t)\in\mathscr{L}^\beta_{p,T}\},$$

*with canonical norms*

$$\|\varphi\|_{\mathscr{M}^\gamma_T\mathscr{C}^\alpha_p}:=\|(t\mapsto t^\gamma\varphi(t))\|_{C_T\mathscr{C}^\alpha_p},\qquad \|\varphi\|_{\mathscr{L}^{\gamma,\beta}_{p,T}}:=\|(t\mapsto t^\gamma\varphi(t))\|_{\mathscr{L}^\beta_{p,T}}.$$

Now we need to translate the ingredients for paracontrolled distributions to this new functional setting. Most of this was worked out by Prömel and Trabs in [47] in more generality than we need it here, and the following lemma is a collection of weaker versions of their results:

**Lemma 6.2. ([47], Lemma 2.1)** *Let $p\in[1,\infty]$ and let $\beta\in\mathbb{R}$ and $u,v\in\mathscr{S}'$. Then we have for all $\alpha>0$*

$$\|u\varolessthan v\|_{\mathscr{C}^\beta_p}\lesssim\min\left\{\|u\|_{L^p}\|v\|_\beta,\|u\|_{L^\infty}\|v\|_{\mathscr{C}^\beta_p}\right\},$$

*and for $\alpha<0$ furthermore*

$$\|u\varolessthan v\|_{\mathscr{C}^{\alpha+\beta}_p}\lesssim\min\left\{\|u\|_{\mathscr{C}^\alpha_p}\|v\|_\beta,\|u\|_\alpha\|v\|_{\mathscr{C}^\beta_p}\right\}.$$

*If $\alpha+\beta>0$ we have*

$$\|u\odot v\|_{\mathscr{C}^{\alpha+\beta}_p}\lesssim\min\left\{\|u\|_{\mathscr{C}^\alpha_p}\|v\|_\beta,\|u\|_\alpha\|v\|_{\mathscr{C}^\beta_p}\right\}.$$

*If $\alpha\in(0,1)$ and $\gamma\in\mathbb{R}$ is such that $\beta+\gamma<0$ but $\alpha+\beta+\gamma>0$, then*

$$\|C(u,v,w)\|_{\mathscr{C}^{\alpha+\beta+\gamma}_p}\lesssim\|u\|_{\mathscr{C}^\alpha_p}\|v\|_\beta\|w\|_\gamma.$$

We also need Schauder estimates in $\mathscr{L}^{\gamma,\alpha}_{p,T}$ spaces:

**Lemma 6.3. ([26], Lemma 6.6)** *Let $\alpha \in (0,2)$ and $\gamma \in [0,1)$, as well as $p \in [1,\infty]$ and $T > 0$. Then we have*

$$\|(t \mapsto P_t u)\|_{\mathscr{L}_{p,T}^{\gamma,\alpha}} \lesssim \|u\|_{\mathscr{C}_p^{\alpha-2\gamma}},$$
$$\left\|\left(t \mapsto \int_0^t P_{t-s} u(s) \mathrm{d}s\right)\right\|_{\mathscr{L}_{p,T}^{\gamma,\alpha}} \lesssim (1+T)\|u\|_{\mathscr{M}_T^{\gamma}\mathscr{C}_p^{\alpha-2}}.$$

Adapting the proof of this lemma to deal with the modified paraproduct leads to the following generalization of Corollary 5.10

**Lemma 6.4.** *Let $\alpha \in (0,1)$, $\beta \in \mathbb{R}$, and $\gamma \in [0,1)$, as well as $p \in [1,\infty]$ and $T > 0$. Then*

$$\|u \prec\!\!\!\prec w - u \circledcirc w\|_{\mathscr{M}_T^{\gamma}\mathscr{C}_p^{\alpha+\beta}} \lesssim (1+T)\|u\|_{\mathscr{L}_{p,T}^{\gamma,\alpha}}(\|w(0)\|_{\beta} + \|\mathscr{L}w\|_{C_T\mathscr{C}^{\beta-2}}).$$

And finally we need an interpolation estimate:

**Lemma 6.5. ([26], Lemma 6.8)** *Let $\alpha \in (0,2)$, $\gamma \in (0,1)$, and $\kappa \in [0, \alpha \wedge 2\gamma)$, as well as $p \in [1,\infty]$ and $T > 0$. Then*

$$\|u\|_{\mathscr{L}_{p,T}^{\gamma-\kappa/2,\alpha-\kappa}} \lesssim \|u\|_{\mathscr{L}_{p,T}^{\gamma,\alpha}}.$$

We now fix $p \in [1,\infty]$ consider an initial condition $u_0 \in \mathscr{C}_p^0$. As before we let $\alpha \in (2/3, 1)$ and $\beta \in (2/3, \alpha]$ We define paracontrolled distributions in our new setting as follows:

**Definition 6.6.** *We say $(u, u', u^\sharp)$ is paracontrolled, $u \in \mathscr{D}_{p,T}^{\beta}$, if*

$$u = u' \prec\!\!\!\prec X + u^\sharp$$

*with*

$$u \in \mathscr{L}_{p,T}^{\beta/2,\beta}, \qquad u' \in \mathscr{L}_{p,T}^{\beta/2,\beta}, \qquad u^\sharp \in \mathscr{L}_{p,T}^{(\alpha+\beta)/2,\alpha+\beta}.$$

Now consider $u \in \mathscr{D}_{p,T}^{\alpha}$, and note that by the interpolation estimate we have $\mathscr{D}_{p,T}^{\beta} \subset \mathscr{D}_{p,T}^{\alpha}$. Thus we get from the estimates by Prömel and Trabs, using that $2\alpha + \beta > 2$,

$$\begin{aligned} u\xi = & \underbrace{u \circledcirc \xi}_{\mathscr{M}_T^{\alpha/2}\mathscr{C}_p^{2\alpha-2}} + \underbrace{u \circledcirc \xi}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{\alpha-2}} + \underbrace{u^\sharp \odot \xi}_{\mathscr{M}_T^{(\alpha+\beta)/2}\mathscr{C}_p^{2\alpha+\beta-2}} + \underbrace{(u' \prec\!\!\!\prec X - u' \circledcirc X) \odot \xi}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{2\alpha+\beta-2}} \\ & + \underbrace{C(u', X, \xi)}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{2\alpha+\beta-2}} + \underbrace{u'(X \diamond \xi)}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{2\alpha-2}}. \end{aligned}$$

Therefore,

$$u\xi - u \circledcirc \xi \in \mathscr{M}_T^{(\alpha+\beta)/2}\mathscr{C}_p^{2\alpha-2},$$

and then

$$\|\mathscr{L}u^\sharp\|_{\mathscr{M}_T^{\alpha}\mathscr{C}_p^{2\alpha-2}} \leqslant \|u \circledcirc X\|_{\mathscr{M}_T^{\alpha}\mathscr{C}_p^{2\alpha-2}} + \|u\xi - u \circledcirc \xi\|_{\mathscr{M}_T^{\alpha}\mathscr{C}_p^{2\alpha-2}} \lesssim T^{(\alpha-\beta)/2},$$

which gives us the small factor for the contraction. We also have

$$t \mapsto P_t u^\sharp(0) = P_t(u_0 - u_0 \prec\!\!\!\prec X) \in \mathscr{M}_T^{\alpha}\mathscr{C}_p^{2\alpha},$$

and thus we can control $u^\sharp \in \mathscr{L}_{p,T}^{\alpha,2\alpha}$. Moreover,

$$\|u\|_{\mathscr{L}_{p,T}^{\alpha/2,\alpha}} \lesssim \|u^\sharp\|_{\mathscr{L}_{p,T}^{\alpha,2\alpha}} + T^{(\alpha-\beta)/2}\|\mathscr{L}(u \prec\!\!\!\prec X)\|_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{\alpha-2}} + \|u_0 \circledcirc X\|_{\mathscr{C}_p^0},$$

and

$$\mathscr{L}(u \prec\!\!\prec X) = \underbrace{u \circledcirc \xi}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{\alpha-2}} - \underbrace{u \circledcirc X}_{\mathscr{M}_T^{\beta/2}\mathscr{C}_p^{\alpha}} .$$

We can now apply the same strategy as in Section 5.3 to obtain the following result:

**Theorem 6.7.** *Let $p \in [1, \infty]$, let $(\xi, X, X \diamond \xi) \in \mathscr{C}^{\alpha-2} \times \mathscr{C}^{\alpha} \times \mathscr{C}^{2\alpha-2}$, and let $u_0 \in \mathscr{C}_p^0$. Then for all $T > 0$ there exists a unique $u \in \mathscr{D}_T^{\alpha}$ such that*

$$\mathscr{L} u = u \xi, \qquad u(0) = u_0.$$

*Moreover, $u$ depends continuously on $(u_0, \xi, X, X \diamond \xi)$, and for all $t > 0$ we have $u(t) \in \mathscr{C}^{\alpha}$.*

**Proof.** Everything is clear by now, only the last point $u(t) \in \mathscr{C}^{\alpha}$ merits some discussion. By construction, have $u(s) \in \mathscr{C}_p^{\alpha}$ for all $s > 0$. So by Besov embedding (Lemma 3.15) we have for $p_1 > p$ with $\alpha - 2\left(\frac{1}{p} - \frac{1}{p_1}\right) = 0$:

$$u(t/2) \in \mathscr{C}_{p_1}^{\alpha - 2\left(\frac{1}{p} - \frac{1}{p_1}\right)} = \mathscr{C}_{p_1}^0.$$

From here we can bootstrap to increase the integrability to the $L^{\infty}$ scale: We now use $u(t/2)$ as new initial condition for the equation on $[t/2, 2t/3]$, so at $u(2t/3)$ we get an even better integrability $\mathscr{C}_{p_2}^0$ with $p_2 > p_1$ such that $\alpha - 2\left(\frac{1}{p_1} - \frac{1}{p_2}\right) = 0$. This bootstrapping ends after finally many steps, when we arrive at the $L^{\infty}$ scale. □

## 6.3 A strong maximum principle

We saw in the previous section that $u(t) \in \mathscr{C}^{\alpha}$ is a continuous function for all $t > 0$, even if the initial condition is only a Dirac delta $\delta \in \mathscr{C}_1^0$. In particular, it makes sense to speak of the sign of $u(t, x)$. By approximation it is not hard to see that whenever $u_0$ is a positive measure in $\mathscr{C}_p^0$ for some $p \in [1, \infty]$, then $u(t, x) \geqslant 0$ for all $t > 0$ and $x \in \mathbb{T}^2$. Here we present a nice argument due to Cannizzaro, Friz, Gassiat [11], inspired by Mueller [41], which shows that in fact the solution becomes instantly strictly positive, as long as $u_0$ can be approximated by positive functions.

We start with a simple observation about the heat kernel:

**Lemma 6.8.** **([11], (5.3))** *Let $p_t$ the Gaussian density on $\mathbb{R}^d$ with variance $2t$. For all $\rho > 0$ there exists $t_{\rho} > 0$ such that whenever $u \geqslant 0$ satisfies $u \geqslant 1$ on the ball $B(x, \kappa)$, then for all $t \in [0, t_{\rho}]$*

$$p_t * u(y) \geqslant \frac{1}{4}, \qquad y \in B(x, \kappa + t\rho).$$

**Proof.** Without loss of generality we take $x = 0$. For any $y \in B(x, \kappa)$ there exists a unique $|z| \leqslant 1$ with $y = z \cdot (\kappa + t\rho)$. So if $Z$ is a standard Gaussian variable, then

$$\begin{aligned} p_t * u(y) \geqslant P_t \mathbb{1}_{B(0,\kappa)}(y) &= \mathbb{P}(|y + \sqrt{2t} Z| \leqslant \kappa) \\ &= \mathbb{P}\left( Z \in B\left( \frac{y}{\sqrt{2t}}, \frac{\kappa}{\sqrt{2t}} \right) \right) \\ &= \mathbb{P}\left( Z \in B\left( z\left( \frac{\kappa}{\sqrt{2t}} + \sqrt{\frac{t}{2}} \rho \right), \frac{\kappa}{\sqrt{2t}} \right) \right), \end{aligned}$$

and for $t\to 0$ we have

$$\begin{aligned}
&\lim_{t\to 0}\inf_{|z|\leqslant 1}\mathbb{P}\left(Z\in B\left(z\left(\frac{\kappa}{\sqrt{2t}}+\sqrt{\frac{t}{2}}\rho\right),\frac{\kappa}{\sqrt{2t}}\right)\right)\\
=&\lim_{t\to 0}\inf_{|z|\leqslant 1}\mathbb{P}\left(Z\in B\left(|z|e_1\left(\frac{\kappa}{\sqrt{2t}}+\sqrt{\frac{t}{2}}\rho\right),\frac{\kappa}{\sqrt{2t}}\right)\right)\\
=&\lim_{t\to 0}\mathbb{P}\left(Z\in B\left(e_1\left(\frac{\kappa}{\sqrt{2t}}+\sqrt{\frac{t}{2}}\rho\right),\frac{\kappa}{\sqrt{2t}}\right)\right)=\mathbb{P}(Z_1>0)=\frac{1}{2}.
\end{aligned}$$

This proves the claim. □

**Theorem 6.9. (Strong maximum principle, [11], Theorem 5.1)** *Let $u_0\in\mathscr{C}_p^0$ be a positive measure. Assume furthermore that $u_0\neq 0$. Then*

$$u(t,x)>0,\qquad t>0, x\in\mathbb{T}^2.$$

**Proof.** By the previous discussion we know that $u(s)\in\mathscr{C}^\alpha$ for all strictly positive times $s>0$. Moreover, from a convolution argument the positive measure $u_0$ can be approximated by a sequence of positive functions $u_0^n$, and therefore $u(s)\geqslant 0$ by approximation, and for sufficiently small $s>0$ we have $u(s)\neq 0$ by continuity. By considering $u(s)$ as a new initial condition, we may assume without loss of generality that $u_0$ is a positive continuous function with $u_0\neq 0$. Moreover, it turns out to be easier to interpret $u$ as a periodic function on $\mathbb{R}^2$.

Then there exists a ball $B(x,\kappa)\subset\mathbb{R}^2$ on which $u_0\geqslant\varepsilon$ for some $\varepsilon>0$. Since the equation for $u$ is linear, $\varepsilon^{-1}u$ solves the same equation but with initial condition $\varepsilon^{-1}u_0$. Since $\varepsilon^{-1}u>0$ if and only if $u>0$, we may forget about the multiplication with $\varepsilon^{-1}$, and thus we assume without loss of generality that

$$u_0\geqslant 1\qquad\text{on}\qquad B(x,\kappa)\subset\mathbb{R}^2.$$

We can decompose

$$u(s)=P_su_0+\int_0^s P_{s-r}(u(r)\xi)\mathrm{d}r=:P_su_0+w(s).$$

By adapting the proof of Theorem 6.7 to initial conditions in $\mathscr{C}^\alpha$ (for which we get a smaller blow-up factor $\gamma$) we see that $w\in C_T^{\alpha/2}L^\infty$. Let now $t>0$. Since $w(0)=0$, there exists $C>0$ such that for all $s\in[0,t]$

$$\|w(s)\|_{L^\infty}\leqslant Cs^{\alpha/2}.$$

Moreover,

$$P_su_0=\sum_{k\in\mathbb{Z}^2}p_s(\cdot+k)*u_0\geqslant p_s*\mathbb{1}_{B(x,\kappa)},$$

so that by Lemma 6.8 we get for all $\rho>0$ and $s<t_\rho$

$$P_su_0\geqslant\frac{1}{4}$$

on $B(x,\kappa+s\rho)$. So if $s\in(0,t_\rho)$ is small enough such that

$$Cs^{\alpha/2}<\frac{1}{8},$$

we get

$$u(s,y)\geqslant\frac{1}{8},\qquad y\in B(x,\kappa+s\rho).$$

Using the linearity of the equation, we can repeat the argument on $[s, 2s]$ and obtain $u(2s, y) \geqslant 1/64$ for $y \in B(x, \kappa + 2s\rho)$, and so on, until we arrive at

$$u(t, y) > 0, \qquad y \in B(x, \kappa + t\rho).$$

Since $\rho > 0$ was arbitrary, this completes the proof. □

This argument is very flexible, essentially we only used that the equation is linear. In particular, it extends to all linear equations that can be solved with paracontrolled distributions or regularity structures.

## 6.4 The Anderson Hamiltonian

The parabolic Anderson model can be written as

$$\partial_t u = \Delta u + u\xi = (\Delta + \xi)u =: \mathscr{H}u,$$

where $\mathscr{H}$ is the Anderson Hamiltonian,

$$\mathscr{H} = \Delta + \xi$$

(or, taking renormalization into account, $\mathscr{H} = \Delta + \xi - \infty$). So we formally have $u(t) = e^{t\mathscr{H}}u_0$, and we hope to obtain information about the behavior of $u$ from $\mathscr{H}$.

To do this, we first have to construct $\mathscr{H}$. In principle, we can define $\mathscr{H}u$ for all *paracontrolled* (with slightly different definition than before) $u$ of the form

$$u = u' \olessthan X + u^\sharp,$$

where $X = (1 - \Delta)^{-1}\xi \in \mathscr{C}^\alpha$, $u, u' \in \mathscr{C}_p^\alpha$, $u^\sharp \in \mathscr{C}_p^{2\alpha}$ for $\alpha \in (2/3, 1)$, because then we have

$$\mathscr{H}u = \Delta(u' \olessthan X) + \Delta u^\sharp + u \olessthan \xi + u \ogreaterthan \xi + u^\sharp \odot \xi + C(u', X, \xi) + u'(X \diamond \xi).$$

The problem is that while the right hand side is well defined, it is still only in $\mathscr{C}_p^{2\alpha-2}$ and thus only a distribution and not a function. Also, we would like to construct $\mathscr{H}$ as a self-adjoint operator on a Hilbert space, so that we can use spectral theory.

The first problem could be overcome very easily by restricting our attention to a subspace of the paracontrolled distributions: If we write $T_t$ for the map that sends $u_0$ to $u(t)$, where $u$ is the solution of the PAM, then for all $u_0 \in \mathscr{C}_p^0$ both $T_s u_0$ and the integral $\int_0^t T_s u_0 \mathrm{d}s$ are paracontrolled, and

$$\mathscr{H}\int_0^t T_s u_0 \mathrm{d}s = \int_0^t \mathscr{H}T_s u_0 \mathrm{d}s = \int_0^t \partial_s T_s u_0 \mathrm{d}s = T_t u_0 - u_0,$$

so if $u_0$ is "nice enough" (depending on the space on which we want to define $\mathscr{H}$), then $\int_0^t T_s u_0 \mathrm{d}s$ is in the domain of $\mathscr{H}$. Since also $t^{-1}\int_0^t T_s u_0 \mathrm{d}s$ converges to $u_0$, we would obtain that the domain is dense. The problem with this approach is that it seems not so easy to obtain information about the spectrum of $\mathscr{H}$ from that construction.

Therefore, we take a different approach which goes back to Allez and Chouk [2]: For $\lambda > 0$ we consider the *resolvent equation*

$$(\lambda - \mathscr{H})u = v$$

for $v \in L^2 \subset \mathscr{C}_2^0$. We can rewrite this equation as a paracontrolled PDE as follows:

$$(\lambda - \Delta)u = u\xi + v \qquad \Leftrightarrow \qquad u = (\lambda - \Delta)^{-1}(u\xi + v)$$

As there is no time variable, we do not need the modified paraproduct and for large $\lambda$ the equation is actually easier to solve than the parabolic Anderson model. The space of paracontrolled distributions for this problem is with $\beta \in (2/3, \alpha)$

$$(u, u', u^\sharp) \in \mathscr{C}_2^\alpha \times \mathscr{C}_2^\beta \times \mathscr{C}_2^{\alpha+\beta}: \qquad u^\sharp = u - u' \olessthan X.$$

We obtain a small factor for the contraction property by choosing $\lambda$ large (which we fix from now on). The solution $u$ is then in $\mathscr{C}_2^{\beta} \subset B_{2,2}^{\beta'}$ for all $\beta' \in (0, \beta)$.

**Exercise 6.3.** For $\gamma \in \mathbb{R}$ we define the $L^2$ Sobolev space

$$H^{\gamma} = \left\{ u \in \mathscr{S}'(\mathbb{T}^d) : \|u\|_{H^{\gamma}}^2 := \sum_{k \in \mathbb{Z}^d} |\hat{u}(k)|^2 (1 + |k|^2)^{\gamma} < \infty \right\}.$$

i. Show that $H^{\gamma} = B_{2,2}^{\gamma}$ with equivalent norms.

ii. Show that $H^0 = L^2$.

iii. Show that bounded sets in $H^{\gamma}$ are relatively compact in $H^{\gamma'}$ whenever $\gamma > \gamma'$.

By the exercise we see that the operator

$$R_{\lambda} : L^2 \ni v \mapsto (\lambda - \mathscr{H})^{-1} v = u \in L^2$$

is compact (i.e. it maps bounded sets to relatively compact sets). Moreover, it is self-adjoint: Assume for the moment that $\xi$ is a bounded function, then $\lambda - \mathscr{H} = \lambda - \Delta - \xi$ is self-adjoint because multiplication operators are trivially self-adjoint and $\Delta$ is self-adjoint as well, and therefore

$$\langle R_{\lambda} v, w \rangle_{L^2} = \langle R_{\lambda} v, (\lambda - \mathscr{H}) R_{\lambda} w \rangle_{L^2} = \langle (\lambda - \mathscr{H}) R_{\lambda} v, R_{\lambda} w \rangle_{L^2} = \langle v, R_{\lambda} w \rangle_{L^2}.$$

By approximation, this carries over to our situation.

So $R_{\Lambda}$ is a compact self-adjoint operator on the Hilbert space $L^2(\mathbb{T}^2)$, and by the spectral theorem for compact operators there exists an orthonormal basis of eigenfunctions $(e_n)_{n \in \mathbb{N}}$ and real valued eigenvalues $(\kappa_n)_{n \in \mathbb{N}}$ such that $|\kappa_1| \geqslant |\kappa_2| \geqslant \cdots$ with $|\kappa_n| \to 0$ and

$$R_{\lambda} e_n = \kappa_n e_n, \qquad n \in \mathbb{N}.$$

We just need one more information about the $\kappa_n$:

**Lemma 6.10.** *We have $\kappa_n > 0$ for all $n \in \mathbb{N}$.*

**Proof.** By definition of $R_{\lambda} = (\lambda - \mathscr{H})^{-1}$ we have $R_{\lambda} v \neq 0$ for all $v \neq 0$, and therefore $|\kappa_n| > 0$ for all $n$. Thus it suffices to show that $\kappa_n \geqslant 0$, which follows immediately once we show that $R_{\lambda}$ is a positive operator, i.e. that

$$\langle R_{\lambda} v, v \rangle_{L^2} \geqslant 0$$

for all $v \in L^2$. Indeed, then

$$\kappa_n = \kappa_n \langle e_n, e_n \rangle_{L^2} = \langle R_{\lambda} e_n, e_n \rangle_{L^2} \geqslant 0.$$

To see that $R_{\lambda}$ is positive, we use the representation

$$R_{\lambda} v = \int_0^{\infty} e^{-\lambda t} T_t v \mathrm{d}t.$$

Indeed, we have for $S > 0$

$$(\lambda - \mathscr{H}) \int_0^S e^{-\lambda t} T_t v \mathrm{d}t = \int_0^S -\partial_t (e^{-\lambda t} T_t v) \mathrm{d}t = v - e^{-\lambda S} T_S v,$$

and since the PAM is linear we have $\|T_S v\|_{L^2} \leqslant K e^{KS} \|v\|_{L^2}$ for some $K > 0$. Without loss of generality we assume that $\lambda > K$ (in fact we already had to take $\lambda > K$ in the construction of $R_{\lambda}$), and then we can send $S \to \infty$ to get

$$(\lambda - \mathscr{H}) \int_0^{\infty} e^{-\lambda t} T_t v \mathrm{d}t = v,$$

which proves the representation for $R_\lambda$. So now it suffices to show that $T_t$ is a positive operator for all $t$. But

$$\langle T_t v, v\rangle_{L^2} = \langle T_{t/2} T_{t/2} v, v\rangle_{L^2} = \langle T_{t/2} v, T_{t/2} v\rangle_{L^2} = \|T_{t/2} v\|_{L^2}^2 \geqslant 0,$$

where we used the semigroup property of $(T_t)$, and that $T_t$ is self-adjoint for all $t$: If $\xi$ is a bounded function, this follows for example from the Feynman-Kac formula, because we may interpret $v$ as a periodic function on $\mathbb{R}^2$ and then obtain with a two-dimensional Brownian motion

$$\begin{aligned}
\int_{\mathbb{T}^2} T_t v(x) w(x) \mathrm{d}x &= \int_{\mathbb{T}^2} \mathbb{E}\left[ v(x+B_t) \exp\left( \int_0^t \xi(x+B_s) \mathrm{d}s \right) \right] w(x) \mathrm{d}x \\
&= \mathbb{E}\left[ \int_{\mathbb{T}^2} v(x+B_t) \exp\left( \int_0^t \xi(x+B_s) \mathrm{d}s \right) w(x) \mathrm{d}x \right] \\
&= \mathbb{E}\left[ \int_{\mathbb{T}^2} v(x) \exp\left( \int_0^t \xi(x+B_s-B_t) \mathrm{d}s \right) w(x-B_t) \mathrm{d}x \right] \\
&= \mathbb{E}\left[ \int_{\mathbb{T}^2} v(x) \exp\left( \int_0^t \xi(x-(B_t-B_{t-s})) \mathrm{d}s \right) w(x-(B_t-B_{t-t})) \mathrm{d}x \right] \\
&= \mathbb{E}\left[ \int_{\mathbb{T}^2} v(x) \exp\left( \int_0^t \xi(x+B_s) \mathrm{d}s \right) w(x+B_t) \mathrm{d}x \right] \\
&= \int_{\mathbb{T}^2} v(x) T_t w(x) \mathrm{d}x,
\end{aligned}$$

where we used that $s \mapsto -(B_t - B_{t-s})$, $s \in [0,t]$, is a Brownian motion. By approximation, $T_t$ is also self-adjoint in the white noise case. $\square$

Consequently, the eigenvalues $\kappa_n$ are not only decreasing in absolute value, they are actually decreasing. We claim that the $e_n$ are also eigenfunctions for $\mathscr{H}$. Indeed,

$$(\lambda - \mathscr{H}) e_n = \kappa_n^{-1} (\lambda - \mathscr{H}) \kappa_n e_n = \kappa_n^{-1} (\lambda - \mathscr{H}) R_\lambda e_n = \kappa_n^{-1} e_n,$$

and thus

$$\mathscr{H} e_n = (\lambda - \kappa_n^{-1}) e_n =: \lambda_n e_n$$

Since the $\kappa_n$ are decreasing, also the $\lambda_n$ are decreasing, and since $\kappa_n \to 0$ we have $\lambda_n \to -\infty$ for $n \to \infty$. We thus obtained a spectral decomposition

$$\mathscr{H} v = \sum_{n=1}^\infty \lambda_n \langle e_n, v\rangle_{L^2} e_n,$$

which is valid whenever $v$ is in the domain of $\mathscr{H}$.

## 6.5 Long time behavior of the periodic parabolic Anderson model

We formally have $T_t = e^{t\mathscr{H}}$, and with spectral calculus we can make this rigorous:

$$T_t u = \sum_{n=1}^\infty e^{t\lambda_n} \langle e_n, u\rangle_{L^2} e_n.$$

For $u$ such that $\langle u, e_n\rangle_{L^2} \neq 0$ only for finitely many $n$ this representation follows immediately from the fact that both $T_t u$ and $\sum_{n=1}^\infty e^{t\lambda_n} \langle e_n, u\rangle_{L^2} e_n$ solve the equation $\partial_t v = \mathscr{H} v$, and then it extends to general $u$ by approximation.

**Lemma 6.11.** *The operator $\mathscr{H}$ has a spectral gap, i.e. $\lambda_1 > \lambda_2$. Moreover, $e_1(x) > 0$ for all $x \in \mathbb{T}^2$.*

**Proof.** Consider the cone

$$K = \{v \in L^2 \colon v \geqslant 0\} \subset L^2.$$

We say $v \geqslant w$ (resp. $v \gg w$) if $v - w$ is in $K$ (resp. in the interior $K^\circ$ of $K$), or in other words if $v - w \geqslant 0$ (resp. $v - w > 0$) almost everywhere. Then $T_t$ is a compact linear operator that is strongly positive, i.e. such that

$$T_t v \gg 0$$

whenever $v \geqslant 0$; indeed, this follows from the strong maximum principle Theorem 6.9.

By (a consequence of) the Krein-Rutman theorem, see Theorem 19.3 in [17], we have $\lambda_1 > \lambda_2$ and $e_1 \gg 0$. Since $e_1 = \kappa_1^{-1} R_\lambda e_1$ we know that $e_1$ is paracontrolled, i.e. there exist $e_1 = e_1' \circledcirc X + e_1^\sharp$ with $e_1' \in \mathscr{C}_2^\beta$ and $e_1^\sharp \in \mathscr{C}_2^{\alpha+\beta}$. The paraproduct estimates together with the Besov embedding theorem thus show that $e_1 \in \mathscr{C}^\varepsilon$ for some $\varepsilon > 0$ and thus $e_1$ is a continuous function and $e_1(x) > 0$ for all $x$. □

We now collected all ingredients needed to describe the long time behavior of the PAM:

**Theorem 6.12.** *There exists $\kappa > 0$ such that for all $u \in L^2$ with $u \geqslant 0$ and $u \neq 0$*

$$\lim_{t \to \infty} \left\| \frac{T_t u}{e^{t\lambda_1} \langle u, e_1 \rangle_{L^2} e_1} - 1 \right\|_{L^\infty} = 0$$

*Consequently, we have for $u' \geqslant 0$ with $u' \neq 0$*

$$\lim_{t \to \infty} \left\| \frac{T_t u}{T_t u'} - \frac{\langle u, e_1 \rangle_{L^2}}{\langle u', e_1 \rangle_{L^2}} \right\|_{L^\infty} = 0,$$

*i.e. the ratio of two solutions for different initial conditions becomes constant for large times.*

**Proof.** First note that $\langle u, e_1 \rangle_{L^2} > 0$ since $e_1(x) > 0$ for all $x$, and therefore the division by $e^{t\lambda_1} \langle u, e_1 \rangle_{L^2} e_1$ is allowed. We also have $e_1(x) \geqslant \varepsilon > 0$ for all $x$, because $e_1$ is continuous and $\mathbb{T}^2$ is compact and thus $e_1$ attains its minimum. Thus

$$\frac{T_t u}{e^{t\lambda_1} \langle u, e_1 \rangle_{L^2} e_1} = 1 + \frac{1}{\langle u, e_1 \rangle_{L^2} e_1} \sum_{n=2}^\infty e^{t(\lambda_n - \lambda_1)} \langle u, e_n \rangle_{L^2} e_n.$$

From here it is trivial to show that $\frac{T_t u}{e^{t\lambda_1} \langle u, e_1 \rangle_{L^2} e_1} - 1$ converges to zero in $L^2$, but we have to work a bit more the get the convergence in $L^\infty$. We have for $\tau > 0$ and $t > \tau$

$$\begin{aligned}
&\left| \frac{T_t u(x)}{e^{t\lambda_1} \langle u, e_1 \rangle_{L^2} e_1(x)} - 1 \right| \\
&\leqslant \left\| \frac{1}{\langle u, e_1 \rangle_{L^2} e_1} \right\|_{L^\infty} \sum_{n=2}^\infty e^{t(\lambda_n - \lambda_1)} |\langle u, e_n \rangle_{L^2}| |e_n(x)| \\
&\leqslant |\langle u, e_1 \rangle_{L^2} \varepsilon|^{-1} \left( \sum_{n=2}^\infty e^{2\tau(\lambda_n - \lambda_1)} |e_n(x)|^2 \right)^{1/2} \left( \sum_{n=2}^\infty e^{2(t-\tau)(\lambda_n - \lambda_1)} |\langle u, e_n \rangle_{L^2}|^2 \right)^{1/2} \\
&\lesssim \left( \sum_{n=1}^\infty e^{2\tau\lambda_n} |e_n(x)|^2 \right)^{1/2} e^{2(t-\tau)(\lambda_2 - \lambda_1)} \sum_{n=2}^\infty |\langle u, e_n \rangle_{L^2}|^2 \\
&\lesssim \left( \sum_{n=1}^\infty e^{2\tau\lambda_n} |e_n(x)|^2 \right)^{1/2} e^{2(t-\tau)(\lambda_2 - \lambda_1)} \|u\|_{L^2}^2.
\end{aligned}$$

Now we use that $\lambda_1 > \lambda_2$ by Lemma 6.11, so the claim follows once we show that $\sum_{n=1}^{\infty} e^{2\tau\lambda_n} |e_n(x)|^2 < \infty$. But if $\rho^{\varepsilon}$ is a smooth approximation of the Dirac delta, then

$$\begin{aligned}\sum_{n=1}^{\infty} e^{2\tau\lambda_n} |e_n(x)|^2 &\leqslant \liminf_{\varepsilon\to 0} \sum_{n=1}^{\infty} e^{2\tau\lambda_n} \langle e_n, \rho^{\varepsilon}(x-\cdot)\rangle_{L^2} \langle e_n, \rho^{\varepsilon}(x-\cdot)\rangle_{L^2}\\ &= \int u^{\rho^{\varepsilon}(x-\cdot)}(2\tau, y)\rho^{\varepsilon}(y)\mathrm{d}y,\end{aligned}$$

where $u^{\rho^{\varepsilon}(x-\cdot)}$ is the solution to PAM with initial condition $\rho^{\varepsilon}$. By our previous results $u^{\rho^{\varepsilon}(x-\cdot)}(2\tau,\cdot)$ converges uniformly to $u^{\delta(x-\cdot)}$ and thus we get

$$\sum_{n=1}^{\infty} e^{2\tau\lambda_n} |e_n(x)|^2 \leqslant u^{\delta(x-\cdot)}(2\tau, x) < \infty.$$

To derive the limiting behavior of $T_t u / T_t u'$, note that

$$\frac{T_t u}{T_t u'} = \frac{T_t u}{e^{t\lambda_1} \langle u, e_1\rangle_{L^2} e_1} \times \frac{e^{t\lambda_1}\langle u', e_1\rangle_{L^2} e_1}{T_t u'} \times \frac{\langle u, e_1\rangle_{L^2}}{\langle u', e_1\rangle_{L^2}},$$

and now note that if $\|f_n - 1\|_{L^\infty} \to 0$, then also $\|1/f_n - 1\|_{L^\infty} \to 0$, from where we deduce the second claimed convergence. □

Consequently, the initial condition does not influence the limiting shape of the solution to the PAM at all and only contributes through the scalar factor $\langle u, e_1\rangle$.

We can also solve the PAM on $\mathbb{R}^2$, and there the situation is much more complicated. Then the spectrum of the operator $\mathscr{H}$ is unbounded from above, and it does not generate a continuous contraction semigroup. While we can still solve the PAM on $\mathbb{R}^2$, the solution at time $t$ lives in a larger space than at time 0, with more permissive weights capturing the growth/decay at infinity; see [30] or [36] for details.

## 6.6 Some related linear equations

The spectral point of view provides us with easy solution theories for some other linear equations, for example the stochastic Schroedinger equation

$$i\partial_t u = \Delta u + u\xi = \mathscr{H}\xi$$

with $i = \sqrt{-1}$, or the stochastic wave equation

$$\partial_{tt} u = (\mathscr{H} - \lambda)\xi$$

for $\lambda \geqslant \lambda_1$. In the first case we can set

$$u(t) = \sum_{n=1}^{\infty} e^{it\lambda_n} \langle u, e_n\rangle_{L^2} e_n,$$

and in the second case we consider the equation as a system,

$$\begin{aligned}\partial_t u &= v\\ \partial_t v &= (\mathscr{H} - \lambda) u,\end{aligned}$$

so that we can set

$$K_t(u_0, v_0),$$

where

$$K_t = \begin{pmatrix} \cos(t(\lambda - \mathscr{H})^{1/2}) & (\lambda - \mathscr{H})^{-1/2}\sin(t(\lambda - \mathscr{H})^{1/2}) \\ -(\lambda - \mathscr{H})^{1/2}\sin(t(\lambda - \mathscr{H})^{1/2}) & \cos(t(\lambda - \mathscr{H})^{1/2}) \end{pmatrix}.$$

Based on this point of view we can then introduce nonlinear perturbations to the equations, for example by considering the mild formulation based on $(e^{it\mathscr{H}})_t$ resp. $(K_t)_t$, see [20] for details.

# 7 Relation with regularity structures

Hairer's regularity structures [28] provide another approach towards dealing with singular SPDEs, and they are based on closely related ideas, although they use very different technical tools. They are based on generalizations of the Taylor expansion and of increment characterizations of regularity. Here we discuss some links between paracontrolled distributions and regularity structures, essentially how the different descriptions of regularity are compatible.

My aim is not to give an introduction to regularity structures, this section addresses mainly readers with some previous exposure and in the beginning we only collect the main notions from [28] without any motivation or intuitive explanation. For nice introductions to regularity structures see for example [29, 14].

**Definition 7.1.** *A regularity structure is a triple $\mathcal{T} = (A, T, G)$, where $A \subset \mathbb{R}$ without accumulation point except possibly at $\infty$, where*

$$T = \bigoplus_{\alpha \in A} T_\alpha$$

*and each $T_\alpha$ is a Banach space, and where $G$ is a group of bounded linear operators on $T$ such that for all $\Gamma \in G$ and all $\tau \in T_\alpha$*

$$\Gamma\tau - \tau \in \bigoplus_{\beta < \alpha} T_\beta, \qquad \Gamma \in G, \tau \in T_\alpha.$$

*We call $T$ the model space and $G$ the structure group.*

A regularity structure is a purely abstract construct that provides a framework in which we can set up new notions of regularity and new function spaces. These function spaces depend on concrete realizations of regularity structures, that are encoded in *models*.

For $\varphi: \mathbb{R}^d \to \mathbb{R}$ we define

$$\varphi_x^\lambda := \lambda^{-d}\varphi(\lambda^{-1}(\cdot - x)), \qquad \lambda > 0, x \in \mathbb{R}^d.$$

For $\alpha \in A$ and $\tau \in T$ we write $\mathcal{Q}_\alpha \tau$ for the projection of $\tau$ onto $T_\alpha$, and

$$\|\tau\|_\alpha := \|\mathcal{Q}_\alpha \tau\|$$

**Definition 7.2.** *Let $\mathcal{T} = (A, T, G)$ be a regularity structure and $d \in \mathbb{N}$. A model for $\mathcal{T}$ on $\mathbb{R}^d$ consists of maps*

$$\Pi: \mathbb{R}^d \to \mathcal{L}(T, \mathscr{S}'(\mathbb{R}^d)), \qquad \Gamma: \mathbb{R}^d \times \mathbb{R}^d \to G,$$

*such the algebraic relations*

$$\Pi_x \Gamma_{xy} = \Pi_y, \qquad \Gamma_{xy}\Gamma_{yz} = \Gamma_{xz}$$

*hold, and such that with $r > -\alpha$ for all $\alpha \in A$ the following analytic relation holds: There exists $C > 0$ such that*

$$|\Pi_x \tau(\varphi_x^\lambda)| \leqslant C\lambda^\alpha \|\tau\|_\alpha, \qquad \|\Gamma_{x,y}\tau\|_\beta \leqslant C|x-y|^{\alpha-\beta}\|\tau\|_\alpha,$$

*for all $\varphi \in C_b^r$ with $\|\varphi\|_{C_b^r} \leqslant 1$, for all $\lambda \in (0,1]$, for all $x, y \in \mathbb{R}^d$ and for all $\alpha, \beta \in A$.*

Models provide the framework in which to model regularity, and given a model we can define new function spaces:

**Definition 7.3.** *Let $\mathcal{T} = (A, T, G)$ be a regularity structure with model $(\Pi, \Gamma)$. For $\gamma > 0$ the space of modelled distributions $\mathscr{D}^\gamma$ consists of the maps $f: \mathbb{R}^d \to \bigoplus_{\alpha<\gamma} T_\alpha \subset T$ such that*

$$\|f_x - \Gamma_{xy} f_y\|_\alpha \leqslant C|x-y|^{\gamma-\alpha}, \qquad \|f_x\|_\alpha \leqslant C$$

*for all $\alpha < \gamma$ and $x, y \in \mathbb{R}^d$, where $C > 0$. We write $\|f\|_{\mathscr{D}^\gamma}$ for the smallest such constant.*

Modelled distributions take values in the abstract Banach space $T$. But we can associate to each modelled distribution an element of $\mathscr{S}'(\mathbb{R}^d)$:

**Theorem 7.4. ([28], Theorem 3.10)** *Let $\gamma > 0$. Then there exists a bounded linear operator*

$$\mathscr{R}: \mathscr{D}^\gamma \to \mathscr{C}^{\alpha_0},$$

*where $\alpha_0 = \min_{\alpha \in A}$, such that for some $C > 0$*

$$|\mathscr{R}f(\varphi_x^\lambda) - \Pi_x f_x(\varphi_x^\lambda)| \leqslant C\lambda^\gamma \tag{7.1}$$

*for all $\varphi \in C_b^r$ with $\|\varphi\|_{C_b^r} \leqslant 1$, for all $\lambda \in (0,1]$, and for all $x \in \mathbb{R}^d$. Here $r$ is as in Definition 7.2. Moreover, $\mathscr{R}f$ is the unique element of $\mathscr{S}'$ that satisfies (7.1).*

We now want to link the theory of regularity structures with paracontrolled distributions. More precisely, we give descriptions of modelled distributions based on paraproducts. The material here is from Section 6 of [22] and from [35].

**Definition 7.5.** *For $f: \mathbb{R}^d \to T$ we define*

$$P(f, \Pi)(x) := \sum_{j \geqslant 1} \iint K_{<j-1}(x-y) K_j(x-z) \Pi_y f_y(z) \mathrm{d}y \mathrm{d}z$$

*and*

$$P(f, \Gamma)(x) := \sum_{j \geqslant 1} \iint K_{<j-1}(x-y) K_j(x-z) \Gamma_{zy} f_y \mathrm{d}y \mathrm{d}z$$

*whenever these are well defined. Note that $P(f, \Pi)$ takes values in $\mathscr{S}'(\mathbb{R}^d)$, while $P(f, \Gamma)$ takes values in $T$.*

We can extend our Besov spaces $\mathscr{C}^\kappa$ easily to distributions with values in a Banach space $X$, by writing

$$\|u\|_{\mathscr{C}^\kappa(X)} = \sup_{j \geqslant -1} 2^{j\kappa} \|K_j * u\|_{L^\infty(X)} =: \sup_{j \geqslant -1} 2^{j\kappa} \|\Delta_j u\|_{L^\infty(X)}$$

**Lemma 7.6.** *Let $\gamma > 0$ and $f \in \mathscr{D}^\gamma$. Then for all $\alpha < \gamma$*

$$\|\mathcal{Q}_\alpha(f - P(f, \Gamma))\|_{\mathscr{C}^{\gamma-\alpha}(T_\alpha)} \lesssim \|f\|_{\mathscr{D}^\gamma},$$

*and also*

$$\|\mathscr{R}f - P(f,\Pi)\|_\gamma \lesssim \|f\|_\gamma.$$

**Proof.** First observe the trivial estimate

$$\|\mathcal{Q}_\alpha \Delta_{\leqslant 0} f\|_{\mathscr{C}^\gamma(T_\alpha)} \lesssim \|f\|_{L^\infty(T_\alpha)} \leqslant \|f\|_{\mathscr{D}^\gamma}.$$

Moreover, since $\int K_{<j-1}(x)\mathrm{d}x = 1$ for all $j \geqslant 1$:

$$\begin{aligned}&\mathcal{Q}_\alpha(f - \Delta_{\leqslant 0} f - P(f,\Gamma))(x)\\ =&\sum_{j\geqslant 1} \iint K_{<j-1}(x-y)K_j(x-z)\mathcal{Q}_\alpha(f_z - \Gamma_{zy}f_y)\mathrm{d}y\mathrm{d}z.\end{aligned}$$

By Lemma 3.17 it suffices to bound each addend of the series, and using that $f \in \mathscr{D}^\gamma$ we have

$$\begin{aligned}&\left|\iint K_{<j-1}(x-y)K_j(x-z)\mathcal{Q}_\alpha(f_z - \Gamma_{zy}f_y)\mathrm{d}y\mathrm{d}z\right|\\ \leqslant&\iint |K_{<j-1}(x-y)K_j(x-z)| \times |z-y|^{\gamma-\alpha}\|f\|_{\mathscr{D}^\gamma}\mathrm{d}y\mathrm{d}z\\ \lesssim&\|f\|_{\mathscr{D}^\gamma}\iint |K_{<j-1}(x-y)K_j(x-z)| \times (|z-x|^{\gamma-\alpha} + |x-y|^{\gamma-\alpha})\mathrm{d}y\mathrm{d}z\\ \lesssim&\|f\|_{\mathscr{D}^\gamma} 2^{-j(\gamma-\alpha)},\end{aligned}$$

which shows that $\|\mathcal{Q}_\alpha(f - P(f,\Gamma))\|_{\mathscr{C}^{\gamma-\alpha}(T_\alpha)} \lesssim \|f\|_{\mathscr{D}^\gamma}$.

The second bound holds by very similar arguments:

$$\|\Delta_{\leqslant 0}\mathscr{R}f\|_\gamma \lesssim \|\mathscr{R}f\|_{\alpha_0} \lesssim \|f\|_{\mathscr{D}^\gamma}$$

by definition of $\Delta_{\leqslant 0}$ and the reconstruction operator, and

$$(\Delta_{>0}\mathscr{R} - P(f,\Pi))(x) = \sum_{j\geqslant 1}\iint K_{<j-1}(x-y)K_j(x-z)(\mathscr{R}f(z) - \Pi_y f_y(z))\mathrm{d}y\mathrm{d}z.$$

Again it suffices to bound each addend individually, and

$$\begin{aligned}&\left|\iint K_{<j-1}(x-y)K_j(x-z)(\mathscr{R}f(z) - \Pi_y f_y(z))\mathrm{d}y\mathrm{d}z\right|\\ =&\left|\int K_{<j-1}(x-y)\big(\mathscr{R}f\big(K_x^{2^{-j}}\big) - \Pi_y f_y\big(K_x^{2^{-j}}\big)\big)\mathrm{d}y\right|\\ \leqslant&\left|\int K_{<j-1}(x-y)\big(\mathscr{R}f\big(K_x^{2^{-j}}\big) - \Pi_x f_x\big(K_x^{2^{-j}}\big)\big)\mathrm{d}y\right|\\ +&\left|\int K_{<j-1}(x-y)\Pi_x(f_x - \Gamma_{xy}f_y)\big(K_x^{2^{-j}}\big)\mathrm{d}y\right|\\ \lesssim&\left|\int K_{<j-1}(x-y)2^{-j\gamma}\|f\|_{\mathscr{D}^\gamma}\mathrm{d}y\right| + \sum_{\alpha<\gamma}\left|\int K_{<j-1}(x-y)2^{-j\alpha}|x-y|^{\gamma-\alpha}\|f\|_{\mathscr{D}^\gamma}\mathrm{d}y\right|\\ \lesssim&2^{-j\gamma}\|f\|_{\mathscr{D}^\gamma},\end{aligned}$$

which completes the proof. □

Note that $\mathscr{R}f$ itself is much more irregular than $\mathscr{C}^\gamma$, in general we only get $\mathscr{R}f \in \mathscr{C}^{\alpha_0}$. So the lemma gives a decomposition of $\mathscr{R}f$ into a paraproduct and a smooth remainder.

To do: paracontrolled distributions are modelled.

# 8 A nonlinear stochastic wave equation

## 8.1 Dimension 2

Here we follow Gubinelli, Koch and Oh [24] and study the equation

$$\partial_{tt}^2 u = \Delta u + u^2 + \xi$$

on $\mathbb{R}_+ \times \mathbb{T}^2$, where $\xi$ is a space-time white noise. While at first sight this looks very similar to the $\Phi_2^3$ equation, it behaves were differently because the wave equation has much worse regularizing properties than the heat heat equation.

We can bring the equation to first order in time by rewriting it as a system, $u = (u, v)$, with

$$\begin{aligned}\partial_t u &= v\\ \partial_t v &= \Delta v + u^2 + \xi,\end{aligned}$$

or

$$\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = A \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ u^2 + \xi \end{pmatrix},$$

where

$$A = \begin{pmatrix} 0 & 1 \\ \Delta & 0 \end{pmatrix}.$$

Then we have with $|\nabla| := (-\Delta)^{1/2}$

$$e^{tA} = \begin{pmatrix} \cos(t|\nabla|) & |\nabla|^{-1}\sin(t|\nabla|) \\ -|\nabla|\sin(t|\nabla|) & \cos(t|\nabla|) \end{pmatrix}, \tag{8.1}$$

where $\cos(t|\nabla|)$ and $\sin(t|\nabla|)$ are defined in terms of spectral calculus, or explicitly through the Fourier transform:

$$\cos(t|\nabla|)u = \sum_{k \in \mathbb{Z}^2} \cos(t|2\pi k|)\hat{u}(k)e^{2\pi i k \cdot x}.$$

We can verify this representation for $e^{tA}$ by differentiating the matrix in (8.1) and by showing that the derivative equals $Ae^{tA}$.

In particular, the variation of constants formula gives

$$\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = e^{tA} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \int_0^t e^{(t-s)A} \begin{pmatrix} 0 \\ u(s)^2 + \xi(s) \end{pmatrix} \mathrm{d}s,$$

and since we are mainly interested in $u$:

$$u(t) = \mathcal{S}(t)(u_0, v_0) + \int_0^t \frac{\sin((t-s)|\nabla|)}{|\nabla|}(u(s)^2 + \xi(s))\mathrm{d}s,$$

where

$$\mathcal{S}(t)(u_0, v_0) := \cos(t|\nabla|)u_0 + \frac{\sin(t|\nabla|)}{|\nabla|}v_0.$$

Here we see an important difference compared to the heat equation: While $P_t = e^{t\Delta} = e^{-t|\nabla|^2}$ is infinitely regularizing (with a blow-up for $t \to 0$), the same is not true for $\sin(t|\nabla|)/|\nabla|$, which only seems to gain one derivative.

Let us proceed anyways, and apply the tools that we developed for dealing with the $\Phi_2^4$ equation: We make the Ansatz

$$u = Z + w,$$

where

$$Z(t) = \int_0^t \frac{\sin((t-s)|\nabla|)}{|\nabla|} \xi(s) \mathrm{d}s.$$

To simplify notation, we also write

$$\mathcal{I} f(t) := \int_0^t \frac{\sin((t-s)|\nabla|)}{|\nabla|} f(s) \mathrm{d}s,$$

so that for example $Z = \mathcal{I}\xi$. Then

$$w = \mathcal{S}(u_0, v_0) + \mathcal{I}(w^2 + 2wZ + Z^2).$$

Of course, there will be problems with defining $Z^2$ because of the irregularity of $Z$. We can derive the regularity of $Z$ along the lines of Lemma 4.3, and this gives

$$Z \in C_T \mathscr{C}^{-\kappa}$$

almost surely for all $\kappa > 0$. This may be surprising, because it is the same regularity that we got for the convolution of $\xi$ with the better behaved heat kernel, but at least an estimate in the Sobolev scale is actually very easy to obtain: We have for a family of complex valued standard Brownian motions $(B^k)_{k \in \mathbb{Z}}$ (i.e. both real and imaginary part of $B^k$ are independent standard Brownian motions) such that $\mathbb{E}[B^k_t B^\ell_s] = 2\delta_{k,-\ell} s \wedge t$ the representation

$$Z(t,x) = \sum_{k \in \mathbb{Z}^2} e^{2\pi i k \cdot x} \int_0^t \frac{\sin((t-s)|2\pi k|)}{|2\pi k|} \mathrm{d}B^k_s,$$

and thus

$$\mathbb{E}[\|Z(t)\|^2_{H^\alpha}] = \sum_{k \in \mathbb{Z}^2} (1+|k|^2)^\alpha 2 \int_0^t \frac{\sin((t-s)|2\pi k|)^2}{|2\pi k|^2} \mathrm{d}s \lesssim \sum_{k \in \mathbb{Z}^2} (1+|k|^2)^{\alpha-1},$$

which is finite as soon as $\alpha < 0$. Since also

$$\int_0^t e^{-2(t-s)|2\pi k|^2} \mathrm{d}s = O(|k|^{-2}),$$

this explains to some extent why we see the same regularity as for the heat equation. See Proposition 2.1 of [24] for the precise derivation of the regularity of $Z$, where it is also shown that for all $n$

$$Z^{:n:} := \lim_{\varepsilon \to 0} Z_\varepsilon^{:n:} := \lim_{\varepsilon \to 0} H_n(Z_\varepsilon, \mathrm{var}(Z_\varepsilon)),$$

for a suitable mollification $Z_\varepsilon$ of $Z$, satisfies $Z^{:n:} \in C_T \mathscr{C}^{-\kappa}$ for all $n$ and all $\kappa > 0$. Hence, we modify the equation for $w$ to take the renormalization into account and try to solve

$$w = \mathcal{S}(u_0, v_0) + \mathcal{I}(w^2 + 2wZ + Z^{:2:}).$$

For that purpose we have to understand the regularizing properties of $\mathcal{I}$ better, which are provided by the Strichartz estimates for $\mathcal{I}$. To formulate them, we need a definition:

**Definition 8.1. ([24], Lemma 3.1)** *Let $s \in (0,1)$. A pair $(q,r) \in (2,\infty] \times [2,\infty)$ is* s-admissible *if*

$$\frac{1}{q} + \frac{2}{r} = 1 - s \qquad \text{and} \qquad 2 \leqslant r \leqslant \begin{cases} \dfrac{6}{3-4s}, & s < \dfrac{3}{4} \\ \infty, & \text{else} \end{cases}.$$

*A pair* $(\tilde{q},\tilde{r})\in[1,2)\times(1,2]$ *is* <u>*dual s-admissible*</u> *if the conjugate exponents* $(\tilde{q}',\tilde{r}')$ *are* $(1-s)$*-admissible, or equivalently if*

$$\frac{1}{\tilde{q}}+\frac{2}{\tilde{r}}=3-s,\qquad \text{and}\qquad \max\left\{1+,\frac{6}{7-4s}\right\}\leqslant\tilde{r}\leqslant\frac{2}{2-s},$$

*where* $\tilde{r}\geqslant 1+$ *means* $\tilde{r}>1$.

We also need suitable function spaces to work in. Unfortunately, the operator $\mathcal{I}$ only has good regularization properties in $L^2$-Sobolev spaces. Therefore, we set

$$\mathcal{H}^s:=H^s\times H^{s-1},\qquad s\in\mathbb{R}.$$

We also write

$$L_T^pX=L^p([0,T],X)$$

for a Banach space $X$

**Lemma 8.2. ([24], Lemma 3.2)** *Let* $s\in(0,1)$ *and let* $(q,r)$ *be s-admissible and* $(\tilde{q},\tilde{r})$ *be dual s-admissible. Then we have for*

$$u=\mathcal{S}(u_0,v_0)+\mathcal{I}f$$

*and* $T\in(0,1]$ *the following* <u>*Strichartz estimates*</u>*:*

$$\|(u,\partial_tu)\|_{L_T^\infty\mathcal{H}^s}+\|u\|_{L_T^qL^r}\lesssim\|(u_0,v_0)\|_{\mathcal{H}^s}+\min\left\{\|f\|_{L_T^{\tilde{q}}L^{\tilde{r}}},\|f\|_{L_T^1H^{s-1}}\right\}.$$

In other words, we gain one derivative on the Sobolev scale $H^\alpha$ for $u$, but we can also gain integrability instead.

Now let us try to set up a Picard iteration for $w$ in $L_T^\infty\mathcal{H}^s$ with $s\in(0,1)$. Then we have to control the right hand side of the equation, which is done in the following lemma:

**Lemma 8.3.** *We have for* $s\in(0,1)$ *and* $\kappa\in(0,s\wedge(1-s))$

$$\|w^2+2wZ+Z^{:2:}\|_{L_T^\infty H^{s-1}}\lesssim\|w\|_{L_T^\infty H^s}^2+\|w\|_{L_T^\infty H^s}\|Z\|_{L_T^\infty\mathscr{C}^{-\kappa}}+\|Z^{:2:}\|_{L_T^\infty\mathscr{C}^{-\kappa}}.$$

**Proof.** We decompose

$$w^2=w\odot w+2w\otimes w$$

and estimate for $\kappa\in(0,s)$ via Besov embedding (and using that $d=2$)

$$\|w\odot w\|_{L_T^\infty H^{s-1}}\lesssim\|w\odot w\|_{L_T^\infty\mathscr{C}_2^{s-1+\kappa}}\lesssim\|w\odot w\|_{L_T^\infty\mathscr{C}_1^{s+\kappa}}\lesssim\|w\|_{L_T^\infty\mathscr{C}_2^{(s+\kappa)/2}}^2\lesssim\|w\|_{L_T^\infty H^s}^2,$$

as well as

$$\begin{aligned}\|w\otimes w\|_{L_T^\infty H^{s-1}}&\lesssim\|w\otimes w\|_{L_T^\infty\mathscr{C}_2^{s-1+\kappa}}\\&\lesssim\|w\|_{L_T^\infty\mathscr{C}_\infty^{-1+\kappa}}\|w\|_{L_T^\infty\mathscr{C}_2^s}\\&\lesssim\|w\|_{L_T^\infty\mathscr{C}_2^\kappa}\|w\|_{L_T^\infty\mathscr{C}_2^s}\\&\lesssim\|w\|_{L_T^\infty H^s}^2.\end{aligned}$$

Next, we let $\kappa\in(0,(1-s)\wedge s)$ and get

$$\|wZ\|_{L_T^\infty H^{s-1}}\lesssim\|wZ\|_{L_T^\infty\mathscr{C}_2^{-\kappa}}\lesssim\|w\|_{L_T^\infty\mathscr{C}_2^s}\|Z\|_{L_T^\infty\mathscr{C}^{-\kappa}}\lesssim\|w\|_{L_T^\infty H^s}\|Z\|_{L_T^\infty\mathscr{C}^{-\kappa}}.$$

And finally

$$\|Z^{:2:}\|_{L_T^\infty H^{s-1}}\lesssim\|Z^{:2:}\|_{L_T^\infty\mathscr{C}^{-\kappa}}.$$

$\square$

Combining this with the Strichartz estimate, we can set up a Picard iteration for $(u, \partial_t u)$ in $C_T \mathcal{H}^s$ and get existence and uniqueness of solutions for small $T > 0$.

It may seem surprising that we stated the Strichartz estimates in such a complicated way, when the regularization effect in $\mathcal{H}^s$ was all we needed. But that is very much due to the fact that here we only considered the nonlinearity $u^2$. If we replace $u^2$ by $u^k$ for $k > 2$, then things become more complicated because we lose more integrability, and then we need to start keeping track also of the integrability. Moreover, we pick up additional constraints for the regularity $s$ of the initial condition.

## 8.2 Dimension 3

Now let us try to see what could be done for the same wave equation

$$\partial_{tt}^2 u = \Delta u + u^2 + \xi$$

in $d = 3$, i.e. $u: \mathbb{R}_+ \times \mathbb{T}^3 \to \mathbb{R}$. Here we follow Gubinelli, Koch and Oh [23]. As before, we start with the ansatz

$$u = \mathsf{I} + w,$$

where we use tree notation now and

$$\mathsf{I} = \mathcal{I}\xi.$$

The same simple computation we did above to derive the regularity of $Z$ also works in this setting, and it suggests (correctly) that $\mathsf{I} \in C_T \mathscr{C}^{-1/2-}$. The next term in the expansion is

$$\mathsf{Y} = \mathcal{I}(\mathsf{I}^{:2:}),$$

and if we apply the usual heuristic for guessing the regularity, we get $\mathsf{I}^{:2:} \in C_T \mathscr{C}^{-1-}$ and then, since $\mathcal{I}$ should gain one derivative because of the factor $|\nabla|^{-1}$, we guess

$$\mathsf{Y} = \mathcal{I}(\mathsf{I}^{:2:}) \in C_T \mathscr{C}^{0-}.$$

But unlike in the parabolic setting, this guess is in fact suboptimal, and we can show that

$$\mathsf{Y} \in C_T \mathscr{C}^{1/2-},$$

i.e. it is “half a derivative” better than expected. But to see this we have to estimate $\mathsf{Y}$ directly and cannot first construct $\mathsf{I}^{:2:}$ and then apply Strichartz estimates. See [23], Proposition 1.6.

We thus make the ansatz

$$u = \mathsf{I} + \mathsf{Y} + w$$

and obtain the following equation for $w$:

$$w = \mathcal{S}(u_0, v_0) + \mathcal{I}((w + \mathsf{Y})^2 + 2(w + \mathsf{Y})\mathsf{I}).$$

Since $\mathsf{I} \in C_T \mathscr{C}^{-1/2-}$, we expect $w$ at best to have regularity $1/2-$, so that $w\mathsf{I}$ is not well defined. To proceed, we use the paracontrolled ansatz, although in a slightly different formulation that goes back to Mourrat and Weber [39]. Namely, we make the Ansatz

$$w = w_1 + w_2$$

with

$$\begin{aligned} w_1 &= \mathcal{I}(2(w_1 + w_2 + \mathsf{Y}) \varolessthan \mathsf{I}) \\ w_2 &= \mathcal{S}(u_0, v_0) + \mathcal{I}((w_1 + w_2 + \mathsf{Y})^2 + 2(w + \mathsf{Y}) \odot \mathsf{I} + 2(w + \mathsf{Y}) \varogreaterthan \mathsf{I}). \end{aligned}$$

By the usual power counting we would guess the regularities $w_1 \in C_T B_?^{1/2-}$ and $w_2 \in C_T B_?^{1-}$, where $B_?$ denotes a suitable Besov space that will be determined later. We also see that we would expect having to control

$$\Upsilon \odot \text{¡} \in C_T \mathscr{C}^{0-},$$

and indeed this is possible. Now it looks like we are in good shape to put our usual paracontrolled machinery in place, provided that $(u_0, v_0)$ are regular enough and we control $(\mathcal{I}\text{¡}) \odot \text{¡}$. However, there is one major problem: When solving parabolic paracontrolled equations, we essentially used that the heat semigroup commutes with the paraproduct up to a smoother remainder (this was hidden in Corollary 5.10), and this is no longer true for the operator $\mathcal{I}$. We can however control the resonant product with stochastic computations, i.e. show that

$$f \mapsto \mathcal{I}(\mathcal{I}(f \olessthan \text{¡}) \odot \text{¡})$$

is a bounded random operator between suitable Sobolev spaces. From here the analysis is similar to what we have seen before.

# Appendix A Regularity for Walsh type SPDEs

**Proof. (of Lemma 1.28)** $Z$ is the stochastic integral of $W$ against the deterministic function $p$, and therefore it is Gaussian (if we replace $p$ by a deterministic bounded elementary function then this follows from the Gaussianity of $W$, and it extends to $p$ by a limiting argument). Since $Z$ is centered Gaussian, the $q$-th moment is bounded by the $q/2$-th power of the second moment and we get

$$\begin{aligned}\mathbb{E}[|Z(t+s,x+y) - Z(t,x)|^q]^{1/q} \lesssim & \mathbb{E}[|Z(t+s,x+y) - Z(t+s,x)|^2]^{1/2} \\ & + \mathbb{E}[|Z(t+s,x) - Z(t,x)|^2]^{1/2}.\end{aligned}$$

In the nonlinear case with nontrivial $f$, $g$ we would use the Burkholder-Davis-Gundy inequality instead of Gaussianity. The first term on the right hand side is

$$\begin{aligned}&\mathbb{E}[|Z(t+s,x+y) - Z(t+s,x)|^2] \\ &= \mathbb{E}\left[\left(\int_{[0,t+s]\times\mathbb{R}^d} (p(t+s-r,x+y-z) - p(t+s-r,x-z)) W(dr,dz)\right)^2\right] \\ &= \int_0^{t+s} \int_{\mathbb{R}^d\times\mathbb{R}^d} (p(r,x+y-z_1) - p(r,x-z_1))(p(r,x+y-z_2) - p(r,x-z_2)) K(dz_1,dz_2) dr,\end{aligned}$$

where in the last line we used the substitution $t+s-r \to r$ for the time integral.

Next, we observe that for $c > 1$

$$\begin{aligned}|p(r,x+y-z_1) - p(r,x-z_1)| &= \left|\int_0^1 \nabla p(r,(x-z_1)+\lambda y) \cdot y \, d\lambda\right| \\ &= \left|\int_0^1 \frac{(x-z_1)+\lambda y}{2r} p(r,(x-z_1)+\lambda y) \cdot y \, d\lambda\right| \\ &\lesssim r^{-1/2} |y| \int_0^1 p(cr,(x-z_1)+\lambda y) d\lambda,\end{aligned}$$

where we used that for all $m \geqslant 0$ (and here we take $m = 1/2$)

$$\left(\frac{|z|^2}{r}\right)^m p(r,z) = (4\pi r)^{-d/2} \left(\frac{|z|^2}{r}\right)^m e^{-\frac{|z|^2}{4r}} \lesssim c^{d/2} (4\pi c r)^{-d/2} e^{-\frac{1}{c}\frac{|z|^2}{4r}} \simeq p(cr,z). \qquad \text{(A.1)}$$

On the other hand we also have the trivial bound

$$|p(r,x+y-z_1)-p(r,x-z_1)|\leqslant p(r,x+y-z_1)+p(r,x-z_1),$$

and therefore by interpolation with $\kappa\in[0,1]$:

$$\begin{aligned}
&|p(r,x+y-z_1)-p(r,x-z_1)|\\
=&|p(r,x+y-z_1)-p(r,x-z_1)|^{\kappa}|p(r,x+y-z_1)-p(r,x-z_1)|^{1-\kappa}\\
\lesssim&(r^{-1/2}|y|)^{\kappa}\bigg(\int_0^1 p(cr,(x-z_1)+\lambda y)d\lambda\bigg)^{\kappa}(p(r,x+y-z_1)+p(r,x-z_1))^{1-\kappa}\\
\lesssim&(r^{-1/2}|y|)^{\kappa}\bigg(\int_0^1 p(cr,(x-z_1)+\lambda y)d\lambda+p(r,x+y-z_1)+p(r,x-z_1)\bigg),
\end{aligned}$$

where the last step follows from Young's inequality for products, $ab\leqslant\frac{1}{p}a^p+\frac{1}{m}b^m$, where $p,m>1$ are such that $\frac{1}{p}+\frac{1}{m}=1$, and which we applied with $p=\frac{1}{\kappa}$. We apply this bound with $\kappa=\beta$ and obtain with some more fiddling and using that $K\in\mathcal{K}^{\alpha}$

$$\begin{aligned}
&\int_0^{t+s}\int_{\mathbb{R}^d\times\mathbb{R}^d}(p(r,x+y-z_1)-p(r,x-z_1))(p(r,x+y-z_2)-p(r,x-z_2))K(dz_1,dz_2)dr\\
&\lesssim\int_0^{t+s}(r^{-1/2}|y|)^{2\beta}r^{-\alpha}dr\lesssim|y|^{2\kappa},
\end{aligned}$$

where we used that $-\beta-\alpha>-1$ and therefore the pole $r^{-\beta-\alpha}$ is integrable in $r=0$.

Similarly, we have for the time difference (substituting $t+s-r\to r$ in the first integral and $t-r\to r$ in the second integral

$$\begin{aligned}
&\mathbb{E}[|Z(t+s,x)-Z(t,x)|^2]\\
&=\int_0^s\int_{\mathbb{R}^d\times\mathbb{R}^d}p(r,x-z_1)p(r,x-z_2)K(dz_1,dz_2)dr\\
&\quad+\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}(p(r+s,x-z_1)-p(r,x-z_1))(p(r+s,x-z_2)-p(r,x-z_2))K(dz_1,dz_2)dr.
\end{aligned}$$

The first term is bounded by

$$\left|\int_0^s\int_{\mathbb{R}^d\times\mathbb{R}^d}p(r,x-z_1)p(r,x-z_2)K(dz_1,dz_2)dr\right|\lesssim\int_0^s r^{-\alpha}dr\lesssim s^{1-\alpha}\leqslant T^{1-\alpha-\beta}s^{\beta}\lesssim s^{\beta}.$$

For the second term we bound for $c>1$

$$\begin{aligned}
|p(r+s,z)-p(r,z)|&=\left|\int_0^1\partial_r p(r+\lambda s,z)s\,d\lambda\right|\\
&=\int_0^1\bigg(\frac{|z|^2}{4(r+\lambda s)^2}-\frac{d}{2}(4\pi(r+\lambda s))^{-1}\bigg)p(r+\lambda s,z)s\,d\lambda\\
&\lesssim\int_0^1\frac{s}{r+\lambda s}p(c(r+\lambda s),z)d\lambda\\
&\leqslant\frac{s}{r}\int_0^1 p(c(r+\lambda s),z)d\lambda,
\end{aligned}$$

where we applied (A.1) with $m=1$. By another interpolation argument this yields

$$\begin{aligned}
&\int_0^t\int_{\mathbb{R}^d\times\mathbb{R}^d}(p(r+s,x-z_1)-p(r,x-z_1))(p(r+s,x-z_2)-p(r,x-z_2))K(dz_1,dz_2)dr\\
&\lesssim\int_0^t\Big(\frac{s}{r}\Big)^{\beta}r^{-\alpha}dr\lesssim s^{\beta}.
\end{aligned}$$

This yields the claim. □

# Appendix B Motivation for the $\Phi_d^4$ model

Consider a domain $\mathcal{O} \subset \mathbb{R}^d$ and the potential

$$V(\phi) = \int_{\mathcal{O}} |\nabla \phi(x)|^2 dx + \int_{\mathcal{O}} \left(\frac{1}{2}\phi(x)^4 - \phi(x)^2\right) dx.$$

We are interested in the measure

$$\mu(\mathrm{d}\phi) = \frac{1}{Z} \exp(-V(\phi)) \prod_{x \in \mathcal{O}} d\phi(x), \tag{B.1}$$

where $\prod_{x \in \mathcal{O}} \mathrm{d}\phi(x)$ should be interpreted as the "Lebesgue measure" on the space of real-valued functions on $\mathcal{O}$ (which of course does not exist). This measure is called the $\Phi_d^4$ measure (4 because of the power $\phi(x)^4$ in the potential, $d$ because of the dimension of the base space), and it is important in quantum field theory, see for example [19].

It is also a sort of continuum version of the Ising model: Recall that the Ising model on the finite lattice $\Lambda \subset \mathbb{Z}^d$ is the measure on $\{-1,1\}^\Lambda$ given by the density

$$\frac{1}{Z} \exp\left(\sum_{i,j:|i-j|=1} \sigma_i \sigma_j\right) \prod_{i \in \Lambda} d\sigma_i = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{i,j:|i-j|=1} (\sigma_i - \sigma_j)^2\right) \prod_{i \in \Lambda} d\sigma_i,$$

where now $\prod_{i \in \Lambda} d\sigma_i$ is the counting measure on $\{-1,1\}^\Lambda$; the value of $Z$ will change throughout this discussion, it is always taken as the positive constant for which the total mass of the measure becomes 1. This is a model for a ferromagnet, and the spin $\sigma_i$ is the magnetization at point $i$ (positive or negative). The Ising model has the tendency of favoring configurations $\sigma \in \{-1,1\}^\Lambda$ for which neighboring spins are aligned, i.e. for which $\sigma_i = \sigma_j$ if $|i-j| = 1$. On the other hand, the counting measure introduces randomness.

In our case the potential $V$ is also small for $\phi$ with small gradient (i.e. for which $\int |\nabla \phi(x)|^2 \mathrm{dx}$ is small), and the "Lebesgue measure" introduces randomness. But now there is the additional contribution $\int_{\mathcal{O}} \left(\frac{1}{2}\phi(x)^4 - \phi(x)^2\right) dx$ to the potential, which can be motivated as follows: For the Ising model the spins $\sigma_i$ only take the values $\pm 1$, while the functions $\phi$ take arbitrary values in $\mathbb{R}$ and shifting $\phi$ by a constant does not change the value of $\int |\nabla \phi(x)|^2 \mathrm{d}x$. So the term $\int_{\mathcal{O}} \left(\frac{1}{2}\phi(x)^4 - \phi(x)^2\right) dx$ in a way anchors $\phi$ around the values $\pm 1$. Indeed, the double well potential is minimized exactly in the points $\pm 1$:

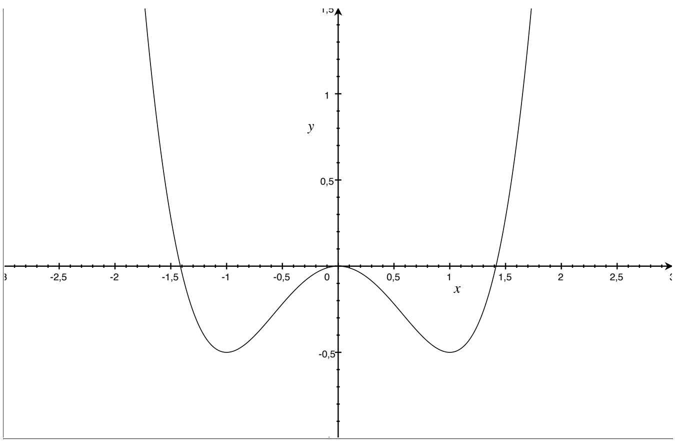


**Figure B.1.**

Of course, this whole discussion is purely formal because the Lebesgue measure $\prod_{x\in\mathcal{O}} d\phi(x)$ does not exist. To attempt making it rigorous, we could note that, at least if we ignore boundary terms from the integration by parts (this is for example ok if $\mathcal{O}=\mathbb{T}^d=(\mathbb{R}/\mathbb{Z})^d$ is the torus or if $\mathcal{O}=\mathbb{R}^d$ and we consider functions with sufficient decay at $\infty$)

$$\exp\bigg(-\int|\nabla\phi(x)|^2dx\bigg)=\exp\bigg(\int\phi(x)\Delta\phi(x)dx\bigg)=\exp\bigg(-\int\phi(x)(-\Delta)\phi(x)dx\bigg).$$

The operator $-\Delta$ is symmetric and positive, so this expression looks very much like a Gaussian density: If $C$ is a symmetric and strictly positive definite matrix on $\mathbb{R}^n$, then the measure

$$\pi(\mathrm{d}x)=\frac{1}{Z}\exp(-\langle x,Cx\rangle)\prod_i dx_i,$$

where $\langle\cdot,\cdot\rangle$ is the inner product on $\mathbb{R}^n$, is centered Gaussian with covariance $\frac{1}{2}C^{-1}$, i.e. for all $a,b$ we have

$$\int\langle x,a\rangle\langle x,b\rangle\pi(dx)=\frac{1}{2}\langle a,C^{-1}b\rangle.$$

So in our case we would expect

$$\nu(d\phi)=\frac{1}{Z}\exp\bigg(-\int|\nabla\phi(x)|^2dx\bigg)\prod_{x\in\mathcal{O}}d\phi(x)$$

to be a centered Gaussian measure with covariance

$$\int\langle\phi,f\rangle\langle\phi,g\rangle\nu(d\phi)=\frac{1}{2}\langle f,(-\Delta)^{-1}g\rangle=\frac{1}{2}\langle(-\Delta)^{-1/2}f,(-\Delta)^{-1/2}g\rangle,$$

where now $\langle\cdot,\cdot\rangle$ is the inner product on $L^2(\mathcal{O})$. Modulo technicalities (it is for example not always possible to invert $-\Delta$, and it may be better to consider $1-\Delta$ instead), such a measure exists, and it is called the *Gaussian free field*.

This brings us a good step closer towards making sense of the $\Phi^4_d$ measure that we formally defined in (B.1). A "rigorous candidate" seems to be

$$\mu(d\phi)=\frac{1}{Z}\exp\bigg(-\int\bigg(\frac{1}{2}\phi(x)^4-\phi(x)^2\bigg)dx\bigg)\nu(d\phi).$$

And indeed this measure is well defined if $d=1$. Unfortunately, for $d>1$ the measure $\nu$ is only supported on generalized functions (Schwartz distributions), and $\nu(L^p(\mathcal{O}))=0$ for all $p\in[1,\infty]$.

**Exercise. (Suggested by Scott Smith)** Consider the "massive Gaussian free field on $\mathbb{R}^d$", i.e. the centered Gaussian process $(\eta(\varphi))_{\varphi\in\mathscr{S}}$ with covariance

$$\mathbb{E}[\eta(\varphi)\eta(\psi)]=\frac{1}{2}\langle\varphi,(m-\Delta)^{-1}\psi\rangle_{L^2(\mathbb{R}^d)},$$

for mass $m>0$. The covariance is well defined, and therefore $\eta$ exists. Show that if $(\rho_n)\subset C^\infty_c$ is an approximation of the identity with $\mathrm{supp}(\rho_n)\subset B(0,1/n)$, then $\mathbb{E}[\eta(\rho_n(x-\cdot))^2]\to\infty$ if and only if $d>1$. This shows that the formal expression $\eta(x)=\eta(\delta(x-\cdot))$ for the Dirac delta does not exist (a sequence of Gaussian random variables with diverging variance cannot even converge in distribution).

**Solution.** We can represent

$$(m-\Delta)^{-1}f(x)=\int_0^\infty e^{-mt}\int_{\mathbb{R}^d}p(t,x-y)f(y)\mathrm{d}y\mathrm{d}t$$

for the centered Gaussian density with covariance $2t\mathbb{I}_d$, i.e.

$$p(t,x)=(4\pi t)^{-d/2}\exp\left(-\frac{|x|^2}{4t}\right).$$

Thus we get

$$\begin{aligned}&\langle\rho_n(x-\cdot),((m-\Delta)^{-1}\rho_n(x-\cdot))\rangle\\&=\int_{\mathbb{R}^d}\rho_n(x-z)\int_0^\infty e^{-mt}\int_{\mathbb{R}^d}p(t,x-z-y)\rho_n(y)\mathrm{d}y\mathrm{d}t\mathrm{d}z\\&=\int_{\mathbb{R}^d}\rho_n(z)\int_0^\infty e^{-mt}\int_{\mathbb{R}^d}p(t,z-y)\rho_n(y)\mathrm{d}y\mathrm{d}t\mathrm{d}z.\end{aligned}$$

We can restrict the integration to $|z|,|y|\leqslant 1/n$ because otherwise $\rho_n$ vanishes. But then $|z-y|\leqslant 2/n$ and thus

$$p(t,z-y)\gtrsim t^{-d/2}\exp\left(-\frac{4}{4tn^2}\right),$$

i.e.

$$\begin{aligned}\langle\rho_n(x-\cdot),((m-\Delta)^{-1}\rho_n(x-\cdot))\rangle&\gtrsim\int_{\mathbb{R}^d}\rho_n(z)\int_0^\infty e^{-mt}\int_{\mathbb{R}^d}t^{-d/2}\exp\left(-\frac{1}{tn^2}\right)\rho_n(y)\mathrm{d}y\mathrm{d}t\mathrm{d}z\\&=\int_0^\infty e^{-mt}t^{-d/2}\exp\left(-\frac{1}{tn^2}\right)\mathrm{d}t\\&\geqslant\int_{n^{-2}}^\infty e^{-mt}t^{-d/2}\exp(-1)\mathrm{d}t\gtrsim\int_{n^{-2}}^1 t^{-d/2}\mathrm{d}t,\end{aligned}$$

which diverges as $n\to\infty$ if $d\geqslant 2$.

On the other hand, for $d=1$ the uniform bound $p(t,x)\lesssim t^{-1/2}$ shows that the variance converges.

Since $\nu$ is only supported on distributions, it is unclear how to interpret the expression

$$\int\left(\frac{1}{2}\phi(x)^4-\phi(x)^2\right)dx$$

for typical $\phi\in\operatorname{supp}(\nu)$, and we are still stuck with the construction of $\phi$. This problem can be overcome for $d<4$ with the help of renormalization group techniques from quantum field theory [19], and one has to suitably renormalize the potential $\frac{1}{2}\phi(x)^4-\phi(x)^2$ by subtracting infinite counterterms. Let us stress that $d$ is the dimension of space-time so we would like to take $d=4$ for physical applications. Unfortunately this case is just out of reach of the existing theories!

Here we will follow a different route towards constructing $\mu$: We will construct a Markov process with invariant measure $\mu$. To understand how to do that, let us consider a smooth potential function $V\colon\mathbb{R}^n\to\mathbb{R}$, such that

$$\int\exp(-V(x))\prod_{i=1}^n dx_i<\infty,$$

for example $V(x)=x^2$. Then $\mu(dx)=\frac{1}{Z}\exp(V(x))\prod_i dx_i$ defines a probability measure, and it is a classical result that the following $n$-dimensional SDE, sometimes called *overdamped Langevin dynamics*, has $\mu$ as a reversible measure:

$$dX_t=-\frac{1}{2}\nabla V(X_t)dt+dW_t.$$

To formally show this, observe that the generator of $X$ is $\mathcal{L} = -\frac{1}{2}\nabla V \cdot \nabla + \frac{1}{2}\Delta$ and $\mathcal{L}$ is self-adjoint in $L^2(\mu)$ (this is not a proof of our claim, but at least it explains intuitively why $\mu$ should be reversible for $X$). In coordinates and formal notation we can also write

$$\partial_t X(t,i) = -\frac{1}{2}\partial_i V(X(t,\cdot)) + \xi(t,i),$$

where we set $\xi(f) = \sum_i \int_{\mathbb{R}_+} f(t,i) dW_t^i$ and thus we can interpret $\xi$ is a "space-time white noise on $\mathbb{R}_+ \times \{1,\ldots,n\}$", i.e. a centered Gaussian process indexed by $L^2(\mathbb{R}_+ \times \{1,\ldots,n\})$ (with product measure of Lebesgue measure and counting measure) such that

$$\mathbb{E}[\xi(f)\xi(g)] = \langle f, g\rangle = \sum_{i=1}^n \int_{\mathbb{R}_+} f(t,i) g(t,i) dt.$$

This provides a way of constructing $\mu$: simply let the dynamics of $X$ run and apply the ergodic theorem (this still requires ergodicity of $X$, which we take for granted here) to get

$$\int f(x)\mu(dx) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(X_s) ds.$$

If we apply the same philosophy to the $\Phi_d^4$ measure, we would formally get a dynamic process that satisfies

$$\partial_t \phi(t,x) = -\frac{1}{2}\delta_x V(\phi(t,\cdot)) + \xi(t,x) = \Delta\phi(t,x) - \phi(t,x)^3 + \phi(t,x) + \xi(t,x), \quad \text{(B.2)}$$

where $\xi$ is now a space-time white noise on $\mathbb{R}_+ \times \mathcal{O}$, i.e. $\xi$ is centered Gaussian and

$$\mathbb{E}[\xi(f)\xi(g)] = \langle f, g\rangle = \int_{\mathbb{R}_+ \times \mathcal{O}} f(t,x) g(t,x) dt dx$$

for all $f, g \in L^2(\mathbb{R}_+ \times \mathcal{O})$. The functional derivative $\delta_x V(\phi)$ is a bit subtle and can be computed by taking first the functional derivative in the direction of some function $\psi$, and then formally setting $\psi = \delta(x - \cdot)$ for the Dirac delta in $x$.

The equation (B.2) is called the $\Phi_d^4$ equation, or the *stochastic quantization equation* for the $\Phi_d^4$ measure [45]. We should not expect a free lunch though, and we cannot overcome the essential difficulties in the construction of the $\Phi_d^4$ measure by simply writing the problem in the language of stochastic differential equations. In fact (B.2) is still badly ill posed, because the solution $\phi(t,\cdot)$ is at fixed times $t$ only a distribution in the space variable, and therefore we have again a problem with the nonlinearity, this time given by $\phi^3$. In fact, at least for $d = 3$ the solution theory for the $\Phi_3^4$ *equation* (B.2) is much younger than the construction of the $\Phi_d^4$ measure, and it was one of the first big successes of Hairer's regularity structures [28] respectively paracontrolled distributions [22, 12]; the $\Phi_2^4$ equation had been previously solved with Dirichlet forms [1] or with the "Da Prato-Debussche method" [16].

**Bibliography:** The discussion is partially inspired by the nice survey papers [14, 40].

# Bibliography

**[1]** Sergio Albeverio and Michael Röckner. Stochastic differential equations in infinite dimensions: solutions via Dirichlet forms. *Probab. Theory Related Fields*, 89(3):347–386, 1991.

**[2]** Romain Allez and Khalil Chouk. The continuous Anderson hamiltonian in dimension two. *ArXiv preprint arXiv:1511.02718*, 2015.

**[3]** Hajer Bahouri, Jean-Yves Chemin, and Raphael Danchin. *Fourier analysis and nonlinear partial differential equations*. Springer, 2011.

**[4]** Ismaël Bailleul and Frédéric Bernicot. Higher order paracontrolled calculus. *ArXiv preprint arXiv:1609.06966*, 2016.

**[5]** Ismaël Bailleul, Frédéric Bernicot, and Dorothee Frey. Space-time paraproducts for paracontrolled calculus, 3d-PAM and multiplicative Burgers equations. *ArXiv preprint arXiv:1506.08773*, 2015.

**[6]** Carlo Bellingeri. *Itô formulae for the stochastic heat equation via the theories of rough paths and regularity structures*. PhD thesis, Sorbonne Université, 2019.

**[7]** Lorenzo Bertini and Giambattista Giacomin. Stochastic Burgers and KPZ equations from particle systems. *Comm. Math. Phys.*, 183(3):571–607, 1997.

**[8]** Jean-Michel Bony. Calcul symbolique et propagation des singularites pour les équations aux dérivées partielles non linéaires. *Ann. Sci. Éc. Norm. Supér. (4)*, 14:209–246, 1981.

**[9]** Y. Bruned, M. Hairer, and L. Zambotti. Algebraic renormalisation of regularity structures. *Invent. Math.*, 215(3):1039–1156, 2019.

**[10]** Yvain Bruned, Ajay Chandra, Ilya Chevyrev, and Martin Hairer. Renormalising SPDEs in regularity structures. *ArXiv preprint arXiv:1711.10239*, 2017.

**[11]** Giuseppe Cannizzaro, Peter K Friz, and Paul Gassiat. Malliavin calculus for regularity structures: the case of gPAM. *J. Funct. Anal.*, 272(1):363–419, 2017.

**[12]** Rémi Catellier and Khalil Chouk. Paracontrolled distributions and the 3-dimensional stochastic quantization equation. *Ann. Probab.*, 46(5):2621–2679, 2018.

**[13]** Ajay Chandra and Martin Hairer. An analytic BPHZ theorem for regularity structures. *ArXiv preprint arXiv:1612.08138*, 2016.

**[14]** Ajay Chandra and Hendrik Weber. Stochastic PDEs, regularity structures, and interacting particle systems. *Ann. Fac. Sci. Toulouse Math. (6)*, 26(4):847–909, 2017.

**[15]** Laure Coutin and Zhongmin Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields*, 122(1):108–140, 2002.

**[16]** Giuseppe Da Prato and Arnaud Debussche. Strong solutions to the stochastic quantization equations. *Ann. Probab.*, 31(4):1900–1916, 2003.

**[17]** Klaus Deimling. *Nonlinear functional analysis*. Springer-Verlag, Berlin, 1985.

**[18]** Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

**[19]** James Glimm and Arthur Jaffe. *Quantum physics. A functional integral point of view*. Springer-Verlag, New York, Second edition, 1987.

**[20]** M. Gubinelli, B. Ugurcan, and I. Zachhuber. Semilinear evolution equations for the Anderson Hamiltonian in two and three dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.*, 8(1):82–149, 2020.

**[21]** Massimiliano Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, nov 2004.

**[22]** Massimiliano Gubinelli, Peter Imkeller, and Nicolas Perkowski. Paracontrolled distributions and singular PDEs. *Forum of Mathematics, Pi*, 3(e6), 2015.

**[23]** Massimiliano Gubinelli, Herbert Koch, and Tadahiro Oh. Paracontrolled approach to the three-dimensional stochastic nonlinear wave equation with quadratic nonlinearity. *ArXiv preprint arXiv:1811.07808*, 2018.

**[24]** Massimiliano Gubinelli, Herbert Koch, and Tadahiro Oh. Renormalization of the two-dimensional stochastic nonlinear wave equations. *Trans. Amer. Math. Soc.*, 370(10):7335–7359, 2018.

**[25]** Massimiliano Gubinelli and Nicolas Perkowski. Lectures on singular stochastic PDEs. *Ensaios Mat.*, 29, 2015.

**[26]** Massimiliano Gubinelli and Nicolas Perkowski. KPZ reloaded. *Comm. Math. Phys.*, 349(1):165–269, 2017.

**[27]** Martin Hairer. Solving the KPZ equation. *Ann. Math.*, 178(2):559–664, 2013.

**[28]** Martin Hairer. A theory of regularity structures. *Invent. Math.*, 198(2):269–504, 2014.

**[29]** Martin Hairer. Introduction to regularity structures. *Braz. J. Probab. Stat.*, 29(2):175–210, 2015.

**[30]** Martin Hairer and Cyril Labbé. A simple construction of the continuum parabolic Anderson model on $\boldsymbol{R}^2$. *Electron. Commun. Probab.*, 20:0, 2015.

**[31]** Svante Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

**[32]** Wei Liu and Michael Röckner. *Stochastic partial differential equations: an introduction*. Universitext. Springer, Cham, 2015.

**[33]** Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.*, 14(2):215–310, 1998.

**[34]** Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

**[35]** Jörg Martin and Nicolas Perkowski. A Littlewood-Paley description of modelled distributions. *ArXiv preprint arXiv:1808.00500*, 2018.

**[36]** Jörg Martin and Nicolas Perkowski. Paracontrolled distributions on Bravais lattices and weak universality of the 2d parabolic Anderson model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):2058–2110, 2019.

**[37]** Jean-Christophe Mourrat and Hendrik Weber. Convergence of the two-dimensional dynamic Ising-Kac model to $\Phi^4_2$. *Comm. Pure Appl. Math.*, 70(4):717–812, 2017.

**[38]** Jean-Christophe Mourrat and Hendrik Weber. Global well-posedness of the dynamic $\Phi^4$ model in the plane. *Ann. Probab.*, 45(4):2398–2476, 2017.

**[39]** Jean-Christophe Mourrat and Hendrik Weber. The dynamic $\Phi^4_3$ model comes down from infinity. *Comm. Math. Phys.*, 356(3):673–753, 2017.

**[40]** Jean-Christophe Mourrat, Hendrik Weber, and Weijun Xu. Construction of $\Phi^4_3$ diagrams for pedestrians. In *From particle systems to partial differential equations*, volume 209 of *Springer Proc. Math. Stat.*, pages 1–46. Springer, Cham, 2017.

**[41]** Carl Mueller. On the support of solutions to the heat equation with noise. *Stochastics Stochastics Rep.*, 37(4):225–245, 1991.

**[42]** David Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, Second edition, 2006.

**[43]** David Nualart and Samy Tindel. A construction of the rough path above fractional Brownian motion using Volterra's representation. *The Annals of Probability*, 39(3):1061–1096, 2011.

**[44]** Etienne Pardoux. Stochastic partial differential equations. Unpublished lecture notes, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.405.4805rep=rep1type=pdf, 2007.

**[45]** G. Parisi and Yong Shi Wu. Perturbation theory without gauge fixing. *Sci. Sinica*, 24(4):483–496, 1981.

**[46]** Nicolas Simon Perkowski. *Studies of robustness in stochastic analysis and mathematical finance*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2014.

**[47]** David J. Prömel and Mathias Trabs. Rough differential equations driven by signals in Besov spaces. *J. Differential Equations*, 260(6):5202–5249, 2016.

**[48]** Jeremy Quastel. Introduction to KPZ. In *Current developments in mathematics, 2011*, pages 125–194. Int. Press, Somerville, MA, 2012.

**[49]** Michael Reed and Barry Simon. *Methods of modern mathematical physics. I*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, Second edition, 1980. Functional analysis.

**[50]** Jérémie Unterberger. A rough path over multidimensional fractional Brownian motion with arbitrary Hurst index by Fourier normal ordering. *Stochastic Processes and their Applications*, 120(8):1444–1472, 2010.

**[51]** John B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.