Lectures on singular stochastic PDEs

M. Gubinelli [*†] N. Perkowski [‡§]

Version 1.00 – January 2015

Abstract

These are the notes for a course at the 18th Brazilian School of Probability held from August 3rd to 9th, 2014 in Mambucaba. The aim of the course is to introduce the basic problems of non–linear PDEs with stochastic and irregular terms. We explain how it is possible to handle them using two main techniques: the notion of energy solutions [GJ10, GJ13] and that of paracontrolled distributions, recently introduced in [GIP13]. In order to maintain a link with physical intuitions, we motivate such singular SPDEs via a homogenization result for a diffusion in a random potential.

Contents

[*] *Affiliation:* CEREMADE & CNRS UMR 7534; Université Paris Dauphine and IUF, France
[†] *Email:* gubinelli@ceremade.dauphine.fr
[‡] *Affiliation:* CEREMADE & CNRS UMR 7534; Université Paris Dauphine
[§] *Email:* perkowski@ceremade.dauphine.fr

1 Introduction

The aim of these lectures is to explain how to apply controlled path ideas [Gub04] to solve simple problems in singular PDEs. The hope is that the insight gained by doing so can inspire new applications or the construction of other more powerful tools to analyze a wider class of problems.

We discuss some problems involving singular stochastic non–linear parabolic equations from the point of view of controlled paths. To understand the origin of such singular equations, we have chosen to present the example of a homogenization problem for a singular potential in a linear parabolic equation. This point of view has the added benefit that it allows us to track back the renormalization needed to handle the singularities as effects living on other scales than those of interest. The basic problem is that of having to handle effects of the microscopic scales and their interaction through non–linearities on the macroscopic behavior of the solution.

Mathematically, this problem translates into the attempt of making Schwartz's theory of distribution coexist with non–linear operations which are notoriously not continuous in the usual topologies on distributions. This is a very old problem of analysis and has been widely studied. The additional input which is not present in the usual approaches is that the singularities which force us to treat the problem in the setting of Schwartz's distributions are of a stochastic nature. So we dispose of two handles on the problem: the analytical one and the probabilistic one. The right mix of the two will provide an effective solution to a wide class of problems.

A first and deep understanding of these problems has been obtained starting from the late '90s by T. Lyons [Lyo98], who introduced a theory of *rough paths* in order to settle the conflict of topology and non–linearity in the context of driven differential equations, or more generally in the context of the non–linear analysis of time–varying signals. Nowadays there are many expositions of this theory [LQ02, FV10, LCL07, FH14] and we refer the reader to the literature for more details.

In [Gub04, Gub10], the notion of *controlled paths* has been introduced in order to extend the applicability of the rough path ideas to a larger class of problems that are not necessarily related to the integration of ODEs but which still retain the one–dimensional nature of the directions in which the irregularity manifest itself. The controlled path approach has been used to make sense of the evolution of irregular objects such as vortex filaments and certain SPDEs. Later Hairer understood how to apply these ideas to the long standing problem of the Kardar–Parisi–Zhang equation [Hai13], and his insights prompted the researchers to try more ambitious approaches to extend rough paths to a multidimensional setting.

In [GIP13], in collaboration with P. Imkeller, we introduced a notion of *paracontrolled distributions* which is suitable suitable to handle a wide class of SPDEs which were well out of reach with previously known methods. Paracontrolled distributions can be understood as an extension of controlled paths to a multidimensional setting, and they are based on new combinations of basic tools from harmonic analysis.

At the same time, Hairer managed to devise a vast generalization of the basic construction of controlled rough paths in the multidimensional and distributional setting, which he called the theory of *regularity structures* [Hai14] and which subsumes standard analysis based on Hölder spaces and controlled rough path theory but goes well beyond that. Just few days after the lectures in Mambucaba took place, it was announced that Martin Hairer was awarded a Fields Medal for his work on SPDEs and in particular for his theory of regularity structures [Hai14] as a tool for dealing with singular SPDEs. This prize witnesses the exciting period we are experiencing: we now understand sound lines of attack to long standing problems, and there are countless opportunities to apply similar ideas to new problems.

The plan of the lectures is the following. We start by explaining "energy solutions" [GJ10, GJ13], a notion of solution to (a particular class of) singular PDEs which has the advantage of being comparably easy to handle but which also has the inconvenience of not having a comprehensive uniqueness theory so far. This will allow us to introduce the reader to SPDEs in a progressive manner, and also to introduce Gaussian tools (Wick products, hypercontractivity) and to present some of the basic phenomena appearing when dealing with singular SPDEs. Next we set up the analytical tools we need in the rest of the lectures: Besov spaces and some basic harmonic analysis based on the Littlewood–Paley decomposition. Next, in order to motivate the reader and to provide a physical ground for the intuition to stand on, we discuss a homogenization problem for the linear heat equation with random potential. This will allow us to derive the need for the weak topologies we shall use and for irregular objects like white–noise from first principles and "concrete" applications. The homogenization problem also allows us to see the renormalization effects appear naturally and to track their mathematical mean-

ing. Starting from these problems we introduce the two–dimensional parabolic Anderson model, the simplest SPDE in which most of the features of more difficult problems are already present, and we explain how to use paraproducts and the paracontrolled ansatz in order to keep the non–linear effect of the singular data under control. Then we go on to discuss the more involved situation of the Stochastic Burgers equation in one dimension, which is one of the avatars of the Kardar–Parisi–Zhang equation.

Conventions and notations We write $a \lesssim b$ if there exists a constant $C > 0$, independent of the variables under consideration, such that $a \leqslant Cb$. Similarly we define $\gtrsim$. We write $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$. If we want to emphasize the dependence of C on the variable x, then we write $a(x) \lesssim_x b(x)$.

If i and j are index variables of Littlewood–Paley blocks (to be defined below), then $i \lesssim j$ is to be interpreted as $2^i \lesssim 2^j$, and similarly for $\simeq$ and $\lesssim$. In other words, $i \lesssim j$ means $i \leqslant j + N$ for some fixed $N \in \mathbb{N}$ that does not depend on i or j.

We use standard multi-index notation: for $\mu \in \mathbb{N}_0^d$ we write $|\mu| = \mu_1 + \ldots + \mu_d$ and $\partial^\mu = \partial^{|\mu|}/\partial_{x_1}^{\mu_1} \ldots \partial_{x_d}^{\mu_d}$, as well as $x^\mu = x_1^{\mu_1} \cdot \ldots \cdot x_d^{\mu_d}$ for $x \in \mathbb{R}^d$.

For $\alpha > 0$ we write C_b^α for the functions $F : \mathbb{R} \to \mathbb{R}$ which are $\lfloor \alpha \rfloor$ times continuously differentiable with $(\alpha - \lfloor \alpha \rfloor)$–Hölder continuous derivatives of order $\lfloor \alpha \rfloor$, equipped with its natural norm $\| \cdot \|_{C_b^\alpha}$.

If we write $u \in \mathscr{C}^{\alpha-}$, then that means that u is in $\mathscr{C}^{\alpha-\varepsilon}$ for all $\varepsilon > 0$. The $\mathscr{C}^\alpha$ spaces will be defined below.

2 Energy solutions

The first issue one encounters when dealing with singular SPDEs is the ill–posed character of the equation, even in a weak sense. Typically, the non–linearity does not make sense in the natural spaces where solutions live and one has to provide a suitable smaller space in which it is possible to give an appropriate interpretation to "ambiguous quantities" that are featured in the equation.

Energy solutions [GJ10, GJ13] are a relatively simple tool in order to come up with well–defined non–linearities. The drawback is that currently the issue of uniqueness is open, at least in the most interesting cases. It is not clear if uniqueness of energy solutions holds or even how to find conditions that ensure uniqueness. On the other hand, proving existence of energy solutions or even convergence to energy solutions is usually a quite simple problem, at least compared to the other approaches like paracontrolled solutions or regularity structures, where existence requires already quite a large amount of computations but where uniqueness can be established quite easily afterwards.

Our aim here is to motivate the ideas leading to the notion of energy solutions. We will not insist on a detailed formulation of all the available results. The reader can always

refer to the original paper [GJ13] for missing details. Applications to the large scale behavior of particle systems are studied in [GJ10, GJ14].

We will study energy solutions for the stochastic Burgers equation on the torus $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$. The unknown $u : \mathbb{R}_+ \times \mathbb{T} \to \mathbb{R}$ should satisfy

$$\partial_t u = \Delta u + \partial_x u^2 + \partial_x \xi,$$

where $\xi : \mathbb{R}_+ \times \mathbb{T} \to \mathbb{R}$ is a space–time white noise defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ fixed once and for all. That is, ξ is a centered Gaussian process indexed by $L^2(\mathbb{R}_+ \times \mathbb{T})$ with covariance

$$\mathbb{E}[\xi(f)\xi(g)] = \int_{\mathbb{R}_+ \times \mathbb{T}} f(t,x)g(t,x)\mathrm{d}t\mathrm{d}x.$$

The equation has to be understood as a relation for processes which are distributions in space with sufficiently regular time dependence (for a short introduction to distributions on $\mathbb{T}$ see Section 3 below). In particular, if we test the above relation with $\varphi \in \mathscr{S} := \mathscr{S}(\mathbb{T}) := C^\infty(\mathbb{T})$, denote with $u_t(\varphi)$ the pairing of the distribution $u(t,\cdot)$ with φ, and integrate in time over the interval $[0,t]$, we get

$$u_t(\varphi) = u_0(\varphi) + \int_0^t u_s(\Delta\varphi)\mathrm{d}s - \int_0^t \langle u_s^2, \partial_x\varphi\rangle \mathrm{d}s - \int_0^t \xi_s(\partial_x\varphi)\mathrm{d}s.$$

Let us discuss the various terms in this equation. In order to make sense of $u_t(\varphi)$ and $\int_0^t u_s(\Delta\varphi)\mathrm{d}s$, it is enough to assume that for all $\varphi \in \mathscr{S}$ the mapping $(t,\omega) \mapsto u_t(\varphi)(\omega)$ is a stochastic process with continuous trajectories. Next, if we denote $M_t(\varphi) = \int_0^t \xi_s(\partial_x\varphi)\mathrm{d}s$ then, at least by a formal computation, we have that $(M_t(\varphi))_{t\geqslant 0, \varphi\in\mathscr{S}}$ is a Gaussian random field with covariance

$$\mathbb{E}[M_t(\varphi)M_s(\psi)] = (t \wedge s)\langle \partial_x\varphi, \partial_x\psi\rangle_{L^2(\mathbb{T})}.$$

In particular, for every $\varphi \in \mathscr{S}$ the stochastic process $(M_t(\varphi))_{t\geqslant 0}$ is a Brownian motion with covariance

$$\|\varphi\|^2_{H^1(\mathbb{T})} := \langle \partial_x\varphi, \partial_x\varphi\rangle_{L^2(\mathbb{T})}.$$

Here we used the notation M in order to stress the fact that $M_t(\varphi)$ is a martingale in its natural filtration and more generally in the filtration $\mathcal{F}_t = \sigma(M_s(\varphi) : s \leqslant t, \varphi \in H^1(\mathbb{T}))$, $t \geqslant 0$.

The most difficult term is of course the nonlinear one: $\int_0^t \langle u_s^2, \partial_x\varphi\rangle \mathrm{d}s$. In order to define it, we need to square the distribution u_t, an operation which in general can be quite dangerous. A natural approach would be to define it as the limit of some regularizations. For example, if $\rho : \mathbb{R} \to \mathbb{R}_+$ is a compactly supported C^∞ function such that $\int_{\mathbb{R}} \rho(x)\mathrm{d}x = 1$, and we set $\rho_\varepsilon(\cdot) = \rho(\cdot/\varepsilon)/\varepsilon$, then we can set $\mathcal{N}_{t,\varepsilon}(u)(x) = \int_0^t (\rho_\varepsilon * u_s)(x)^2\mathrm{d}s$ and define $\mathcal{N}_t(u) = \lim_{\varepsilon\to 0} \mathcal{N}_{t,\varepsilon}(u)$ whenever the limit exists in $\mathscr{S}' := \mathscr{S}'(\mathbb{T})$, the space of distributions on $\mathbb{T}$. Then the question arises which properties u should have for this convergence to occur.

2.1 The Ornstein–Uhlenbeck process

Let us simplify the problem and start by studying the linearized equation obtained by neglecting the non–linear term. Let X be a solution to

$$X_t(\varphi) = X_0(\varphi) + \int_0^t X_s(\Delta\varphi)\mathrm{d}s + M_t(\varphi) \tag{1}$$

for all $t \geqslant 0$ and $\varphi \in \mathscr{S}$. This equation has *at most* one solution (for fixed X_0). Indeed, the difference D between two solutions should satisfy $D_t(\varphi) = \int_0^t D_s(\Delta\varphi)\mathrm{d}s$, which means that D is a distributional solution to the heat equation. Taking $\varphi(x) = e_k(x)$, where

$$e_k(x) := \exp(-ikx)/\sqrt{2\pi}, \qquad k \in \mathbb{Z},$$

we get $D_t(e_k) = -k^2 \int_0^t D_s(e_k)\mathrm{d}s$ and then by Gronwall's inequality $D_t(e_k) = 0$ for all $t \geqslant 0$. This easily implies that $D_t = 0$ in $\mathscr{S}'$ for all $t \geqslant 0$.

To obtain the *existence* of a solution, observe that

$$X_t(e_k) = X_0(e_k) - k^2 \int_0^t X_s(e_k)\mathrm{d}s + M_t(e_k)$$

and that $M_t(e_0) = 0$, while for all $k \neq 0$ the process $\beta_t(k) = M_t(e_k)/(-ik)$ is a complex valued Brownian motion (i.e. real and imaginary part are independent Brownian motions with the same variance). The covariance of β is given by

$$\mathbb{E}[\beta_t(k)\beta_s(m)] = (t \wedge s)\delta_{k+m=0}$$

and moreover $\beta_t(k)^* = \beta_t(-k)$ for all $k \neq 0$ (where $\cdot^*$ denotes complex conjugation), as well as $\beta_t(0) = 0$. In other words, $(X_t(e_k))$ is a complex–valued Ornstein–Uhlenbeck process which solves a linear one–dimensional SDE and has an explicit representation given by the variation of constants formula

$$X_t(e_k) = e^{-k^2 t} X_0(e_k) + ik \int_0^t e^{-k^2(t-s)} \mathrm{d}_s\beta_s(k).$$

This is enough to determine $X_t(\varphi)$ for all $t \geqslant 0$ and $\varphi \in \mathscr{S}$.

Exercise 1 Show that $(X_t(e_k) : t \in \mathbb{R}_+, k \in \mathbb{Z})$ is a complex Gaussian random field, that is for all $n \in \mathbb{N}$, for all $t_1, \ldots, t_n \in \mathbb{R}_+$, $k_1, \ldots, k_n \in \mathbb{Z}$, the vector

$$(\mathrm{Re}(X_{t_1}(k_1)), \ldots, \mathrm{Re}(X_{t_n}(k_n)), \mathrm{Im}(X_{t_1}(k_1)), \ldots, \mathrm{Im}(X_{t_n}(k_n)))$$

is multivariate Gaussian. Show that X has mean $\mathbb{E}[X_t(e_k)] = e^{-k^2 t} X_0(e_k)$ and covariance

$$\mathbb{E}[(X_t(e_k) - \mathbb{E}[X_t(e_k)])(X_s(e_m) - \mathbb{E}[X_s(e_m)])] = k^2\delta_{k+m=0} \int_0^{t\wedge s} e^{-k^2(t-r)-k^2(s-r)}\mathrm{d}r$$

as well as

$$\mathbb{E}[(X_t(e_k) - \mathbb{E}[X_t(e_k)])(X_s(e_m) - \mathbb{E}[X_s(e_m)])^*] = k^2\delta_{k=m}\int_0^{t\wedge s} e^{-k^2(t-r)-k^2(s-r)}\mathrm{d}r.$$

In particular,

$$\mathbb{E}[|X_t(e_k) - \mathbb{E}[X_t(e_k)]|^2] = \frac{1-e^{-2k^2t}}{2}.$$

Next we examine the Sobolev regularity of X. For this purpose, we need the following definition.

Definition 1 *Let $\alpha \in \mathbb{R}$. Then the Sobolev space H^α is defined as*

$$H^\alpha := H^\alpha(\mathbb{T}) := \left\{\rho \in \mathscr{S}' : \|\rho\|_{H^\alpha} := \sum_{k\in\mathbb{Z}}(1+|k|^2)^\alpha|\rho(e_k)|^2 < \infty\right\}.$$

We also write CH^α for the space of continuous functions from $\mathbb{R}_+$ to H^α.

Lemma 1 *Let $\gamma \leqslant -1/2$ and assume that $X_0 \in H^\gamma$. Then almost surely $X \in CH^{\gamma-}$.*

Proof Let $\alpha = \gamma - \varepsilon$ and consider

$$\|X_t - X_s\|_{H^\alpha}^2 = \sum_{k\in\mathbb{Z}}(1+|k|^2)^\alpha|X_t(e_k) - X_s(e_k)|^2.$$

Let us estimate the $L^{2p}(\Omega)$ norm of this quantity for $p \in \mathbb{N}$ by writing

$$\mathbb{E}\|X_t - X_s\|_{H^\alpha}^{2p} = \sum_{k_1,\dots,k_p\in\mathbb{Z}}\prod_{i=1}^p(1+|k_i|^2)^\alpha\mathbb{E}\prod_{i=1}^p|X_t(e_{k_i}) - X_s(e_{k_i})|^2.$$

By Cauchy–Schwartz, we get

$$\mathbb{E}\|X_t - X_s\|_{H^\alpha}^{2p} \lesssim \sum_{k_1,\dots,k_p\in\mathbb{Z}}\prod_{i=1}^p(1+|k_i|^2)^\alpha\prod_{i=1}^p(\mathbb{E}|X_t(e_{k_i}) - X_s(e_{k_i})|^{2p})^{1/p}.$$

Note now that $X_t(e_{k_i}) - X_s(e_{k_i})$ is a Gaussian random variable, so that there exists a universal constant C_p for which

$$\mathbb{E}|X_t(e_{k_i}) - X_s(e_{k_i})|^{2p} \leqslant C_p(\mathbb{E}|X_t(e_{k_i}) - X_s(e_{k_i})|^2)^p.$$

Moreover,

$$X_t(e_k) - X_s(e_k) = (e^{-k^2(t-s)} - 1)X_s(e_k) + ik\int_s^t e^{-k^2(t-r)}\mathrm{d}_r\beta_r(k),$$

leading to

$$\begin{aligned}\mathbb{E}|X_t(e_k) - X_s(e_k)|^2 &= (e^{-k^2(t-s)} - 1)^2 \mathbb{E}|X_s(e_k)|^2 + k^2 \int_s^t e^{-2k^2(t-r)} \mathrm{d}r \\ &= (e^{-k^2(t-s)} - 1)^2 e^{-2k^2 s} |X_0(e_k)|^2 + (e^{-k^2(t-s)} - 1)^2 k^2 \int_0^s e^{-2k^2(s-r)} \mathrm{d}r \\ &\quad + k^2 \int_s^t e^{-2k^2(t-r)} \mathrm{d}r \\ &= (e^{-k^2 t} - e^{-k^2 s})^2 |X_0(e_k)|^2 + \frac{1}{2}(e^{-k^2(t-s)} - 1)^2 (1 - e^{-2k^2 s}) + \frac{1}{2}(1 - e^{-2k^2(t-s)}).\end{aligned}$$

For any $\kappa \in [0, 1]$ and $k \neq 0$, we thus have

$$\mathbb{E}|X_t(e_k) - X_s(e_k)|^2 \lesssim (k^2(t-s))^\kappa (|X_0(e_k)|^2 + 1),$$

while for $k = 0$ we have $\mathbb{E}|X_t(e_0) - X_s(e_0)|^2 = 0$. Therefore,

$$\begin{aligned}\mathbb{E}\|X_t - X_s\|_{H^\alpha}^{2p} &\lesssim \sum_{k_1,\dots,k_p \in \mathbb{Z}_0} \prod_{i=1}^p (1 + |k_i|^2)^\alpha \prod_{i=1}^p \mathbb{E}|X_t(e_{k_i}) - X_s(e_{k_i})|^2 \\ &\lesssim (t-s)^{\kappa p} \sum_{k_1,\dots,k_p \in \mathbb{Z}_0} \prod_{i=1}^p (1 + |k_i|^2)^\alpha (k_i^2)^\kappa (|X_0(e_{k_i})|^2 + 1) \\ &\lesssim (t-s)^{\kappa p} \Big[\sum_{k \in \mathbb{Z}_0} (1 + |k|^2)^\alpha (k^2)^\kappa (|X_0(e_k)|^2 + 1) \Big]^p \\ &\lesssim (t-s)^{\kappa p} \Big(\|X_0\|_{H^{\alpha+\kappa}(\mathbb{T})}^{2p} + \Big[\sum_{k \in \mathbb{Z}_0} (1 + |k|^2)^\alpha (k^2)^\kappa \Big]^p \Big).\end{aligned}$$

If $\alpha < -1/2 - \kappa$, the sum on the right hand side is finite and we obtain an estimation for the modulus of continuity of $t \mapsto X_t$ in $L^{2p}(\Omega; H^\alpha)$:

$$\mathbb{E}\|X_t - X_s\|_{H^\alpha}^{2p} \lesssim (t-s)^{\kappa p} [1 + \|X_0\|_{H^{\alpha+\kappa}}^{2p}].$$

Now Kolmogorov's continuity criterion allows us to conclude that almost surely $X \in CH^\alpha$ whenever $X_0 \in H^{\alpha+\kappa}$. □

Now note that the regularity of the Ornstein–Uhlenbeck process does not allow us to form the quantity X_t^2 point–wise in time since by Fourier inversion $X_t = \sum_k X_t(e_k) e_k^*$, and therefore we should have

$$X_t^2(e_k) = (2\pi)^{-1/2} \sum_{\ell+m=k} X_t(e_\ell) X_t(e_m).$$

Of course, at the moment this expression is purely formal since we cannot guarantee that the infinite sum converges. A reasonable thing to try is to approximate the square

by regularizing the distribution, taking the square, and then trying to remove the regularization. Let Π_N be the projector of a distribution onto a finite number of Fourier modes:

$$(\Pi_N \rho)(x) = \sum_{|k| \leqslant N} \rho(e_k) e_k^*(x).$$

Then $\Pi_N X_t(x)$ is a smooth function of x and we can consider $(\Pi_N X_t)^2$ which satisfies

$$(\Pi_N X_t)^2(e_k) = (2\pi)^{-1/2} \sum_{\ell+m=k} \mathbb{I}_{|\ell| \leqslant N, |m| \leqslant N} X_t(e_\ell) X_t(e_m).$$

We would then like to take the limit $N \to +\infty$. For convenience, we will perform the computations below in the limit $N = +\infty$, but one has to come back to the case of finite N in order to make it rigorous.

Let us introduce the notation $\mathbb{Z}_0 = \mathbb{Z} \setminus \{0\}$. Then

$$\begin{aligned}\mathbb{E}[X_t^2(e_k)] &= (2\pi)^{-1/2} \delta_{k=0} \sum_{m \in \mathbb{Z}_0} \mathbb{E}[X_t(e_{-m}) X_t(e_m)] \\ &= (2\pi)^{-1/2} \delta_{k=0} \sum_{m \in \mathbb{Z}_0} e^{-2m^2 t} |X_0(e_m)|^2 + (2\pi)^{-1/2} \delta_{k=0} \sum_{m \in \mathbb{Z}_0} m^2 \int_0^t e^{-2m^2(t-s)} \mathrm{d}s\end{aligned}$$

and

$$\sum_{m \in \mathbb{Z}_0} m^2 \int_0^t e^{-2m^2(t-s)} \mathrm{d}s = \frac{1}{2} \sum_{m \in \mathbb{Z}_0} (1 - e^{-2m^2 t}) = +\infty.$$

This is not really a problem since in the equation only components of $u_t^2(e_k)$ with $k \neq 0$ appear. However, $X_t^2(e_k)$ is not even a well–defined random variable. For the moment, let us assume that $X_0 = 0$, which will slightly simplify the computation. If $k \neq 0$, we have

$$\mathbb{E}[|X_t^2(e_k)|^2] = \mathbb{E}[X_t^2(e_k) X_t^2(e_{-k})] = (2\pi)^{-1} \sum_{\ell+m=k} \sum_{\ell'+m'=-k} \mathbb{E}[X_t(e_\ell) X_t(e_m) X_t(e_{\ell'}) X_t(e_{m'})].$$

By Wick's theorem (see [Jan97], Theorem 1.28), the expectation can be computed in terms of the covariances of all possible pairings of the four Gaussian random variables (3 possible combinations):

$$\begin{aligned}\mathbb{E}[X_t(e_\ell) X_t(e_m) X_t(e_{\ell'}) X_t(e_{m'})] = {} & \mathbb{E}[X_t(e_\ell) X_t(e_m)] \mathbb{E}[X_t(e_{\ell'}) X_t(e_{m'})] \\ & + \mathbb{E}[X_t(e_\ell) X_t(e_{\ell'})] \mathbb{E}[X_t(e_m) X_t(e_{m'})] \\ & + \mathbb{E}[X_t(e_\ell) X_t(e_{m'})] \mathbb{E}[X_t(e_m) X_t(e_{\ell'})].\end{aligned}$$

Since $k \neq 0$, we have $\ell + m \neq 0$ and $\ell' + m' \neq 0$ which allows us to neglect the first term since it is zero. By symmetry of the summations, the two other give the same

contribution and we remain with

$$\begin{aligned}\mathbb{E}[|X_t^2(e_k)|^2] &= 2\sum_{\ell+m=k}\sum_{\ell'+m'=k}\mathbb{E}[X_t(e_\ell)X_t(e_{\ell'})]\mathbb{E}[X_t(e_m)X_t(e_{m'})] \\ &= 2\sum_{\ell+m=k}\mathbb{E}[X_t(e_\ell)X_t(e_{-\ell})]\mathbb{E}[X_t(e_m)X_t(e_{-m})] \\ &= \frac{1}{2}\sum_{\ell+m=k}(1-e^{-2\ell^2 t})(1-e^{-2m^2 t}) = +\infty.\end{aligned} \tag{2}$$

This shows that even when tested against smooth test functions, X_t^2 is not in $L^2(\Omega)$. This indicates that there are problems with X_t^2 and indeed one can show that $X_t^2(e_k)$ does not make sense as a random variable.

To understand this better, observe that the Ornstein–Uhlenbeck process can be decomposed as

$$X_t(e_k) = ik\int_{-\infty}^t e^{-k^2(t-s)}\mathrm{d}\beta_s(k) - ike^{-k^2 t}\int_{-\infty}^0 e^{k^2 s}\mathrm{d}\beta_s(k),$$

where we extended the Brownian motions $(\beta_s(k))_{s\geqslant 0}$ to two sided complex Brownian motions by considering independent copies. It is not difficult to show that the second term gives rise to a smooth function if $t > 0$, so all the irregularity of X_t is described by the first term which we call $Y_t(e_k)$. Note that $Y_t(e_k) \sim \mathcal{N}_{\mathbb{C}}(0, 1/2)$ for all $k \in \mathbb{Z}_0$ and $t \in \mathbb{R}$, where we write

$$U \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$$

if $U = V + iW$, where V and W are independent random variables with distribution $\mathcal{N}(0, \sigma^2/2)$. The random distribution Y_t then satisfies $Y_t(\varphi) \sim \mathcal{N}(0, \|\varphi\|^2_{L^2(\mathbb{T})}/2)$, or in other words it is ($1/\sqrt{2}$ times) the white noise on $\mathbb{T}$. It is also possible to deduce that the white noise on $\mathbb{T}$ is indeed the invariant measure of the Ornstein–Uhlenbeck process, that it is the only one, and that it is approached quite fast.

So we should expect that, at fixed time, the regularity of the Ornstein–Uhlenbeck process is like that of the space white noise and this is a way of understanding our difficulties in defining X_t^2 since this will be, modulo smooth terms, the square of the space white noise.

A different matter is to make sense of the time–integral of $\partial_x X_t^2$. Let us give it a name and call it $J_t(\varphi) = \int_0^t \partial_x X_s^2(\varphi)\mathrm{d}s$. For $J_t(e_k)$, the computation of its variance gives a quite different result.

Lemma 2 *Almost surely, $J \in C^{1/2-}H^{-1/2-}$.*

Proof Proceeding as in (2), we have now

$$\mathbb{E}[|J_t(e_k)|^2] = 2k^2 \int_0^t \int_0^t \sum_{\ell+m=k} \mathbb{E}[X_s(e_\ell)X_{s'}(e_{-\ell})]\mathbb{E}[X_s(e_m)X_{s'}(e_{-m})]\mathrm{d}s\mathrm{d}s'.$$

If $s > s'$, we have

$$\mathbb{E}[X_s(e_\ell)X_{s'}(e_{-\ell})] = \frac{1}{2}e^{-\ell^2(s-s')}(1 - e^{-2\ell^2 s'}),$$

and therefore

$$\begin{aligned}
\mathbb{E}[|J_t(e_k)|^2] &= \frac{k^2}{2} \int_0^t \int_0^t \sum_{\ell+m=k} e^{-(\ell^2+m^2)|s-s'|}(1 - e^{-2\ell^2(s'\wedge s)})(1 - e^{-2m^2(s'\wedge s)})\mathrm{d}s\mathrm{d}s' \\
&\leqslant \frac{k^2}{2} \int_0^t \int_0^t \sum_{\ell+m=k} e^{-(\ell^2+m^2)|s-s'|}\mathrm{d}s\mathrm{d}s' \leqslant k^2 t \sum_{\ell+m=k} \int_0^\infty e^{-(\ell^2+m^2)r}\mathrm{d}r \\
&= k^2 t \sum_{\ell+m=k} \frac{1}{\ell^2 + m^2}.
\end{aligned}$$

Now for $k \neq 0$

$$\sum_{\ell+m=k} \frac{1}{\ell^2 + m^2} \lesssim \int_{\mathbb{R}} \frac{\mathrm{d}x}{x^2 + (k-x)^2} \lesssim \frac{1}{|k|},$$

so finally $\mathbb{E}[|J_t(e_k)|^2] \lesssim |k|t$. Redoing a similar computation in the case $J_t(e_k) - J_s(e_k)$, we obtain $\mathbb{E}[|J_t(e_k) - J_s(e_k)|^2] \lesssim |k| \times |t - s|$. To go from this estimate to a path–wise regularity result of the distribution $(J_t)_t$, following the line of reasoning of Lemma 1, we need to estimate the p-th moment of $J_t(e_k) - J_s(e_k)$. Gaussian hypercontractivity (see Theorem 3.50 of [Jan97]) tells us that all L^p moments of polynomials of Gaussian random variables are equivalent and in particular that

$$\mathbb{E}[|J_t(e_k) - J_s(e_k)|^{2p}] \lesssim_p (\mathbb{E}[|J_t(e_k) - J_s(e_k)|^2])^p.$$

From here we easily derive that almost surely $J \in C^{1/2-}H^{-1/2-}$. □

This shows that $\partial_x X_t^2$ exists as a space–time distribution but not as a continuous function of time with values in distributions in space. The key point in the proof of Lemma 2 is the fact that the Ornstein–Uhlenbeck process decorrelates quite rapidly in time.

The construction of the process J does not solve our problem since we need similar properties for the full solution u of the non–linear dynamics (or for some approximations thereof), and all we have done so far relies on explicit computations and the specific Gaussian features of the Ornstein–Uhlenbeck process. But at least this give us a hint that indeed there could exist a way of making sense of the term $\partial_x u(t,x)^2$, even if only as a space–time distribution, and that in doing so we should exploit some decorrelation properties of the dynamics.

So when dealing with the full solution u, we need a replacement for the Gaussian computations used above. This will be provided, in the current setting, by stochastic calculus along the time direction. Indeed, note that for each $\varphi \in \mathscr{S}$ the process $(X_t(\varphi))_{t\geqslant 0}$ is a semimartingale in the filtration $(\mathcal{F}_t)_{t\geqslant 0}$.

Before proceeding with these computations, we need to develop some tools to describe the Itô formula for functions of the Ornstein–Uhlenbeck process. This will be also serve us as an opportunity to set up some analysis on Gaussian spaces.

2.2 Gaussian computations

For cylindrical functions $F : \mathscr{S}' \to \mathbb{R}$ of the form $F(\rho) = f(\rho(\varphi_1), \dots, \rho(\varphi_n))$ with $\varphi_1, \dots, \varphi_n \in \mathscr{S}$ and $f : \mathbb{R}^n \to \mathbb{R}$ at least C_b^2, we have by Itô's formula

$$\mathrm{d}_t F(X_t) = \sum_{i=1}^n F_i(X_t)\mathrm{d}X_t(\varphi_i) + \frac{1}{2}\sum_{i,j=1}^n F_{i,j}(X_t)\mathrm{d}\langle X(\varphi_i), X(\varphi_j)\rangle_t,$$

where $\langle\rangle_t$ denotes the quadratic covariation of two continuous semimartingales and where $F_i(\rho) = \partial_i f(\rho(\varphi_1), \dots, \rho(\varphi_n))$ and $F_{i,j}(\rho) = \partial^2_{i,j} f(\rho(\varphi_1), \dots, \rho(\varphi_n))$, with ∂_i denoting the derivative with respect to the i-th argument. Now

$$\mathrm{d}\langle X(\varphi_i), X(\varphi_j)\rangle_t = \mathrm{d}\langle M(\varphi_i), M(\varphi_j)\rangle_t = \langle \partial_x\varphi_i, \partial_x\varphi_j\rangle_{L^2(\mathbb{T})}\mathrm{d}t,$$

and then

$$\mathrm{d}_t F(X_t) = \sum_{i=1}^n F_i(X_t)\mathrm{d}M_t(\varphi_i) + L_0F(X_t)\mathrm{d}t,$$

where L_0 is the second–order differential operator defined on cylindrical functions F as

$$L_0F(\rho) = \sum_{i=1}^n F_i(\rho)\rho(\Delta\varphi_i) + \sum_{i,j=1}^n \frac{1}{2}F_{i,j}(\rho)\langle \partial_x\varphi_i, \partial_x\varphi_j\rangle_{L^2(\mathbb{T})}. \tag{3}$$

Another way to describe the generator L_0 is to give its value on the functions $\rho \mapsto \exp(\rho(\psi))$ for $\psi \in \mathscr{S}$, which is

$$L_0 e^{\rho(\psi)} = e^{\rho(\psi)}(\rho(\Delta\psi) - \frac{1}{2}\langle \psi, \Delta\psi\rangle_{L^2(\mathbb{T})}).$$

If F, G are two cylindrical functions (which we can take of the form $F(\rho) = f(\rho(\varphi_1), \dots, \rho(\varphi_n))$ and $G(\rho) = g(\rho(\varphi_1), \dots, \rho(\varphi_n))$ for the same $\varphi_1, \dots, \varphi_n \in \mathscr{S}$), we can check that

$$L_0(FG) = (L_0F)G + F(L_0G) + \mathcal{E}(F, G), \tag{4}$$

where the quadratic form $\mathcal{E}$ is given by

$$\mathcal{E}(F, G)(\rho) = \sum_{i,j} F_i(\rho)G_j(\rho)\langle \partial_x\varphi_i, \partial_x\varphi_j\rangle_{L^2(\mathbb{T})}. \tag{5}$$

Lemma 3 *(Gaussian integration by parts) Let $(Z_i)_{i=1,\dots,M}$ be a M-dimensional Gaussian vector with zero mean and covariance $(C_{i,j})_{i,j=1,\dots,M}$. Then for all $g \in C_b^1(\mathbb{R}^M)$ we have*

$$\mathbb{E}[Z_k g(Z)] = \sum_\ell C_{k,\ell} \mathbb{E}\left[\frac{\partial g(Z)}{\partial Z_\ell}\right].$$

Proof Use that $\mathbb{E}[e^{i\langle Z,\lambda\rangle}] = e^{-\langle\lambda, C\lambda\rangle/2}$ and moreover that

$$\begin{aligned}\mathbb{E}[Z_k e^{i\langle Z,\lambda\rangle}] &= (-i)\frac{\partial}{\partial\lambda_k}\mathbb{E}[e^{i\langle Z,\lambda\rangle}] = (-i)\frac{\partial}{\partial\lambda_k}e^{-\langle\lambda,C\lambda\rangle/2} = i(C\lambda)_k e^{-\langle\lambda,C\lambda\rangle/2}\\ &= i\sum_\ell C_{k,\ell}\lambda_\ell \mathbb{E}[e^{i\langle Z,\lambda\rangle}] = \sum_\ell C_{k,\ell}\mathbb{E}[\frac{\partial}{\partial Z_\ell}e^{i\langle Z,\lambda\rangle}].\end{aligned}$$

The relation is true for trigonometric functions and taking Fourier transforms we see that it holds for all $g \in \mathscr{S}$. Is then a matter of taking limits to show that we can extend it to any $g \in C_b^1(\mathbb{R}^M)$. □

As a first application of this formula let us show that $\mathbb{E}[L_0 F(\eta)] = 0$ for every cylindrical function, where $\eta(\varphi) \sim \mathcal{N}(0, \|\varphi\|^2_{L^2(\mathbb{T})}/2)$ is a space white noise. Indeed, note that $\mathbb{E}[\eta(\varphi_i)\eta(\Delta\varphi_j)] = \frac{1}{2}\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}$, leading to

$$\begin{aligned}\mathbb{E}\sum_{i,j=1}^n \frac{1}{2}F_{i,j}(\eta)\langle\partial_x\varphi_i, \partial_x\varphi_j\rangle_{L^2(\mathbb{T})} &= -\mathbb{E}\sum_{i,j=1}^n \frac{1}{2}F_{i,j}(\eta)\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}\\ &= -\frac{1}{2}\sum_{i,j=1}^n \langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}\mathbb{E}\frac{\partial}{\partial\eta(\varphi_i)}F_j(\eta)\\ &= -\sum_{j=1}^n \mathbb{E}[\eta(\Delta\varphi_j)F_j(\eta)],\end{aligned}$$

so that $\mathbb{E}[L_0 F(\eta)] = 0$. In combination with Itô's formula, this shows that the white noise law is indeed a stationary distribution for X (check it!). From now on we fix the initial distribution $X_0 \sim \eta$, which means that $X_t \sim \eta$ for all $t \geqslant 0$.

As another application of the Gaussian integration by parts formula, we get

$$\begin{aligned}\frac{1}{2}\mathbb{E}[\mathcal{E}(F,G)(\eta)] &= -\frac{1}{2}\sum_{i,j}\mathbb{E}[F_i(\eta)G_j(\eta)]\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}.\\ &= -\frac{1}{2}\sum_{i,j}\mathbb{E}[(F(\eta)G_j(\eta))_i]\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}\\ &\quad + \frac{1}{2}\sum_{i,j}\mathbb{E}[F(\eta)G_{ij}(\eta)]\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}\\ &= -\sum_j \mathbb{E}[F(\eta)G_j(\eta)\eta(\Delta\varphi_j)] + \frac{1}{2}\sum_{i,j}\mathbb{E}[F(\eta)G_{ij}(\eta)]\langle\varphi_i, \Delta\varphi_j\rangle_{L^2(\mathbb{T})}\\ &= -\mathbb{E}[(FL_0G)(\eta)].\end{aligned}$$

Combining this with (4) and with $\mathbb{E}[L_0(FG)(\eta)] = 0$, we obtain $\mathbb{E}[(FL_0G)(\eta)] = \mathbb{E}[(GL_0F)(\eta)]$. That is, L_0 is a symmetric operator with respect to the law of η.

Consider now the operator D, defined on cylindrical functions F by

$$\mathrm{D}F(\rho) = \sum_i F_i(\rho)\varphi_i \tag{6}$$

so that DF takes values in $\mathscr{S}'$, the continuous linear functionals on $\mathscr{S}$.

Exercise 2 Show that D is independent of the specific representation of F, that is if

$$F(\rho) = f(\rho(\varphi_1), \dots, \rho(\varphi_n)) = g(\rho(\psi_1), \dots, \rho(\psi_m))$$

for all $\rho \in \mathscr{S}'$, then

$$\sum_i \partial_i f(\rho(\varphi_1), \dots, \rho(\varphi_n))\varphi_i = \sum_j \partial_j g(\rho(\psi_1), \dots, \rho(\psi_m))\psi_m.$$

Hint: One possible strategy is to show that for all $\theta \in \mathscr{S}$,

$$\langle \mathrm{D}F(\rho), \theta\rangle = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} F(\rho + \varepsilon\theta)|_{\varepsilon=0}.$$

We have

$$\mathbb{E}[F(\eta)\langle\psi, \mathrm{D}G(\eta)\rangle] + \mathbb{E}[G(\eta)\langle\psi, \mathrm{D}F(\eta)\rangle] = \sum_i \mathbb{E}[(FG)_i(\eta)\langle\psi, \varphi_i\rangle] = 2\mathbb{E}[\eta(\psi)(FG)(\eta)],$$

and therefore

$$\mathbb{E}[F(\eta)\langle\psi, \mathrm{D}G(\eta)\rangle] = \mathbb{E}[G(\eta)\langle\psi, \mathrm{D}^*F(\eta)\rangle]$$

with $\mathrm{D}^*F(\rho) = -\mathrm{D}F(\rho) + 2\rho F(\rho)$ being the adjoint of D for the scalar product in $L^2(\mathrm{law}(\eta))$. Let $\mathrm{D}_\psi F = \langle\psi, \mathrm{D}F\rangle$ and similarly for $\mathrm{D}^*_\psi F = -\mathrm{D}_\psi F + 2\rho(\psi)$.

Exercise 3 Let $(e_n)_{n\geqslant 1}$ be an orthonormal basis of $L^2(\mathbb{T})$. Show that

$$L_0 = \frac{1}{2}\sum_n \mathrm{D}^*_{e_n}\mathrm{D}_{\Delta e_n}.$$

The commutator between D_θ and D^*_ψ is given by

$$[\mathrm{D}_\theta, \mathrm{D}^*_\psi]F(\rho) := (\mathrm{D}_\theta\mathrm{D}^*_\psi - \mathrm{D}^*_\psi\mathrm{D}_\theta)F(\rho) = 2\langle\psi, \theta\rangle_{L^2(\mathbb{T})}F(\rho),$$

whereas $[\mathrm{D}^*_\theta, \mathrm{D}^*_\psi] = 0$. Therefore,

$$[L_0, \mathrm{D}^*_\psi] = \frac{1}{2}\sum_n [\mathrm{D}^*_{e_n}\mathrm{D}_{\Delta e_n}, \mathrm{D}^*_\psi] = \frac{1}{2}\sum_n \mathrm{D}^*_{e_n}[\mathrm{D}_{\Delta e_n}, \mathrm{D}^*_\psi] + \frac{1}{2}\sum_n [\mathrm{D}^*_{e_n}, \mathrm{D}^*_\psi]\mathrm{D}_{\Delta e_n}$$

$$= \sum_n \mathrm{D}^*_{e_n} \langle \psi, \Delta e_n \rangle_{L^2(\mathbb{T})} = \mathrm{D}^*_{\Delta\psi}.$$

So if ψ is an eigenvector of Δ with eigenvalue λ, then $[L_0, D^*_\psi] = \lambda \mathrm{D}^*_\psi$. Let now $(\psi_n)_{n\in\mathbb{N}}$ be an orthonormal eigenbasis for Δ with eigenvalues $\Delta\psi_n = \lambda_n$ and consider the functions

$$H(\psi_{i_1}, \dots, \psi_{i_n}) : \mathscr{S} \to \mathbb{R}, \qquad H(\psi_{i_1}, \dots, \psi_{i_n})(\rho) = \left(\mathrm{D}^*_{\psi_{i_1}} \cdots \mathrm{D}^*_{\psi_{i_n}} 1\right)(\rho).$$

Then

$$\begin{aligned} L_0 H(\psi_{i_1}, \dots, \psi_{i_n}) &= L_0 \mathrm{D}^*_{\psi_{i_1}} \cdots \mathrm{D}^*_{\psi_{i_n}} 1 = \mathrm{D}^*_{\psi_{i_1}} L_0 \mathrm{D}^*_{\psi_{i_2}} \cdots \mathrm{D}^*_{\psi_{i_n}} 1 + \lambda_{i_1} \mathrm{D}^*_{\psi_{i_1}} \cdots \mathrm{D}^*_{\psi_{i_n}} 1 \qquad (7) \\ &= \cdots = (\lambda_{i_1} + \cdots + \lambda_{i_n}) H(\psi_{i_1}, \dots, \psi_{i_n}), \end{aligned}$$

where we used that $L_0 1 = 0$. So these functions are eigenfunctions for L_0 and the eigenvalues are all the possible combinations of $\lambda_{i_1} + \cdots + \lambda_{i_n}$ for $i_1, \dots, i_n \in \mathbb{N}$. We have immediately that for different n these functions are orthogonal in $L^2(\mathrm{law}(\eta))$. They are actually orthogonal as soon as the indices i differ since in that case there is an index j which is in one but not in the other and using the fact that $\mathrm{D}^*_{\psi_j}$ is adjoint to D_{ψ_j} and that $D_{\psi_j} G = 0$ if G does not depend on ψ_j we get the orthogonality. The functions $H(\psi_{i_1}, \dots, \psi_{i_n})$ are polynomials and they are called *Wick polynomials*.

Lemma 4 *For all $\psi \in \mathscr{S}$, almost surely*

$$(e^{\mathrm{D}^*_\psi} 1)(\eta) = e^{2\eta(\psi) - \|\psi\|^2}.$$

Proof If F is a cylindrical function of the form $F(\rho) = f(\rho(\varphi_1), \dots, \rho(\varphi_m))$ with $f \in \mathscr{S}(\mathbb{R}^m)$, then

$$\mathbb{E}[F(\eta)(e^{\mathrm{D}^*_\psi} 1)(\eta)] = \mathbb{E}[e^{\mathrm{D}_\psi} F(\eta)] = \mathbb{E}[F(\eta + \psi)] = \mathbb{E}[F(\eta) e^{2\eta(\psi) - \|\psi\|^2}],$$

where the second step follows from the fact that if we note $\Psi_t(\eta) = F(\eta + t\psi)$ we have $\partial_t \Psi_t(\eta) = \mathrm{D}_\psi \Psi_t(\eta)$ and $\Psi_0(\eta) = F(\eta)$ so that $\Psi_t(\eta) = (e^{t\mathrm{D}_\psi} F)(\eta)$ for all $t \geqslant 0$ and in particular for $t = 1$. The last step is simply a Gaussian change of variables. Indeed if we take $\varphi_1 = \psi$ and $\varphi_k \perp \psi$ for $k \geqslant 2$ we have

$$\mathbb{E}[F(\eta + \psi)] = \mathbb{E}[f(\eta(\psi) + \langle \psi, \psi \rangle, \eta(\varphi_2), \dots, \eta(\varphi_m))]$$

since $(\eta + \psi)(\varphi_k) = \eta(\varphi_k)$ for $k \geqslant 2$. Now observe that $\eta(\psi)$ is independent of $(\eta(\varphi_2), \dots, \eta(\varphi_m))$ so that

$$\mathbb{E}[f(\eta(\psi) + \langle \psi, \psi \rangle, \eta(\varphi_2), \dots, \eta(\varphi_m))] = \int_{\mathbb{R}} \frac{e^{-z^2/\|\psi\|^2}}{\sqrt{\pi \|\psi\|^2}} \mathbb{E}[f(z + \langle \psi, \psi \rangle, \eta(\varphi_2), \dots, \eta(\varphi_m))]$$

$$= \int_{\mathbb{R}} \frac{e^{-z^2/\|\psi\|^2}}{\sqrt{\pi \|\psi\|^2}} e^{2z - \|\psi\|^2} \mathbb{E}[f(z, \eta(\varphi_2), \dots, \eta(\varphi_m))] = \mathbb{E}[F(\eta) e^{2\eta(\psi) - \|\psi\|^2}].$$

□

Theorem 1 *The Wick polynomials* $\{H(\psi_{i_1}, \dots, \psi_{i_n})(\eta) : n \geqslant 0, i_1, \dots, i_n \in \mathbb{N}\}$ *form an orthogonal basis of* $L^2(\mathrm{law}(\eta))$.

Proof Taking $\psi = \sum_i \sigma_i \psi_i$ in Lemma 4, we get

$$e^{2\sum_i \sigma_i \eta(\psi_i) - \sum_i \sigma_i^2 \|\psi_i\|^2} = (e^{\mathrm{D}^*_\psi} 1)(\eta) = \sum_{n \geqslant 0} \frac{((\mathrm{D}^*_\psi)^n 1)(\eta)}{n!}$$
$$= \sum_{n \geqslant 0} \sum_{i_1, \dots, i_n} \frac{\sigma_{i_1} \cdots \sigma_{i_n}}{n!} H(\underbrace{\psi_{i_1}, \dots, \psi_{i_n}}_{n \text{ times}})(\eta),$$

which is enough to show that any random variable in L^2 can be expanded in a series of Wick polynomials showing that the Wick polynomials are an orthogonal basis of $L^2(\mathrm{law}(\eta))$ (but they are still not normalized). Indeed assume that $Z \in L^2(\mathrm{law}(\eta))$ but $Z \perp H(\psi_{i_1}, \dots, \psi_{i_n})(\eta)$ for all $n \geqslant 0$, $i_1, \dots, i_n \in \mathbb{N}$, then

$$0 = e^{\sum_i \sigma_i^2 \|\psi_i\|^2} \mathbb{E}[Z(e^{\mathrm{D}^*_\psi} 1)(\eta)] = e^{\sum_i \sigma_i^2 \|\psi_i\|^2} \mathbb{E}[Z e^{2\sum_i \sigma_i \eta(\psi_i) - \sum_i \sigma_i^2 \|\psi_i\|^2}] = \mathbb{E}[Z e^{2\sum_i \sigma_i \eta(\psi_i)}].$$

Since the σ_i are arbitrary, this means that Z is orthogonal to any polynomial in η and then that it is orthogonal also to $\exp\left(i \sum_i \sigma_i \eta(\psi_i)\right)$. So let $f \in \mathscr{S}(\mathbb{R}^M)$ and $\sigma_i = 0$ for $i > m$, and observe that

$$0 = (2\pi)^{-m/2} \int \mathrm{d}\sigma_1 \cdots \mathrm{d}\sigma_m \mathscr{F} f(\sigma_1, \dots, \sigma_m) \mathbb{E}[Z e^{i \sum_i \sigma_i \eta(\psi_i)}] = \mathbb{E}[Z f(\eta(\psi_1), \dots, \eta(\psi_M))],$$

which means that Z is orthogonal to all the random variables in L^2 which are measurable with respect to the σ–field generated by $(\eta(\psi_n))_{n \geqslant 0}$. This implies $Z = 0$. That is, Wick polynomials form a basis for $L^2(\mathrm{law}(\eta))$. □

Example 1 The first few (un–normalized) Wick polynomials are

$$H(\psi_i)(\rho) = \mathrm{D}^*_{\psi_i} 1 = 2\rho(\psi_i),$$

$$H(\psi_i, \psi_j)(\rho) = \mathrm{D}^*_{\psi_i} \mathrm{D}^*_{\psi_j} 1 = 2\mathrm{D}^*_{\psi_i} \rho(\psi_j) = -2\delta_{i=j} + 4\rho(\psi_i)\rho(\psi_j),$$

and

$$H(\psi_i, \psi_j, \psi_k)(\rho) = \mathrm{D}^*_{\psi_i}(-2\delta_{j=k} + 4\rho(\psi_j)\rho(\psi_k))$$
$$= -4\delta_{j=k}\rho(\psi_i) - 4\delta_{i=j}\rho(\psi_k) - 4\delta_{i=k}\rho(\psi_j) + 8\rho(\psi_i)\rho(\psi_j)\rho(\psi_k).$$

Some other properties of Wick polynomials can be derived using the commutation relation between D and D^*. By linearity $\mathrm{D}^*_{\varphi+\psi} = \mathrm{D}^*_\varphi + \mathrm{D}^*_\psi$, so that using the symmetry of H we get

$$H_n(\varphi + \psi) := H(\underbrace{\varphi + \psi, \dots, \varphi + \psi}_{n}) = \sum_{0 \leqslant k \leqslant n} \binom{n}{k} H(\underbrace{\varphi, \dots, \varphi}_{k}, \underbrace{\psi, \dots, \psi}_{n-k}).$$

Then note that by Lemma 4 we have

$$(e^{\mathrm{D}^*_\varphi}1)(\eta)(e^{\mathrm{D}^*_\psi}1)(\eta) = e^{\eta(\varphi)-\frac{1}{4}\|\varphi\|^2}e^{\eta(\psi)-\frac{1}{4}\|\psi\|^2} = e^{\eta(\varphi+\psi)-\frac{1}{4}\|\varphi+\psi\|^2+\frac{1}{2}\langle\varphi,\psi\rangle}$$
$$= (e^{\mathrm{D}^*_{\varphi+\psi}}1)(\eta)e^{\frac{1}{2}\langle\varphi,\psi\rangle}.$$

Expanding the exponentials,

$$\sum_{m,n}\frac{H_m(\varphi)}{m!}\frac{H_n(\psi)}{n!} = \sum_{r,\ell}\frac{H_r(\varphi+\psi)}{r!}\frac{\left(\frac{1}{2}\langle\varphi,\psi\rangle\right)^\ell}{\ell!} = \sum_{p,q,\ell}\frac{H(\overbrace{\varphi,\dots,\varphi}^{p},\overbrace{\psi,\dots,\psi}^{q})}{p!q!}\frac{\left(\frac{1}{2}\langle\varphi,\psi\rangle\right)^\ell}{\ell!},$$

and identifying the terms of the same homogeneity in φ and ψ respectively we get

$$H_m(\varphi)H_n(\psi) = \sum_{p+\ell=m}\sum_{q+\ell=n}\frac{m!n!}{p!q!\ell!}H(\overbrace{\varphi,\dots,\varphi}^{p},\overbrace{\psi,\dots,\psi}^{q})\left(\frac{1}{2}\langle\varphi,\psi\rangle\right)^\ell. \tag{8}$$

This gives a general formula for such products. By polarization of this multilinear form, we can also get a general formula for the products of general Wick polynomials. Indeed taking $\varphi = \sum_{i=1}^m \kappa_i\varphi_i$ and $\psi = \sum_{j=1}^n \lambda_j\psi_j$ for arbitrary real coefficients $\kappa_1,\dots,\kappa_m$ and $\lambda_1,\dots,\lambda_n$, we have

$$H_m(\sum_{i=1}^m \kappa_i\varphi_i)H_n(\sum_{j=1}^n \lambda_j\psi_j)$$
$$= \sum_{i_1,\dots,i_m}\sum_{j_1,\dots,j_n}\kappa_{i_1}\cdots\kappa_{i_m}\lambda_{j_1}\cdots\lambda_{j_m}H(\varphi_1,\dots,\varphi_m)H(\psi_{j_1},\dots,\psi_{j_n}).$$

Deriving this with respect to all the κ,λ parameters and setting them to zero, we single out the term

$$\sum_{\sigma\in S_m,\omega\in S_n}H(\varphi_{\sigma(1)},\dots,\varphi_{\sigma(m)})H(\psi_{\omega(1)},\dots,\psi_{\omega(n)}) = m!n!H(\varphi_1,\dots,\varphi_m)H(\psi_1,\dots,\psi_n),$$

where S_k denotes the symmetric group on $\{1,\dots,k\}$, and where we used the symmetry of the Wick polynomials. Doing the same also for the right hand side of (8) we get

$$H(\varphi_1,\dots,\varphi_m)H(\psi_1,\dots,\psi_n)$$
$$= \sum_{p+\ell=m}\sum_{q+\ell=n}\frac{1}{p!q!\ell!}\sum_{i,j}H(\overbrace{\varphi_{i_1},\dots,\varphi_{i_p}}^{p},\overbrace{\psi_{j_1},\dots,\psi_{j_q}}^{q})\prod_{r=1}^{\ell}\langle\varphi_{i_{p+r}},\psi_{j_{q+r}}\rangle,$$

where the sum over i,j runs over $i_1,\dots,i_n$ permutation of $1,\dots,n$ and similarly for $j_1,\dots,j_m$. In particular, we obtain

$$\mathbb{E}[H(\psi_1,\dots,\psi_n)H(\psi_1,\dots,\psi_n)] = \frac{1}{n!}\sum_{i,j}\prod_{r=1}^n\langle\psi_{i_r},\psi_{j_r}\rangle = \sum_{\sigma\in S_n}\prod_{r=1}^n\langle\psi_r,\psi_{\sigma(r)}\rangle.$$

Remark 1 In our problem it will be convenient to take the Fourier basis as basis in the above computations. Let $e_k(x) = \exp(ikx)/\sqrt{2\pi} = a_k(x) + ib_k(x)$ where $\left(\sqrt{2}a_k\right)_{k\in\mathbb{N}}$ and $\left(\sqrt{2}b_k\right)_{k\in\mathbb{N}}$ form together a real valued orthonormal basis for $L^2(\mathbb{T})$. Then $\rho(e_k)^* = \rho(e_{-k})$ whenever ρ is real valued, and we will denote $\mathrm{D}_k = \mathrm{D}_{e_k} = \mathrm{D}_{a_k} + i\mathrm{D}_{b_k}$ and similarly for $\mathrm{D}_k^* = \mathrm{D}_{a_k}^* - i\mathrm{D}_{b_k}^* = -\mathrm{D}_{-k} + 2\rho(e_{-k})$. In this way, D_k^* is the adjoint of D_k with respect to the Hermitian scalar product on $L^2(\Omega;\mathbb{C})$ and the Ornstein–Uhlenbeck generator takes the form

$$L_0 = \frac{1}{2}\sum_{k\in\mathbb{N}}(\mathrm{D}^*_{\partial_x a_k}\mathrm{D}_{\partial_x a_k} + \mathrm{D}^*_{\partial_x b_k}\mathrm{D}_{\partial_x b_k}) = \frac{1}{2}\sum_{k\in\mathbb{Z}} k^2\mathrm{D}_k^*\mathrm{D}_k \tag{9}$$

and

$$\mathcal{E}(F,G) = \sum_{k\in\mathbb{Z}} k^2(\mathrm{D}_k F)^*(\mathrm{D}_k G). \tag{10}$$

2.3 The Itô trick

We are ready now to start our computations. Recall that we want to analyse $J_t(\varphi) = \int_0^t \partial_x X_s^2(\varphi)\mathrm{d}s$ using Itô calculus with respect to the Ornstein–Uhlenbeck process. We want to understand J_t as a correction term in Itô's formula, so we have to find a function F such that $L_0F(X_t) = \partial_x X_t^2$. Of course, F will not be a cylindrical function but we only defined L_0 on cylindrical functions. So to make the following calculations rigorous we would again have to replace $\partial_x X_t^2$ by $\partial_x \Pi_n X_t^2$ and then pass to the limit, see the paper [GJ13] for details. As before we will perform the calculations already in the limit $N = +\infty$, in order to simplify the computations and not to obscure the ideas through technicalities.

Note that

$$\partial_x X_t^2(e_k) = ik\sum_{\ell+m=k} X_t(e_\ell)X_t(e_m) = ik\sum_{\ell+m=k} H_{\ell,m}(X_t),$$

where $H_{\ell,m}(\rho) = \frac{1}{4}(\mathrm{D}^*_{-\ell}\mathrm{D}^*_{-m}1)(\rho) = \rho(e_\ell)\rho(e_m) - \frac{1}{2}\delta_{\ell+m=0}$ is a second order Hermite polynomial so that $L_0H_{\ell,m} = -(\ell^2+m^2)H_{\ell,m}$ by (7). Therefore, it is enough to take

$$F(X_t)(e_k) = -ik\sum_{\ell+m=k}\frac{H_{\ell,m}(X_t)}{\ell^2+m^2}. \tag{11}$$

This corresponds to the distribution $F(X_t)(\varphi) = -\int_0^\infty \partial_x(e^{s\Delta}X_t)^2(\varphi)\mathrm{d}s$ (check it!). Then

$$F(X_t)(\varphi) = F(X_0)(\varphi) + M_{F,t}(\varphi) + J_t(\varphi),$$

where $M_{F,t}(\varphi)$ is a martingale with quadratic variation

$$\mathrm{d}\langle M_{F,*}(\varphi), M_{F,*}(\varphi)\rangle_t = \mathcal{E}(F(*)(\varphi), F(*)(\varphi))(X_t)\mathrm{d}t.$$

We can estimate

$$\mathbb{E}[|J_t(\varphi) - J_s(\varphi)|^{2p}] \lesssim_p \mathbb{E}[|M_{F,t}(\varphi) - M_{F,s}(\varphi)|^{2p}] + \mathbb{E}[|F(X_t)(\varphi) - F(X_s)(\varphi)|^{2p}].$$

To bound the martingale expectation, we will use the following Burkholder inequality:

Lemma 5 *Let m be a continuous local martingale with $m_0 = 0$. Then for all $T \geqslant 0$ and $p > 1$,*

$$\mathbb{E}[\sup_{t \leqslant T} |m_t|^{2p}] \leqslant C_p \mathbb{E}[\langle m \rangle_T^p].$$

Proof Start by assuming that m and $\langle m \rangle$ are bounded. Itô's formula yields

$$\mathrm{d}|m_t|^{2p} = (2p)|m_t|^{2p-1}\mathrm{d}m_t + \frac{1}{2}(2p)(2p-1)|m_t|^{2p-2}\mathrm{d}\langle m \rangle_t,$$

and therefore

$$\mathbb{E}[|m_T|^{2p}] = C_p \mathbb{E}\left[\int_0^T |m_s|^{2p-2}\mathrm{d}\langle m \rangle_s\right] \leqslant C_p \mathbb{E}[\sup_{t \leqslant T} |m_t|^{2p-2}\langle m \rangle_T].$$

By Cauchy–Schwartz we get

$$\mathbb{E}[|m_T|^{2p}] \leqslant C_p \mathbb{E}[\sup_{t \leqslant T} |m_t|^{2p}]^{(2p-2)/2p} \mathbb{E}[\langle m \rangle_T^p]^{1/p}.$$

But now Doob's L^p inequality yields $\mathbb{E}[\sup_{t \leqslant T} |m_t|^{2p}] \leqslant C_p' \mathbb{E}[|m_T|^{2p}]$, and this implies the claim in the bounded case. The unbounded case can be treated with a localization argument. □

Applying Burkholder's inequality, we obtain

$$\begin{aligned}
\mathbb{E}[|J_t(\varphi) - J_s(\varphi)|^{2p}] &\lesssim_p \mathbb{E}\left[\left|\int_s^t \mathcal{E}(F(*)(\varphi), F(*)(\varphi))(X_r)\mathrm{d}r\right|^p\right] + \mathbb{E}[|F(X_t)(\varphi) - F(X_s)(\varphi)|^{2p}] \\
&\leqslant (t-s)^{p-1} \int_s^t \mathbb{E}[|\mathcal{E}(F(*)(\varphi), F(*)(\varphi))(X_r)|^p]\mathrm{d}r \\
&\quad + \mathbb{E}[|F(X_t)(\varphi) - F(X_s)(\varphi)|^{2p}] \\
&= (t-s)^p \mathbb{E}[|\mathcal{E}(F(*)(\varphi), F(*)(\varphi))(\eta)|^p] + \mathbb{E}[|F(X_t)(\varphi) - F(X_s)(\varphi)|^{2p}],
\end{aligned}$$

using that $X_r \sim \eta$. Now

$$\mathrm{D}_m F(\rho)(e_k) = -2ik \frac{\rho(e_{k+m})}{(k+m)^2 + (-m)^2},$$

and therefore

$$\begin{aligned}
\mathcal{E}(F(*)(e_k), F(*)(e_k))(\rho) &= \sum_m m^2 \mathrm{D}_{-m}F(\rho)(e_{-k})\mathrm{D}_m F(\rho)(e_k) \\
&= 4k^2 \sum_{\ell+m=k} m^2 \frac{|\rho(e_\ell)|^2}{(\ell^2+m^2)^2} \lesssim k^2 \sum_{\ell+m=k} \frac{|\rho(e_\ell)|^2}{\ell^2+m^2},
\end{aligned}$$

which implies that

$$\mathbb{E}[|\mathcal{E}(F(*)(e_k), F(*)(e_k))(\eta)|] \lesssim k^2 \mathbb{E} \sum_{\ell+m=k} \frac{|\eta(e_\ell)|^2}{\ell^2+m^2} \lesssim k^2 \sum_{\ell+m=k} \frac{1}{\ell^2+m^2} \lesssim |k|.$$

A similar computation gives also that

$$\mathbb{E}[|\mathcal{E}(F(*)(e_k), F(*)(e_k))(\eta)|^p] \lesssim |k|^p.$$

Further, we have

$$\begin{aligned}\mathbb{E}[|F(X_t)(e_k) - F(X_s)(e_k)|^2] &\lesssim k^2 \sum_{\ell+m=k} \mathbb{E}\left[\frac{(H_{\ell,m}(X_t) - H_{\ell,m}(X_s))^2}{(\ell^2+m^2)^2}\right] \\ &\lesssim k^2|t-s| \sum_{\ell+m=k} \frac{m^2}{(\ell^2+m^2)^2} \lesssim |k||t-s|.\end{aligned}$$

And finally, this computation lets us recover the result that

$$\mathbb{E}[|J_t(e_k) - J_s(e_k)|^{2p}] \lesssim_p (t-s)^p |k|^p.$$

The advantage of the Itô trick with respect to the explicit Gaussian computation is that it goes over to the non–Gaussian case. Indeed note that while the boundary term $F(X_t)(\varphi) - F(X_s)(\varphi)$ has been estimated using a lot of the Gaussian information about X, we used only the law at a fixed time to handle the term $\int_s^t \mathcal{E}(F(*)(\varphi), F(*)(\varphi))(X_r)\mathrm{d}r$.

In order to carry over these computation to the full process u solution of the non–linear dynamics we need to replace the generator of X with that of u and to have a way to handle the boundary terms. The idea is now to reverse the Markov process u in time, which will allow us to kill the antisymmetric part of the generator and at the same time kill the boundary terms. Indeed observe that if u solves the stochastic Burgers equation, then formally we have the Itô formula

$$\mathrm{d}_t F(u_t) = \sum_{i=1}^n F_i(u_t)\mathrm{d}M_t(\varphi_i) + LF(u_t)\mathrm{d}t,$$

where L is now the full generator of the non–linear dynamics, given by

$$LF(\rho) = L_0 F(\rho) + \sum_i F_i(\rho)\langle \partial_x \rho^2, \varphi_i\rangle = L_0 F(\rho) + BF(\rho),$$

where

$$BF(\rho) = \sum_k (\partial_x \rho^2)(e_k)\mathrm{D}_k F(\rho).$$

Formally, the non–linear term is antisymmetric with respect to the invariant measure of L_0. Indeed since B is a first order operator

$$\mathbb{E}[(BF(\eta))G(\eta)] = \mathbb{E}[(B(FG)(\eta))] - \mathbb{E}[F(\eta)(BG(\eta))] = -\mathbb{E}[F(\eta)(BG(\eta))] \tag{12}$$

provided $\mathbb{E}[BF(\eta)] = 0$ for any cylinder function F. Let us show this.

$$\begin{aligned}\mathbb{E}[BF(\eta)] &= \sum_k \mathbb{E}[(\partial_x \eta^2)(e_k) \mathrm{D}_k F(\eta)] \\ &= \sum_k \mathbb{E}[(\mathrm{D}_k(\partial_x \eta^2)(e_k))F(\eta)] + \sum_k \mathbb{E}[\mathrm{D}_k[(\partial_x \eta^2)(e_k)F(\eta)]]\end{aligned}$$

But now, using that $e_k^* e_k = 1/(2\pi)$ and that $\langle \eta, 1\rangle$ we have

$$\mathrm{D}_k(\partial_x \eta^2)(e_k) = 2\langle \partial_x(\eta e_k), e_k\rangle = -2ik\langle \eta e_k, e_k\rangle = -2ik\langle \eta, e_k^* e_k\rangle = -\frac{2ik}{2\pi}\langle \eta, 1\rangle = 0$$

and using the Gaussian integration by parts

$$\begin{aligned}\mathbb{E}[BF(\eta)] &= \sum_k \mathbb{E}[\mathrm{D}_k[(\partial_x \eta^2)(e_k)F(\eta)]] = -\sum_k \mathbb{E}[\eta(e_k)(\partial_x \eta^2)(e_k)F(\eta)] \\ &= -\mathbb{E}[\langle \eta, \partial_x \eta^2\rangle F(\eta)] = -\frac{1}{3}\mathbb{E}[\langle 1, \partial_x \eta^3\rangle F(\eta)] = 0\end{aligned}$$

since $-\langle 1, \partial_x \eta^3\rangle = \langle \partial_x 1, \eta^3\rangle = 0$.

The dynamics of u backwards in time has a Markovian description which is the subject of the next exercise.

Exercise 4 Let $(y_t)_{t\geqslant 0}$ be a stationary Markov process on a Polish space, with semigroup $(P_t)_{t\geqslant 0}$ and stationary distribution μ. Show that if P_t^* is the adjoint of P_t in $L^2(\mu)$, then (P_t^*) is a semigroup of operators on $L^2(\mu)$ (that is $P_0^* = \mathrm{id}$ and $P_{s+t}^* = P_s^* P_t^*$ as operators on $L^2(\mu)$). Show that if $y_0 \sim \mu$, then for all $T > 0$ the process $\hat{y}_t = y_{T-t}$, $t \in [0, T]$, is also Markov, with semigroup $(P_t^*)_{t\in[0,T]}$, and that μ is also an invariant distribution for (P_t^*). Show also that if (P_t) has generator L then (P_t^*) has generator L^* which is the adjoint of L with respect to $L^2(\mu)$.

Now if we reverse the process in time letting $\hat{u}_t = u_{T-t}$, we have by stationarity

$$\mathbb{E}[F(\hat{u}_t)G(\hat{u}_0)] = \mathbb{E}[F(u_{T-t})G(u_T)] = \mathbb{E}[F(u_0)G(u_t)].$$

So if we denote by $\hat{L}$ the generator of $\hat{u}$:

$$\mathbb{E}[\hat{L}F(\hat{u}_0)G(\hat{u}_0)] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathbb{E}[F(\hat{u}_t)G(\hat{u}_0)] = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathbb{E}[F(u_0)G(u_t)] = \mathbb{E}[LG(u_0)F(u_0)],$$

which means that $\hat{L}$ is the adjoint of L, that is

$$\hat{L}F(\rho) = L_0 F(\rho) - BF(\rho) = L_0 F(\rho) - \sum_k (\partial_x \rho^2)(e_k)\mathrm{D}_k F(\rho).$$

Then Itô's formula for the reversed process gives

$$\mathrm{d}_t F(\hat{u}_t) = \sum_{i=1}^n F_i(\hat{u}_t)\mathrm{d}\hat{M}_t(\varphi_i) + \hat{L}F(\hat{u}_t)\mathrm{d}t,$$

where for all test functions φ, the process $\hat{M}(\varphi)$ is a martingale in the filtration of $\hat{u}$ with covariance

$$\mathrm{d}\langle \hat{M}(\varphi), \hat{M}(\psi)\rangle_t = \langle \partial_x\varphi, \partial_x\psi\rangle_{L^2(\mathbb{T})}\mathrm{d}t.$$

Combining the Itô formulas for u and $\hat{u}$, we get

$$F(u_T)(\varphi) = F(u_0)(\varphi) + M_{F,T}(\varphi) + \int_0^T LF(u_s)(\varphi)\mathrm{d}s$$

and

$$\begin{aligned} F(u_0)(\varphi) = F(\hat{u}_T)(\varphi) &= F(\hat{u}_0)(\varphi) + \hat{M}_{F,T}(\varphi) + \int_0^T \hat{L}F(\hat{u}_s)(\varphi)\mathrm{d}s \\ &= F(u_T)(\varphi) + \hat{M}_{F,T}(\varphi) + \int_0^T \hat{L}F(u_s)(\varphi)\mathrm{d}s, \end{aligned}$$

and summing up these two equalities gives

$$0 = M_{F,T}(\varphi) + \hat{M}_{F,T}(\varphi) + \int_0^T (\hat{L}+L)F(u_s)(\varphi)\mathrm{d}s,$$

that is

$$2\int_0^T L_0F(u_s)(\varphi)\mathrm{d}s = -M_{F,T}(\varphi) - \hat{M}_{F,T}(\varphi).$$

An added benefit of this forward–backward representation is that only term which required quite a lot of informations about X, that is the boundary term $F(X_t)(\varphi)-F(X_s)(\varphi)$ does not appear at all now. As above if $2L_0F(\rho) = \partial_x\rho^2$, we end up with

$$\int_0^T \partial_x u_s^2(\varphi)\mathrm{d}s = -M_{F,T}(\varphi) - \hat{M}_{F,T}(\varphi). \tag{13}$$

Exercise 5 Perform a similar formal calculation as in (12) to see that $\mathbb{E}[LF(\eta)] = 0$ for all cylindrical functions F, so that η is also invariant for the stochastic Burgers equation. Combine this with (13) to show that setting $\mathcal{N}_t^N(\varphi) = \int_0^t \partial_x(\Pi_N u_s)^2(\varphi)\mathrm{d}s$ we have

$$\mathbb{E}[|\mathcal{N}_t^N(e_k) - \mathcal{N}_s^N(e_k)|^{2p}] \lesssim_p (t-s)^p|k|^p$$

and letting $\mathcal{N}_t^{N,M} = \mathcal{N}_t^N - \mathcal{N}_t^M$ we get

$$\mathbb{E}[|\mathcal{N}_t^{N,M}(e_k) - \mathcal{N}_s^{N,M}(e_k)|^{2p}] \lesssim_p (|k|/N)^{\varepsilon p}(t-s)^p|k|^p$$

for all $1 \leqslant N \leqslant M$. Use this to derive that

$$(\mathbb{E}[\|\mathcal{N}_t^{N,M} - \mathcal{N}_s^{N,M}\|_{H^\alpha}^{2p}])^{1/2p} \lesssim_{p,\alpha} N^{-\varepsilon/2}(t-s)^{1/2}$$

for all $\alpha < -1-\varepsilon$, and realize that this estimate allows you to prove compactness of the approximations $\mathcal{N}^N$ and then convergence to a limit $\mathcal{N}$ in $L^{2p}(\Omega; C^{1/2-}H^{-1-})$.

2.4 Controlled distributions

Let us cook up a definition which will allow us to rigorously perform the formal computations above in a general setting.

Definition 2 *Let $u, \mathcal{A} : \mathbb{R}_+ \times \mathbb{T} \to \mathbb{R}$ be a couple of generalized process such that*

i. For all $\varphi \in \mathscr{S}(\mathbb{T})$ the process $t \mapsto u_t(\varphi)$ is a continuous semimartingale satisfying

$$u_t(\varphi) = u_0(\varphi) + \int_0^t u_s(\Delta\varphi)\mathrm{d}s + \mathcal{A}_t(\varphi) + M_t(\varphi),$$

where $t \mapsto M_t(\varphi)$ is a martingale with quadratic variation $\langle M_t(\varphi), M_t(\psi)\rangle = \langle \partial_x\varphi, \partial_x\psi\rangle_{L^2(\mathbb{T})}$ and $t \mapsto \mathcal{A}_t(\varphi)$ is a finite variation process with $\mathcal{A}_0 = 0$.

ii. For all $t \geqslant 0$ the random distribution $\varphi \mapsto u_t(\varphi)$ is a zero mean space white noise with variance $\|\varphi\|^2_{L^2_0}/2$.

iii. For any $T > 0$ the reversed process $\hat{u}_t = u_{T-t}$ has again properties i, ii with martingale $\hat{M}$ and finite variation part $\hat{\mathcal{A}}$ such that $\hat{\mathcal{A}}_t(\varphi) = -(\mathcal{A}_T(\varphi) - \mathcal{A}_{T-t}(\varphi))$.

Any pair of processes $(u, \mathcal{A})$ satisfying these condition will be called controlled by the Ornstein–Uhlenbeck process and we will denote the set of all such processes with $\mathcal{Q}_{\mathrm{ou}}$.

Theorem 2 *Assume that $(u, \mathcal{A}) \in \mathcal{Q}_{\mathrm{ou}}$ and for any $N \geqslant 1$, $t \geqslant 0$, $\varphi \in \mathscr{S}$ let*

$$\mathcal{N}^N_t(\varphi) = \int_0^t \partial_x(\Pi_N u_s)^2(\varphi)\mathrm{d}s$$

Then for any $p \geqslant 1$ $(\mathcal{N}^N)_{N\geqslant 1}$ converges in probability to a space–time distribution $\mathcal{N} \in C^{1/2-}H^{-1-}$.

We are now at a point were we can give a meaning to our original equation.

Definition 3 *A pair of random distribution $(u, \mathcal{A}) \in \mathcal{Q}_{\mathrm{ou}}$ is an* energy solution *to the stochastic Burgers equation if it satisfies*

$$u_t(\varphi) = u_0(\varphi) + \int_0^t u_s(\Delta\varphi)\mathrm{d}s + \mathcal{N}_t(\varphi) + M_t(\varphi)$$

for all $t \geqslant 0$ and $\varphi \in \mathscr{S}$. That is if $\mathcal{A} = \mathcal{N}$.

Now we are in a relatively standard setting of needing to prove existence and uniqueness of such energy solutions. Note that in general the solutions are *pairs* of processes $(u, \mathcal{A})$.

Remark 2 The notion of energy solution has been introduced (in a slightly different way) in the work of Gonçalves and Jara [GJ10] on macroscopic universal fluctuations of weakly asymmetric interacting particle systems.

2.5 Existence of solutions

For the existence the way to proceed is quite standard. We approximate the equation, construct approximate solutions and then try to have enough compactness to have limiting points which then naturally will satisfy the requirements for energy solutions. For any $N \geqslant 1$ consider solutions u^N to

$$\partial_t u^N = \Delta u^N + \partial_x \Pi_N (\Pi_N u^N)^2 + \partial_x \xi$$

These are generalized functions such that

$$\mathrm{d}u_t^N(e_k) = -k^2 u_t^N(e_k)\mathrm{d}t + [\partial_x \Pi_N (\Pi_N u^N)^2](e_k)\mathrm{d}t + ik\mathrm{d}\beta_t(k)$$

for $k \in \mathbb{Z}$ and $t \geqslant 0$. We take u_0 to be the white noise with covariance $u_t(\varphi) \sim \mathcal{N}(0, \|\varphi\|^2/2)$. The point of our choice of the non–linearity is that this (infinite–dimensional) system of equations decomposes into a finite dimensional system for $(v^N(k) = \Pi_N u^N(e_k))_{k:|k|\leqslant N}$ and an infinite number of one–dimensional equations for each $u^N(e_k)$ with $|k| > N$. Indeed if $|k| > N$ we have $[\partial_x \Pi_N (\Pi_N u^N)^2](e_k) = 0$ so $u_t(e_k) = X_t(e_k)$ the Ornstein–Uhlenbeck process with initial condition $X_0 = u_0$ which renders it stationary in time (check it). The equation for $(v^N(k))_{|k|\leqslant N}$ reads

$$\mathrm{d}v_t^N(k) = -k^2 v_t^N(k)\mathrm{d}t + b_k(v_t^N)\mathrm{d}t + ik\mathrm{d}\beta_t(k), \qquad |k| \leqslant N, t \geqslant 0$$

where

$$b_k(v_t^N) = ik \sum_{\ell+m=k} \mathbb{I}_{|\ell|,|k|,|m|\leqslant N} v_t^N(\ell) v_t^N(m).$$

This is a standard finite–dimensional ODE having global solutions for all initial conditions which gives rise to a nice Markov process. The fact that solutions do not blow up even if the interaction is quadratic can be seen by computing the evolution of the norm

$$A_t = \sum_{|k|\leqslant N} |v_t^N(k)|^2$$

and by showing that

$$\mathrm{d}A_t = 2\sum_{|k|\leqslant N} v_t^N(-k)\mathrm{d}v_t^N(-k) = -2k^2 A_t \mathrm{d}t + 2\sum_{|k|\leqslant N} v_t^N(-k) b_k(v_t^N)\mathrm{d}t + 2\iota k \sum_{|k|\leqslant N} v_t^N(-k)\mathrm{d}\beta_t(k).$$

But now

$$\begin{aligned}\sum_{|k|\leqslant N} v_t^N(-k) b_k(v_t^N) &= 2\iota \sum_{k,\ell,m:\ell+m=k} \mathbb{I}_{|\ell|,|k|,|m|\leqslant N} k v_t^N(\ell) v_t^N(m) v_t^N(-k) \\ &= -2\iota \sum_{k,\ell,m:\ell+m+k=0} \mathbb{I}_{|\ell|,|k|,|m|\leqslant N}(k) v_t^N(\ell) v_t^N(m) v_t^N(k)\end{aligned}$$

and by symmetry of this expression it is equal to

$$= -\frac{2}{3}\iota \sum_{k,\ell,m:\ell+m+k=0} \mathbb{I}_{|\ell|,|k|,|m|\leqslant N}(k+\ell+m)v_t^N(\ell)v_t^N(m)v_t^N(k) = 0,$$

so $A_t = A_0 + M_t$ where $\mathrm{d}M_t = 2\sum_{|k|\leqslant N} \mathbb{I}_{|k|\leqslant N}(\iota k)v_t^N(-k)\mathrm{d}\beta_t(k)$. Now

$$\mathbb{E}[M_T^2] \lesssim \int_0^T \sum_{|k|\leqslant N} k^2 |v_t^N(k)|^2 \mathrm{d}t \lesssim N^2 \int_0^T A_t \mathrm{d}t$$

and then by martingales inequalities

$$\begin{aligned}\mathbb{E}[\sup_{t\in[0,T]}(A_t)^2] &\leqslant 2\mathbb{E}[A_0^2] + 2\mathbb{E}[\sup_{t\in[0,T]}(M_t)^2] \leqslant 2\mathbb{E}[A_0^2] + 2\mathbb{E}[M_T^2] \\ &\leqslant 2\mathbb{E}[A_0^2] + CN^2 \int_0^T \mathbb{E}(A_t)\mathrm{d}t.\end{aligned}$$

Now Gronwall's inequality gives

$$\mathbb{E}[\sup_{t\in[0,T]}(A_t)^2] \lesssim e^{CN^2T}\mathbb{E}[A_0^2],$$

from where we can deduce (by a continuation argument) that almost surely there is no blowup at finite time for the dynamics. From the Galerkin approximations the Itô trick will provide enough compactness in order to pass to the limit and build an energy solution to the Stochastic Burgers equation; see [RVW01] for details on how to implement the Itô trick on the level of diffusions.

3 Distributions and Besov spaces

Here we collect some classical results from harmonic analysis which we will need in the following. We concentrate on distributions and SPDEs on the torus, but everything in this Section applies mutatis mutandis on the full space $\mathbb{R}^d$, see [GIP13]. The only problem is that then the stochastic terms will no longer be in the Besov spaces $\mathscr{C}^\alpha$ which we encounter below but rather in weighted Besov spaces. Handling SPDEs in weighted function spaces is more delicate and we prefer here to concentrate to the simpler situation of the torus.

The space of distributions $\mathscr{S}' = \mathscr{S}'(\mathbb{T}^d)$ is the set of linear maps f from $\mathscr{S} = C^\infty(\mathbb{T}^d, \mathbb{C})$ to $\mathbb{C}$, such that there exist $k \in \mathbb{N}$ and $C > 0$ with

$$|\langle f, \varphi\rangle| := |f(\varphi)| \leqslant C \sup_{|\mu|\leqslant k} \|\partial^\mu \varphi\|_{L^\infty(\mathbb{T}^d)}$$

for all $\varphi \in \mathscr{S}$.

Example 2 Clearly $L^p = L^p(\mathbb{T}^d) \subset \mathscr{S}'$ for all $p \geqslant 1$, and more generally the space of finite signed measures on $\left(\mathbb{T}^d, \mathscr{B}(\mathbb{T}^d)\right)$ is contained in $\mathscr{S}'$. Another example of a distribution is $\varphi \mapsto \partial^\mu \varphi(x)$ for $\mu \in \mathbb{N}_0^d$ and $x \in \mathbb{T}$.

In particular, the Fourier transform $\mathscr{F} f : \mathbb{Z}^d \to \mathbb{C}$,

$$\mathscr{F} f(k) = \hat{f}(k) = \langle f, e_k \rangle$$

with $e_k = e^{-i\langle k, \cdot \rangle}/(2\pi)^{d/2}$ is defined for all $f \in \mathscr{S}'$, and it satisfies $\left|\mathscr{F} f(k)\right| \leqslant |P(k)|$ for a suitable polynomial P. Conversely, if $(g(k))_{k \in \mathbb{Z}^d}$ is at most of polynomial growth, then its inverse Fourier transform

$$\mathscr{F}^{-1} g = \sum_{k \in \mathbb{Z}^d} g(k) e_k^*$$

defines a distribution.

Exercise 6 Show that the Fourier transform of $\varphi \in \mathscr{S}$ decays faster than any rational function (we say that it is of *rapid decay*). Combine this with the fact that $\mathscr{F}$ defines a bijection from $L^2(\mathbb{T}^d)$ to $\ell^2(\mathbb{Z}^d)$ with inverse $\mathscr{F}^{-1}$ to show that $\mathscr{F}^{-1}\mathscr{F} f = f$ for all $f \in \mathscr{S}'$ and $\mathscr{F}\mathscr{F}^{-1} g = g$ for all g of polynomial growth. Extend the Parseval formula

$$\langle f, \varphi^* \rangle_{L^2(\mathbb{T}^d)} = \int_{\mathbb{T}^d} f(x) g(x)^* \mathrm{d}x = \sum_k \hat{f}(k) \hat{\varphi}(k)^*$$

from $f, \varphi \in L^2(\mathbb{T}^d)$ to $f \in \mathscr{S}'$ and $\varphi \in \mathscr{S}$.

Exercise 7 Fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. On that space let ξ be a spatial white noise on $\mathbb{T}^d$, i.e. ξ is a centered Gaussian process indexed by $L^2(\mathbb{T}^d)$, with covariance

$$\mathbb{E}[\xi(f)\xi(g)] = \int_{\mathbb{T}^d} f(x) g(x) \mathrm{d}x.$$

Show that there exists $\tilde{\xi}$ with $\mathbb{P}(\tilde{\xi}(f) = \xi(f)) = 1$ for all $f \in L^2$, such that $\tilde{\xi}(\omega) \in \mathscr{S}'$ for all $\omega \in \Omega$.

Hint: Show that $\mathbb{E}\left[\sum_{k \in \mathbb{Z}^d} \exp(\lambda \|\xi(e_k)\|^2)/(1 + |k|^{d+1})\right] < \infty$ for some suitable $\lambda > 0$.

Linear maps on $\mathscr{S}'$ can be defined by duality: if $A : \mathscr{S} \to \mathscr{S}$ is such that for all $k \in \mathbb{N}$ there exists $n \in \mathbb{N}$ and $C > 0$ with $\sup_{|\mu| \leqslant k} \|\partial^\mu (A\varphi)\|_{L^\infty} \leqslant C \sup_{|\mu| \leqslant n} \|\partial^\mu \varphi\|_{L^\infty}$, then we set $\langle {}^t A f, \varphi \rangle = \langle f, A\varphi \rangle$. Differential operators are defined by $\langle \partial^\mu f, \varphi \rangle = (-1)^{|\mu|} \langle f, \partial^\mu \varphi \rangle$. If $\varphi : \mathbb{Z}^d \to \mathbb{C}$ grows at most polynomially, then it defines a *Fourier multiplier*

$$\varphi(\mathrm{D}) : \mathscr{S}' \to \mathscr{S}', \qquad \varphi(\mathrm{D}) f = \mathscr{F}^{-1}(\varphi \mathscr{F} f).$$

Exercise 8 Use the Fourier inversion formula of Exercise 6 to show that for $f \in \mathscr{S}'$, $\varphi \in \mathscr{S}$ and for $u, v : \mathbb{Z}^d \to \mathbb{C}$ with u of polynomial growth and v of rapid decay

$$\mathscr{F}(f\varphi)(k) = (2\pi)^{-d/2} \sum_\ell \hat{f}(k - \ell) \hat{\varphi}(\ell) \quad \text{and} \quad \mathscr{F}^{-1}(uv)(x) = (2\pi)^{d/2} \left\langle \mathscr{F}^{-1} u, \mathscr{F}^{-1} v(x - \cdot) \right\rangle.$$

We will now use Littlewood–Paley blocks to obtain a decomposition of distributions into an infinite series of smooth functions. Of course, we have already such a decomposition at our disposal: $f = \sum_k \hat{f}(k) e_k^*$. But it turns out to be convenient not to consider each Fourier coefficient separately, but to work with projections on dyadic Fourier blocks.

Definition 4 *A dyadic partition of unity consists of two nonnegative radial functions $\chi, \rho \in C^\infty(\mathbb{R}^d, \mathbb{R})$, where ρ is supported in a ball $\mathscr{B} = \{|x| \leqslant c\}$ and ρ is supported in an annulus $\mathscr{A} = \{a \leqslant |x| \leqslant b\}$ for suitable $a, b, c > 0$, such that*

1. *$\chi + \sum_{j \geqslant 0} \rho(2^{-j}\cdot) \equiv 1$ and*
2. *$\operatorname{supp}(\chi) \cap \operatorname{supp}(\rho(2^{-j}\cdot)) \equiv 0$ for $j \geqslant 1$ and $\operatorname{supp}(\rho(2^{-i}\cdot)) \cap \operatorname{supp}(\rho(2^{-j}\cdot)) \equiv 0$ for all $i, j \geqslant 0$ with $|i - j| \geqslant 1$.*

We will often write $\rho_{-1} = \chi$ and $\rho_j = \rho(2^{-j}\cdot)$ for $j \geqslant 0$.

Dyadic partitions of unity exist, see [BCD11]. The reason for considering smooth partitions rather than indicator functions is that indicator functions do not have good Fourier multiplier properties. For example we only have $\|\mathbb{I}_{[2^j, 2^{j+1})}(|\mathrm{D}|) f\|_{L^\infty} \lesssim j \|f\|_{L^\infty}$, whereas $\|\rho_j(\mathrm{D}) f\|_{L^\infty} \lesssim \|f\|_{L^\infty}$ uniformly in j. From now on we fix a dyadic partition of unity (χ, ρ) and define the dyadic blocks

$$\Delta_j f = \rho_j(\mathrm{D}) f = \mathscr{F}^{-1}(\rho_j \hat{f}), \quad j \geqslant -1.$$

We also use the notation

$$S_j f = \sum_{i \leqslant j-1} \Delta_i f$$

as well as $K_i = (2\pi)^{-d/2} \mathscr{F}^{-1} \rho_i$ so that

$$K_i * f = \mathscr{F}^{-1}\left(\rho_j \mathscr{F} f\right) = \Delta_i f.$$

Every dyadic block has a compactly supported Fourier transform and is therefore in $\mathscr{S}$. It is easy to see that $f = \sum_{j \geqslant -1} \Delta_j f = \lim_{j \to \infty} S_j f$ for all $f \in \mathscr{S}'$.

For $\alpha \in \mathbb{R}$, the Hölder-Besov space $\mathscr{C}^\alpha$ is given by $\mathscr{C}^\alpha = B^\alpha_{\infty,\infty}(\mathbb{T}^d, \mathbb{R})$, where for $p, q \in [1, \infty]$ we define

$$B^\alpha_{p,q} = B^\alpha_{p,q}(\mathbb{T}^d, \mathbb{R}) = \Big\{ f \in \mathscr{S}' : \|f\|_{B^\alpha_{p,q}} = \Big(\sum_{j \geqslant -1} (2^{j\alpha} \|\Delta_j f\|_{L^p})^q \Big)^{1/q} < \infty \Big\},$$

with the usual interpretation as ℓ^∞ norm if $q = \infty$. Then $B^\alpha_{p,q}$ is a Banach space and while the norm $\|\cdot\|_{B^\alpha_{p,q}}$ depends on (χ, ρ), the space $B^\alpha_{p,q}$ does not and any other dyadic partition of unity corresponds to an equivalent norm (this follows from Lemma 10 below). We write $\|\cdot\|_\alpha$ instead of $\|\cdot\|_{B^\alpha_{\infty,\infty}}$.

Exercise 9 Let δ_0 denote the Dirac delta in 0. Show that $\delta_0 \in \mathscr{C}^{-d}$.

If $\alpha \in (0,\infty) \setminus \mathbb{N}$, then $\mathscr{C}^\alpha$ is the space of $\lfloor \alpha \rfloor$ times differentiable functions whose partial derivatives of order $\lfloor \alpha \rfloor$ are $(\alpha - \lfloor \alpha \rfloor)$-Hölder continuous (see page 99 of [BCD11]). Note however, that for $k \in \mathbb{N}$ the space $\mathscr{C}^k$ is strictly larger than C^k, the space of k times continuously differentiable functions. Below we will give the proof for $\alpha \in (0,1)$, but before we still need some tools.

Lemma 6 *(Poisson summation) Let $\varphi : \mathbb{R}^d \to \mathbb{C}$ be a Schwartz function. Then*

$$\mathscr{F}^{-1}\varphi(x) = \sum_{k\in\mathbb{Z}^d} \mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(x+2\pi k),$$

for all $x \in \mathbb{T}^d$, where $\mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\varphi(y)e^{i\langle x,y\rangle}\mathrm{d}y$.

Proof Let $g(x) = \sum_{k\in\mathbb{Z}^d} \mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(x+2\pi k)$. The function $\mathscr{F}^{-1}_{\mathbb{R}^d}\varphi$ is of rapid decay since $\varphi \in \mathscr{S}$ so the sum converges absolutely and defines a continuous function $g : \mathbb{R}^d \to \mathbb{R}$ which is periodic of period 2π in every direction. The Fourier transform over the torus $\mathbb{T}^d$ of this function is

$$\mathscr{F}g(y) = \int_{\mathbb{T}^d} e^{-i\langle x,y\rangle} g(x)\frac{\mathrm{d}x}{(2\pi)^{d/2}} = \int_{\mathbb{T}^d}\sum_{k\in\mathbb{Z}^d} \mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(x+2\pi k)e^{-i\langle x+2\pi k,y\rangle}\frac{\mathrm{d}x}{(2\pi)^{d/2}}$$

since $e^{-i\langle 2\pi k,y\rangle} = 1$. By dominated convergence the sum and the integral can be combined in a overall integration over $\mathbb{R}^d$:

$$\mathscr{F}g(y) = \int_{\mathbb{R}^d} \mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(x)e^{-i\langle x,y\rangle}\frac{\mathrm{d}x}{(2\pi)^{d/2}} = \mathscr{F}_{\mathbb{R}^d}\mathscr{F}^{-1}_{\mathbb{R}^d}\varphi(y) = \varphi(y)$$

so we deduce that $g(x) = \mathscr{F}^{-1}\varphi(x)$. □

Exercise 10 Show that $\|\cdot\|_\alpha \leqslant \|\cdot\|_\beta$ for $\alpha \leqslant \beta$, that $\|\cdot\|_{L^\infty} \lesssim \|\cdot\|_\alpha$ for $\alpha > 0$, that $\|\cdot\|_\alpha \lesssim \|\cdot\|_{L^\infty}$ for $\alpha \leqslant 0$, and that $\|S_j \cdot\|_{L^\infty} \lesssim 2^{j\alpha}\|\cdot\|_\alpha$ for $\alpha < 0$.

Hint: When proving $\|\cdot\|_\alpha \lesssim \|\cdot\|_{L^\infty}$ for $\alpha \leqslant 0$, you might need Poisson's summation formula.

The following Bernstein inequality is extremely useful when dealing with functions with compactly supported Fourier transform.

Lemma 7 *(Bernstein inequality) Let $\mathscr{B}$ be a ball and $k \in \mathbb{N}_0$. For any $\lambda \geqslant 1$, $1 \leqslant p \leqslant q \leqslant \infty$, and $f \in L^p$ with $\operatorname{supp}(\mathscr{F}f) \subseteq \lambda\mathscr{B}$ we have*

$$\max_{\mu\in\mathbb{N}^d:|\mu|=k} \|\partial^\mu f\|_{L^q} \lesssim_{k,\mathscr{B}} \lambda^{k+d\left(\frac{1}{p}-\frac{1}{q}\right)}\|f\|_{L^p}.$$

Proof Let ψ be a compactly supported C^∞ function on $\mathbb{R}^d$ such that $\psi \equiv 1$ on $\mathscr{B}$ and write $\psi_\lambda(x) = \psi(\lambda^{-1}x)$. Then

$$\begin{aligned}\partial^\mu f(x) &= \partial^\mu \mathscr{F}^{-1}\left(\psi_\lambda \mathscr{F}f\right)(x) = (2\pi)^{d/2}\left\langle f, \partial^\mu\left(\mathscr{F}^{-1}\psi_\lambda\right)(x-\cdot)\right\rangle \\ &= (2\pi)^{d/2}\left(f * \partial^\mu\left(\mathscr{F}^{-1}\psi_\lambda\right)\right)(x).\end{aligned}$$

By Young's inequality, we get

$$\|\partial^\mu f\|_{L^q} \lesssim \|f\|_{L^p} \left\|\partial^\mu \left(\mathscr{F}^{-1}\psi_\lambda\right)\right\|_{L^r},$$

where $1 + 1/q = 1/p + 1/r$. Now it is a simple exercise to verify $\|\cdot\|_{L^r} \leqslant \|\cdot\|_{L^1}^{1/r} \|\cdot\|_{L^\infty}^{1-1/r}$, and

$$\begin{aligned}\left\|\partial^\mu \left(\mathscr{F}^{-1}\psi_\lambda\right)\right\|_{L^1} &= \int_{\mathbb{T}^d} \left|\sum_k \partial^\mu \left(\mathscr{F}^{-1}_{\mathbb{R}^d}\psi_\lambda\right)(x + 2\pi k)\right| \mathrm{d}x \leqslant \int_{\mathbb{R}^d} \left|\partial^\mu \left(\mathscr{F}^{-1}_{\mathbb{R}^d}\psi_\lambda\right)(x)\right| \mathrm{d}x \\ &= \lambda^{|\mu|} \int_{\mathbb{R}^d} \lambda^d \left|\left(\partial^\mu \mathscr{F}^{-1}_{\mathbb{R}^d}\psi\right)(\lambda x)\right| \mathrm{d}x \simeq \lambda^{|\mu|},\end{aligned}$$

whereas

$$\begin{aligned}\sup_{x\in\mathbb{T}^d} \Big|\sum_k \partial^\mu (\mathscr{F}^{-1}_{\mathbb{R}^d}\psi_\lambda)(x + 2\pi k)\Big| &= \lambda^{d+|\mu|} \sup_{x\in\mathbb{T}^d} \Big|\sum_k (\partial^\mu \mathscr{F}^{-1}_{\mathbb{R}^d}\psi)(\lambda(x + 2\pi k))\Big| \\ &\lesssim \lambda^{d+|\mu|} \sup_{x\in\mathbb{T}^d} \sum_k (1 + \lambda|x + 2\pi k|)^{-2d} \\ &\lesssim \lambda^{d+|\mu|} \sup_{x\in\mathbb{T}^d} \sum_k (1 + |x + 2\pi k|)^{-2d} \lesssim \lambda^{d+|\mu|}.\end{aligned}$$

We end up with

$$\|\partial^\mu f\|_{L^q} \lesssim \|f\|_{L^p} \left\|\partial^\mu \left(\mathscr{F}^{-1}\psi_\lambda\right)\right\|_{L^r} \lesssim \|f\|_{L^p} \lambda^{|\mu|/r} \lambda^{(d+|\mu|)(1-1/r)} = \|f\|_{L^p} \lambda^{|\mu|+d(1/p-1/q)}.$$

□

It then follows immediately that for $\alpha \in \mathbb{R}$, $f \in \mathscr{C}^\alpha$, $\mu \in \mathbb{N}_0^d$, we have $\partial^\mu f \in \mathscr{C}^{\alpha-|\mu|}$. Another simple application of the Bernstein inequalities is the Besov embedding theorem, the proof of which we leave as an exercise.

Lemma 8 *(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 $B^\alpha_{p_1,q_1}$ is continuously embedded into $B^{\alpha-d(1/p_1-1/p_2)}_{p_2,q_2}$.*

Exercise 11 In the setting of Exercise 7, use Besov embedding to show that $\mathbb{E}[\|\tilde{\xi}\|^p_{-d/2-\varepsilon}] < \infty$ for all $p \geqslant 1$ and $\varepsilon > 0$ (in particular $\tilde{\xi} \in \mathscr{C}^{-d/2-}$ almost surely).

Hint: Estimate $\mathbb{E}[\|\tilde{\xi}\|^{2p}_{B^\alpha_{2p,2p}}]$ using Gaussian hypercontractivity (equivalence of moments).

As another application of the Bernstein inequality, let us show that $\mathscr{C}^\alpha = C^\alpha$ for $\alpha \in (0, 1)$.

Lemma 9 *For $\alpha \in (0, 1)$ we have $\mathscr{C}^\alpha = C^\alpha$, the space of α-Hölder continuous functions, and*

$$\|f\|_\alpha \simeq \|f\|_{C^\alpha} = \|f\|_{L^\infty} + \sup_{x\neq y} \frac{|f(x) - f(y)|}{d_{\mathbb{T}^d}(x, y)^\alpha},$$

where $d_{\mathbb{T}^d}(x, y)$ denotes the canonical distance on $\mathbb{T}^d$.

Proof Start by noting that for $f \in \mathscr{C}^\alpha$ we have $\|f\|_{L^\infty} \leqslant \sum_j \|\Delta_j f\|_{L^\infty} \leqslant \sum_j 2^{-j\alpha}\|f\|_\alpha \lesssim \|f\|_\alpha$. Let now $x \neq y \in \mathbb{T}^d$ and choose j_0 with $2^{-j_0} \simeq d_{\mathbb{T}^d}(x,y)$. For $j \leqslant j_0$ we use Bernstein's inequality to obtain

$$|\Delta_j f(x) - \Delta_j f(y)| \lesssim \|\mathrm{D}\Delta_j f\|_{L^\infty} d_{\mathbb{T}}(x,y) \lesssim 2^j \|\Delta_j f\|_{L^\infty} d_{\mathbb{T}}(x,y) \leqslant 2^{j(1-\alpha)}\|f\|_\alpha d_{\mathbb{T}}(x,y),$$

whereas for $j > j_0$ we simply estimate

$$|\Delta_j f(x) - \Delta_j f(y)| \lesssim \|\Delta_j f\|_{L^\infty} \lesssim 2^{-j\alpha}\|f\|_\alpha.$$

Summing over j, we get

$$\begin{aligned}|f(x) - f(y)| &\leqslant \sum_{j \leqslant j_0} 2^{j(1-\alpha)}\|f\|_\alpha d_{\mathbb{T}}(x,y) + \sum_{j > j_0} 2^{-j\alpha}\|f\|_\alpha \\ &\simeq \|f\|_\alpha (2^{j_0(1-\alpha)} d_{\mathbb{T}}(x,y) + 2^{-j_0\alpha}) \simeq \|f\|_\alpha d_{\mathbb{T}^d}(x,y)^\alpha.\end{aligned}$$

Conversely, if $f \in C^\alpha$, then we estimate $\|\Delta_{-1} f\|_{L^\infty} \lesssim \|f\|_{L^\infty}$. For $j \geqslant 0$, the function ρ_j satisfies $\int \rho_j \mathrm{d}x = 0$, and therefore

$$\begin{aligned}|\Delta_j f(x)| &= \Big| \int_{\mathbb{T}^d} \mathscr{F}^{-1}\rho_j(x-y)(f(y) - f(x))\mathrm{d}y \Big| \\ &= \Big| \int_{\mathbb{T}^d} \sum_k \mathscr{F}^{-1}_{\mathbb{R}^d}\rho_j(x - y + 2\pi k)(f(y) - f(x))\mathrm{d}y \Big| \\ &= \Big| \int_{\mathbb{R}^d} \mathscr{F}^{-1}_{\mathbb{R}^d}\rho_j(x-y)(f(y) - f(x))\mathrm{d}y \Big|.\end{aligned}$$

Now $|f(y) - f(x)| \leqslant \|f\|_{C^\alpha} d_{\mathbb{T}}(x,y)^\alpha \leqslant \|f\|_{C^\alpha}|x-y|^\alpha$, and thus we end up with

$$|\Delta_j f(x)| \leqslant \|f\|_{C^\alpha} \left| 2^{jd} \int_{\mathbb{R}^d} \left| \left(\mathscr{F}^{-1}_{\mathbb{R}^d}\rho\right)(2^j(x-y)) \right| |x-y|^\alpha \mathrm{d}y \right| \lesssim \|f\|_{C^\alpha} 2^{-j\alpha}.$$

□

The following lemma, a characterization of Besov regularity for functions that can be decomposed into pieces which are localized in Fourier space, will be immensely useful in what follows.

Lemma 10

1. *Let $\mathscr{A}$ be an annulus, let $\alpha \in \mathbb{R}$, and let (u_j) be a sequence of smooth functions such that $\mathscr{F}u_j$ has its support in $2^j\mathscr{A}$, and such that $\|u_j\|_{L^\infty} \lesssim 2^{-j\alpha}$ for all j. Then*

$$u = \sum_{j \geqslant -1} u_j \in \mathscr{C}^\alpha \qquad \text{and} \qquad \|u\|_\alpha \lesssim \sup_{j \geqslant -1}\{2^{j\alpha}\|u_j\|_{L^\infty}\}.$$

2. *Let $\mathscr{B}$ be a ball, let $\alpha > 0$, and let (u_j) be a sequence of smooth functions such that $\mathscr{F}u_j$ has its support in $2^j\mathscr{B}$, 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 If $\mathscr{F}u_j$ is supported in $2^j\mathscr{A}$, then $\Delta_i u_j \neq 0$ only for $i \sim j$. Hence, we obtain

$$\|\Delta_i u\|_{L^\infty} \leqslant \sum_{j:j\sim i} \|\Delta_i u_j\|_{L^\infty} \leqslant \sup_{k\geqslant -1}\{2^{k\alpha}\|u_k\|_{L^\infty}\} \sum_{j:j\sim i} 2^{-j\alpha} \simeq \sup_{k\geqslant -1}\{2^{k\alpha}\|u_k\|_{L^\infty}\}2^{-i\alpha}.$$

If $\mathscr{F}u_j$ is supported in $2^j\mathscr{B}$, then $\Delta_i u_j \neq 0$ only for $i \lesssim j$. Therefore,

$$\|\Delta_i u\|_{L^\infty} \leqslant \sum_{j:j\gtrsim i} \|\Delta_i u_j\|_{L^\infty} \leqslant \sup_{k\geqslant -1}\{2^{k\alpha}\|u_k\|_{L^\infty}\} \sum_{j:j\gtrsim i} 2^{-j\alpha} \lesssim \sup_{k\geqslant -1}\{2^{k\alpha}\|u_k\|_{L^\infty}\}2^{-i\alpha},$$

using $\alpha > 0$ in the last step. □

When solving SPDEs, we will need the smoothing properties of the heat semigroup. For that purpose we study functions of time with values in distribution spaces. If $\mathbb{X}$ is a Banach space with norm $\|\cdot\|_{\mathbb{X}}$ and if $T > 0$, then we define $C\mathbb{X}$ and $C_T\mathbb{X}$ as the spaces of continuous functions from $[0,\infty)$ respectively $[0,T]$ to $\mathbb{X}$, and $C_T\mathbb{X}$ is equipped with the supremum norm $\|\cdot\|_{C_T\mathbb{X}}$. If $\alpha \in (0,1)$ then we write $C^\alpha\mathbb{X}$ for the functions in $C\mathbb{X}$ that are α–Hölder continuous on every interval $[0,T]$, and we write

$$\|f\|_{C_T^\alpha\mathbb{X}} = \sup_{0\leqslant s<t\leqslant T} \frac{\|f(t)-f(s)\|}{|t-s|^\alpha}.$$

We then define $\mathscr{L}^\alpha = C\mathscr{C}^\alpha \cap C^{\alpha/2}L^\infty$ for $\alpha \in (0,2)$. For $T > 0$ we set $\mathscr{L}_T^\alpha = C_T\mathscr{C}^\alpha \cap C_T^{\alpha/2}L^\infty$ and we equip $\mathscr{L}_T^\alpha$ with the norm

$$\|\cdot\|_{\mathscr{L}_T^\alpha} = \max\{\|\cdot\|_{C_T\mathscr{C}^\alpha}, \|\cdot\|_{C_T^{\alpha/2}L^\infty}\}.$$

The notation $\mathscr{L}^\alpha$ is chosen to be reminiscent of the operator $\mathscr{L} = \partial_t - \Delta$ and indeed the parabolic spaces $\mathscr{L}^\alpha$ are adapted to $\mathscr{L}$ in the sense that the temporal regularity "counts twice", which is due to the fact that $\mathscr{L}$ contains a first order temporal but a second order spatial derivative. If we would replace Δ by a fractional Laplacian $-(-\Delta)^\sigma$, then we would have to consider the space $C\mathscr{C}^\alpha \cap C^{\alpha/(2\sigma)}L^\infty$ instead of $\mathscr{L}^\alpha$.

We have the following Schauder estimate on the scale of $(\mathscr{L}^\alpha)_\alpha$ spaces:

Lemma 11 *Let $\alpha \in (0,2)$ and let $(P_t)_{t\geqslant 0}$ be the semigroup generated by the periodic Laplacian, $\mathscr{F}(P_t f)(k) = e^{-t|k|^2}\mathscr{F}f(k)$. For $f \in C\mathscr{C}^{\alpha-2}$ define $Jf(t) = \int_0^t P_{t-s}f_s \mathrm{d}s$. Then*

$$\|Jf\|_{\mathscr{L}_T^\alpha} \lesssim (1+T)\|f\|_{C_T\mathscr{C}^{\alpha-2}}$$

for all $T > 0$. If $u \in \mathscr{C}^\alpha$, then $t \mapsto P_t u \in \mathscr{L}^\alpha$ and

$$\|t \mapsto P_t u\|_{\mathscr{L}_T^\alpha} \lesssim \|u\|_\alpha.$$

Bibliographic notes For a gentle introduction to Littlewood–Paley theory and Besov spaces see the recent monograph [BCD11], where most of our results are taken from. There the case of tempered distributions on $\mathbb{R}^d$ is considered. The theory on the torus is developed in [ST87]. The Schauder estimates for the heat semigroup are classical and can be found in [GIP13, GP14].

4 Diffusion in a random environment

Let us consider the following d-dimensional homogenization problem. Fix $\varepsilon > 0$ and let $u^\varepsilon : \mathbb{R}_+ \times \mathbb{T}^d \to \mathbb{R}$ be the solution to the Cauchy problem

$$\partial_t u^\varepsilon(t,x) = \Delta u^\varepsilon(t,x) + \varepsilon^{-\alpha} V(x/\varepsilon) u^\varepsilon(t,x), \qquad u^\varepsilon(0) = u_0,$$

where $V : \mathbb{T}^d_\varepsilon \to \mathbb{R}$ is a random field defined on the rescaled torus $\mathbb{T}^d_\varepsilon = (\mathbb{R}/(2\pi\varepsilon^{-1}\mathbb{Z}))^d$. This model describes the diffusion of particles in a random medium (replacing ∂_t by $i\partial_t$ gives the Schrödinger equation of a quantum particle evolving in a random potential). For a review of related results the reader can give a look at the recent paper of Bal and Gu [BG13]. The limit $\varepsilon \to 0$ corresponds to looking at the large scale behavior of the model since it can be understood as the equation for the *macroscopic* density $u^\varepsilon(t,x) = u(t/\varepsilon^2, x/\varepsilon)$ which corresponds to a *microscopic* density $u : \mathbb{R}_+ \times \mathbb{T}^d_\varepsilon \to \mathbb{R}$ evolving according to the parabolic equation

$$\partial_t u(t,x) = \Delta u(t,x) + \varepsilon^{2-\alpha} V(x) u(t,x), \qquad u^\varepsilon(0,\cdot) = u_0(\varepsilon\cdot).$$

We assume that $V : \mathbb{T}^d_\varepsilon \to \mathbb{R}$ is Gaussian and has mean zero and homogeneous correlation function C_ε given by

$$C_\varepsilon(x-y) = \mathbb{E}[V(x)V(y)] = (\varepsilon/\sqrt{2\pi})^d \sum_{k \in \varepsilon\mathbb{Z}^d_0} e^{i\langle x-y,k\rangle} R(k)$$

where $\mathbb{Z}^d_0 = \mathbb{Z}^d \backslash \{0\}$. On $R : \mathbb{R}^d \to \mathbb{R}_+$ we make the following hypothesis: for some $\beta \in (0,d]$ we have $R(k) = |k|^{\beta-d}\tilde{R}(k)$ where $\tilde{R} \in \mathscr{S}(\mathbb{R}^d)$ is a smooth radial function of rapid decay. For $\beta < d$ it would be equivalent to require that spatial correlations (in the limit $\varepsilon \to 0$) decay as $|x|^{-\beta}$. For $\beta = d$ this hypothesis means that spatial correlations are of rapid decay. Indeed by dominated convergence

$$\begin{aligned}\lim_{\varepsilon\to 0} C_\varepsilon(x) &= \int_{\mathbb{R}^d} \frac{\mathrm{d}k}{(2\pi)^{d/2}} e^{i\langle x,k\rangle} R(k) = \int_{\mathbb{R}^d} \frac{\mathrm{d}k}{(2\pi)^{d/2}} e^{i\langle x,k\rangle} |k|^{\beta-d} \tilde{R}(k) \\ &= (2\pi)^{d/2} \left(\mathscr{F}^{-1}_{\mathbb{R}^d}(|\cdot|^{\beta-d}) * \mathscr{F}^{-1}_{\mathbb{R}^d}(\tilde{R}) \right)(x).\end{aligned}$$

Now $\mathscr{F}^{-1}_{\mathbb{R}^d}(\tilde{R}) \in \mathscr{S}(\mathbb{R}^d)$ and $\mathscr{F}^{-1}_{\mathbb{R}^d}(|\cdot|^{\beta-d})(x) \simeq |x|^{-\beta}$ if $0 < \beta < d$ (see for example Proposition 1.29 of [BCD11]), so $\lim_{\varepsilon\to 0} |C_\varepsilon(x)| \lesssim |x|^{-\beta}$ for $|x| \to +\infty$.

Let us write $V_\varepsilon(x) = \varepsilon^{-\alpha} V(x/\varepsilon)$ so that $\partial_t u^\varepsilon = \Delta u^\varepsilon + V_\varepsilon u^\varepsilon$, and let us compute the variance of the Littlewood–Paley blocks of V_ε.

In order to perform more easily some computations we can introduce a family of centered complex Gaussian random variables $\{g(k)\}_{k\in\varepsilon\mathbb{Z}_0}$ such that $g(k)^* = g(-k)$ and $\mathbb{E}[g(k)g(k')] = \delta_{k+k'=0}$ and represent $V_\varepsilon(x)$ as

$$V_\varepsilon(x) = \frac{\varepsilon^{d/2-\alpha}}{\left(\sqrt{2\pi}\right)^{d/2}} \sum_{k\in\varepsilon\mathbb{Z}_0^d} e^{i\langle x,k/\varepsilon\rangle}\sqrt{R(k)}g(k)$$

Lemma 12 *Assume $\beta - 2\alpha \geqslant 0$.We have for any $\varepsilon > 0$ and $i \geqslant 0$ and any $0 \leqslant \kappa \leqslant \beta - 2\alpha$:*

$$\mathbb{E}[|\Delta_i V_\varepsilon(x)|^2] \lesssim 2^{(2\alpha+\kappa)i}\varepsilon^\kappa.$$

This estimate implies that if $\beta > 2\alpha$, there exists $\kappa > 0$ with $V_\varepsilon \to 0$ in $L^2(\Omega; B_{2,2}^{-\alpha-\kappa/2}(\mathbb{T}^d))$ as $\varepsilon \to 0$.

Proof A spectral computation gives

$$\Delta_i V_\varepsilon(x) = \frac{\varepsilon^{d/2-\alpha}}{(\sqrt{2\pi})^{d/2}} \sum_{k\in\varepsilon\mathbb{Z}_0^d} e^{i\langle x,k/\varepsilon\rangle}\rho_i(k/\varepsilon)\sqrt{R(k)}g(k)$$

so

$$\begin{aligned}\mathbb{E}[|\Delta_i V_\varepsilon(x)|^2] &= \varepsilon^d(\sqrt{2\pi})^d\varepsilon^{-2\alpha}\textstyle\sum_{k\in\varepsilon\mathbb{Z}_0^d}\rho_i(k/\varepsilon)^2 e^{i\langle x,k/\varepsilon\rangle}R(k)\\ &= (\sqrt{2\pi})^d\varepsilon^{d-2\alpha}\textstyle\sum_{k\in\varepsilon\mathbb{Z}_0^d}\rho(k/(\varepsilon 2^i))^2R(k) \qquad (14)\\ &\lesssim \varepsilon^{d-2\alpha}2^{id}\sup_{k\in\varepsilon 2^i\mathscr{A}}R(k),\end{aligned}$$

where $\mathscr{A}$ is the annulus in which ρ is supported. Now if $\varepsilon 2^i \leqslant 1$ we have $\mathbb{E}[|\Delta_i V_\varepsilon(x)|^2] \lesssim 2^{id}\varepsilon^{d-2\alpha}(\varepsilon 2^i)^{\beta-d} = \varepsilon^{\beta-2\alpha}2^{i\beta}$. The assumption $\beta - 2\alpha \geqslant 0$ then implies $\mathbb{E}[|\Delta_i V_\varepsilon(x)|^2] \lesssim 2^{(2\alpha+\kappa)i}\varepsilon^\kappa$ for any $0 \leqslant \kappa \leqslant \beta - 2\alpha$. In the case $\varepsilon 2^i > 1$ we use that $\int_{B(0,1)^c} R(k)\mathrm{d}k < +\infty$ to estimate

$$\varepsilon^d \sum_{k\in\varepsilon\mathbb{Z}^d}\rho(k/(\varepsilon 2^i))^2R(k) \lesssim \varepsilon^d\sum_{k\in\mathbb{Z}^d}R(\varepsilon k) \lesssim \int_{\mathbb{R}^d}R(k)\mathrm{d}k < +\infty,$$

and then $\mathbb{E}[|\Delta_i V_\varepsilon(x)|^2] \lesssim \varepsilon^{-2\alpha} \lesssim 2^{2\alpha i}(\varepsilon 2^i)^\kappa$ for any small $\kappa > 0$. □

Note that the computation carried out in equation (14) implies also that if $\beta - 2\alpha < 0$, then essentially V_ε does not converge in any reasonable sense since the variance of the Littlewood–Paley blocks explodes.

Remark 3 The same calculation as in (14) shows that

$$\mathbb{E}[\Delta_i V_\varepsilon(x)\Delta_j V_\varepsilon(x)] = 0$$

whenever $|i - j| > 1$, because in that case $\rho_i\rho_j \equiv 0$.

The previous analysis shows that it is reasonable to take $\alpha \leqslant \beta/2$ in order to have some hope of obtaining a well defined limit as $\varepsilon \to 0$. In this case V_ε stays bounded (at least) in spaces of distributions of regularity $\alpha-$. This brings us to the problem of obtaining estimates for the parabolic PDE

$$\mathscr{L}u^\varepsilon(t,x) = (\partial_t - \Delta)u^\varepsilon(t,x) = V_\varepsilon(x)u^\varepsilon(t,x), \qquad (t,x) \in [0,T] \times \mathbb{T}^d,$$

depending only on negative regularity norms of V_ε. On one side the regularity of u^ε is then limited by the regularity of the right hand side which cannot be better than that of V_ε. On the other side the product of V_ε with u_ε can cause problems since we try to multiply an (a priori) irregular object with one of limited regularity.

Assume that $V_\varepsilon \in \mathscr{C}^{\gamma-2}$ with $\gamma > 0$. It is then reasonable to assume that also $V_\varepsilon u^\varepsilon \in C_T\mathscr{C}^{\gamma-2}$ and that $u^\varepsilon \in C_T\mathscr{C}^\gamma$ as a consequence of the regularising effect of the heat operator (Lemma 11). We will see in Section 5.1 below that the product $V_\varepsilon u^\varepsilon$ is under control only if $\gamma + \gamma - 2 > 0$, that is if $\gamma > 1$. Since $V_\varepsilon \to 0$ in $\mathscr{C}^{-1+}$, it is not difficult to show that u^ε converges as $\varepsilon \to 0$ to the solution u of the linear equation $\mathscr{L}u = 0$. In this case the random potential will not have any effect in the limit.

The interesting situation then is when $\gamma \leqslant 1$. To understand what could happen in this case let us use a simple transformation of the solution. Write $u^\varepsilon = \exp(X^\varepsilon)v^\varepsilon$ where X^ε satisfies the equation $\mathscr{L}X^\varepsilon = V^\varepsilon$ with initial condition $X^\varepsilon(0,\cdot) = 0$. Then

$$\mathscr{L}u^\varepsilon = \exp(X^\varepsilon)\left(v^\varepsilon \mathscr{L}X^\varepsilon + \mathscr{L}v^\varepsilon - v^\varepsilon(\partial_x X^\varepsilon)^2 - 2\langle \partial_x X^\varepsilon, \partial_x v^\varepsilon\rangle_{\mathbb{R}^d}\right) = \exp(X^\varepsilon)v^\varepsilon V_\varepsilon.$$

Since $\exp(X^\varepsilon) > 0$ on $[0,T] \times \mathbb{T}^d$, this implies that v^ε satisfies

$$\mathscr{L}v^\varepsilon - v^\varepsilon|\partial_x X^\varepsilon|^2 - 2\langle \partial_x X^\varepsilon, \partial_x v^\varepsilon\rangle_{\mathbb{R}^d} = 0, \qquad (t,x) \in [0,T] \times \mathbb{T}^d.$$

Our Schauder estimates imply that $X^\varepsilon \in C_T\mathscr{C}^\gamma$ with uniform bounds in $\varepsilon > 0$, so that the problematic term is $|\partial_x X^\varepsilon|^2$ for which this estimate does not guarantee existence.

Note that

$$\partial_x X_\varepsilon(t,x) = \frac{\varepsilon^{d/2-\alpha}}{(\sqrt{2\pi})^{d/2}} \sum_{k \in \varepsilon\mathbb{Z}^d_0} e^{i\langle x, k/\varepsilon\rangle} G_\varepsilon(t,k)g(k) \tag{15}$$

with

$$G_\varepsilon(t,k) = i\frac{k}{\varepsilon}\frac{[1 - e^{-t|k/\varepsilon|^2}]}{|k/\varepsilon|^2}\sqrt{R(k)}.$$

Lemma 13 *Assume that*

$$\sigma^2 = (\sqrt{2\pi})^d \int_{\mathbb{R}^d} \frac{R(k)}{k^2}\mathrm{d}k < +\infty.$$

Then if $\alpha = 1$ and $t > 0$ we have

$$\lim_{\varepsilon \to 0} \mathbb{E}[|\partial_x X^\varepsilon|^2(t,x)] = \sigma^2,$$

and if $\alpha < 1$ *and* $t > 0$

$$\lim_{\varepsilon \to 0} \mathbb{E}[(|\partial_x X^\varepsilon|)^2(t,x)] = 0.$$

Moreover

$$\mathrm{Var}[\Delta_q(|\partial_x X^\varepsilon|^2)(t,x)] \lesssim \varepsilon^{4-4\alpha} \min(\sigma^4, (\varepsilon 2^q)^{\beta-2} \|\tilde{R}\|_\infty \sigma^2).$$

Proof A computation similar to that leading to equation (14) gives

$$\begin{aligned}\mathbb{E}[|\partial_x X^\varepsilon|^2(t,x)] &= \varepsilon^d (\sqrt{2\pi})^d \varepsilon^{-2\alpha} \sum_{k \in \varepsilon \mathbb{Z}_0^d} |k/\varepsilon|^2 \Big[\int_0^t e^{-(t-s)|k/\varepsilon|^2} \mathrm{d}s \Big]^2 R(k) \\ &= \varepsilon^d (\sqrt{2\pi})^d \varepsilon^{2-2\alpha} \sum_{k \in \varepsilon \mathbb{Z}_0^d} \frac{[1 - e^{-t(k/\varepsilon)^2}]^2}{k^2} R(k),\end{aligned}$$

which for $\varepsilon \to 0$ and $t > 0$ tends to

$$\lim_{\varepsilon \to 0} \mathbb{E}[|\partial_x X^\varepsilon|^2(t,x)] = \mathbb{I}_{\alpha<1} (\sqrt{2\pi})^d \int_{\mathbb{R}^d} \frac{R(k)}{k^2} \mathrm{d}k = \mathbb{I}_{\alpha<1} \sigma^2.$$

Let us now study the variance of $|\partial_x X^\varepsilon|^2(t,x)$. Using equation (15) we have

$$\Delta_q(|\partial_x X^\varepsilon|^2)(t,x) = \frac{\varepsilon^{d-2\alpha}}{(2\pi)^{d/2}} \sum_{k_1,k_2 \in \varepsilon \mathbb{Z}_0^d} e^{i\langle k_1+k_2, x/\varepsilon \rangle} \rho_q((k_1+k_2)/\varepsilon) G_\varepsilon(t,k_1) G_\varepsilon(t,k_2) g(k_1) g(k_2).$$

By Wick's theorem ([Jan97], Theorem 1.28)

$$\begin{aligned}\mathrm{Cov}(g(k_1)g(k_2), g(k_1')g(k_2')) &= \mathbb{E}[g(k_1)g(k_1')]\mathbb{E}[g(k_2)g(k_2')] + \mathbb{E}[g(k_1)g(k_2')]\mathbb{E}[g(k_2)g(k_1')] \\ &= \mathbb{I}_{k_1+k_1'=k_2+k_2'=0} + \mathbb{I}_{k_1+k_2'=k_2+k_1'=0},\end{aligned}$$

which implies

$$\mathrm{Var}[\Delta_q(|\partial_x X^\varepsilon|^2)(t,x)] = \frac{\varepsilon^{2d-4\alpha}}{(2\pi)^d} \sum_{k_1,k_2 \in \varepsilon \mathbb{Z}_0^d} (\rho_q((k_1+k_2)/\varepsilon))^2 |G_\varepsilon(t,k_1)|^2 |G_\varepsilon(t,k_2)|^2.$$

For any $q \geqslant 0$ (the case $q = -1$ is left to the reader), the variables k_1 and k_2 are bounded away from 0 and we have

$$\mathrm{Var}[\Delta_q(|\partial_x X^\varepsilon|^2)(t,x)] \lesssim \varepsilon^{2d+4-4\alpha} \sum_{k_1,k_2 \in \varepsilon \mathbb{Z}_0^d} (\rho_q((k_1+k_2)/\varepsilon))^2 \frac{|R(k_1)||R(k_2)|}{|k_1|^2 |k_2|^2}.$$

A first estimate is obtained by just dropping the factor $\rho_q((k_1+k_2)/\varepsilon)$ and results in the bound

$$\mathrm{Var}[\Delta_q(|\partial_x X^\varepsilon|^2)(t,x)] \lesssim \varepsilon^{2d+4-4\alpha} \sum_{k_1,k_2 \in \varepsilon \mathbb{Z}_0^d} \frac{|R(k_1)||R(k_2)|}{|k_1|^2 |k_2|^2} \lesssim \varepsilon^{4-4\alpha} \sigma^4$$

Another estimate proceed by taking into account the constraint given by the support of $\rho_q((k_1+k_2)/\varepsilon)$. In order to satisfy $k_1+k_2 \sim \varepsilon 2^q$ we must have $k_2 \lesssim k_1 \sim \varepsilon 2^q$ or $\varepsilon 2^q \lesssim k_1 \sim k_2$. In the first case

$$\varepsilon^{2d+4-4\alpha} \sum_{k_1,k_2\in\varepsilon\mathbb{Z}_0^d} \mathbb{I}_{k_2\lesssim k_1\sim\varepsilon 2^q} \frac{|R(k_1)||R(k_2)|}{|k_1|^2|k_2|^2} \lesssim 2^{q(\beta-2)}\varepsilon^{d+\beta+2-4\alpha}\|\tilde{R}\|_\infty \sum_{k_2\in\varepsilon\mathbb{Z}_0^d} \mathbb{I}_{k_2\lesssim\varepsilon 2^q}\frac{|R(k_2)|}{|k_2|^2}$$
$$\lesssim (\varepsilon 2^q)^{\beta-2}\|\tilde{R}\|_\infty \int \mathrm{d}k \frac{|R(k)|}{|k|^2} \lesssim (\varepsilon 2^q)^{\beta-2}\varepsilon^{4-4\alpha}\|\tilde{R}\|_\infty\sigma^2$$

since $|R(k_1)|/|k_1|^2 \lesssim \|\tilde{R}\|_\infty(\varepsilon 2^q)^{\beta-d-2}$. If $\varepsilon 2^q \lesssim k_1 \sim k_2$ we similarly have

$$\varepsilon^{2d+4-4\alpha} \sum_{k_1,k_2\in\varepsilon\mathbb{Z}_0^d} \mathbb{I}_{\varepsilon 2^q\lesssim k_1\sim k_2} \frac{|R(k_1)||R(k_2)|}{|k_1|^2|k_2|^2} \lesssim 2^{q(\beta-2)}\varepsilon^{d+\beta+2-4\alpha}\|\tilde{R}\|_\infty \sum_{k_2\in\varepsilon\mathbb{Z}_0^d} \mathbb{I}_{\varepsilon 2^q\lesssim k_2}\frac{|R(k_2)|}{|k_2|^2}$$
$$\lesssim (\varepsilon 2^q)^{\beta-2}\varepsilon^{4-4\alpha}\|\tilde{R}\|_\infty \int \mathrm{d}k \frac{|R(k)|}{|k|^2} \lesssim (\varepsilon 2^q)^{\beta-2}\varepsilon^{4-4\alpha}\|\tilde{R}\|_\infty\sigma^2.$$

□

This lemma shows that the interesting situation is $\alpha = 1$. Then provided $\sigma^2 < +\infty$ and $\beta > 2$ we have $|\partial_x X^\varepsilon|^2(t) \to \sigma^2$ in $\mathscr{C}^{0-}$ for all $t > 0$, and in fact the convergence is uniform on $[c, C]$ for all $0 < c < C$. An easy consequence of this is that v^ε converges to the solution of the PDE

$$\mathscr{L}v = \sigma^2 v \tag{16}$$

and since $X^\varepsilon \to 0$ in $C_T\mathscr{C}^\gamma$ we finally obtain the convergence of $(u^\varepsilon)_{\varepsilon>0}$ to the same v.

4.1 The 2d generalized parabolic Anderson model

The case $\alpha = 1$ and $\beta = 2$ remains open in the previous analysis. When $\beta = 2$ we cannot expect σ^2 to be finite and moreover from the above computations we see that the variance of $|\partial_x X^\varepsilon|^2$ remains finite and does not go to zero so the limiting object should satisfy a stochastic PDE rather than a deterministic one. If we let $\sigma_\varepsilon^2(t) = \mathbb{E}[|\partial_x X^\varepsilon|^2(t,x)]$ (which depends on time but which is easily shown to be independent of $x \in \mathbb{T}^2$), then we expect that solving the *renormalized* equation

$$\mathscr{L}\tilde{u}^\varepsilon = V_\varepsilon \tilde{u}^\varepsilon - \sigma_\varepsilon^2\tilde{u}^\varepsilon$$

should give rise in the limit to a well defined random field $\tilde{u}$ satisfying $\tilde{u} = e^X\tilde{v}$, where

$$\mathscr{L}\tilde{v} = \tilde{v}\eta + 2\langle\partial_x X, \partial_x\tilde{v}\rangle_{\mathbb{R}^d}$$

and where X is the limit of X^ε as $\varepsilon \to 0$ while η is the limit of $(\partial_x X^\varepsilon)^2 - \sigma_\varepsilon^2(t)$. The relation of u^ε with $\tilde{u}^\varepsilon$ is easily seen to be $\tilde{u}^\varepsilon(t,x) = e^{-\int_0^t \sigma_\varepsilon^2(s)\mathrm{d}s}u^\varepsilon(t,x)$. The renormalization procedure is therefore equivalent to a time–dependent rescaling of the solution to the initial problem.

We will study the renormalization and convergence problem for a more general equation of the form

$$\mathscr{L}u^\varepsilon = F(u^\varepsilon)V_\varepsilon, \tag{17}$$

where $F : \mathbb{R} \to \mathbb{R}$ is a sufficiently smooth non–linearity. One possible motivation is that if z^ε solves the linear PDE and we set $u^\varepsilon = \varphi(z^\varepsilon)$ for some invertible $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi' > 0$, then

$$\mathscr{L}u^\varepsilon = \varphi'(z^\varepsilon)\mathscr{L}z^\varepsilon - \varphi''(z^\varepsilon)|\partial_x z^\varepsilon|^2 = \varphi'(z^\varepsilon)z^\varepsilon V_\varepsilon - \varphi''(z^\varepsilon)(\varphi'(z^\varepsilon))^{-2}|\partial_x u^\varepsilon|^2$$

and thus u^ε satisfies the PDE

$$\mathscr{L}u^\varepsilon = F_1(u^\varepsilon)V_\varepsilon + F_2(u^\varepsilon)(\partial_x u^\varepsilon)^2$$

where $F_1(x) = \varphi'(\varphi^{-1}(x))\varphi^{-1}(x)$ and $F_2(x) = -\varphi''(\varphi^{-1}(x))(\varphi'(\varphi^{-1}(x)))^{-2}$. In the situation we are interested in, the second term in the right hand side is simpler to treat than the first term so, for the time being, we will drop it and we will concentrate on the equation (17) in $d = 2$ with $\alpha = 1$ and short ranged ($\beta = d$) potential V which we refer to as *generalized parabolic Anderson model* (GPAM).

Under these conditions V_ε converges to the white noise in space which we usually denote with ξ and our aim will be to set up a theory in which the non–linear operations involved in the definition of the dynamics of the GPAM are well defined, including the possibility of the renormalization which already appears in the linear case as hinted above.

While the reader should always have in mind a limiting procedure from a well defined model like the ones we were considering so far, in the following we will mostly discuss the limiting equation. The specific phenomena appearing when trying to track the oscillations of the term $F(u^\varepsilon)V_\varepsilon$ as $\varepsilon \to 0$ will be described by a *renormalized product* $F(u) \diamond \xi$ and so we write the GPAM as

$$\mathscr{L}u(t,x) = F(u(t,x)) \diamond \xi(x), \qquad u(0) = u_0. \tag{18}$$

In the linear case $F(u) = u$, the problem of the renormalization can be solved along the lines suggested above. Another possible line of attack comes from the theory of Gaussian spaces and in particular from Wick products, see for example [Hu02]. However, the definition of the Wick product relies on the concrete chaos expansion of its factors, and since nonlinear functions change the chaos expansion in a complicated way, there is little hope of directly extending the Wick product approach to the nonlinear case and moreover using these non–local (in the probability space) objects can deliver solutions which are not physically acceptable [Cha00].

Equation (18) is structurally very similar to the stochastic differential equation

$$\partial_t v(t) = F(v(t))\partial_t B^H(t), \qquad v(0) = v_0, \tag{19}$$

where B^H denotes a fractional Brownian motion with Hurst index $H \in (0,1)$. There are many ways to solve (19) in the Brownian case. Since we are interested in a way that might

extend to (18) where the irregularity appears along the two–dimensional spatial variable x, we should exclude all approaches based on information, filtrations, and a direction of time; in particular, any approach that works for $H \neq 1/2$ might seem promising. But Lyons' theory of rough paths [Lyo98] equips us exactly with the techniques we need to solve (19) for general H. More precisely, if for $H > 1/3$ we are given $\int_0^\cdot B_s^H \mathrm{d}B_s^H$, then we can use the controlled rough path integral [Gub04] to make sense of $\int_0^\cdot f_s \mathrm{d}B_s^H$ for any f which "looks like" B^H, and this allows us to solve (19). So the main ingredients required for controlled rough paths are the integral $\int_0^\cdot B_s^H \mathrm{d}B_s^H$ for the reference path B^H, and the fact that we can describe paths which look like B^H. It is worthwhile to note that while we need probability theory to construct $\int_0^\cdot B_s^H \mathrm{d}B_s^H$, the construction of $\int_0^\cdot f_s \mathrm{d}B_s^H$ is achieved using pathwise arguments and it is given as a continuous map of f and $\left(B^H, \int_0^\cdot B_s^H \mathrm{d}B_s^H\right)$. As a consequence, the solution to the SDE (19) depends pathwise continuously on $\left(B^H, \int_0^\cdot B_s^H \mathrm{d}B^H\right)$.

By the structural similarity of (18) and (19), we might hope to extend the rough path approach to (18). The equivalent of B^H is given by the solution ϑ to $\mathscr{L}\vartheta = \xi$, $\vartheta(0) = 0$, so that we should assume the renormalized product $\vartheta \diamond \xi$ as given. Then we might hope to define $f \diamond \xi$ for all f that "look like ϑ", however this is to be interpreted. Of course, rough paths can only be applied to functions of a one–dimensional index variable, while for (18) the problem lies in the irregularity of ξ in the spatial variable $x \in \mathbb{T}^2$.

In the following we combine the ideas from controlled rough paths with Bony's paraproduct, a tool from functional analysis that allows us to extend them to functions of a multidimensional parameter. Using the paraproduct, we are able to make precise in a simple way what we mean by "distributions looking like a reference distribution". We can then define products of suitable distributions and solve (18) as well as many other interesting singular SPDEs.

4.2 More singular problems

Keeping the homogenization problem as leitmotiv for these lectures, we could consider also space–time varying environments $V_\varepsilon(t,x) = \varepsilon^{-\alpha} V(t/\varepsilon^2, x/\varepsilon)$. The scaling of the temporal variable is chosen so that it is compatible with the diffusive scaling from a microscopic description, where $V(t,x)$ has typical variation in space and time in scales of order 1. Assume that $d = 1$, then when the random field V is Gaussian, zero mean, and with short–range space–time correlations, the natural choice for the magnitude of the macroscopic fluctuations is $\alpha = 3/2$. In this case V_ε converges as $\varepsilon \to 0$ to a space–time white noise ξ. Understanding the limit dynamics as $\varepsilon \to 0$ of the solution u^ε to the linear equation $\mathscr{L}u^\varepsilon = V_\varepsilon u^\varepsilon$ represents now a more difficult problem than in the time independent situation. A Gaussian computation shows that the random field X^ε, solution to $\mathscr{L}X^\varepsilon = V_\varepsilon$ (e.g. with zero initial condition), stays bounded in $C_T\mathscr{C}^{1/2-}$ as $\varepsilon \to 0$. Since $\mathscr{L}$ is a second order operator (if we use an appropriate parabolic weighting of the time and space regularities), ξ is expected to live in a space of distributions of regularity $-3/2-$. This is to be compared with the $-1-$ of the space white noise which

had to be dealt with in the GPAM. Renormalization effects are then expected to be stronger in this setting and the limiting object, which we denote with w, should satisfy a (suitably renormalized) linear stochastic heat equation with multiplicative noise (SHE)

$$\mathscr{L}w(t,x) = w(t,x) \diamond \xi(t,x), \qquad w(0) = w_0. \tag{20}$$

As hinted by the computations in the more regular case, it is useful to consider the change of variables $w = e^h$ which is called Cole–Hopf transformation. Here $h : [0,\infty) \times \mathbb{T} \to \mathbb{R}$ is a new unknown which satisfies now the Kardar–Parisi–Zhang (KPZ) equation:

$$\mathscr{L}h(t,x) = (\partial_x h(t,x))^{\diamond 2} + \xi(t,x), \qquad h(0) = h_0 \tag{21}$$

where the difficulty comes now from the squaring of the derivative but which has the nice feature to be additively perturbed by the space–time white noise, a feature which simplifies many considerations. Another relevant model in applications is obtained by taking the space derivative of KPZ and letting $u(t,x) = \partial_x h(t,x)$ in order to obtain the stochastic conservation law

$$\mathscr{L}u(t,x) = \partial_x (u(t,x))^{\diamond 2} + \partial_x \xi(t,x), \qquad u(0) = u_0, \tag{22}$$

which we will refer to as the stochastic Burgers equation (SBE). In all these cases, $\diamond$ denotes a suitably renormalized product.

The KPZ equation was derived by Kardar–Parisi–Zhang in 1986 as a universal model for the random growth of an interface [KPZ86]. For a long time it could not be solved due to the fact that there was no way to make sense of the nonlinearity $(\partial_x h)^{\diamond 2}$ in (21). The only way to make sense of KPZ was to apply the Cole-Hopf transform [BG97]: solve SHE (20) (which is accessible to Itô integration) and set $h = \log w$. But there was no intrinsic interpretation of what it means to solve (21). Finally in 2012, Hairer [Hai13] used rough paths to give a meaning to the equation and to obtain solutions directly at the KPZ level. In Section 6 we will sketch how to recover his solution in the paracontrolled setting. Applications of the techniques used by Hairer to solve the KPZ problem to a more general homogenization problem with ergodic potentials (not necessarily Gaussian) have been studied in [HPP13].

4.3 Hairer's regularity structures

In [Hai14], Hairer introduces a theory of regularity structures which can also be considered a generalization of the theory of controlled rough paths to functions of a multidimensional index variable. Hairer fundamentally rethinks the notion of regularity. Usually a function is called smooth if it can be approximated around every point by a polynomial of a given degree (the Taylor polynomial). Naturally, the solution to an SPDE driven by –say– Gaussian space-time white noise is not smooth in that sense. So in Hairer's theory, a function is called smooth if locally it can be approximated by the noise (and higher order terms constructed from the noise). This induces a natural topology in which the solutions to semilinear SPDEs depend continuously on the driving signal.

At this date it seems that the theory of regularity structures has a wider range of applicability than the paracontrolled approach described in [GIP13], but also at the expense of a very deep conceptual sophistication. There are problems (like the one–dimensional heat equation with multiplicative noise and general nonlinearity) that cannot be solved using paracontrolled distributions, but these problems seem also quite difficult (even if doable and work in progress) to tackle with regularity structures. Moreover, equations of a more general kind, say dispersive equations or wave equations, are still poorly (or not at all) understood in both approaches.

5 The paracontrolled PAM

As we have tried to motivate in the previous sections we are looking for a theory for PAM which describes the possible limits of the equation

$$\mathscr{L}u = F(u)\eta \tag{23}$$

driven by sufficiently regular η but as η is converging to the space white noise ξ. From this point of view we are looking for a priori estimates on the solution u to (23) which depend only on distributional norms of η. So in the following we will assume that we have at hand only a uniform control of η in $C_T\mathscr{C}^{\gamma-2}$ for some $\gamma > 0$. For the application to the 2d space white noise we could take $\gamma = 1-$, but we will not use this specific information in order to probe the range of applicability of our approach and we will only assume that the exponent γ is such that $3\gamma - 2 > 0$.

Assume for a moment that we are in the simpler situation $\gamma > 1$ and $u_0 \in \mathscr{C}^\gamma$ and let us try to solve equation (23) via Picard iterations $(u^n)_{n\geqslant 0}$ starting from $u^0 \equiv u_0$. Since F preserves the $C\mathscr{C}^\gamma$-regularity (which can be seen by identifying $C\mathscr{C}^\gamma$ with the classical space of bounded Hölder–continuous functions of space), the product $F(u^0(t))\eta$ is well defined as an element of $\mathscr{C}^{\gamma-2}$ for all $t \geqslant 0$ since $2\gamma - 2 > 0$ and we are in condition to apply Corollary 1 below on the product of elements in Hölder–Besov spaces. Now by Lemma 11, the Laplacian gains two degrees of regularity so that the solution u^1 to $\mathscr{L}u^1 = F(u^0)\eta$, $u^1(0) = u_0$, is in $C\mathscr{C}^\gamma$. From here we obtain a contraction on $C_T\mathscr{C}^\gamma$ for some small $T > 0$ whose value does not depend on u_0, which gives us global in time existence and uniqueness of solutions. Note that in one dimension the space white noise has regularity $\mathscr{C}^{-1/2-}$ (see Exercise 11) so taking $\gamma = 3/2-$ we have determined that the one–dimensional PAM can be solved globally in time with standard techniques.

When the condition $2\gamma - 2 > 0$ is not satisfied we still have that if $\eta \in C_T\mathscr{C}^{\gamma-2}$ then $u \in \mathscr{L}^\gamma = C_T\mathscr{C}^{\gamma-2} \cap C_T^{\gamma/2}L^\infty$ by the standard parabolic estimates of Lemma 11. However with the regularities at hand we cannot use Corollary 1 anymore to guarantee the continuity of the operator $(u, \eta) \mapsto F(u)\eta$. Moreover, as already seen in the simpler homogenization problems above this is not a technical difficulty but a real issue of the regime $\gamma \leqslant 1$. We expect that controlling the model in this regime can be quite tricky

since limits exists when $\eta \to 0$ but the limiting solution still feels residual order one effects from the vanishing driving signal η. This situation cannot be improved from the point of view of standard analytic considerations. What is needed is a finer control of the solution u which allows to analyse in more detail the possible resonances between the fluctuations of u and those of η.

Before going on we will revise the problem of multiplication of distributions in the scale of Hölder–Besov spaces, introducing the basic tool of our general analysis: Bony's paraproduct.

5.1 The paraproduct and the resonant term

Paraproducts are bilinear operations introduced by Bony [Bon81] to linearize a class of nonlinear hyperbolic PDEs in order to analyse the regularity of their solutions. In terms of Littlewood–Paley blocks, a general product fg of two distributions can be (at least formally) decomposed as

$$fg = \sum_{j\geqslant -1}\sum_{i\geqslant -1} \Delta_i f \Delta_j g = f \prec g + f \succ g + f \circ g.$$

Here $f \prec g$ is the part of the double sum with $i < j-1$, $f \succ g$ is the part with $i > j+1$, and $f \circ g$ is the "diagonal" part, where $|i-j| \leqslant 1$. More precisely, we define

$$f \prec g = g \succ f = \sum_{j\geqslant -1}\sum_{i=-1}^{j-2} \Delta_i f \Delta_j g \qquad \text{and} \qquad f \circ g = \sum_{|i-j|\leqslant 1} \Delta_i f \Delta_j g.$$

Of course, the decomposition depends on the dyadic partition of unity used to define the blocks Δ_j, and also on the particular choice of the pairs (i,j) in the diagonal part. The choice of taking all (i,j) with $|i-j| \leqslant 1$ into the diagonal part corresponds to the fact that the partition of unity can be chosen such that $\operatorname{supp} \mathscr{F}(\Delta_i f \Delta_j g) \subseteq 2^j \mathscr{A}$ if $i < j-1$, where $\mathscr{A}$ is a suitable annulus. If $|i-j| \leqslant 1$, the only apriori information on the spectral support of the various term in the double sum is $\operatorname{supp} \mathscr{F}(\Delta_i f \Delta_j g) \subseteq 2^j \mathscr{B}$, that is they are supported in balls and in particular they can have non–zero contributions to very low wave vectors. We call $f \prec g$ and $f \succ g$ *paraproducts*, and $f \circ g$ the *resonant* term.

Bony's crucial observation is that $f \prec g$ (and thus $f \succ g$) is always a well-defined distribution. Heuristically, $f \prec g$ behaves at large frequencies like g (and thus retains the same regularity), and f provides only a frequency modulation of g. The only difficulty in constructing fg for arbitrary distributions lies in handling the diagonal term $f \circ g$. The basic result about these bilinear operations is given by the following estimates.

Theorem 3 *(Paraproduct estimates) For any $\beta \in \mathbb{R}$ and $f, g \in \mathscr{S}'$ we have*

$$\|f \prec g\|_\beta \lesssim_\beta \|f\|_{L^\infty} \|g\|_\beta, \tag{24}$$

and for $\alpha < 0$ furthermore

$$\|f \prec g\|_{\alpha+\beta} \lesssim_{\alpha,\beta} \|f\|_\alpha \|g\|_\beta. \tag{25}$$

For $\alpha + \beta > 0$ we have

$$\|f \circ g\|_{\alpha+\beta} \lesssim_{\alpha,\beta} \|f\|_\alpha \|g\|_\beta. \tag{26}$$

Proof There exists an annulus $\mathscr{A}$ such that $S_{j-1} f \Delta_j g$ has Fourier transform supported in $2^j \mathscr{A}$, and for $f \in L^\infty$ we have

$$\|S_{j-1} f \Delta_j g\|_{L^\infty} \leqslant \|S_{j-1} f\|_{L^\infty} \|\Delta_j g\|_{L^\infty} \lesssim \|f\|_{L^\infty} 2^{-j\beta} \|g\|_\beta.$$

By Lemma 10, we thus obtain (24). The proof of of (25) and (26) works in the same way, where for estimating $f \circ g$ we need $\alpha + \beta > 0$ because the terms of the series are supported in a ball and not in an annulus. □

A simple corollary is then the following:

Corollary 1 *Let $f \in \mathscr{C}^\alpha$ and $g \in \mathscr{C}^\beta$ with $\alpha + \beta > 0$, then the product $(f, g) \mapsto fg$ is a bounded bilinear map from $\mathscr{C}^\alpha \times \mathscr{C}^\beta$ to $\mathscr{C}^{\alpha \wedge \beta}$. While $f \prec g$, $f \succ g$, and $f \circ g$ depend on the specific dyadic partition of unity, the product fg does not.*

The independence of the product from the dyadic partition of unity easily follows by taking smooth approximations.

The ill–posedness of $f \circ g$ for $\alpha + \beta \leqslant 0$ can be interpreted as a resonance effect since $f \circ g$ contains exactly those part of the double series where f and g are in the same frequency range. The paraproduct $f \prec g$ can be interpreted as frequency modulation of g, which should become more clear in the following example.

Example 3 In Figure 1 we see a slowly oscillating positive function u, while Figure 2 depicts a fast sine curve v. The product uv, which here equals the paraproduct $u \prec v$ since u has no rapidly oscillating components, is shown in Figure 3. We see that the local fluctuations of uv are due to v, and that uv is essentially oscillating with the same speed as v.

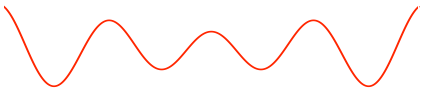

Figure 1: The function u

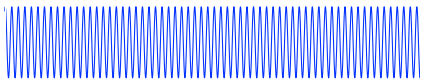

Figure 2: The function v

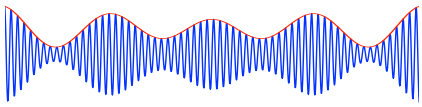

Figure 3: The function $u \prec v$

Example 4 If $f \in \mathscr{C}^\gamma(\mathbb{T})$ and $g \in \mathscr{C}^\delta(\mathbb{T})$ with $\gamma + \delta > 1$, then we can define $\int f \mathrm{d}g := \int (f \partial_t g)$, which is well defined since $\partial_t g \in \mathscr{C}^{\delta-1}$ and $\gamma + \delta - 1 > 0$, and since integration is a linear map. In this way we recover the Young integral [You36].

Example 5 Let B^H be a fractional Brownian bridge on $\mathbb{T}$ (or simply a fractional Brownian motion on $[0, \pi]$, reflected on $[\pi, 2\pi]$) and assume that $H > 1/2$. We have $\varphi(B^H) \in \mathscr{C}^{H-}$ for all Lipschitz continuous φ, and $\partial_t B^H \in \mathscr{C}^{(H-1)-}$, and in particular $\varphi(B^H)\partial_t B^H$ is well-defined. This can be used to solve SDEs driven by B^H in a pathwise sense.

The condition $\alpha + \beta > 0$ is essentially sharp, at least at this level of generality, see [You36] for counterexamples. It excludes of course the Brownian case: if B is a Brownian motion, then almost surely $B \in \mathscr{C}^\alpha_{\mathrm{loc}}$ for all $\alpha < 1/2$, so that $\partial_t B \in \mathscr{C}^{\alpha-1}_{\mathrm{loc}}$ and thus $B \circ \partial_t B$ fails to be well defined. See also [LCL07], Proposition 1.29 for an instructive example which shows that this is not a shortcoming of our description of regularity, but that it is indeed impossible to define the product $B\partial_t B$ as a continuous bilinear operation on distribution spaces.

Other counterexamples are given by our discussion of the homogenization problem above. More simply, one can consider the following situation.

Example 6 Consider the sequence of functions $f_n : \mathbb{T} \to \mathbb{C}$ given by $f_n(t) = e^{in^2 t}/n$. Then it is easy to show that $\|f_n\|_\gamma \to 0$ for all $\gamma < 1/2$. However let $g_n(t) = \mathrm{Re}\, f_n(t)\, \mathrm{Im}\, \partial_t f_n(t) = \cos(n^2 t)^2$. Then $g_n \to 1/2$ in $\mathscr{C}^{0-}$ which shows that the map $f \mapsto (\mathrm{Re}\, f)(\partial_t\, \mathrm{Im}\, f)$ cannot be continuous in $\mathscr{C}^\gamma$ if $\gamma < 1/2$. Pictorially the situation is summarized in Figure 4, where we sketched the three dimensional curve given by $t \mapsto \left(\mathrm{Re}\, f_n(t), \mathrm{Im}\, f_n(t), \int_0^t g_n(s)\mathrm{d}s\right)$ for various values of n and in the limit.

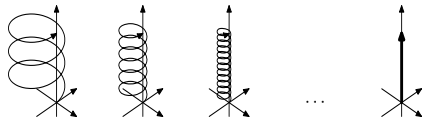

Figure 4: Resonances give macroscopic effects

5.2 Commutator estimates and paralinearization

The product $F(u)\eta$ appearing in the right hand side of PAM can be decomposed via the paraproduct $\prec$ as a sum of three terms

$$F(u)\eta = F(u) \prec \eta + F(u) \circ \eta + F(u) \succ \eta.$$

The first and the last of these terms are continuous in any topology we choose for $F(u)$ and η. The *resonant term* $F(u) \circ \eta$ however is problematic. It gathers the products of the oscillations of $F(u)$ and η on comparable dyadic scales and these products can contribute to all larger scales in such a way that microscopic oscillations might build up to

a macroscopic effect which does not disappear in the limit (as we already have seen). If the function F is smooth enough, then we expect the resonances between $F(u)$ and η to correspond to the resonances between u and η, and as we will see this is justified.

The expected regularity of the different terms is

$$\underbrace{F(u) \prec \eta}_{\gamma-2} + \underbrace{F(u) \circ \eta}_{2\gamma-2} + \underbrace{F(u) \succ \eta}_{2\gamma-2}, \tag{27}$$

but unless $2\gamma-2 > 0$ the resonant term $F(u) \circ \eta$ cannot be controlled using only the $C\mathscr{C}^\gamma$–norm of u and the $C\mathscr{C}^{\gamma-2}$–norm of η. If F is at least C^2, we can use a *paralinearization* result (stated precisely in Lemma 16 below) to rewrite this term as

$$F(u) \circ \eta = F'(u)(u \circ \eta) + \Pi_F(u, \eta), \tag{28}$$

with a remainder $\Pi_F(u, \eta) \in \mathscr{C}^{3\gamma-2}$ provided $3\gamma - 2 > 0$. The difficulty is now localized in the linearized resonant product $u \circ \eta$. In order to control this term, we would like to exploit the fact that the function u is not a generic element of $C\mathscr{C}^\gamma$ but that it has a specific structure, since $\mathscr{L}u$ has to match the paraproduct decomposition given in (27) where the least regular term is expected to be $F(u) \prec \eta \in C\mathscr{C}^{\gamma-2}$.

In order to do so, we postulate that the solution u is given by the following *paracontrolled ansatz:*

$$u = u^X \prec X + u^\sharp,$$

for functions $u^X, X, u^\sharp$ such that $u^X, X \in \mathscr{C}^\gamma$ and the remainder $u^\sharp \in C\mathscr{C}^{2\gamma}$. This decomposition allows for a finer analysis of the resonant term $u \circ \eta$: indeed, we have

$$u \circ \eta = (u^X \prec X) \circ \eta + u^\sharp \circ \eta = u^X(X \circ \eta) + C(u^X, X, \eta) + u^\sharp \circ \eta, \tag{29}$$

where the *commutator* is defined by $C(u^X, X, \eta) = (u^X \prec X) \circ \eta - u^X(X \circ \eta)$. Observe now that the term $u^\sharp \circ \eta$ does not pose any further problem, as it is bounded in $C\mathscr{C}^{3\gamma-2}$. The key point is now that the commutator is a bounded multilinear function of its arguments as long as the sum of their regularities is strictly positive, see Lemma 14 below. By assumption, we have $3\gamma - 2 > 0$, and therefore $C(u^X, X, \eta) \in C\mathscr{C}^{3\gamma-2}$.

The only problematic term which remains to be handled is thus the bilinear functional of the noise given by $X \circ \eta$. Here we need to make the assumption that $X \circ \eta \in C\mathscr{C}^{2\gamma-2}$ in order for the product $u^X(X \circ \eta)$ to be well defined. This assumption is not guaranteed by the analytical estimates at hand, and it has to be added as a further requirement to our construction.

Granting this last step, we have obtained that the right hand side of equation (23) is well defined and a continuous function of $(u, u^X, u^\sharp, X, \eta, X \circ \eta)$.

It remains to check that the paracontrolled ansatz is coherent with the equation satisfied by solutions to PAM. Let us first consider the linear example $F(u) = u$. Here we saw that the solution is of the form $u = e^X v$ with

$$\mathscr{L}v = v|\partial_x X|^2 + 2\langle \partial_x v, \partial_x X \rangle_{\mathbb{R}^2},$$

where $|\partial_x X|^2 \in C\mathscr{C}^{2\gamma-2}$ and $\partial_x X \in C\mathscr{C}^{\gamma-1}$ and therefore $v \in C\mathscr{C}^{2\gamma}$. In particular

$$u = e^X v = v \prec e^X + C\mathscr{C}^{2\gamma} = v \prec (e^X \prec X) + C\mathscr{C}^{2\gamma},$$

using a paralinearization result in last step (see Lemma 15 below). Now the double paraproduct $f \prec (g \prec h)$ satisfies

$$\|f \prec (g \prec h) - (fg) \prec h\|_{\alpha+\beta} \lesssim \|f\|_\alpha \|g\|_\alpha \|h\|_\beta,$$

see [Bon81], and therefore $u = (ve^X) \prec X + C\mathscr{C}^{2\gamma} = u \prec X + C\mathscr{C}^{2\gamma}$ which shows that the paracontrolled ansatz is at least justified in the linear case.

In the nonlinear case, the paracontrolled ansatz and the Leibniz rule for the paraproduct imply that (23) can be rewritten as

$$\mathscr{L}u = \mathscr{L}(u^X \prec X + u^\sharp) = u^X \prec \mathscr{L}X + \left[\mathscr{L}, u^X \prec\right] X + \mathscr{L}u^\sharp = F(u) \prec \eta + F(u) \circ \eta + F(u) \succ \eta,$$

where we recall that $\left[\mathscr{L}, u^X \prec\right] X = \mathscr{L}(u^X \prec X) - u^X \prec \mathscr{L}X$ denotes the commutator. If we choose X such that $\mathscr{L}X = \eta$ and we set $u^X = F(u)$, then we can use (28) and (29) to obtain the following equation for the remainder $u^\sharp$:

$$\begin{aligned} \mathscr{L}u^\sharp \quad &= F'(u)F(u)(X \circ \eta) + F(u) \succ \eta - [\mathscr{L}, F(u) \prec] X \\ &\quad + F'(u)C(F(u), X, \eta) + F'(u)(u^\sharp \circ \eta) + \Pi_F(u, \eta). \end{aligned} \tag{30}$$

Lemma 18 below ensures that $\mathscr{L}^{-1}\left[\mathscr{L}, F(u) \prec\right] X \in C\mathscr{C}^{2\gamma}$ whenever $u \in \mathscr{L}^\gamma$, and we have already seen that all the other terms on the right hand side are in $C\mathscr{C}^{2\gamma-2}$ so that this allows us to control $u^\sharp$ in $C\mathscr{C}^{2\gamma}$. Together with $u = F(u) \prec X + u^\sharp$, equation (30) gives an equivalent description of the solution and allows us to obtain a priori estimates on u and $u^\sharp$ in terms of $(u_0, \|\eta\|_{\gamma-2}, \|X \circ \eta\|_{2\gamma-2})$. It is now easy to show that if $F \in C_b^3$, then u depends continuously on the data $(u_0, \eta, X \circ \eta)$, so that we have a robust strategy to pass to the limit in (17) and to make sense of the solution to (23) also for irregular $\eta \in C\mathscr{C}^{\gamma-2}$ as long as $\gamma > 2/3$.

In the remainder of this section we will prove the results (paralinearization and various key commutators) which we used in the discussion above, before going on to gather the consequences of our analysis in the next section. When the time dependence does not play any role we state the results for distributions depending only on the space variable as the extension to time varying functions will not add further difficulty.

Lemma 14 *Assume that $\alpha, \beta, \gamma \in \mathbb{R}$ are such that $\alpha + \beta + \gamma > 0$ and $\beta + \gamma \neq 0$. Then for $f, g, h \in C^\infty$ the trilinear operator*

$$C(f, g, h) = ((f \prec g) \circ h) - f(g \circ h)$$

allows for the bound

$$\|C(f, g, h)\|_{\beta+\gamma} \lesssim \|f\|_\alpha \|g\|_\beta \|h\|_\gamma, \tag{31}$$

and can thus be uniquely extended to a bounded trilinear operator from $\mathscr{C}^\alpha \times \mathscr{C}^\beta \times \mathscr{C}^\alpha$ to $\mathscr{C}^{\beta+\gamma}$.

Proof For $\beta+\gamma>0$ this follows from the paraproduct estimates, so let $\beta+\gamma<0$. By definition

$$\begin{aligned} C(f,g,h) &= \sum_{i,j,k,\ell} \Delta_i(\Delta_j f \Delta_k g)\Delta_\ell h(\mathbb{I}_{j<k-1}\mathbb{I}_{|i-\ell|\leqslant 1} - \mathbb{I}_{|k-\ell|\leqslant 1}) \\ &= \sum_{i,j,k,\ell} \Delta_i(\Delta_j f \Delta_k g)\Delta_\ell h(\mathbb{I}_{j<k-1}\mathbb{I}_{|i-\ell|\leqslant 1}\mathbb{I}_{|k-\ell|\leqslant N} - \mathbb{I}_{|k-\ell|\leqslant 1}), \end{aligned}$$

where we used that $S_{k-1}f\Delta_k g$ has support in an annulus $2^k\mathscr{A}$, so that $\Delta_i(S_{k-1}f\Delta_k g)\neq 0$ only if $|i-k|\leqslant N-1$ for some fixed $N\in\mathbb{N}$, which in combination with $|i-\ell|\leqslant 1$ yields $|k-\ell|\leqslant N$. Now for fixed k, the term $\sum_\ell \mathbb{I}_{2\leqslant|k-\ell|\leqslant N}\Delta_k g\Delta_\ell h$ is spectrally supported in an annulus $2^k\mathscr{A}$, so that $\sum_{k,\ell}\mathbb{I}_{2\leqslant|k-\ell|\leqslant N}\Delta_k g\Delta_\ell h\in\mathscr{C}^{\beta+\gamma}$ and we may add and subtract $f\sum_{k,\ell}\mathbb{I}_{2\leqslant|k-\ell|\leqslant N}\Delta_k g\Delta_\ell h$ to $C(f,g,h)$ while maintaining the bound (31). It remains to treat

$$\begin{aligned} &\sum_{i,j,k,\ell} \Delta_i(\Delta_j f\Delta_k g)\Delta_\ell h\mathbb{I}_{|k-\ell|\leqslant N}(\mathbb{I}_{j<k-1}\mathbb{I}_{|i-\ell|\leqslant 1}-1) \\ &\qquad = -\sum_{i,j,k,\ell}\Delta_i(\Delta_j f\Delta_k g)\Delta_\ell h\mathbb{I}_{|k-\ell|\leqslant N}(\mathbb{I}_{j\geqslant k-1}+\mathbb{I}_{j<k-1}\mathbb{I}_{|i-\ell|>1}). \end{aligned} \tag{32}$$

We estimate both terms on the right hand side separately. For $m\geqslant -1$ we have

$$\begin{aligned} &\Big\|\Delta_m\Big(\sum_{i,j,k,\ell}\Delta_i(\Delta_j f\Delta_k g)\Delta_\ell h\mathbb{I}_{|k-\ell|\leqslant N}\mathbb{I}_{j\geqslant k-1}\Big)\Big\|_{L^\infty} \\ &\quad\leqslant \sum_{j,k,\ell}\mathbb{I}_{|k-\ell|\leqslant N}\mathbb{I}_{j\geqslant k-1}\|\Delta_m(\Delta_j f\Delta_k g\Delta_\ell h)\|_{L^\infty} \lesssim \sum_{j\gtrsim m}\sum_{k\lesssim j}2^{-j\alpha}\|f\|_\alpha 2^{-k\beta}\|g\|_\beta 2^{-k\gamma}\|h\|_\gamma \\ &\quad\lesssim \sum_{j\gtrsim m}2^{-j(\alpha+\beta+\gamma)}\|f\|_\alpha\|g\|_\beta\|h\|_\gamma \lesssim 2^{-m(\alpha+\beta+\gamma)}\|f\|_\alpha\|g\|_\beta\|h\|_\gamma, \end{aligned}$$

using $\beta+\gamma<0$. It remains to estimate the second term in (32). For $|i-\ell|>1$ and $i\sim k\sim\ell$, any term of the form $\Delta_i()\Delta_\ell()$ is spectrally supported in an annulus $2^\ell\mathscr{A}$, and therefore

$$\begin{aligned} &\Big\|\Delta_m\Big(\sum_{i,j,k,\ell}\Delta_i(\Delta_j f\Delta_k g)\Delta_\ell h\mathbb{I}_{|k-\ell|\leqslant N}\mathbb{I}_{j<k-1}\mathbb{I}_{|i-\ell|>1}\Big)\Big\|_{L^\infty} \\ &\qquad\lesssim \sum_{i,j,k,\ell}\mathbb{I}_{j<k-1}\mathbb{I}_{i\sim k\sim\ell\sim m}\|\Delta_i(\Delta_j f\Delta_k g)\Delta_\ell h\|_{L^\infty} \\ &\qquad\lesssim \sum_{j\lesssim m}2^{-j\alpha}\|f\|_\alpha 2^{-m\beta}\|g\|_\beta 2^{-m\gamma}\|h\|_\gamma \lesssim 2^{-m(\beta+\gamma)}\|f\|_\alpha\|g\|_\beta\|h\|_\gamma. \end{aligned}$$

□

Remark 4 For $\beta+\gamma=0$ we can apply the commutator estimate with $\gamma'<\gamma$, as long as $\alpha+\beta+\gamma'>0$.

Our next result is a simple paralinearization lemma for non–linear operators.

Lemma 15 (see also [BCD11], Theorem 2.92) *Let $\alpha \in (0,1)$, $\beta \in (0,\alpha]$, and let $F \in C_b^{1+\beta/\alpha}$. There exists a locally bounded map $R_F : \mathscr{C}^\alpha \to \mathscr{C}^{\alpha+\beta}$ such that*

$$F(f) = F'(f) \prec f + R_F(f) \tag{33}$$

for all $f \in \mathscr{C}^\alpha$. More precisely, we have

$$\|R_F(f)\|_{\alpha+\beta} \lesssim \|F\|_{C_b^{1+\beta/\alpha}}(1 + \|f\|_\alpha^{1+\beta/\alpha}).$$

If $F \in C_b^{2+\beta/\alpha}$, then R_F is locally Lipschitz continuous:

$$\|R_F(f) - R_F(g)\|_{\alpha+\beta} \lesssim \|F\|_{C_b^{2+\beta/\alpha}}(1 + \|f\|_\alpha + \|g\|_\alpha)^{1+\beta/\alpha}\|f - g\|_\alpha.$$

Proof The difference $F(f) - F'(f) \prec f$ is given by

$$R_F(f) = F(f) - F'(f) \prec f = \sum_{i \geqslant -1} [\Delta_i F(f) - S_{i-1}F'(f)\Delta_i f] = \sum_{i \geqslant -1} u_i,$$

and every u_i is spectrally supported in a ball $2^i\mathscr{B}$. For $i < 1$, we simply estimate $\|u_i\|_{L^\infty} \lesssim \|F\|_{C_b^1}(1 + \|f\|_\alpha)$. For $i \geqslant 1$ we use the fact that f is a bounded function to write the Littlewood–Paley projections as convolutions and obtain

$$\begin{aligned} u_i(x) &= \int K_i(x-y)K_{<i-1}(x-z)[F(f(y)) - F'(f(z))f(y)]\mathrm{d}y\mathrm{d}z \\ &= \int K_i(x-y)K_{<i-1}(x-z)[F(f(y)) - F(f(z)) - F'(f(z))(f(y) - f(z))]\mathrm{d}y\mathrm{d}z, \end{aligned}$$

where $K_i = \mathscr{F}^{-1}\rho_i$, $K_{<i-1} = \sum_{j<i-1} K_j$, and where we used that $\int K_i(y)\mathrm{d}y = \rho_i(0) = 0$ for $i \geqslant 0$ and $\int K_{<i-1}(z)\mathrm{d}z = 1$ for $i \geqslant 1$. Now we can apply a first order Taylor expansion to F and use the β/α–Hölder continuity of F' in combination with the α–Hölder continuity of f, to deduce

$$\begin{aligned} |u_i(x)| &\lesssim \|F\|_{C_b^{1+\beta/\alpha}}\|f\|_\alpha^{1+\beta/\alpha} \int |K_i(x-y)K_{<i-1}(x-z)| \times |z-y|^{\alpha+\beta}\mathrm{d}y\mathrm{d}z \\ &\lesssim \|F\|_{C_b^{1+\beta/\alpha}}\|f\|_\alpha^{1+\beta/\alpha} 2^{-i(\alpha+\beta)}. \end{aligned}$$

Therefore, the estimate for $R_F(f)$ follows from Lemma 10. The estimate for $R_F(f) - R_F(g)$ is shown in the same way. □

Let g be a distribution belonging to $\mathscr{C}^\beta$ for some $\beta < 0$. Then the map $f \mapsto f \circ g$ behaves, modulo smoother correction terms, like a derivative operator:

Lemma 16 *Let $\alpha \in (0,1)$, $\beta \in (0,\alpha]$, $\gamma \in \mathbb{R}$ be such that $\alpha+\beta+\gamma > 0$ and $\alpha+\gamma \neq 0$. Let $F \in C_b^{1+\beta/\alpha}$. Then there exists a locally bounded map $\Pi_F : \mathscr{C}^\alpha \times \mathscr{C}^\gamma \to \mathscr{C}^{\alpha+\gamma}$ such that*

$$F(f)\circ g = F'(f)(f\circ g) + \Pi_F(f,g) \tag{34}$$

for all $f \in \mathscr{C}^\alpha$ and all smooth g. More precisely, we have

$$\|\Pi_F(f,g)\|_{\alpha+\gamma} \lesssim \|F\|_{C_b^{1+\beta/\alpha}}(1+\|f\|_\alpha^{1+\beta/\alpha})\|g\|_\gamma.$$

If $F \in C_b^{2+\beta/\alpha}$, then Π_F is locally Lipschitz continuous:

$$\begin{aligned}\|\Pi_F(f,g)-\Pi_F(u,v)\|_{\alpha+\gamma}&\\ \lesssim \|F\|_{C_b^{2+\beta/\alpha}}&(1+\|f\|_\alpha+\|u\|_\alpha)^{1+\beta/\alpha}(1+\|v\|_\gamma)(\|f-u\|_\alpha+\|g-v\|_\gamma).\end{aligned}$$

Proof Use the paralinearization and commutator lemmas above to deduce that

$$\begin{aligned}\Pi(f,g) &= F(f)\circ g - F'(f)(f\circ g) = R_F(f)\circ g + (F'(f)\prec f)\circ g - F'(f)(f\circ g)\\ &= R_F(f)\circ g + C(F'(f),f,g),\end{aligned}$$

so that the claimed bounds easily follow from Lemma 14 and Lemma 15. □

Besides this sort of chain rule, we also have a Leibniz rule for $f \mapsto f\circ g$:

Lemma 17 *Let $\alpha \in (0,1)$ and $\gamma < 0$ be such that $2\alpha+\gamma > 0$ and $\alpha+\gamma \neq 0$. Then there exists a bounded trilinear operator $\Pi_\times : \mathscr{C}^\alpha \times \mathscr{C}^\alpha \times \mathscr{C}^\gamma \to \mathscr{C}^{\alpha+\gamma}$, such that*

$$(fu)\circ g = f(u\circ g) + u(f\circ g) + \Pi_\times(f,u,g)$$

for all $f,u \in \mathscr{C}^\alpha(\mathbb{R})$ and all smooth g.

Proof It suffices to note that $fu = f\prec u + f\succ u + f\circ u$, which leads to

$$\Pi_\times(f,u,g) = (fu)\circ g - f(u\circ g) + u(f\circ g) = C(f,u,g) + C(u,f,g) + (f\circ u)\circ g.$$

□

Lemma 18 *Let $\beta < 1$, $\alpha \in \mathbb{R}$, and let $f \in \mathscr{L}^\beta$ and $G \in C\mathscr{C}^\alpha$ with $\mathscr{L}G \in C\mathscr{C}^{\alpha-2}$. There exists $H = H(f,G)$ such that $\mathscr{L}H = \big[\mathscr{L}, f\prec\big]G$ and $H(0)=0$. Moreover $H \in C\mathscr{C}^{\alpha+\beta} \cap C^{(\alpha\wedge\beta)/2}L^\infty$ and for all $T>0$*

$$\|H\|_{C_T^{(\alpha\wedge\beta)/2}L^\infty} + \|H\|_{C_T\mathscr{C}^{\alpha+\beta}} \lesssim \|f\|_{\mathscr{L}_T^\beta}\left(\|G\|_{C_T\mathscr{C}^\alpha} + \|\mathscr{L}G\|_{C_T\mathscr{C}^{\alpha-2}}\right).$$

Proof Let $T > 0$ and let f_ε be a time mollification of f such that $\|\partial_t f_\varepsilon\|_{C_T L^\infty} \lesssim \varepsilon^{\beta/2-1}\|f\|_{\mathscr{L}^\beta_T}$ and $\|f_\varepsilon - f\|_{C_T L^\infty} \lesssim \varepsilon^{\beta/2}\|f\|_{\mathscr{L}^\beta}$ for all $\varepsilon > 0$. For example we can take $f_\varepsilon = \rho_\varepsilon * f$ with $\rho_\varepsilon(t) = \rho(t/\varepsilon)/\varepsilon$ and $\rho : \mathbb{R} \to \mathbb{R}$ compactly supported, smooth, and of unit integral. For $i \geqslant -1$ we have

$$\mathscr{L}\Delta_i H = \Delta_i\left[\mathscr{L}((f - f_\varepsilon) \prec G) - (f - f_\varepsilon) \prec \mathscr{L}G\right] + \Delta_i\left[\mathscr{L}(f_\varepsilon \prec G) - f_\varepsilon \prec \mathscr{L}G\right],$$

so that

$$\begin{aligned}\mathscr{L}\Delta_i(H - (f - f_\varepsilon) \prec G) &= -\Delta_i\left[(f - f_\varepsilon) \prec \mathscr{L}G\right] + \Delta_i\left[\mathscr{L}(f_\varepsilon \prec G) - f_\varepsilon \prec \mathscr{L}G\right] \\ &= \Delta_i\left[(f_\varepsilon - f) \prec \mathscr{L}G\right] + \Delta_i\left[\mathscr{L}f_\varepsilon \prec G - 2\partial_x f_\varepsilon \prec \partial_x G\right],\end{aligned}$$

with initial condition $\Delta_i(H - (f - f_\varepsilon) \prec G)(0) = -(\Delta_i(f - f_\varepsilon) \prec G)(0)$. The Schauder estimates for $\mathscr{L}$ (Lemma 11) give

$$\begin{aligned}&\|\Delta_i(H + (f - f_\varepsilon) \prec G)\|_{\mathscr{L}^{\alpha+\beta}_T} \\ &\qquad \lesssim \left\|\Delta_i\left[(f - f_\varepsilon) \prec \mathscr{L}G\right] + \Delta_i\left[\left(\mathscr{L}f_\varepsilon\right) \prec G + \partial_x f_\varepsilon \prec \partial_x G\right]\right\|_{C_T\mathscr{C}^{\alpha+\beta-2}} \\ &\qquad\quad + \|(\Delta_i(f - f_\varepsilon) \prec G)(0)\|_{\alpha+\beta}.\end{aligned}$$

Choosing $\varepsilon = 2^{-2i}$, we have

$$\begin{aligned}\|\Delta_i((f - f_\varepsilon) \prec G)\|_{C_T\mathscr{C}^{\alpha+\beta}} &\lesssim 2^{\beta i}\|\Delta_i((f - f_\varepsilon) \prec G)\|_{C_T\mathscr{C}^\alpha} \lesssim 2^{\beta i}\|f - f_\varepsilon\|_{C_T L^\infty}\|G\|_{C_T\mathscr{C}^\alpha} \\ &\lesssim \|f\|_{\mathscr{L}^\beta_T}\|G\|_{C_T\mathscr{C}^\alpha}\end{aligned}$$

and exactly the same argument also gives

$$\left\|\Delta_i\left[(f - f_\varepsilon) \prec \mathscr{L}G\right]\right\|_{C_T\mathscr{C}^{\alpha+\beta-2}} \lesssim \|f\|_{\mathscr{L}^\beta_T}\|\mathscr{L}G\|_{C_T\mathscr{C}^{\alpha-2}}.$$

Since $\beta < 1$, we further get

$$\begin{aligned}\left\|\Delta_i\left[\mathscr{L}f_\varepsilon \prec G + \partial_x f_\varepsilon \prec \partial_x G\right]\right\|_{C_T\mathscr{C}^{\alpha+\beta-2}} &\lesssim 2^{i(\beta-2)}\|\partial_t f_\varepsilon\|_{C_T L^\infty}\|G\|_{C_T\mathscr{C}^\alpha} + \|f_\varepsilon\|_{C_T\mathscr{C}^\beta}\|G\|_{C_T\mathscr{C}^\alpha} \\ &\lesssim \|f\|_{\mathscr{L}^\beta_T}\|G\|_{C_T\mathscr{C}^\alpha} + \|f\|_{C_T\mathscr{C}^\beta}\|G\|_{C_T\mathscr{C}^\alpha}.\end{aligned}$$

Combining everything, we end up with

$$\|\Delta_i H\|_{C_T\mathscr{C}^{\alpha+\beta}} \lesssim \|f\|_{\mathscr{L}^\beta_T}\left(\|G\|_{C_T\mathscr{C}^\alpha} + \|\mathscr{L}G\|_{C_T\mathscr{C}^{\alpha-2}}\right),$$

which gives the estimate for the space regularity of H since $\|\Delta_i H\|_{C_T L^\infty} \lesssim 2^{-(\alpha+\beta)i}\|\Delta_i H\|_{C_T\mathscr{C}^{\alpha+\beta}}$. The time regularity of H can be controlled similarly by noting that $(f - f_\varepsilon) \prec G \in C_T^{(\alpha\wedge\beta)/2}L^\infty$, uniformly in ε. □

5.3 Paracontrolled distributions

Here we build a calculus of distributions satisfying a paracontrolled ansatz. We start by defining a suitable space of such objects.

Definition 5 *Let $\alpha > 0$ and $\beta \in (0, \alpha]$ be such that $\alpha + \beta \in (0, 2)$, and let $u \in \mathscr{L}^\alpha$. A pair of distributions $(f, f^u) \in \mathscr{L}^\alpha \times \mathscr{L}^\beta$ is called paracontrolled by g if $f^\sharp = f - f^u \prec u \in C\mathscr{C}^{\alpha+\beta} \cap \mathscr{L}^\beta$. In that case we write $f \in \mathscr{D}^\beta = \mathscr{D}^\beta(u)$, and for all $T > 0$ we define the norm*

$$\|f\|_{\mathscr{D}^\beta_T} = \|f\|_{C^{\alpha/2}_T} + \|f^u\|_{\mathscr{L}^\beta_T} + \|f^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta}} + \|f^\sharp\|_{C^{\beta/2}_T L^\infty}.$$

If $\tilde{u} \in \mathscr{L}^\alpha$ and $(\tilde{f}, \tilde{f}^{\tilde{u}}) \in \mathscr{D}^\beta(\tilde{u})$, then we also write

$$d_{\mathscr{D}^\beta_T}(f, \tilde{f}) = \|f^u - \tilde{f}^{\tilde{u}}\|_{\mathscr{L}^\beta_T} + \|f^\sharp - \tilde{f}^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta}} + \|f^\sharp - \tilde{f}^\sharp\|_{C^{\beta/2}_T L^\infty}.$$

Of course we should really write $(f, f^u) \in \mathscr{D}^\beta$ since given f and g, the derivative f^u is usually not uniquely determined. But in the applications there will always be an obvious candidate for the derivative, and no confusion will arise.

Remark 5 The space $\mathscr{D}^\beta$ does not depend on the specific dyadic partition of unity. Indeed, Bony [Bon81] has shown that if $\tilde{\prec}$ is the paraproduct constructed from another partition of unity, then $\|f^u \prec u - f^u \tilde{\prec} u\|_{C_T\mathscr{C}^{\alpha+\beta}} \lesssim \|f^u\|_{C_T\mathscr{C}^\beta}\|u\|_{C_T\mathscr{C}^\alpha}$.

Nonlinear operations As an immediate consequence of Lemma 14 we can multiply paracontrolled distributions provided that we know how to multiply the reference distributions.

Theorem 4 *Let $\alpha, \beta \in \mathbb{R}$, $\gamma < 0$, with $\alpha + \beta + \gamma > 0$ and $\alpha + \gamma \neq 0$. Let $u \in C\mathscr{C}^\alpha$, $v \in C\mathscr{C}^\gamma$, and let $\eta \in C\mathscr{C}^{\alpha+\gamma}$. Then*

$$\mathscr{D}^\beta(u) \ni f \mapsto f \cdot v := f \prec v + f \succ v + f^\sharp \circ v + C(f^u, u, v) + f^u \eta \in C\mathscr{C}^\gamma$$

defines a linear operator which for all $T > 0$ admits the bound

$$\|(fv)^\sharp\|_{C_T\mathscr{C}^{\alpha+\gamma}} := \|f \cdot v - f \prec v\|_{C_T\mathscr{C}^{\alpha+\gamma}} \lesssim \|f\|_{\mathscr{D}^\beta_T}\left(\|v\|_{C_T\mathscr{C}^\gamma} + \|u\|_{C_T\mathscr{C}^\alpha}\|v\|_{C_T\mathscr{C}^\gamma} + \|\eta\|_{C_T\mathscr{C}^{\alpha+\gamma}}\right).$$

If there exist sequences of smooth functions (u_n) and (v_n) converging to u and v respectively for which $(u_n \circ v_n)$ converges to η, then $f \cdot v$ does not depend on the dyadic partition of unity used to construct it.

Furthermore, there exists a quadratic polynomial P so that if $\tilde{u}, \tilde{v}, \tilde{\eta}$ satisfy the same assumptions as u, v, η respectively, if $\tilde{f} \in \mathscr{D}^\beta(\tilde{u})$, and if

$$M = \max\left\{\|u\|_{C_T\mathscr{C}^\alpha}, \|v\|_{C_T\mathscr{C}^\gamma}, \|\eta\|_{C_T\mathscr{C}^{\alpha+\gamma}}, \|\tilde{u}\|_{C_T\mathscr{C}^\alpha}, \|\tilde{v}\|_{C_T\mathscr{C}^\gamma}, \|\tilde{\eta}\|_{C_T\mathscr{C}^{\alpha+\gamma}}, \|f\|_{\mathscr{D}^\beta_T(u)}, \|\tilde{f}\|_{\mathscr{D}^\beta_T(\tilde{u})}\right\},$$

then

$$\|(fv)^\sharp - (\tilde{f}\tilde{v})^\sharp\|_{\alpha+\gamma} \leqslant P(M)\left(d_{\mathscr{D}^\beta}(f, \tilde{f}) + \|u - \tilde{u}\|_\alpha + \|v - \tilde{v}\|_\gamma + \|\eta - \tilde{\eta}\|_{\alpha+\gamma}\right).$$

Given Lemma 14 (and the paraproduct estimates Theorem 3), the proof is straightforward and we leave it as an exercise. From now on we will usually write fv rather than $f \cdot v$.

To solve equations involving general nonlinear functions, we need to examine the stability of paracontrolled distributions under smooth functions.

Theorem 5 *Let $\alpha \in (0,1)$ and $\beta \in (0,\alpha]$. Let $u \in \mathscr{L}^\alpha$, $f \in \mathscr{D}^\alpha(u)$, and $F \in C_b^{1+\beta/\alpha}$. Then $F(f) \in \mathscr{D}^\beta$ with derivative $(F(f))^u = F'(f)f^u$, and for all $T > 0$*

$$\|F(f)\|_{\mathscr{D}_T^\beta} \lesssim \|F\|_{C_b^{1+\beta/\alpha}}(1 + \|f\|_{\mathscr{D}_T^\alpha}^2)(1 + \|u\|_{\mathscr{L}_T^\alpha}^2).$$

Moreover, there exists a polynomial P which satisfies for all $F \in C_b^{2+\beta/\alpha}$, $\tilde{u} \in \mathscr{L}^\alpha$, $\tilde{f} \in \mathscr{D}^\alpha(\tilde{u})$, and

$$M := \max\left\{\|u\|_{\mathscr{L}_T^\alpha}, \|\tilde{u}\|_{\mathscr{L}_T^\alpha}, \|f\|_{\mathscr{D}_T^\alpha(u)}, \|\tilde{f}\|_{\mathscr{D}_T^\alpha(\tilde{u})}\right\}$$

the bound

$$d_{\mathscr{D}_T^\beta}(F(f), F(\tilde{f})) \leqslant P(M)\|F\|_{C_T^{2+\beta/\alpha}}(d_{\mathscr{D}_T^\alpha}(f,\tilde{f}) + \|u - \tilde{u}\|_{\mathscr{L}_T^\alpha}).$$

The proof is not very complicated but rather lengthy, and we do not present it here.

Schauder estimate for paracontrolled distributions The Schauder estimate Lemma 11 is not quite sufficient: we also need to understand how the heat kernel acts on the paracontrolled structure.

Theorem 6 *Let $\alpha \in (0,1)$ and $\beta \in (0,\alpha]$. Let $u \in C\mathscr{C}^{\alpha-2}$ and $\mathscr{L}U = u$ with $U(0) = 0$. Let $f^u \in \mathscr{L}^\beta$, $f^\sharp \in C\mathscr{C}^{\alpha+\beta-2}$, and $g_0 \in \mathscr{C}^{\alpha+\beta}$. Then $(g, f^u) \in \mathscr{D}^\beta(U)$, where g solves*

$$\mathscr{L}g = f^u \prec u + f^\sharp, \qquad g(0) = g_0,$$

and we have the bound

$$\|g\|_{\mathscr{D}_T^\beta(U)} \lesssim \|g_0\|_{\alpha+\beta} + (1+T)(\|f^u\|_{\mathscr{L}_T^\beta}(1 + \|u\|_{C_T\mathscr{C}^{\alpha-2}}) + \|f^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta-2}})$$

for all $T > 0$. If furthermore $\tilde{u}, \tilde{U}, \tilde{f}^u, \tilde{f}^\sharp, \tilde{g}_0, \tilde{g}$ satisfy the same assumptions as $u, U, f^u, f^\sharp, g_0, g$ respectively, and if $M = \max\{\|f^u\|_{\mathscr{L}_T^\beta}, \|\tilde{u}\|_{C_T\mathscr{C}^{\alpha-2}}, 1\}$, then

$$d_{\mathscr{D}_T^\beta}(g,\tilde{g}) \lesssim \quad \|g_0 - \tilde{g}_0\|_{\alpha+\beta} + (1+T)M(\|f^u - \tilde{f}^u\|_{\mathscr{L}_T^\beta} + \|u - \tilde{u}\|_{C_T\mathscr{C}^{\alpha-2}} + \|f^\sharp - \tilde{f}^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta-2}}).$$

Proof Let us derive an equation for the remainder $g^\sharp$. We have

$$\begin{aligned}\mathscr{L}g^\sharp &= \mathscr{L}(f' \prec U) - \mathscr{L}g = \left[\mathscr{L}(f^u \prec U) - f^u \prec \mathscr{L}U\right] + f^u \prec \mathscr{L}U - [f^u \prec u + f^\sharp] \\ &= \left[\mathscr{L}, f^u \prec\right] U - f^\sharp.\end{aligned}$$

By Lemma 18 we know that there exists $H \in C\mathscr{C}^{\alpha+\beta} \cap C^{\beta/2}L^\infty$ such that $\mathscr{L}H = \left[\mathscr{L}, f^u \prec\right] U$, so we can apply the standard Schauder estimates of Lemma 11 to $\mathscr{L}(g^\sharp - H) = -f^\sharp$ to get

$$\|g^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta}} + \|g^\sharp\|_{C_T^{(\alpha\wedge\beta)/2}L^\infty} \lesssim \|f^u\|_{\mathscr{L}_T^\beta}(\|U\|_{C_T\mathscr{C}^\alpha} + \|\mathscr{L}U\|_{C_T\mathscr{C}^{\alpha-2}}) + \|f^\sharp\|_{C_T\mathscr{C}^{\alpha+\beta-2}}.$$

The estimate for $g^\sharp - \tilde{g}^\sharp$ can be derived in the same way. □

Bibliographic notes Paraproducts were introduced in [Bon81], for a nice introduction see [BCD11]. The commutator estimate Lemma 14 is from [GIP13], but the proof here is new and the statement is slightly different. In [GIP13], we require the additional assumption $\alpha \in (0,1)$ under which C maps $\mathscr{C}^\alpha \times \mathscr{C}^\beta \times \mathscr{C}^\gamma$ to $\mathscr{C}^{\alpha+\beta+\gamma}$ and not only to $\mathscr{C}^{\beta+\gamma}$. Theorem 5 is from [GIP13].

Theorem 6 is new, but it is implicitly used in [GIP13]. The estimates presented here will only allow us to consider regular initial conditions. More general situations can be covered by working on "explosive spaces" of the type

$$\left\{f \in C((0,\infty), \mathscr{C}^\alpha) : \sup_{t\in(0,T]} \|t^{-\gamma} f(t)\|_{\mathscr{C}^\alpha} < \infty \text{ for all } T > 0\right\}$$

and similar for the temporal regularity. This is also done in [GIP13].

Of course it is easily possible to replace the Laplacian by more general pseudo-differential operators. We only used two properties of Δ: the fact that $\Delta(f' \prec U) - f' \prec (\Delta U)$ is smooth, and that the semigroup generated by Δ has a sufficiently strong regularization effect. This is also true for fractional Laplacians and more generally for a wide range of pseudo-differential operators.

5.4 Fixpoint

Let us now give the details for the solution to PAM in the space of paracontrolled distributions. Assume that $F : \mathbb{R} \to \mathbb{R}$ is in $C_b^{1+\varepsilon}$ for some $\varepsilon > 0$ such that $(2+\varepsilon)\gamma > 2$.

We know from Theorem 5 that if $Y \in C\mathscr{C}^\gamma$ and $u \in \mathscr{D}^\gamma(Y)$, then $F(u) \in \mathscr{D}^\gamma(Y)$:

$$\mathscr{D}^\gamma(Y) \xrightarrow{u\mapsto F(u)} \mathscr{D}^{\varepsilon\gamma}(Y). \tag{35}$$

If now $Y \circ \eta \in C\mathscr{C}^{2\gamma-2}$ is given, then Theorem 4 shows that for all $f \in \mathscr{D}^{\varepsilon\gamma}$ we have $f\eta = (f\eta)^\sharp + F(u) \prec \eta$ with $(F(u)\eta)^\sharp \in C\mathscr{C}^{2\gamma-2}$ – it is here that we use $(2+\varepsilon)\gamma > 0$. Integrating against the heat kernel and assuming that $u_0 \in \mathscr{C}^{2\gamma}$, we obtain from Theorem 6 that the solution $(J(f\eta)(t) + P_t u_0)_{t\geqslant 0}$ to $\mathscr{L}J(f\eta) + P_\cdot u_0 = (f\eta)$, $J(f\eta)(0) + P_0 u_0 = u_0$, is in $\mathscr{D}^\gamma(X)$, where X solves $\mathscr{L}X = \eta$ and $X(0) = 0$. In other words, we have a map

$$\mathscr{D}^{\varepsilon\gamma}(Y) \xrightarrow{f\mapsto u_0 + J(f\eta)} \mathscr{D}^\gamma(X), \tag{36}$$

and combining (35), (36) we get

$$\mathscr{D}^\gamma(Y) \xrightarrow{u\mapsto F(u)} \mathscr{D}^{\varepsilon\gamma}(Y) \xrightarrow{F(u)\mapsto u_0 + J(F(u)\eta)} \mathscr{D}^\gamma(X),$$

so that for all $T > 0$ we can define

$$\Gamma_T : \mathscr{D}_T^\gamma(Y) \to \mathscr{D}_T^{\varepsilon\gamma}(X), \qquad \Gamma_T(u) = (u_0 + J(F(u)\eta))|_{[0,T]}.$$

To set up a Picard iteration domain and image space should coincide which means we should take $Y = X$. Refining the analysis, we obtain a scaling factor T^δ when estimating the $\mathscr{D}_T^\gamma(X)$–norm of $\Gamma_T(u)$. This allows us to show that for small $T > 0$, the map Γ_T leaves suitable balls in $\mathscr{D}_T^\gamma(X)$ invariant, and therefore we obtain the (local in time) *existence* of solutions to the equation under the assumption $X \circ \eta \in C\mathscr{C}^{2\gamma-2}$.

To obtain *uniqueness* we need to suppose that $F \in C_b^{2+\varepsilon}$. In that case Theorem 5 gives the local Lipschitz continuity of the map $u \mapsto F(u)$ from $\mathscr{D}_T^\gamma(X)$ to $\mathscr{D}_T^{\varepsilon\gamma}(X)$, while Theorem 4 and Theorem 6 show that $f \mapsto u_0 + J(f\eta)$ defines a Lipschitz continuous map from $\mathscr{D}_T^{\varepsilon\gamma}(X)$ to $\mathscr{D}_T^\gamma(X)$. Again we can obtain a scaling factor T^δ, so that Γ_T defines a contraction on a suitable ball of $\mathscr{D}_T^\gamma(X)$ for some small $T > 0$.

Even better, Γ_T not only depends locally Lipschitz continuously on u, but also on the extended data $(u_0, \eta, X \circ \eta)$, and therefore the solution to (23) depends locally Lipschitz continuously on $(u_0, \eta, X \circ \eta)$.

5.5 Renormalization

So far we argued under the assumption that $X \circ \eta$ exists and has a sufficient regularity. This should be understood via approximations as the existence of a sequence of smooth functions (η_n) that converges to η, such that $(X_n \circ \eta_n)$ converges to $X \circ \eta$. However, this hypothesis is questionable and, recalling our homogenization setting, actually not satisfied at all in the problem we are interested in. More concretely this can be checked by considering an approximation sequence, for example if $\eta = \xi$ is the two–dimensional space white noise. Indeed, if φ is a Schwartz function on $\mathbb{R}^2$ and if $\varphi_n = n\varphi(n\cdot)$ and

$$\eta_n(x) = \varphi_n * \xi(x) = \int_{\mathbb{R}^2} \varphi_n(x-y)\xi(y)\mathrm{d}y = \sum_{k\in\mathbb{Z}^2} \langle \xi, \varphi_n(x + 2\pi k - \cdot)\rangle,$$

then we will see below that there exist constants (c_n) with $\lim_n c_n = \infty$, such that $(X_n \circ \eta_n - c_n)$ converges in $C_T\mathscr{C}^{2\gamma-2}$ for all $T > 0$.

This is not a problem with our specific approximation. The homogenization setting shows that even for $\eta \to 0$ there are cases where the limiting equation is nontrivial. In the paracontrolled setting we have continuous dependence on the data $(\eta, X \circ \eta)$, so this non–triviality of the limit can only mean that it is $X \circ \eta$ which does not converge to zero.

Another way to see that there is a problem is to consider the following representation of the resonant term: use $\mathscr{L}X = \eta$ to write

$$X \circ \eta = X \circ \mathscr{L}X = \frac{1}{2}\mathscr{L}(X \circ X) + \partial_x X \circ \partial_x X = |\partial_x X|^2 + \frac{1}{2}\mathscr{L}(X \circ X) - 2\partial_x X \prec \partial_x X.$$

Integrating this equation over the torus and over $t \in [0, T]$, we get

$$\int_0^T \int_{\mathbb{T}} X \circ \eta \mathrm{d}x\mathrm{d}t = \int_0^T \int_{\mathbb{T}} |\partial_x X|^2 \mathrm{d}x\mathrm{d}t + \frac{1}{2}\int_{\mathbb{T}} (X(T) \circ X(T))\mathrm{d}x - 2\int_0^T \int_{\mathbb{T}} (\partial_x X \prec \partial_x X)\mathrm{d}x\mathrm{d}t$$

but now if $X \circ \eta \in C_T\mathscr{C}^{2\gamma-2}$ and $X \in C_T\mathscr{C}^{\gamma}$ then all the terms should be well defined and finite but this is cannot be since would mean that $\int_0^T \int_{\mathbb{T}} |\partial_x X|^2 \mathrm{d}x\mathrm{d}t < +\infty$. By direct computation however we can check that

$$\int_{\mathbb{T}} |\partial_x X(t, \cdot)|^2 \mathrm{d}x = +\infty$$

for any $t > 0$ almost surely if η is the space white noise. Note also that the problematic term $|\partial_x X|^2$ is exactly the correction term appearing in the analysis of the linear homogenization problem above.

In order to obtain convergence of the smooth solutions in general, we should introduce corrections to the equation to remove the divergent constant c_n. Let us see where the resonant product $X \circ \eta$ appears. We have

$$(F(u)\eta)^\sharp = F(u) \succ \eta + (F(u))^\sharp \circ \eta + C((F(u))^X, X, \eta) + (F(u))^X (X \circ \eta).$$

Now $(F(u))^X = F'(u)u^X$ by Theorem 5, and if u solves the equation, then $u^X = F(u)$ by Theorem 6. So we should really consider the renormalized equation

$$\mathscr{L} u_n = F(u_n) \diamond \eta_n = F(u_n)\xi_n - F'(u_n)F(u_n)c_n$$

and in the limit

$$\mathscr{L} u = F(u) \diamond \eta := (F(u) \diamond \eta)^\sharp + F(u) \prec \eta, \tag{37}$$

where the paracontrolled product $(F(u) \diamond \eta)^\sharp$ is calculated using $X \diamond \eta = \lim_n (X_n \circ \eta_n - c_n)$ in the place of $X \circ \eta$. Formally, we also denote this product by $F(u) \diamond \eta = F(u)\eta - F'(u)u^X \cdot \infty$, so that the solution u will satisfy

$$\mathscr{L} u = F(u) - F'(u)F(u) \cdot \infty.$$

Note that the correction term has exactly the same form as the Itô/Stratonovich corrector for SDEs and indeed we should consider (37) as the "Itô form" of the equation.

Remark 6 The convergence properties of $(X_n \circ \eta_n)$ are in stark contrast to the ODE setting: if we consider the equation $\partial_t u = F(u)\zeta$ rather than PAM, then we should replace X by Z with $\partial_t Z = \zeta$. But then we have in one dimension $Z \circ \zeta = 1/2\partial_t(Z \circ Z)$, so that the convergence of $(Z_n \circ \zeta_n)$ to $Z \circ \zeta$ comes for free with the convergence of (Z_n) to Z. This is due to the Leibniz rule for ∂_t and it is the reason why rough path theory is trivial in

one dimension. As we have discussed, for the second order differential operator $\mathscr{L}$ we have different rules and

$$(X \circ \eta) = (X \circ \mathscr{L}X) = \frac{1}{2}\mathscr{L}(X \circ X) + (\partial_x X \circ \partial_x X),$$

so that in our setting the nontrivial term is $\partial_x X \circ \partial_x X$.

These considerations lead naturally to the following definition.

Definition 6 *(*pam*–enhancement) Let $\gamma \in (2/3, 1)$ and let*

$$\mathcal{X}_{\text{pam}} \subseteq \mathscr{C}^{\gamma-2} \times C\mathscr{C}^{2\gamma-2}$$

be the closure of the image of the map

$$\Theta_{\text{pam}} : C^\infty \times C([0,\infty), \mathbb{R}) \to \mathcal{X}_{\text{pam}},$$

given by

$$\Theta_{\text{pam}}(\theta, f) = (\theta, \Phi \diamond \theta) := (\theta, \Phi \circ \theta - f), \tag{38}$$

where $\Phi = J\theta$, that is $\mathscr{L}\Phi = \theta$ and $\Phi(0) = 0$. We will call $\Theta_{\text{pam}}(\theta, f)$ the renormalized pam*–enhancement of the driving distribution θ. For $T > 0$ we define $\mathcal{X}_{\text{pam}}(T) = \mathcal{X}_{\text{pam}}|_{[0,T]}$ and we write $\|\mathbb{X}\|_{\mathcal{X}_{\text{pam}}(T)}$ for the norm of $\mathbb{X} \in \mathcal{X}_{\text{pam}}(T)$ in the Banach space $\mathscr{C}^{\gamma-2} \times C_T\mathscr{C}^{2\gamma-2}$. Moreover, we define the distance $d_{\mathcal{X}_{\text{pam}}(T)}(\mathbb{X}, \tilde{\mathbb{X}}) = \|\mathbb{X} - \tilde{\mathbb{X}}\|_{\mathcal{X}_{\text{pam}}(T)}$.*

Remark 7 It would be more elegant to renormalize $\Phi \circ \theta$ with a constant and not with a time-dependent function. But since we chose $\Phi(0) = 0$, we have $\Phi(0) \circ \theta = 0$ and therefore $(\Phi_n(0) \circ \theta_n - c_n)$ diverges for any diverging sequence of constants (c_n). A simple way of avoiding this problem is to consider the stationary version $\tilde{\Phi}$ given by

$$\tilde{\Phi}(x) = \int_0^\infty P_t \Pi_{\neq 0}\theta(x)\mathrm{d}t,$$

where $\Pi_{\neq 0}$ denotes the projection on the non-zero Fourier modes, $\Pi_{\neq 0}u = u - \hat{u}(0)$. But then $\tilde{\Phi}$ does not depend on time and in particular $\tilde{\Phi}(0) \neq 0$, so that we have to consider irregular initial conditions in the paracontrolled approach which complicates the presentation. Alternatively, we could observe that in the white noise case there exist constants (c_n) so that $(X_n(t) \circ \xi_n - c_n)$ converges for all $t > 0$, and while the limit $(X(t) \diamond \xi)$ diverges as $t \to 0$, it can be integrated against the heat kernel. Again, this would complicate the presentation and here we choose the simple (and cheap) solution of taking a time-dependent renormalization.

Theorem 7 *Let $\gamma \in (2/3, 1)$ and $\varepsilon > 0$ be such that $(2 + \varepsilon)\gamma > 2$. Let $\mathbb{X} = (X, \eta) \in \mathcal{X}_{\text{pam}}$, $F \in C_b^{2+\varepsilon}$, and $u_0 \in \mathscr{C}^{2\gamma}$. Then there exists a unique solution $u \in \mathscr{D}^\gamma(X)$ to the equation*

$$\mathscr{L}u = F(u) \diamond \eta, \qquad u(0) = u_0,$$

up to the (possibly finite) explosion time $\tau = \tau(u) = \inf\left\{t \geqslant 0 : \|u\|_{\mathscr{D}^\gamma_t} = \infty\right\} > 0$.

Moreover, u *depends on* $(u_0, \mathbb{X}) \in \mathscr{C}^{2\gamma} \times \mathcal{X}_{\mathrm{pam}}$ *in a locally Lipschitz continuous way: if* $M, T > 0$ *are such that for all* $(u_0, \mathbb{X})$ *with* $\|u_0\|_{2\gamma} \vee \|\mathbb{X}\|_{\mathcal{X}_{\mathrm{pam}}(T)} \leqslant M$*, the solution* u *to the equation driven by* $(u_0, \mathbb{X})$ *satisfies* $\tau(u) > T$*, and if* $(\tilde{u}_0, \tilde{\mathbb{X}})$ *is another set of data bounded in the above sense by* M*, then there exists* $C(F, M) > 0$ *for which*

$$d_{\mathscr{D}^\gamma_T}(u, \tilde{u}) \leqslant C(F, M)(\|u_0 - \tilde{u}_0\|_{2\gamma} + d_{\mathcal{X}_{\mathrm{pam}}(T)}(\mathbb{X}, \tilde{\mathbb{X}})).$$

Proof We only have to turn the formal discussion of Section 5.4 into rigorous mathematics. The small factor T^δ is obtained from a scaling argument and while this does not require any new insights it is somewhat lengthy and we refer to [GIP13, GP14] for details.

Let us just indicate how to iterate the construction to obtain the existence of solutions up to the explosion time τ. Let us assume that we constructed u on $[0, T_0]$ for some $T_0 > 0$. Now we no longer have $X(T_0) = 0$, and also the initial condition $u(T_0)$ is no longer in $\mathscr{C}^{2\gamma}$. But we only used $X(0) = 0$ to derive the initial conditions for u' and $u^\sharp$, and we only used $u_0 \in \mathscr{C}^\gamma$ to obtain $\mathscr{C}^{2\gamma}$ initial conditions for $u^\sharp$. Since now we already know $(u'(T_0), u^\sharp(T_0)) \in \mathscr{C}^\gamma \times \mathscr{C}^{2\gamma}$, we do not need this anymore.

As for the continuity in $(u_0, \mathbb{X})$, let $(\tilde{u}_0, \tilde{\mathbb{X}})$ be another set of data also bounded by M. Then the solutions u and $\tilde{u}$ both are bounded in $\mathscr{D}^\gamma_T$ by some constant $C = C(F, M) > 0$. So by the continuity properties of the paracontrolled product (and the other operations involved), we can estimate

$$d_{\mathscr{D}^\gamma_T}(u, \tilde{u}) \leqslant P(C)\left(\|u_0 - \tilde{u}_0\|_{2\gamma} + d_{\mathcal{X}_{\mathrm{pam}}(T)}(\mathbb{X}, \tilde{\mathbb{X}}) + T^\delta d_{\mathscr{D}^\gamma_T}(u, \tilde{u})\right)$$

for a polynomial P. The local Lipschitz continuity on $[0, T]$ immediately follows if we choose $T > 0$ small enough. This can be iterated to obtain the local Lipschitz continuity on "macroscopic" intervals. □

Remark 8 For the local in time existence it is not necessary to assume $F \in C^{2+\varepsilon}_b$, it suffices if $F \in C^{2+\varepsilon}$. This can be seen by considering a ball containing $u_0(x)$ for all $x \in \mathbb{T}^d$, a function $\tilde{F} \in C^{2+\varepsilon}_b$ which coincides with F on this ball, and by stopping u upon exiting the ball.

In the linear case $F(u) = u$ we have global in time solutions: in general we only get local in time solutions because we pick up a superlinear estimate when applying the paralinearization result Theorem 5. This step is not necessary if F is linear, and all the other estimates are linear in u.

5.6 Construction of the extended data

In order to apply Theorem 7 to equation (23) with white noise perturbation, it remains to show that if ξ is a spatial white noise on $\mathbb{T}^2$, then ξ defines an element of $\mathcal{X}_{\mathrm{pam}}$. In other words, we need to construct $X \diamond \xi$.

Since $P_t\xi$ is a smooth function for every $t > 0$, the resonant term $P_t\xi \circ \xi$ is a smooth function, and therefore we could formally set $X \circ \xi = \int_0^\infty (P_t\xi \circ \xi)\mathrm{d}t$. But we will see that this expression does not make sense.

Recall that $(\hat{\xi}(k))_{k\in\mathbb{Z}^2}$ is a complex valued, centered Gaussian process with covariance

$$\mathbb{E}[\hat{\xi}(k)\hat{\xi}(k')] = \delta_{k+k'=0}, \tag{39}$$

and such that $\hat{\xi}(k)^* = \hat{\xi}(-k)$.

Lemma 19 *For any $x \in \mathbb{T}^2$ and $t > 0$ we have*

$$g_t = \mathbb{E}[(P_t\xi \circ \xi)(x)] = \mathbb{E}[\Delta_{-1}(P_t\xi \circ \xi)(x)] = (2\pi)^{-2} \sum_{k\in\mathbb{Z}^2} e^{-t|k|^2}.$$

In particular, g_t does not depend on the partition of unity used to define the $\circ$ operator, and $\int_0^t g_s\mathrm{d}s = \infty$ for all $t > 0$.

Proof Let $x \in \mathbb{T}^2$, $t > 0$, and $\ell \geq -1$. Then

$$\mathbb{E}[\Delta_\ell(P_t\xi \circ \xi)(x)] = \sum_{|i-j|\leqslant 1} \mathbb{E}[\Delta_\ell(\Delta_i(P_t\xi)\Delta_j\xi)(x)],$$

where exchanging summation and expectation is justified because it can be easily verified that the partial sums of $\Delta_\ell(P_t\xi \circ \xi)(x)$ are uniformly L^p–bounded for any $p \geq 1$. Now $P_t = e^{-t|\cdot|^2}(\mathrm{D})$, and therefore we get from (39)

$$\begin{aligned}\mathbb{E}[\Delta_\ell(\Delta_i(P_t\xi)\Delta_j\xi)(x)] &= (2\pi)^{-1} \sum_{k,k'\in\mathbb{Z}^2} e^*_{k+k'}(x)\rho_\ell(k+k')\rho_i(k)e^{-t|k|^2}\rho_j(k')\mathbb{E}[\hat{\xi}(k)\hat{\xi}(k')] \\ &= (2\pi)^{-2} \sum_{k\in\mathbb{Z}^2} \rho_\ell(0)\rho_i(k)e^{-t|k|^2}\rho_j(k) = \frac{\delta_{\ell=-1}}{2\pi} \sum_{k\in\mathbb{Z}^2} \rho_i(k)\rho_j(k)e^{-t|k|^2}.\end{aligned}$$

For $|i - j| > 1$ we have $\rho_i(k)\rho_j(k) = 0$ and therefore

$$g_t = \mathbb{E}[(P_t\xi \circ \xi)(x)] = (2\pi)^{-2} \sum_{k\in\mathbb{Z}^2} \sum_{i,j} \rho_i(k)\rho_j(k)e^{-t|k|^2} = (2\pi)^{-2} \sum_{k\in\mathbb{Z}^2} e^{-t|k|^2},$$

while $\mathbb{E}[(P_t\xi \circ \xi)(x) - \Delta_{-1}(P_t\xi \circ \xi))(x)] = 0$. □

Exercise 12 Let φ be a Schwartz function on $\mathbb{R}^2$ and set

$$\xi_n(x) = ((n^2\varphi(n\cdot)) * \xi)(x) = \int_{\mathbb{R}^2} n^2\varphi(n(x-y))\xi(y)\mathrm{d}y = \sum_{k\in\mathbb{Z}^2} \langle \xi, n^2\varphi(n(x+2\pi k - \cdot))\rangle$$

for $x \in \mathbb{T}^2$. Write $\mathscr{F}_{\mathbb{R}^2}\varphi(z) = \int_{\mathbb{R}^2} e^{-i\langle z,x\rangle}\varphi(x)\mathrm{d}x$. Show that

$$\mathbb{E}[(P_t\xi_n \circ \xi_n)(x)] = \mathbb{E}[\Delta_{-1}(P_t\xi_n \circ \xi_n)(x)] = (2\pi)^{-2} \sum_{k\in\mathbb{Z}^2} e^{-t|k|^2}|\mathscr{F}_{\mathbb{R}^2}\varphi(k/n)|^2.$$

Hint: Use Poisson summation.

The diverging time integral motivates us to study the renormalized product $X \circ \xi - \int_0^\cdot g_s \mathrm{d}s$, where $\int_0^\cdot g_s \mathrm{d}s$ is an "infinite function":

Lemma 20 *Set*

$$(X \diamond \xi)(t) = \int_0^t (P_s \xi \circ \xi - g_s) \mathrm{d}s.$$

Then $\mathbb{E}[\|X \diamond \xi\|^p_{C_T \mathscr{C}^{2\gamma-2}(\mathbb{T}^2)}] < \infty$ *for all* $\gamma < 1$, $p \geq 1$, $T > 0$. *Moreover, if* φ *is a Schwartz function on* $\mathbb{R}^2$ *with* $\int \varphi(x) \mathrm{d}x = 1$, *if* $\xi_n = \varphi_n * \xi$ *with* $\varphi_n = n^2 \varphi(n \cdot)$ *for* $n \in \mathbb{N}$, *and* $X_n(t) = \int_0^\infty P_t \xi_n \mathrm{d}t$, *then*

$$\lim_{n \to \infty} \mathbb{E}[\|X \diamond \xi - (X_n \circ \xi_n - f_n)\|^p_{C_T \mathscr{C}^{2\gamma-2}(\mathbb{T}^2)}] = 0$$

for all $p \geq 1$, *where for all* $x \in \mathbb{T}^2$

$$\begin{aligned} f_n(t) &= \mathbb{E}[X_n(t,x)\xi_n(x)] = \mathbb{E}[(X_n(t) \circ \xi_n)(x)] \\ &= (2\pi)^{-2} \sum_{k \in \mathbb{Z}^2 \setminus \{0\}} \frac{|\mathscr{F}_{\mathbb{R}^2} \varphi(k/n)|^2}{|k|^2} (1 - e^{-t|k|^2}) + (2\pi)^{-2} t. \end{aligned}$$

Proof To lighten the notation, we will only show that $\mathbb{E}[\|X \diamond \xi\|^p_{C_T \mathscr{C}^{2\gamma-2}}] < \infty$. The convergence of $(X_n \circ \xi_n - f_n)$ to $X \diamond \xi$ is shown by applying dominated convergence, and we leave it as an exercise. Let $t > 0$ and define $\Xi_t = P_t \xi \circ \xi - g_t$. Using Gaussian hypercontractivity we will be able to reduce everything to estimating $\mathbb{E}[|\Delta_\ell \Xi_t(x)|^2]$ for $\ell \geqslant -1$ and $x \in \mathbb{T}^2$. Lemma 19 yields $\Delta_\ell g_t = 0 = \mathbb{E}[\Delta_\ell (P_t \xi \circ \xi)(x)]$ for $\ell \geq 0$ and $x \in \mathbb{T}^2$, and $\Delta_{-1} g_t = g_t = \mathbb{E}[\Delta_{-1}(P_t \xi \circ \xi)(x)]$, so that $\mathbb{E}[|\Delta_\ell \Xi_t(x)|^2] = \mathrm{Var}(\Delta_\ell (P_t \xi \circ \xi)(x))$. We have

$$\begin{aligned} \Delta_\ell (P_t \xi \circ \xi)(x) &= \sum_{k \in \mathbb{Z}^2} e_k^*(x) \rho_\ell(k) \mathscr{F}(P_t \xi \circ \xi))(k) \\ &= (2\pi)^{-1} \sum_{k_1, k_2 \in \mathbb{Z}^2} \sum_{|i-j| \leqslant 1} e^*_{k_1+k_2}(x) \rho_\ell(k_1 + k_2) \rho_i(k_1) e^{-t|k_1|^2} \hat{\xi}(k_1) \rho_j(k_2) \hat{\xi}(k_2). \end{aligned}$$

Hence

$$\begin{aligned} &\mathrm{Var}(\Delta_\ell (P_t \xi \circ \xi)(x)) \\ &\quad = (2\pi)^{-2} \sum_{k_1, k_2} \sum_{k_1', k_2'} \sum_{|i-j| \leqslant 1} \sum_{|i'-j'| \leqslant 1} e^*_{k_1+k_2}(x) \rho_\ell(k_1 + k_2) \rho_i(k_1) e^{-t|k_1|^2} \rho_j(k_2) \\ &\qquad \times e^*_{k_1'+k_2'}(x) \rho_\ell(k_1' + k_2') \rho_{i'}(k_1') e^{-t|k_1'|^2} \rho_{j'}(k_2') \, \mathrm{Cov}(\hat{\xi}(k_1)\hat{\xi}(k_2), \hat{\xi}(k_1')\hat{\xi}(k_2')), \end{aligned}$$

where exchanging summation and expectation can be justified a posteriori by the uniform L^p–boundedness of the partial sums. Now Wick's theorem ([Jan97], Theorem 1.28)

gives

$$\begin{aligned}\mathrm{Cov}(\hat{\xi}(k_1)\hat{\xi}(k_2),\hat{\xi}(k_1')\hat{\xi}(k_2')) &= \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_2)\hat{\xi}(k_1')\hat{\xi}(k_2')] - \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_2)]\mathbb{E}[\hat{\xi}(k_1')\hat{\xi}(k_2')]\\ &= \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_2)]\mathbb{E}[\hat{\xi}(k_1')\hat{\xi}(k_2')] + \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_1')]\mathbb{E}[\hat{\xi}(k_2)\hat{\xi}(k_2')]\\ &\quad + \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_2')]\mathbb{E}[\hat{\xi}(k_2)\hat{\xi}(k_1')] - \mathbb{E}[\hat{\xi}(k_1)\hat{\xi}(k_2)]\mathbb{E}[\hat{\xi}(k_1')\hat{\xi}(k_2')]\\ &= (\delta_{k_1+k_1'=0}\delta_{k_2+k_2'=0} + \delta_{k_1+k_2'=0}\delta_{k_2+k_1'=0}),\end{aligned}$$

and therefore

$$\begin{aligned}\mathrm{Var}(\Delta_\ell(P_t\xi\circ\xi)(x)) = (2\pi)^{-4}\sum_{k_1,k_2}\sum_{|i-j|\leqslant 1}\sum_{|i'-j'|\leqslant 1}&\mathbb{I}_{\ell\lesssim i}\mathbb{I}_{\ell\lesssim i'}\rho_\ell^2(k_1+k_2)\rho_i(k_1)\rho_j(k_2)\\ &\times[\rho_{i'}(k_1)\rho_{j'}(k_2)e^{-2t|k_1|^2} + \rho_{i'}(k_2)\rho_{j'}(k_1)e^{-t|k_1|^2-t|k_2|^2}].\end{aligned}$$

Now there exists $c>0$ such that $e^{-2t|k|^2}\lesssim e^{-tc2^{2i}}$ for all $k\in\mathrm{supp}(\rho_i)$ and for all $i\geq -1$. In the remainder of the proof the value of this strictly positive c may change from line to line. If $|i-j|\leqslant 1$, then we also have $e^{-t|k|^2}\lesssim e^{-tc2^{2i}}$ for all $k\in\mathrm{supp}(\rho_j)$. Thus

$$\begin{aligned}&\mathrm{Var}(\Delta_\ell(P_t\xi\circ\xi)(x))\\ &\quad\lesssim \sum_{i,j,i',j'}\mathbb{I}_{\ell\lesssim i}\mathbb{I}_{i\sim j\sim i'\sim j'}\sum_{k_1,k_2}\mathbb{I}_{\mathrm{supp}(\rho_\ell)}(k_1+k_2)\mathbb{I}_{\mathrm{supp}(\rho_i)}(k_1)\mathbb{I}_{\mathrm{supp}(\rho_j)}(k_2)e^{-2tc2^{2i}}\\ &\quad\lesssim \sum_{i:i\gtrsim\ell}2^{2i}2^{2\ell}e^{-tc2^{2i}} \lesssim \frac{2^{2\ell}}{t}\sum_{i:i\gtrsim\ell}e^{-tc2^{2i}} \lesssim \frac{2^{2\ell}}{t}e^{-tc2^{2\ell}},\end{aligned}$$

where we used that $t2^{2i}\lesssim e^{t(c-c')2^{2i}}$ for all $c'<c$.

Consider now $X\diamond\xi(t)=\int_0^t \Xi_s\mathrm{d}s$. We have for all $0\leqslant s<t$

$$\mathbb{E}[\|X\diamond\xi(t)-X\diamond\xi(s)\|^{2p}_{B^{2\gamma-2}_{2p,2p}}] = \sum_\ell 2^{2p\ell(2\gamma-2)}\int_{\mathbb{T}^2}\mathbb{E}[|\Delta_\ell(X\diamond\xi(t)-X\diamond\xi(s))(x)|^{2p}]\mathrm{d}x.$$

Since the random variable $_\ell(X\diamond\xi(t)-X\diamond\xi(s))(x)$ lives in the second non-homogeneous chaos generated by the Gaussian white noise ξ, we may use Gaussian hypercontractivity ([Jan97], Theorem 3.50) to bound

$$\begin{aligned}\mathbb{E}[|\Delta_\ell(X\diamond\xi(t)-X\diamond\xi(s))(x)|^{2p}] &\lesssim \mathbb{E}[|\Delta_\ell(X\diamond\xi(t)-X\diamond\xi(s))(x)|]^{2p}\\ &\leqslant \Big(\int_s^t\mathbb{E}[|\Delta_\ell\Xi_r(x)|]\mathrm{d}r\Big)^{2p}.\end{aligned}$$

But we just showed that

$$\begin{aligned}\mathbb{E}[|\Delta_\ell\Xi_r(x)|] &\leqslant \mathbb{E}[|\Delta_\ell\Xi_r(x)|^2]^{1/2} = (\mathrm{Var}(\Delta_\ell(P_r\xi\circ\xi)(x)))^{1/2}\\ &\lesssim r^{-1/2}2^\ell e^{-\frac{1}{2}rc2^{2\ell}} \leqslant r^{-1/2}2^\ell e^{-rc2^{2\ell}}\end{aligned}$$

(changing again the value of c), and therefore

$$\Big(\mathbb{E}\Big[\|X \diamond \xi(t) - X \diamond \xi(s)\|^{2p}_{B^{2\gamma-2}_{2p,2p}}\Big]\Big)^{1/2p} \lesssim \Big(\sum_\ell \Big(2^{\ell(2\gamma-2)} \int_s^t r^{-1/2} 2^\ell e^{-rc2^{2\ell}} \mathrm{d}r\Big)^{2p}\Big)^{1/2p}$$
$$\leqslant \sum_\ell 2^{\ell(2\gamma-1)} \int_s^t r^{-1/2} e^{-rc2^{2\ell}} \mathrm{d}r$$
$$\lesssim \Big(\int_s^t r^{-1/2} \int_{-1}^\infty (2^x)^{2\gamma-1} e^{-rc2^{2x}} \mathrm{d}x \mathrm{d}r\Big)^{2p}.$$

The change of variable $y = \sqrt{r}2^x$ leads to

$$\lesssim \Big(\int_s^t r^{-1/2} r^{-(2\gamma-1)/2} \int_0^\infty y^{2\gamma-2} e^{-cy^2} \mathrm{d}y \mathrm{d}r\Big)^{2p}.$$

For $\gamma > 1/2$, the integral in y is finite and we end up with

$$\lesssim \Big(\int_s^t r^{-\gamma} \mathrm{d}r\Big)^{2p} \lesssim |t-s|^{2p(1-\gamma)}$$

provided that $\gamma \in (1/2, 1)$. So for large enough p we can use Kolmogorov's continuity criterion to deduce that $\mathbb{E}[\|X \diamond \xi\|^{2p}_{C_T B^{2\gamma-2}_{2p,2p}}] < \infty$ for all $T > 0$. The claim now follows from Besov embedding, Lemma 8. □

Combining Theorem 7 and Lemma 20, we are finally able to solve (23) driven by a space white noise.

Corollary 2 *Let $\varepsilon > 0$ and let $F \in C_b^{2+\varepsilon}$ and assume that u_0 is a random variable that almost surely takes its values in $\mathscr{C}^{2\gamma}$ for some $\gamma \in (2/3, 1)$ with $(2+\varepsilon)\gamma > 2$. Let ξ be a spatial white noise on $\mathbb{T}^2$. Then there exists a unique solution u to*

$$\mathscr{L} u = F(u) \diamond \xi, \qquad u(0) = u_0,$$

up to the (possibly finite) explosion time $\tau = \tau(u) = \inf\big\{t \geqslant 0 : \|u\|_{\mathscr{D}^\gamma_t} = \infty\big\}$ which is almost surely strictly positive.

If (φ_n) and (ξ_n) are as described in Lemma 20, and if $(u_{0,n})$ converges in probability in $\mathscr{C}^{2\gamma}$ to u_0, then u is the limit in probability of the solutions u_n to

$$\mathscr{L} u_n = F(u_n) \diamond \xi_n, \qquad u_n(0) = u_{n,0}.$$

Remark 9 We even have a stronger result: We can fix a null set outside of which $X \diamond \xi$ is regular enough, and once we dispose of that null set we can solve all equations for any regular enough u_0 and F simultaneously, without ever having to worry about null sets again. This is for example interesting when studying stochastic flows or when studying equations with random u_0 and F.

The pathwise continuous dependence on the signal is also powerful in several other applications, for example support theorems and large deviations. For examples in the theory of rough paths see [FV10].

6 The stochastic Burgers equation

Let us now get to our main example, the "KPZ family" of equations. We concentrate here on the stochastics Burgers equation SBE, but essentially the same analysis works for the KPZ equation. We can also treat the heat equation with the same arguments, although in that case we need to set up the equation in the right way by applying a suitable transformation.

Recall that the stochastic Burgers equation SBE is

$$\mathscr{L}u = \partial_x u^2 + \partial_x \xi, \qquad u(0) = u_0, \tag{40}$$

where $u : [0,\infty) \times \mathbb{T} \to \mathbb{R}$, ξ is a space-time white noise, and ∂_x denotes the spatial derivative. As we argued before, the solution u cannot be expected to behave better than the Ornstein–Uhlenbeck process X, the solution of the linear equation $\mathscr{L}X = \partial_x \xi$, and as we saw in Section 2 $X(t)$ is for all $t > 0$ a smooth function of the space variable plus a space white noise. By Exercise 11, the white noise in dimension 1 has regularity $\mathscr{C}^{-1/2-}$. Thus $X \in C\mathscr{C}^{-1/2-}$, and in particular u^2 is the square of a distribution and a priori not well defined.

What raises some hope is that in Lemma 2 we were able to show that $\partial_x X^2$ exists as a space–time distribution. So as in the previous examples there are stochastic cancellations going into $\partial_x X^2$. The energy solution approach was designed to take those cancellations into account in the full solution u, but while it allowed us to work under rather weak assumptions which easily gave us existence of solutions, it did not give us sufficient control to have uniqueness of solutions. On the other side, a suitable paracontrolled ansatz for the solution u will allow us to transfer the cancellation properties of X to u and it will allow us to construct $\partial_x u^2$ as a *continuous* bilinear map, from where existence and uniqueness of solutions easily follows.

6.1 Structure of the solution

In this discussion we consider the case of zero initial condition and smooth noise ξ, and we analyze the structure of the solution. Let us expand u around the Ornstein–Uhlenbeck process X with $\mathscr{L}X = \partial_x \xi$, $X(0) = 0$. Setting $u = X + u^{\geqslant 1}$, we have

$$\mathscr{L}u^{\geqslant 1} = \partial_x(u^2) = \partial_x(X^2) + 2\partial_x(Xu^{\geqslant 1}) + \partial_x((u^{\geqslant 1})^2).$$

Let us define the bilinear map

$$B(f,g) = J\partial_x(fg) = \int_0^{\cdot} P_{\cdot - s}\partial_x(f(s)g(s))\mathrm{d}s.$$

Then we can proceed by performing a further change of variables in order to remove the term $\partial_x(X^2)$ from the equation by setting

$$u = X + B(X,X) + u^{\geqslant 2}. \tag{41}$$

Now $u^{\geqslant 2}$ satisfies

$$\begin{aligned} \mathscr{L} u^{\geqslant 2} &= 2\partial_x(XB(X,X)) + \partial_x(B(X,X)B(X,X)) \\ &\quad +2\partial_x(Xu^{\geqslant 2}) + 2\partial_x(B(X,X)u^{\geqslant 2}) + \partial_x((u^{\geqslant 2})^2). \end{aligned} \tag{42}$$

We can imagine to make a similar change of variables to get rid of the term

$$2\partial_x(XB(X,X)) = \mathscr{L} B(X, B(X,X)).$$

As we proceed in this inductive expansion, we generate a number of explicit terms, obtained by various combinations of X and B. Since we will have to deal explicitly with at least some of these terms, it is convenient to represent them with a compact notation involving binary trees. A binary tree $\tau \in \mathcal{T}$ is either the root $\bullet$ or the combination of two smaller binary trees $\tau = (\tau_1 \tau_2)$, where the two edges of the root of τ are attached to τ_1 and τ_2 respectively. For example

$$(\bullet\bullet) = \vee, \quad (\vee\bullet) = \begin{array}{c}\vee\\ \bullet\end{array}, \quad (\bullet\begin{array}{c}\vee\\ \bullet\end{array}) = \begin{array}{c}\vee\\ \vee\\ \bullet\end{array}, \quad (\vee\vee) = \begin{array}{c}\vee\\ \vee\vee\end{array}, \quad \ldots$$

Then we define recursively

$$X^{\bullet} = X, \qquad X^{(\tau_1\tau_2)} = B(X^{\tau_1}, X^{\tau_2}),$$

giving

$$X^{\vee} = B(X,X), \quad X^{\begin{array}{c}\vee\\ \bullet\end{array}} = B(X, X^{\vee}), \quad X^{\begin{array}{c}\vee\\ \vee\\ \bullet\end{array}} = B(X, X^{\begin{array}{c}\vee\\ \bullet\end{array}}), \quad X^{\begin{array}{c}\vee\\ \vee\vee\end{array}} = B(X^{\vee}, X^{\vee}),$$

and so on. In this notation the expansion (41)–(42) reads

$$u = X + X^{\vee} + u^{\geqslant 2}, \tag{43}$$

$$u^{\geqslant 2} = 2X^{\begin{array}{c}\vee\\ \bullet\end{array}} + X^{\begin{array}{c}\vee\\ \vee\vee\end{array}} + 2B(X, u^{\geqslant 2}) + 2B(X^{\vee}, u^{\geqslant 2}) + B(u^{\geqslant 2}, u^{\geqslant 2}). \tag{44}$$

Remark 10 We observe that formally the solution u of SBE can be expanded as an infinite sum of terms labelled by binary trees:

$$u = \sum_{\tau \in \mathcal{T}} c(\tau) X^{\tau},$$

where $c(\tau)$ is a combinatorial factor counting the number of planar trees which are isomorphic (as graphs) to τ. For example $c(\bullet) = 1$, $c\,(\vee) = 1$, $c\,\left(\begin{array}{c}\vee\\ \bullet\end{array}\right) = 2$, $c(\begin{array}{c}\vee\\ \vee\\ \bullet\end{array}) = 4$, $c\,\left(\begin{array}{c}\vee\\ \vee\vee\end{array}\right) = 1$ and in general $c(\tau) = \sum_{\tau_1,\tau_2 \in \mathcal{T}} \mathbb{I}_{(\tau_1\tau_2)=\tau} c(\tau_1) c(\tau_2)$. Alternatively, we may truncate the summation at trees of degree at most n and set

$$u = \sum_{\tau \in \mathcal{T}, d(\tau) < n} c(\tau) X^{\tau} + u^{\geqslant n},$$

where we denote by $d(\tau) \in \mathbb{N}_0$ the degree of the tree τ, given by $d(\bullet) = 0$ and then inductively $d((\tau_1\tau_2)) = 1 + d(\tau_1) + d(\tau_2)$. For example $d(\mathbf{V}) = 1$, $d\left(\mathbf{V}_{\mathbf{V}}\right) = 2$, $d(\mathbf{V}_{\mathbf{V}_{\mathbf{V}}}) = 3$, $d\left(\mathbf{V}_{\mathbf{V}\mathbf{V}}\right) = 3$. We then obtain for the remainder

$$u^{\geqslant n} = \sum_{\substack{\tau_1,\tau_2 \,:\, d(\tau_1) < n,\, d(\tau_2) < n \\ d((\tau_1\tau_2)) \geqslant n}} c(\tau_1)c(\tau_2)X^{(\tau_1\tau_2)} + \sum_{\tau : d(\tau) < n} c(\tau)B(X^\tau, u^{\geqslant n}) + B(u^{\geqslant n}, u^{\geqslant n}). \tag{45}$$

Our aim is to control the truncated expansion under the natural regularity assumptions in the white noise case, $X \in C\mathscr{C}^{-1/2-}$. Since (45) contains the term $B(X, u^{\geqslant n})$ which in turn contains the paraproduct $J\partial_x(u^{\geqslant n} \prec X)$, the remainder $u^{\geqslant n}$ will be at best in $C\mathscr{C}^{1/2-}$. But then the sum of the regularities of X and $u^{\geqslant n}$ is negative, and the term $B(X, u^{\geqslant n})$ is not well defined. We therefore continue the expansion up to the point (turning out to be $u^{\geqslant 3}$) where we can set up a paracontrolled ansatz for the remainder, which will allow us to make sense of $\partial_x(X \circ u^{\geqslant n})$ and thus of $B(X, u^{\geqslant n})$.

6.2 Paracontrolled solution

Inspired by the partial tree series expansion of u, we set up a paracontrolled ansatz of the form

$$u = X + X^{\mathbf{V}} + 2X^{\mathbf{V}_{\mathbf{V}}} + u^{Q}, \qquad u^{Q} = u' \prec Q + u^{\sharp}, \tag{46}$$

where the functions u', Q and $u^\sharp$ are for the moment arbitrary, but we assume $u', Q \in \mathscr{L}^\gamma$ and $u^\sharp \in \mathscr{L}^{2\gamma}$, where from now on we fix $\gamma \in (1/3, 1/2)$. For such u, the nonlinear term takes the form

$$\begin{aligned}\partial_x u^2 = \partial_x(X^2 + 2X^{\mathbf{V}}X + (X^{\mathbf{V}})^2 + 4X^{\mathbf{V}_{\mathbf{V}}}X) + 2\partial_x(u^Q X) \\ + 2\partial_x(X^{\mathbf{V}}(u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})) + \partial_x((u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})^2),\end{aligned} \tag{47}$$

which gives us an equation for u^Q:

$$\begin{aligned}\mathscr{L}u^Q &= \partial_x((X^{\mathbf{V}})^2 + 4X^{\mathbf{V}_{\mathbf{V}}}X) + 2\partial_x(u^Q X) + 2\partial_x(X^{\mathbf{V}}(u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})) + \partial_x((u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})^2) \\ &= \mathscr{L}X^{\mathbf{V}\mathbf{V}} + 4\mathscr{L}X^{\mathbf{V}_{\mathbf{V}}} + 2\partial_x(u^Q X) + 2\partial_x(X^{\mathbf{V}}(u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})) + \partial_x((u^Q + 2X^{\mathbf{V}_{\mathbf{V}}})^2).\end{aligned} \tag{48}$$

If we formally apply the paraproduct estimate Theorem 3 (which is of course not possible since the regularity requirements for the resonant term are not satisfied), we derive the following natural regularities for the driving terms: $X \in C\mathscr{C}^{-1/2-}$, $X^{\mathbf{V}} \in C\mathscr{C}^{0-}$, $X^{\mathbf{V}_{\mathbf{V}}} \in \mathscr{L}^{1/2-}$, $X^{\mathbf{V}_{\mathbf{V}}} \in \mathscr{L}^{1/2-}$, and $X^{\mathbf{V}\mathbf{V}} \in \mathscr{L}^{1-}$. In terms of γ, we can encode this as

$$X \in C\mathscr{C}^{\gamma-1}, \quad X^{\mathbf{V}} \in C\mathscr{C}^{2\gamma-1}, \quad X^{\mathbf{V}_{\mathbf{V}}} \in \mathscr{L}^{\gamma}, \quad X^{\mathbf{V}_{\mathbf{V}}} \in \mathscr{L}^{\gamma}, \quad X^{\mathbf{V}\mathbf{V}} \in \mathscr{L}^{2\gamma}.$$

Under these regularity assumptions the term $2\partial_x(X^{\mathbf{V}}(u^Q + X^{\mathbf{V}_{\mathbf{V}}})) + \partial_x((u^Q + X^{\mathbf{V}_{\mathbf{V}}})^2)$ is well defined and the only problematic term in (48) is $\partial_x(u^Q X)$. Using the paracontrolled

structure of u^Q, we can make sense of $\partial_x(u^Q X)$ as a bounded operator provided that $Q \circ X \in C\mathscr{C}^{2\gamma-1}$ is given. In other words, the right hand side of (48) is well defined for paracontrolled distributions.

Next, we should specify how to choose Q and which form u' will take for the solution u^Q. We have formally

$$\begin{aligned}\mathscr{L}u^Q &= \mathscr{L}X^{\mathbf{V}\mathbf{V}} + 4\mathscr{L}X^{\mathbf{V}_\mathbf{V}} + 2\partial_x(u^Q X) + 2\partial_x(X^{\mathbf{V}}(u^Q + 2X^{\mathbf{V}_\mathbf{V}})) + \partial_x((u^Q + 2X^{\mathbf{V}_\mathbf{V}})^2)\\ &= 4\partial_x(X^{\mathbf{V}_\mathbf{V}}X) + 2\partial_x(u^Q X) + C\mathscr{C}^{2\gamma-2}\\ &= 4X^{\mathbf{V}_\mathbf{V}} \prec \partial_x X + 2u^Q \prec \partial_x X + C\mathscr{C}^{2\gamma-2},\end{aligned}$$

where we assumed that not only $\mathscr{L}X^{\mathbf{V}_\mathbf{V}} \in C\mathscr{C}^{\gamma-2}$, but that $\partial_x(X^{\mathbf{V}_\mathbf{V}} \circ X) \in C\mathscr{C}^{2\gamma-1}$ (which implies $\mathscr{L}X^{\mathbf{V}_\mathbf{V}} \in C\mathscr{C}^{\gamma-2}$, but also the stronger statement $\mathscr{L}X^{\mathbf{V}_\mathbf{V}} - X^{\mathbf{V}_\mathbf{V}} \prec \partial_x X \in C\mathscr{C}^{2\gamma-2}$). By Theorem 6, u^Q is paracontrolled by $J(\partial_x X)$, and in other words we should set $Q = J(\partial_x X)$. The derivative u' of the solution u^Q will then be given by $u' = 4X^{\mathbf{V}_\mathbf{V}} + 2u^Q$.

Unlike for PAM, here we do not need to introduce a renormalization. This is due to the fact that we differentiate after taking the square: to construct u^2, we would have to subtract an infinite constant and formally consider $u^{\diamond 2} = u^2 - \infty$, or at the level of the approximation $u_n^2 - c_n$. But then

$$\partial_x u^{\diamond 2} = \lim_{n\to\infty} \partial_x(u_n^2 - c_n) = \lim_{n\to\infty} \partial_x u_n^2 = \partial_x u^2.$$

So we obtain the following description of the driving data for the stochastic Burgers equation.

Definition 7 *(*sbe*–enhancement) Let $\gamma \in (1/3, 1/2)$ and let*

$$\mathcal{X}_{\text{sbe}} \subseteq C\mathscr{C}^{\gamma-1} \times C\mathscr{C}^{2\gamma-1} \times \mathscr{L}^\gamma \times \mathscr{L}^{2\gamma} \times C\mathscr{C}^{2\gamma-1} \times C\mathscr{C}^{2\gamma-1}$$

be the closure of the image of the map $\Theta_{\text{sbe}} : C(\mathbb{R}_+, C^\infty(\mathbb{T})) \to \mathcal{X}_{\text{sbe}}$ given by

$$\Theta_{\text{sbe}}(\theta) = (X(\theta), X^{\mathbf{V}}(\theta), X^{\mathbf{V}_\mathbf{V}}(\theta), X^{\mathbf{V}\mathbf{V}}(\theta), (X^{\mathbf{V}_\mathbf{V}} \circ X)(\theta), (Q \circ X)(\theta)), \tag{49}$$

where

$$\begin{aligned} X(\theta) &= J(\partial_x \theta),\\ X^{\mathbf{V}}(\theta) &= B(X(\theta), X(\theta)),\\ X^{\mathbf{V}_\mathbf{V}}(\theta) &= B(X^{\mathbf{V}}(\theta), X(\theta)),\\ X^{\mathbf{V}\mathbf{V}}(\theta) &= B(X^{\mathbf{V}}(\theta), X^{\mathbf{V}}(\theta)),\\ Q(\theta) &= J(\partial_x X(\theta)).\end{aligned} \tag{50}$$

We will call $\Theta_{\text{sbe}}(\theta)$ the sbe*–enhancement of the driving distribution θ. For $T > 0$ we define $\mathcal{X}_{\text{sbe}}(T) = \mathcal{X}_{\text{sbe}}|_{[0,T]}$ and we write $\|\mathbb{X}\|_{\mathcal{X}_{\text{sbe}}(T)}$ for the norm of $\mathbb{X}$ in the Banach space $C_T\mathscr{C}^{\gamma-1} \times C_T\mathscr{C}^{2\gamma-1} \times \mathscr{L}_T^\gamma \times \mathscr{L}_T^{2\gamma} \times C_T\mathscr{C}^{2\gamma-1} \times C_T\mathscr{C}^{2\gamma-1}$. Moreover, we define the distance $d_{\mathcal{X}_{\text{sbe}}(T)}(\mathbb{X}, \tilde{\mathbb{X}}) = \|\mathbb{X} - \tilde{\mathbb{X}}\|_{\mathcal{X}_{\text{sbe}}(T)}$.*

For every $\mathbb{X} \in \mathcal{X}_{\text{sbe}}$, there is an associated space of paracontrolled distributions:

Definition 8 *Let $\mathbb{X} \in \mathcal{X}_{\text{sbe}}$. Then the space of paracontrolled distributions $\mathscr{D}^\gamma(\mathbb{X})$ is defined as the set of all $(u, u') \in C\mathscr{C}^{\gamma-1} \times \mathscr{L}^\gamma$ with*

$$u = X + X^{\mathbf{v}} + 2X^{\mathbf{v}} + u' \prec Q + u^\sharp,$$

where $u^\sharp \in \mathscr{L}^{2\gamma}$. For $T > 0$ we define

$$\|u\|_{\mathscr{D}^\gamma_T} = \|u'\|_{\mathscr{L}^\gamma_T} + \|u^\sharp\|_{C_T\mathscr{C}^{2\gamma}}.$$

If $\tilde{\mathbb{X}} \in \mathcal{X}_{\text{rbe}}$ and $(\tilde{u}, \tilde{u}') \in \mathscr{D}^\gamma(\tilde{\mathbb{X}})$, then we also write

$$d_{\mathscr{D}^\gamma_T}(u, \tilde{u}) = \|u' - \tilde{u}'\|_{\mathscr{L}^\gamma_T} + \|u^\sharp - \tilde{u}^\sharp\|_{C_T\mathscr{C}^{2\gamma}_T}.$$

We now have everything in place to solve SBE driven by $\mathbb{X} \in \mathcal{X}_{\text{sbe}}$.

Theorem 8 *Let $\gamma \in (1/3, 1/2)$. Let $\mathbb{X} \in \mathcal{X}_{\text{sbe}}$, write $\partial_x\theta = \mathscr{L}X$, and let $u_0 \in \mathscr{C}^{2\gamma}$. Then there exists a unique solution $u \in \mathscr{D}^\gamma(\mathbb{X})$ to the equation*

$$\mathscr{L}u = \partial_x u^2 + \partial_x\theta, \qquad u(0) = u_0, \tag{51}$$

up to the (possibly finite) explosion time $\tau = \tau(u) = \inf\left\{t \geqslant 0 : \|u\|_{\mathscr{D}^\gamma_t} = \infty\right\} > 0$.

Moreover, u depends on $(u_0, \mathbb{X}) \in \mathscr{C}^{2\gamma} \times \mathcal{X}_{\text{sbe}}$ in a locally Lipschitz continuous way: if $M, T > 0$ are such that for all $(u_0, \mathbb{X})$ with $\|u_0\|_{2\gamma} \vee \|\mathbb{X}\|_{\mathcal{X}_{\text{sbe}}(T)} \leqslant M$, the solution u to the equation driven by $(u_0, \mathbb{X})$ satisfies $\tau(u) > T$, and if $(\tilde{u}_0, \tilde{\mathbb{X}})$ is another set of data bounded in the above sense by M, then there exists $C(M) > 0$ for which

$$d_{\mathscr{D}^\gamma_T}(u, \tilde{u}) \leqslant C(M)(\|u_0 - \tilde{u}_0\|_{2\gamma} + d_{\mathcal{X}_{\text{sbe}}(T)}(\mathbb{X}, \tilde{\mathbb{X}})).$$

Proof By definition of the term $\partial_x u^2$, the distribution $u \in \mathscr{D}^\gamma(\mathbb{X})$ solves (51) if and only if $u^Q = u - X - X^{\mathbf{v}} - 2X^{\mathbf{v}}$ solves

$$\mathscr{L}u^Q = \mathscr{L}X^{\mathbf{v}} + 4\partial_x(X^{\mathbf{v}}X) + 2\partial_x(u^Q X) + 2\partial_x(X^{\mathbf{v}}(u^Q + 2X^{\mathbf{v}})) + \partial_x((u^Q + 2X^{\mathbf{v}})^2)$$

with initial condition $u^Q(0) = u_0$. This equation is structurally very similar to PAM (23) and can be solved using the same arguments, which we do not reproduce here. □

For this result to be of any use we still have to show that if ξ is the space-time white noise, then there is almost surely an element of $\mathcal{X}_{\text{sbe}}$ associated to $\partial_x\xi$. While for PAM we needed to construct only one term, here we have to construct five terms: $X^{\mathbf{v}}, X^{\mathbf{v}}, X^{\mathbf{v}}, X^{\mathbf{v}} \circ X, Q \circ X$. For details we refer to [GP14]. Alternatively we can simply differentiate the extended data which Hairer constructed for the KPZ equation in Chapter 5 of [Hai13].

The same approach allows us to solve the KPZ equation $\mathscr{L}h = (\partial_x h)^{\diamond 2} + \xi$, and if we are careful how to interpret the product $w \diamond \xi$, then also the linear heat equation $\mathscr{L}w = w \diamond \xi$. In both cases the solution will depend continuously on some suitably extended data that is constructed from ξ in a similar way as described in Definition 7. Moreover, the formal links between the three equations that we discussed in Section 4.2 can be made rigorous. These results will be included in [GP14].

References

[BCD11] Hajer Bahouri, Jean-Yves Chemin and Raphael Danchin. *Fourier analysis and nonlinear partial differential equations*. Springer, Berlin, 2011.

[BG97] Lorenzo Bertini and Giambattista Giacomin. Stochastic Burgers and KPZ equations from particle systems. *Comm. Math. Phys.*, 183(3):571–607, 1997.

[BG13] Guillaume Bal and Yu Gu. Limiting models for equations with large random potential; a review. Preprint available at `http://www.columbia.edu/~gb2030/PAPERS/Review-CMS.pdf`, 2013.

[Bon81] 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.

[Cha00] Terence Chan. Scaling limits of Wick ordered KPZ equation. *Comm. Math. Phys.*, 209(3):671–690, 2000.

[FH14] Peter Friz and Martin Hairer. *A Course on Rough Paths*. Springer, Berlin, 2014.

[FV10] Peter Friz and Nicolas Victoir. *Multidimensional stochastic processes as rough paths. Theory and applications*. Cambridge University Press, Cambridge, 2010.

[GIP13] Massimiliano Gubinelli, Peter Imkeller and Nicolas Perkowski. Paracontrolled distributions and singular PDEs. *ArXiv:1210.2684*, 2014.

[GJ10] Patricia Goncalves and Milton Jara. Universality of KPZ equation. *ArXiv:1003.4478*, 2010.

[GJ13] Massimiliano Gubinelli and Milton Jara. Regularization by noise and stochastic Burgers equations. *Stochastic Partial Differential Equations: Analysis and Computations*, 1(2):325–350, 2013.

[GJ14] Patricia Goncalves and Milton Jara. Nonlinear fluctuations of weakly asymmetric interacting particle systems. *Arch. Ration. Mech. Anal.*, 212(2), 597–644, 2014.

[GP14] Massimiliano Gubinelli and Nicolas Perkowski. KPZ reloaded. In preparation, 2015.

[Gub04] Massimiliano Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

[Gub10] Massimiliano Gubinelli. Ramification of rough paths. *J. Differential Equations*, 248(4):693–721, 2010.

[Hai13] Martin Hairer. Solving the KPZ equation. *Ann. Math.*, 178(2):559–664, 2013.

[Hai14] Martin Hairer. A theory of regularity structures. *Invent. math.*, 198(2):269–504, 2014.

[HPP13] Martin Hairer, Etienne Pardoux and Andrey Piatnitski. Random homogenisation of a highly oscillatory singular potential. *Stochastic Partial Differential Equations: Analysis and Computations*, 1(4):571–605, 2013.

[Hu02] Yaozhong Hu. Chaos expansion of heat equations with white noise potentials. *Potential Anal.*, 16(1):45–66, 2002.

[Jan97] Svante Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[KPZ86] Mehran Kardar, Giorgio Parisi and Yi-Cheng Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889–892, 1986.

[LCL07] Terry J. Lyons, Michael Caruana and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[LQ02] Terry J. Lyons and Zhongmin Qian. *System control and rough paths*. Oxford University Press, 2002.

[Lyo98] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.*, 14(2):215–310, 1998.

[RVW01] Francesco Russo, Pierre Vallois and Jochen Wolf. A generalized class of Lyons-Zheng processes. *Bernoulli*, 7(2):363–379, 2001.

[ST87] Hans-Jürgen Schmeisser and Hans Triebel. *Topics in Fourier analysis and function spaces*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1987.

[You36] Laurence C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.