Singular stochastic PDEs

Martin Hairer*

Abstract. We present a series of recent results on the well-posedness of very singular parabolic stochastic partial differential equations. These equations are such that the question of what it even means to be a solution is highly non-trivial. This problem can be addressed within the framework of the recently developed theory of "regularity structures", which allows to describe candidate solutions locally by a "jet", but where the usual Taylor polynomials are replaced by a sequence of custom-built objects. In order to illustrate the theory, we focus on the particular example of the Kardar-Parisi-Zhang equation, a popular model for interface propagation.

Mathematics Subject Classification (2010). 60H15; 81S20, 82C28.

Keywords. Regularity structures, renormalisation, stochastic PDEs.

1. Introduction

In this article, we report on a recently developed theory [23] allowing to give a robust meaning to a large class of stochastic partial differential equations (SPDEs) that have traditionally been considered to be ill-posed. The general structure of these equations is

$$\mathcal{L}u = F(u) + G(u)\xi \,, \tag{1}$$

where the dominant linear operator $\mathcal{L}$ is of parabolic (or possibly elliptic) type, F and G are local nonlinearities depending on u and its derivatives of sufficiently low order, and ξ is some driving noise. Problems arise when ξ (and therefore also u) is so singular that some of the terms appearing in F and / or the product between G and ξ are ill-posed. For simplicity, we will consider all of our equations in a *bounded* spatial region with periodic boundary conditions.

One relatively simple example of an ill-posed equation of the type (1) is that of a system of equations with a nonlinearity of Burgers type driven by space-time white noise:

$$\partial_t u = \partial_x^2 u + F(u)\,\partial_x u + \xi \,. \tag{2}$$

(See Section 2.2 below for a definition of the space-time white noise ξ.) Here, $u(x,t) \in \mathbf{R}^n$ and F is a smooth matrix-valued function, so that one can in general

*I am delighted to thank the Institute for Advanced Study for its warm hospitality and the 'The Fund for Math' for funding my stay there. This work was supported by the Leverhulme trust through a leadership award, the Royal Society through a Wolfson research award, and the ERC through a consolidator award.

not rewrite the nonlinearity as a total derivative. In this example, which was originally studied in [20] but then further analysed in the series of articles [24, 29, 25], solutions at any fixed instant of time have exactly the same regularity (in space) as Brownian motion. As a consequence, $\partial_x u$ is expected to "look like" white noise. It is of course very well-known from the study of ordinary stochastic differential equations (SDEs) that in this case the product $F(u)\,\partial_x u$ is "unstable": one can get different answers depending on the type of limiting procedure used to define it. This is the reason why one has different solution theories for SDEs: one obtains different answers, depending on whether they are interpreted in the Itô or in the Stratonovich sense [30, 43, 44].

Another example is given by the KPZ equation [32] which can formally be written as

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 - C + \xi \;, \tag{3}$$

and is a very popular model of one-dimensional interface propagation. As in the case of (2), one expects solutions to this equation to "look like" Brownian motion (in space) for any fixed instant of time. Now the situation is much worse however: the nonlinearity looks like the square of white noise, which really shouldn't make any sense! In this particular case however, one can use a "trick", the Cole-Hopf transform, to reduce the problem to an equation that has an interpretation within the framework of classical SPDE theory [4]. Furthermore, this "Cole-Hopf solution" was shown in [4] to be the physically relevant solution since it describes the mesoscopic fluctuations of a certain microscopic interface growth model, see also [17]. On the other hand, the problem of interpreting these solutions directly at the level of (3) and to show their stability under suitable approximations had been open for a long time, before being addressed in [21].

Both examples mentioned so far have only one space dimension. This particular feature (together with some additional structure in the case of the KPZ equation, see Remark 5.17 below) allowed to treat them by borrowing estimates and techniques from the theory of (controlled) rough paths [34, 18, 15]. This approach breaks down in higher spatial dimensions. More recently, a general theory of "regularity structures" was developed in [23], which unifies many previous approaches and allows in particular to treat higher dimensional problems.

Two nice examples of equations that can be treated with this new approach are given by

$$\partial_t \Phi = \Delta\Phi + C\Phi - \Phi^3 + \xi \;, \tag{4a}$$

$$\partial_t \Psi = -\Delta\big(\Delta\Psi + C\Psi - \Psi^3\big) + \operatorname{div}\xi \;, \tag{4b}$$

in space dimension $d = 3$. These equations can be interpreted as the natural "Glauber" and "Kawasaki" dynamics associated to Euclidean Φ^4 field theory in the context of stochastic quantisation [40]. It is also expected to describe the dynamical mesoscale fluctuations for phase coexistence models that are "almost mean-field", see [5]. These equations cease to have function-valued solutions in dimension $d \geq 2$, so that the classical interpretation of the cubic nonlinearity loses its meaning there. In two dimensions, a solution theory for these equations

was developed in [1], which was later improved in [10, 11, 12], see Section 3.1 below. The case $d = 3$ (which is the physically relevant one in the interpretation as dynamical fluctuations for phase coexistence models) had remained open and was eventually addressed in [23].

A final example of the kind of equations that can be addressed by the theory exposed in these notes (but this list is of course not exhaustive) is a continuous analogue to the classical parabolic Anderson model [8]:

$$\partial_t u = \Delta u + u\,\eta + Cu\;, \tag{5}$$

in dimensions $d \in \{2, 3\}$. In this equation, η denotes a noise term that is white in space, but constant in time. This time, the problem is that in dimension $d \geq 2$, the product $u\,\eta$ ceases to make sense classically, as a consequence of the lack of regularity of u.

The following "meta-theorem" (formulated in a somewhat vague sense, precise formulations differ slightly from problem to problem and can be found in the abovementioned articles) shows in which sense one can give meaning to all of these equations.

Theorem 1.1. *Consider the sequence of classical solutions to any of the equations* (2)–(5) *with* ξ *(resp.* η*) replaced by a smooth regularised noise* ξ_ε *and* $C = C_\varepsilon$ *depending on* ε*. Then, there exists a choice* $C_\varepsilon \to \infty$ *such that this sequence of solutions converges to a limit in probability, locally in time. Furthermore, this limit is universal, i.e. does not depend on the details of the regularisation* ξ_ε*.*

Besides these convergence results, the important fact here is that the limit is *independent* of the precise details of the regularisation mechanism. In addition, the theory of regularity structures also yields rates of convergence, as well as an intrinsic description of these limits. It also provides automatically a very detailed local description of these limits.

The aim of this article is to give an overview of the ingredients involved in the proof of a result like Theorem 1.1. We structure this as follows. In Section 2, we recall a number of properties and definitions of Hölder spaces of positive (and negative!) order that will be useful for our argument. In Section 3, we then explain how, using only standard tools, it is possible to provide a robust solution theory for not-so-singular SPDEs, like for example (4) in dimension $d = 2$. Section 4 is devoted to a short overview of the main definitions and concepts of the abstract theory of regularity structures which is a completely general way of formalising the properties of objects that behave "like Taylor polynomials". Section 5 then finally shows how one can apply this general theory to the specific context of the type of parabolic SPDEs considered above, how renormalisation procedures can be built into the theory, and how this affects the equations.

Throughout the whole article, our argumentation will remain mostly at the heuristic level, but we will make the statements and definitions as precise as possible.

1.1. An alternative approach.

A different approach to building solution theories for singular PDEs was developed simultaneously to the one presented

here by Gubinelli & Al in [19]. That approach is based on the properties of Bony's paraproduct [7, 3, 2], in particular on the paralinearisation formula. One advantage is that in the paraproduct-based approach one generally deals with globally defined objects rather than the "jets" used in the theory of regularity structures. This comes at the expense of achieving a less clean break between the analytical and the algebraic aspects of a given problem and obtaining less detailed information about the solutions. Furthermore, its scope is not as wide as that of the theory of regularity structures, see also Remark 5.17 below for more details.

2. Some properties of Hölder spaces

We recall in this section a few standard results from harmonic analysis that are very useful to have in mind. Note first that the linear part of all of the equations described in the introduction is invariant under some space-time scaling. In the case of the heat equation, this is the parabolic scaling. In other words, if u is a solution to the heat equation, then $\tilde{u}(t,x) = u(\lambda^{-2}t, \lambda^{-1}x)$ is also a solution to the heat equation.

This suggests that we should look for solutions in function / distribution spaces respecting this scaling. Given a smooth compactly supported test function φ and a space-time coordinate $z = (t,x)$, we henceforth write $\varphi_z^\lambda(s,y) = \lambda^{-d-2}\varphi\big(\lambda^{-2}(s-t), \lambda^{-1}(y-x)\big)$, where d denotes the spatial dimension and the factor λ^{-d-2} is chosen so that the integral of φ_z^λ is the same as that of φ. In the case of the stochastic Cahn-Hilliard equation (4b), we would naturally use instead a temporal scaling of λ^{-4} and the prefactor would then be λ^{-d-4}.

With these notations at hand, we define spaces of distributions $\mathcal{C}^\alpha$ for $\alpha < 0$ in the following way. Denoting by $\mathcal{B}_\alpha$ the set of smooth test functions $\varphi\colon \mathbf{R}^{d+1} \to \mathbf{R}$ that are supported in the centred ball of radius 1 and such that their derivatives of order up to $1 + |\alpha|$ are uniformly bounded by 1, we set

Definition 2.1. Let η be a distribution on $d+1$-dimensional space-time and let $\alpha < 0$. We say that $\eta \in \mathcal{C}^\alpha$ if the bound $\big|\eta(\varphi_z^\lambda)\big| \lesssim \lambda^\alpha$ holds uniformly over all $\lambda \in (0,1]$, all $\varphi \in \mathcal{B}_\alpha$, and locally uniformly over $z \in \mathbf{R}^{d+1}$.

For $\alpha \geq 0$, we say that a function $f\colon \mathbf{R}^{d+1} \to \mathbf{R}$ belongs to $\mathcal{C}^\alpha$ if, for every $z \in \mathbf{R}^{d+1}$ there exists a polynomial P_z of (parabolic) degree at most α and such that the bound

$$|f(z') - P_z(z')| \lesssim |z - z'|^\alpha \,,$$

holds locally uniformly over z and uniformly over all z' with $|z' - z| \leq 1$. Here, we say that a polynomial P in $z = (t,x)$ is of parabolic degree n if each monomial is of the form z^k with $|k| = 2|k_0| + \sum_{i \neq 0} k_i| \leq n$. In other words, the degree of the time variable "counts double". For $z = (t,x)$, we furthermore write $|z| = |t|^{1/2} + |x|$. (When treating (4b), powers of t count four times and one writes $|z| = |t|^{1/4} + |x|$.)

We now collect a few important properties of the spaces $\mathcal{C}^\alpha$.

2.1. Analytical properties. First, given a function and a distribution (or two distributions) it is natural to ask under what regularity assumptions one can give an unambiguous meaning to their product. It is well-known, at least in the Euclidean case but the extension to the parabolic case is straightforward, that the following result yields a sharp criterion for when, in the absence of any other structural knowledge, one can multiply a function and distribution of prescribed regularity [2, Thm 2.52].

Theorem 2.2. *Let $\alpha, \beta \neq 0$. Then, the map $(f, g) \mapsto f \cdot g$ defined on all pairs of continuous functions extends to a continuous bilinear map from $\mathcal{C}^\alpha \times \mathcal{C}^\beta$ to the space of all distributions if and only if $\alpha + \beta > 0$. Furthermore, if $\alpha + \beta > 0$, the image of the multiplication operator is $\mathcal{C}^{\alpha \wedge \beta}$.*

Another important property of these spaces is given by how they transform under convolution with singular kernels. Let $K \colon \mathbf{R}^{d+1} \to \mathbf{R}$ be a function that is smooth away from the origin and supported in the centred ball of radius 1. One should think of K as being a truncation of the heat kernel $\mathcal{G}$ in the sense that $\mathcal{G} = K + R$ where R is a smooth space-time function. We then say that K is of order β (in the case of a truncation of the heat kernel one has $\beta = 2$) if one can write $K = \sum_{n \geq 0} K_n$ for kernels K_n which are supported in the centred ball of radius 2^{-n} and such that

$$\sup_z |D^k K_n(z)| \lesssim 2^{((d+2)+|k|-\beta)n} \,, \tag{6}$$

for any fixed multiindex k, uniformly in n. Multiplying the heat kernel with a suitable partition of the identity, it is straightforward to verify that this bound is indeed satisfied.

With these notations at hand, one has the following very general Schauder estimate, see for example [41, 42] for special cases.

Theorem 2.3. *Let $\beta > 0$, let K be a kernel of order β, and let $\alpha \in \mathbf{R}$ be such that $\alpha + \beta \notin \mathbf{N}$. Then, the convolution operator $\eta \mapsto K \star \eta$ is continuous from $\mathcal{C}^\alpha$ into $\mathcal{C}^{\alpha+\beta}$.*

Remark 2.4. The condition $\alpha + \beta \notin \mathbf{N}$ seems somewhat artificial. It can actually be dispensed with by slightly changing the definition of $\mathcal{C}^\alpha$.

2.2. Probabilistic properties. Let now η be a random distribution, which we define in general as a continuous linear map $\varphi \mapsto \eta(\varphi)$ from the space of compactly supported smooth test functions into the space of square integrable random variables on some fixed probability space $(\Omega, \mathbf{P})$. We say that it satisfies *equivalence of moments* if, for every $p \geq 1$ there exists a constant C_p such that the bound

$$\mathbf{E}|\eta(\varphi)|^{2p} \leq C_p \big(\mathbf{E}|\eta(\varphi)|^2\big)^p \,,$$

holds for uniformly over all test functions φ. This is of course the case if the random variables $\eta(\varphi)$ are Gaussian, but it also holds if they take values in an inhomogeneous Wiener chaos of fixed order [39].

Given a stationary random distribution η and a (deterministic) distribution C, we say that η has covariance C if $\mathbf{E}\eta(\varphi)\eta(\psi) = \langle C \star \varphi, \psi\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the L^2-scalar product. With this notation at hand, space-time white noise ξ is the Gaussian random distribution on $\mathbf{R}^{d+1}$ with covariance given by the delta distribution. In other words, $\xi(\varphi)$ is centred Gaussian for every φ and $\mathbf{E}\xi(\varphi)\xi(\psi) = \langle\varphi,\psi\rangle_{L^2}$.

Similarly to the case of stochastic processes, a random distribution $\tilde\eta$ is said to be a *version* of η if, for every fixed test function φ, $\tilde\eta(\varphi) = \eta(\varphi)$ almost surely. One then has the following Kolmogorov criterion, a proof of which can be found for example in [23].

Theorem 2.5. *Let η be a stationary random distribution satisfying equivalence of moments and such that, for some $\alpha < 0$, the bound*

$$\mathbf{E}|\eta(\varphi_z^\lambda)|^2 \lesssim \lambda^{2\alpha}\,,$$

holds uniformly over $\lambda \in (0,1]$ and $\varphi \in \mathcal{B}_\alpha$. Then, for any $\kappa > 0$, there exists a $\mathcal{C}^{\alpha-\kappa}$-valued random variable $\tilde\eta$ which is a version of η.

From now on, we will make the usual abuse of terminology and not distinguish between different versions of a random distribution.

Remark 2.6. It follows immediately from the scaling properties of the L^2 norm that one can realise space-time white noise as a random variable in $\mathcal{C}^{-\frac{d}{2}-1-\kappa}$ for every $\kappa > 0$. This is sharp in the sense that it can *not* be realised as a random variable in $\mathcal{C}^{-\frac{d}{2}-1}$. This is akin to the fact that Brownian motion has sample paths belonging to $\mathcal{C}^\alpha$ for every $\alpha < \frac{1}{2}$, but *not* for $\alpha = \frac{1}{2}$.

Let now K be a kernel of order β as before, let ξ be space-time white noise, and set $\eta = K \star \xi$. It then follows from either Theorem 2.5 directly, or from Theorem 2.3 combined with Remark 2.6, that η belongs almost surely to $\mathcal{C}^\alpha$ for every $\alpha < \beta - \frac{d}{2} - 1$. We now turn to the question of how to define powers of η. If $\beta \le \frac{d}{2} + 1$, η is not a random function, so that its powers are in general undefined.

Recall that if ξ is space-time white noise and $L^2(\xi)$ denotes the space of square-integrable random variables that are measurable with respect to the σ-algebra generated by ξ, then $L^2(\xi)$ can be decomposed into a direct sum $L^2(\xi) = \bigoplus_{m\ge 0}\mathcal{H}^m(\xi)$ so that $\mathcal{H}^0$ contains constants, $\mathcal{H}^1$ contains random variables of the form $\xi(\varphi)$ with $\varphi \in L^2$, and $\mathcal{H}^m$ contains suitable generalised Hermite polynomials of order m in the elements of $\mathcal{H}^1$, see [37, 39] for details. Elements of $\mathcal{H}^m$ have a representation by square-integrable kernels of m variables, and this representation is unique if we impose that the kernel is symmetric under permutation of its arguments. In other words, one has a surjection $I^{(m)}\colon L^2(\mathbf{R}^{d+1})^{\otimes m} \to \mathcal{H}^m$ and $I^{(m)}(L) = I^{(m)}(L')$ if and only if the symmetrisations of L and L' coincide.

In the particular case where K is non-singular, η is a random function and its nth power η^n can be represented as

$$\eta^n(\varphi) = \sum_{2m<n} P_{m,n} C^m\, I^{(n-2m)}(K_\varphi^{(n-2m)})\,, \tag{7}$$

where

$$K_\varphi^{(r)}(z_1,\dots,z_r) := \int K(z-z_1)\cdots K(z-z_r)\,\varphi(z)\,dz\;,$$

for some combinatorial factors $P_{m,n}$. Here we have set $C = \int K^2(z)\,dz$. A simple calculation then shows that

Proposition 2.7. *If K is compactly supported, then $K_\varphi^{(n)}$ is square integrable if the function $(K\star\hat K)^n$, where $\hat K(z) = K(-z)$, is integrable.*

We now *define* the nth Wick power $\eta^{\diamond n}$ of η as the random distribution given by only keeping the dominant term in (7):

$$\eta^{\diamond n}(\varphi) = I^{(n)}(K_\varphi^{(n)})\;.$$

By Proposition 2.8, this makes sense as soon as $K\star\hat K \in L^n(\mathbf{R}^{d+1})$. One then has the following result, a version of which can be found for example in [14].

Proposition 2.8. *Let K be a compactly supported kernel of order $\beta \in (\frac{d+2}{2}(1-\frac{1}{n}), \frac{d+2}{2})$ and let $\eta = K\star\xi$ as above. Then, $\eta^{\diamond n}$ is well-defined and belongs almost surely to $\mathcal{C}^\alpha$ for every $\alpha < (2\beta-d-2)\frac{n}{2}$.*

Proof. A simple calculation shows that

$$\big|\big(K\star\hat K\big)(z)\big|^n \lesssim |z|^{(2\beta-d-2)n}\;,$$

so that $\|K_{\varphi_z^\lambda}^{(n)}\|_{L^2}^2 \lesssim \lambda^{(2\beta-d-2)n}$. The claim then follows from Theorem 2.5, noting that random variables belonging to a Wiener-Itô chaos of finite order satisfy the equivalence of moments. □

It is important to note that this result is stable: replacing K by a smoothened kernel K_ε and letting $\varepsilon\to 0$ yields convergence in probability of $\eta_\varepsilon^{\diamond n}$ to $\eta^{\diamond n}$ in $\mathcal{C}^\alpha$ (with α as in the statement of the proposition) for most "reasonable" choices of K_ε. Furthermore, for fixed $\varepsilon > 0$, one has an explicit formula relating $\eta_\varepsilon^{\diamond n}$ to η_ε:

$$\eta_\varepsilon^{\diamond n}(z) = H_n(\eta_\varepsilon(z), C_\varepsilon)\;, \tag{8}$$

where the rescaled Hermite polynomials $H_n(\cdot, C)$ are related to the standard Hermite polynomials by $H_n(u,C) = C^{n/2}H_n(C^{-1/2}u)$ and we have set $C_\varepsilon = \int K_\varepsilon^2(z)\,dz$.

3. General methodology

The general methodology for providing a robust meaning to equations of the type presented in the introduction is as follows. We remark that the main reason why these equations seem to be ill-posed is that there is no canonical way of multiplying arbitrary distributions. The distributions appearing in our setting are however

not arbitrary. For instance, one would expect solutions to semilinear equations of this type to locally "look like" the solutions to the corresponding linear problems. This is because, unlike hyperbolic or dispersive equations, parabolic (or elliptic) equations to not transport singularities. This gives hope that if one could somehow make sense of the nonlinearity, when applied to the solution to the linearised equation (which is a Gaussian process and therefore amenable to explicit calculations), then one could maybe give meaning to the equations themselves.

3.1. The Da Prato-Debussche trick. In some situations, one can apply this idea directly, and this was originally exploited in the series of articles [10, 11, 12]. Let us focus on the example of the dynamical Φ^4 model in dimension 2, which is formally given by

$$\partial_t \Phi = \Delta \Phi + C\Phi - \Phi^3 + \xi \;,$$

where ξ is (spatially periodic) space-time white noise in space dimension 2.

Let now ξ_ε denote a smoothened version of ξ given for example by $\xi_\varepsilon = \rho_\varepsilon \star \xi$, where $\rho_\varepsilon(t,x) = \varepsilon^{-4}\rho(\varepsilon^{-2}t, \varepsilon^{-1}x)$, for some smooth compactly supported space-time mollifier ρ. In this case, denoting again by K a cut-off version of the heat kernel and noting that K is of order 2 (and therefore also of every order less than 2), it is immediate that $\eta = K \star \xi$ satisfies the assumptions of Proposition 2.8 for every integer n.

In view of (8), this suggests that it might be possible to show that the solutions to

$$\begin{aligned} \partial_t \Phi_\varepsilon &= \Delta \Phi_\varepsilon + 3C_\varepsilon \Phi_\varepsilon - \Phi_\varepsilon^3 + \xi_\varepsilon \\ &= \Delta \Phi_\varepsilon - H_3(\Phi_\varepsilon, C_\varepsilon) + \xi_\varepsilon \;, \end{aligned} \tag{9}$$

with $C_\varepsilon = \int K_\varepsilon^2(z)\,dz$ as above, where $K_\varepsilon = \rho_\varepsilon \star K$, converge to a distributional limit as $\varepsilon \to 0$. This is indeed the case, and the argument goes as follows. Writing $\eta_\varepsilon = K_\varepsilon \star \xi$ and $v_\varepsilon = \Phi_\varepsilon - \eta_\varepsilon$ with Φ_ε the solution to (9), we deduce that v_ε solves the equation

$$\partial_t v_\varepsilon = \Delta v_\varepsilon - H_3(\eta_\varepsilon + v_\varepsilon, C_\varepsilon) + R_\varepsilon \;,$$

for some smooth function R_ε that converges to a smooth limit R as $\varepsilon \to 0$. We then use elementary properties of Hermite polynomials to rewrite this as

$$\begin{aligned} \partial_t v_\varepsilon &= \Delta v_\varepsilon - \big(H_3(\eta_\varepsilon, C_\varepsilon) + 3v_\varepsilon H_2(\eta_\varepsilon, C_\varepsilon) + 3v_\varepsilon^2\,\eta_\varepsilon + v_\varepsilon^3\big) + R_\varepsilon \\ &= \Delta v_\varepsilon - \big(\eta_\varepsilon^{\diamond 3} + 3v_\varepsilon \eta_\varepsilon^{\diamond 2} + 3v_\varepsilon^2\,\eta_\varepsilon + v_\varepsilon^3\big) + R_\varepsilon \;. \end{aligned}$$

By Proposition 2.8 (and the remarks that follow), we see that $\eta_\varepsilon^{\diamond n}$ converges in probability to a limit $\eta^{\diamond n}$ in every space $\mathcal{C}^\alpha$ for $\alpha < 0$. We can then *define* a random distribution Φ by $\Phi = \eta + v$, where v is the solution to

$$\partial_t v = \Delta v - \big(\eta^{\diamond 3} + 3v\eta^{\diamond 2} + 3v^2\,\eta + v^3\big) + R \;. \tag{10}$$

As a consequence of Theorem 2.3 (combined with additional estimates showing that the $\mathcal{C}^\gamma$-norm of $K \star (f\mathbf{1}_{t>0})$ is small over short times provided that $f \in \mathcal{C}^\alpha$ for $\alpha \in (-2, 0)$ and $\gamma < \alpha + \beta$), it is relatively easy to show that (10) has local

solutions, and that these solutions are robust with respect to approximations of $\eta^{\diamond n}$ in $\mathcal{C}^\alpha$ for α sufficiently close to 0. In particular, this shows that one has $\Phi_\varepsilon \to \Phi$ in probability, at least locally in time for short times.

Remark 3.1. The dynamical Φ^4 model in dimension 2 was previously constructed in [1] (see also the earlier work [31] where a related but different process was constructed), but that construction relied heavily on *a priori* knowledge about its invariant measure and it was not clear how robust the construction was with respect to perturbations.

3.2. Breakdown of the argument and a strategy to rescue it.

While the argument outlined above works very well for a number of equations, it unfortunately breaks down for the equations mentioned in the introduction. Indeed, consider again (4a), but this time in space dimension $d = 3$. In this case, one has $\eta \in \mathcal{C}^{-\frac{1}{2}-\kappa}$ for every $\kappa > 0$ and, by Proposition 2.8, one can still make sense of $\eta^{\diamond n}$ for $n < 5$. One could therefore hope to define again a solution Φ by setting $\Phi = \eta + v$ with v the solution to (10). Unfortunately, this is doomed to failure: since $\eta^{\diamond 3} \in \mathcal{C}^{-\frac{3}{2}-\kappa}$ (but no better), one can at best hope to have $v \in \mathcal{C}^{\frac{1}{2}-\kappa}$. As a consequence, both products $v \cdot \eta^{\diamond 2}$ and $v^2 \cdot \eta$ fall outside of the scope of Theorem 2.2 and we cannot make sense of (10).

One might hope at this stage that the Da Prato-Debussche trick could be iterated to improve things: identify the "worst" term in the right hand side of (10), make sense of it "by hand", and try to obtain a well-posed equation for the remainder. While this strategy can indeed be fruitful and allows us to deal with slightly more singular problems, it turns out to fail in this situation. Indeed, no matter how many times we iterate this trick, the right hand side of the equation for the remainder v will *always* contain a term proportional to $v \cdot \eta^{\diamond 2}$. As a consequence, one can *never* hope to obtain a remainder of regularity better than $\mathcal{C}^{1-\kappa}$ which, since $\eta^{\diamond 2} \in \mathcal{C}^{-1-\kappa}$, shows that it is not possible to obtain a well-posed equation by this method. See also Remark 5.17 below for a more systematic explanation of when this trick fails.

In some cases, one does not even know how to get started: consider the class of "classical" one-dimensional stochastic PDEs given by

$$\partial_t u = \partial_x^2 u + f(u) + g(u)\xi \,, \tag{11}$$

where ξ denotes space-time white noise, f and g are fixed smooth functions from $\mathbf{R}$ to $\mathbf{R}$, and the spatial variable x takes values on the circle. Then, we know in principle how to use Itô calculus to make sense of (11) by rewriting it as an integral equation and interpreting the integral against ξ as an Itô integral, see [13]. However, this notion of solution is not very robust under approximations since space-time regularisations of the driving noise ξ typically destroy the probabilistic structure required for Itô integration. This is in contrast to the solution theory sketched in Section 3.1 which was very stable under approximations of the driving noise, even though it required suitable adjustments to the equation itself. Unfortunately, the argument of Section 3.1 (try to find some function / distribution η

so that $v = u - \eta$ has better regularity properties and then obtain a well-posed equation for v) appears to break down completely.

The main idea now is that even though we may not be able to find a global object η so that $u - \eta$ has better regularity, it might be possible to find a *local* object that does the trick at any one point. More precisely, setting $\eta = K \star \xi$ as above (this time η is a Hölder continuous function in $\mathcal{C}^{\frac{1}{2}-\kappa}$ for every $\kappa > 0$ by Theorems 2.3 and 2.5), one would expect solutions to (11) to be well approximated by

$$u(z') \approx u(z) + g(u(z))\big(\eta(z') - \eta(z)\big) \,. \tag{12}$$

The intuition is that since K is regular everywhere except at the origin, convolution with K is "almost" a local operator, modulo more regular parts. Since, near any fixed point z, we would expect $g(u)\xi$ to "look like" $g(u(z))\xi$ this suggests that near that point z, the function $K \star (g(u)\xi)$ should "look like" $g(u(z))\eta$, which is what (12) formalises.

Note that this looks very much like a first-order Taylor expansion, but with $\eta(z') - \eta(z)$ playing the role of the linear part $z' - z$. If we assume that (12) yields a good approximation to u, then one would also expect that

$$g(u(z')) \approx g(u(z)) + g'(u(z))g(u(z))\big(\eta(z') - \eta(z)\big) \,,$$

so that $g(u)$ has again a "first-order Taylor expansion" of the same type as the one for u. One could then hope that if we know somehow how to multiply η with ξ, this knowledge could be leveraged to define the product between $g(u)$ and ξ in a robust way. It turns out that this is *not* quite enough for the situation considered here. However, this general strategy turns out to be very fruitful, provided that we also control higher-order local expansions of u, and this is precisely what the theory of regularity structures formalises [23, 26]. In particular, besides being applicable to (11), it also applies to all of the equations mentioned in the introduction.

4. Regularity structures

We now describe a very general framework in which one can formulate "Taylor expansions" of the type (12). We would like to formalise the following features of Taylor expansions. First, the coefficients of a Taylor expansion (i.e. the value and derivatives of a given function in the classical case or the coefficients $u(z)$ and $g(u(z))$ in the case (12)) correspond to terms of different degree / homogeneity and should therefore naturally be thought of as elements in some graded vector space. Second, an expansion around a given point can be reexpanded around a different point at the expense of changing coefficients, like so:

$$a \cdot 1 + b \cdot x + c \cdot x^2 = \big(a + bh + ch^2\big) \cdot 1 + \big(b + 2ch\big) \cdot (x - h) + c \cdot (x - h)^2 \,,$$
$$u \cdot 1 + g(u) \cdot \big(\eta(z') - \eta(z)\big) = \big(u + g(u)(\eta(z'') - \eta(z))\big) \cdot 1 + g(u) \cdot \big(\eta(z') - \eta(z'')\big) \,.$$

Lastly, we see from these expressions that if we order coefficients by increasing homogeneity, then the linear transformation performing the reexpansion has an upper triangular structure with the identity on the diagonal.

4.1. Basic definitions. The properties just discussed are reflected in the following algebraic structure.

Definition 4.1. A *regularity structure* $\mathscr{T} = (A, T, G)$ consists of the following elements:

1. A discrete index set $A \subset \mathbf{R}$ such that $0 \in A$ and A is bounded from below.
2. A *model space* $T = \bigoplus_{\alpha \in A} T_\alpha$, with each T_α a Banach space; elements in T_α are said to have *homogeneity* α. Furthermore T_0 is one-dimensional and has a distinguished basis vector $\mathbf{1}$. Given $\tau \in T$, we write $\|\tau\|_\alpha$ for the norm of its component in T_α.
3. A *structure group* G of (continuous) linear operators acting on T such that, for every $\Gamma \in G$, every $\alpha \in A$, and every $\tau_\alpha \in T_\alpha$, one has
$$\Gamma \tau_\alpha - \tau_\alpha \in T_{<\alpha} := \bigoplus_{\beta < \alpha} T_\beta \, . \tag{13}$$
Furthermore, $\Gamma \mathbf{1} = \mathbf{1}$ for every $\Gamma \in G$.

The prime example of a regularity structure one should keep in mind is the one associated to Taylor polynomials on space-time $\mathbf{R}^{d+1}$. In this case, the space T is given by all polynomials in $d+1$ indeterminates $X_0, \dots, X_d$, with X_0 representing the "time" coordinate. It comes with a canonical basis given by all monomials of the type $X^k = X_0^{k_0} \cdots X_d^{k_d}$ with k an arbitrary multiindex. The basis vector $\mathbf{1}$ is the one corresponding to the zero multiindex. The space T has a natural grading by postulating that the homogeneity of X^k is $|k| = 2k_0 + \sum_{i \neq 0} k_i$ and a natural norm by postulating that $\|X^k\| = 1$. In the case of the polynomial regularity structure, the structure group G is simply given by $\mathbf{R}^{d+1}$, endowed with addition, and acting on monomials by

$$\hat{\Gamma}_h X^k = (X - h)^k = (X_0 - h_0)^{k_0} \cdots (X_d - h_d)^{k_d} \, . \tag{14}$$

It is immediate that all axioms of a regularity structure are satisfied in this case.

In the case of polynomials, there is a natural "realisation" of the structure $\mathscr{T}$ at each space-time point z, which is obtained by turning an abstract polynomial into the corresponding concrete polynomial (viewed now as a real-valued function on $\mathbf{R}^{d+1}$) based at z. In other words, we naturally have a family of linear maps $\Pi_z \colon T \to \mathcal{C}^\infty(\mathbf{R}^d)$ given by

$$\big(\Pi_z X^k\big)(z') = (z'_0 - z_0)^{k_0} \cdots (z'_d - z_d)^{k_d} \, . \tag{15}$$

It is immediate that the group G transforms these maps into each other in the sense that $\Pi_z \hat{\Gamma}_h = \Pi_{z+h}$. It is furthermore an immediate consequence of the scaling properties of monomials that the maps Π_z and the representation $h \mapsto \hat{\Gamma}_h$ of $\mathbf{R}^{d+1}$ are "compatible" with our grading for the model space T. More precisely, one has

$$\langle \varphi_z^\lambda, \Pi_z X^k \rangle = \lambda^{|k|} \langle \varphi, \Pi_0 X^k \rangle \, , \qquad \|\hat{\Gamma}_h X^k\|_\ell = C_{k,\ell} |h|^{|k| - \ell} \, ,$$

for some constants $C_{k,\ell}$ and every $\ell \le |k|$. Here, $\langle \cdot, \cdot \rangle$ denotes again the usual L^2-scalar product.

These observations suggest the following definition of a "model" for $\mathscr{T}$, where we impose properties similar to the ones we just found for the polynomial model. A model always requires the specification of an ambient space, together with a possibly inhomogeneous scaling. For definiteness, we will fix our ambient space to be $\mathbf{R}^{d+1}$ endowed with the parabolic scaling as above. We also denote by $\mathcal{S}'$ the space of all distributions (the letter $\mathcal{D}$ is reserved for a different usage below). We also denote by $L(E,F)$ the set of all continuous linear maps between the topological vector spaces E and F.

Definition 4.2. Given a regularity structure $\mathscr{T}$, a *model* for $\mathscr{T}$ consists of maps

$$\mathbf{R}^{d+1} \ni z \mapsto \Pi_z \in L(T, \mathcal{S}') \;, \qquad \mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \ni (z, z') \mapsto \Gamma_{zz'} \in G \;,$$

satisfying the algebraic compatibility conditions

$$\Pi_z \Gamma_{zz'} = \Pi_{z'} \;, \qquad \Gamma_{zz'} \circ \Gamma_{z'z''} = \Gamma_{zz''} \;, \tag{16}$$

as well as the analytical bounds

$$|\langle \Pi_z \tau, \varphi_z^\lambda \rangle| \lesssim \lambda^\alpha \|\tau\| \;, \qquad \|\Gamma_{zz'}\tau\|_\beta \lesssim |z - z'|^{\alpha-\beta} \|\tau\| \;. \tag{17}$$

Here, the bounds are imposed uniformly over all $\tau \in T_\alpha$, all $\beta < \alpha \in A$, and all test functions $\varphi \in \mathcal{B}_r$ with $r = \inf A$, and locally uniformly in z and z'.

Remark 4.3. These definitions suggest a natural topology for the space $\mathscr{M}$ of all models for a given regularity structure, generated by the following family of pseudo-metrics indexed by compact sets K:

$$\sup_{z \in K} \Big(\sup_{\varphi, \lambda, \alpha, \tau} \lambda^{-\alpha} |\langle \Pi_z \tau - \bar{\Pi}_z \tau, \varphi_z^\lambda \rangle| + \sup_{|z-z'| \le 1} \sup_{\alpha, \beta, \tau} |z-z'|^{\beta-\alpha} \|\Gamma_{zz'}\tau - \bar{\Gamma}_{zz'}\tau\|_\beta \Big) \;. \tag{18}$$

Here the inner suprema run over the same sets as before, but with $\|\tau\| = 1$.

4.2. Hölder classes.

It is clear from the above discussion that if $\mathscr{T}$ is the polynomial structure, Π is defined as in (25), and $\Gamma_{zz'} = \hat{\Gamma}_{z'-z}$ with $\hat{\Gamma}_h$ as in (14), then (Π, Γ) is a model for $\mathscr{T}$ in the sense of Definition 4.2. Given an arbitrary regularity structure $\mathscr{T}$ and an arbitrary model (Π, Γ), it is now natural to define the corresponding "Hölder spaces" as spaces of distributions that can locally (near any space-time point z) be approximated by $\Pi_z \tau$ for some $\tau \in T$. This would be the analogue to the statement that a smooth function is one that can locally be approximated by a polynomial.

There is however one major difference with the case of smooth functions. It is of course the case that if f is smooth, then the coefficients of the Taylor expansion of f at any point are uniquely determined by the behaviour of f in the vicinity of that point. This is in general *not* the case anymore in the context of the framework we just described. To appreciate this fact, consider the following example. Fix

$\alpha \in (0,1)$ and $m \in \mathbf{N}$, and take for $\mathscr{T}$ the regularity structure where $A = \{0, \alpha\}$, $T_0 \cong \mathbf{R}$ with basis vector $\mathbf{1}$, $T_\alpha \cong \mathbf{R}^m$ with basis vectors $(e_i)_{i \le m}$, and structure group $G \cong \mathbf{R}^m$ acting on T via $\hat{\Gamma}_h e_i = e_i - h_i \mathbf{1}$. Let then W be an $\mathbf{R}^m$-valued α-Hölder continuous function defined on the ambient space and set

$$\Pi_z \mathbf{1} = 1 \,, \qquad \big(\Pi_z e_i\big)(z') = W_i(z') - W_i(z) \,, \qquad \Gamma_{zz'} = \hat{\Gamma}_{W(z) - W(z')} \,.$$

Again, it is straightforward to verify that this does indeed define a model for $\mathscr{T}$. In fact, setting $m = 1$ and $W = \eta$, this is precisely the structure one would use to formalise the expansion (12).

Let now $F \colon \mathbf{R}^m \to \mathbf{R}$ be a smooth function and consider the function f on the ambient space given by $f(z) = F(W(z))$. For any z, we furthermore set

$$T \ni \hat{f}(z) = F(W(z))\, \mathbf{1} + \sum_{i=1}^m (\partial_i F)(W(z))\, e_i \,.$$

It then follows immediately from the usual Taylor expansion of F and the definition of the model (Π, Γ) that one has the bound

$$\big| f(z') - \big(\Pi_z \hat{f}(z)\big)(z') \big| \lesssim |z - z'|^{2\alpha} \,, \tag{19}$$

so that in this context and with respect to this specific model, the function f behaves as if it were of class $\mathcal{C}^{2\alpha}$ with "Taylor series" given by $\hat{f}$. In the case where the underlying space is one-dimensional, this is precisely the insight exploited in the theory of rough paths [35, 36, 16] in order to develop a pathwise approach to stochastic calculus. More specifically, the perspective given here (i.e. controlling functions via analogues to Taylor expansion) is that of the theory of controlled rough paths developed in [18].

It is now very natural to ask whether, just like in the case of smooth functions, a bound of the type (19) is sufficient to uniquely specify $\hat{f}(z)$ for every point z. Unfortunately, the answer to this question is that "it depends". The reason is that while (17) imposes an upper bound on the behaviour of Π_z in the vicinity of z, it does *not* impose any corresponding lower bound. For example, $W \equiv 0$ is an α-Hölder continuous function that we could have used to build our model. In that case, the value of the e_i-component in $\hat{f}$ is completely irrelevant for (19), so that uniqueness of the "Taylor series" fails. Suppose on the other hand that the underlying space is one-dimensional, that $\alpha \in (\frac{1}{4}, \frac{1}{2})$, and that W is a typical sample path of a Brownian trajectory. In this case it was shown in [27, Thm 3.4] that a bound of the type (19) is indeed sufficient to uniquely determine all the coefficients of $\hat{f}$ (at least for almost all Brownian trajectories).

Remark 4.4. The fact that $\hat{f}$ is uniquely determined by f in the Brownian case can be interpreted as an analogue to the fact that the Doob-Meyer decomposition of a semimartingale is unique. Since the statement given in [27] is quantitative, it can be interpreted as a deterministic analogue to Norris's lemma, of which various incarnations can be found in [6, 33, 38].

Consider now a sequence W^ε of smooth (random) functions so that W^ε converges to Brownian motion in $\mathcal{C}^\alpha$ as $\varepsilon \to 0$. For definiteness, take for W_ε piecewise linear interpolations on a grid of size ε. Then, if we know *a priori* that we have a bound of the type (19) with a proportionality constant of order 1, this determines the coefficients of $\hat{f}$ "almost uniquely" up to an error of order about $\varepsilon^{2\alpha-\frac{1}{2}}$.

What this discussion suggests is that we should really reverse our point of view from what we are used to: instead of fixing a function and asking whether it has a certain Hölder regularity by checking whether it is possible to find a "Taylor expansion" at each point satisfying a bound of the type (19), we should take the candidate expansion as our fundamental object and ask under which condition it does indeed approximate one single function / distribution around each point at the prescribed order. More precisely, fix some $\gamma > 0$ (the order of our "Taylor expansion") and consider a function $f\colon \mathbf{R}^{d+1} \to T_{<\gamma}$. Under which assumptions can we find a distribution ζ such that ζ "looks like" the distribution $\Pi_z f(z)$ (in a suitable sense) near every point z? We claim that the "right" answer is given by the following definition.

Definition 4.5. Given a regularity structure $\mathscr{T}$ and a model (Π, Γ) as above, we define $\mathcal{D}^\gamma$ as the space of functions $f\colon \mathbf{R}^{d+1} \to T_{<\gamma}$ such that the bound

$$\|f(z) - \Gamma_{zz'} f(z')\|_\alpha \lesssim |z - z'|^{\gamma-\alpha} \,. \tag{20}$$

holds for every $\alpha < \gamma$, locally uniformly in z and z'.

Remark 4.6. This definition makes sense and is non-empty even for negative γ, as long as $\gamma > \inf A$.

Remark 4.7. The notation $\mathcal{D}^\gamma$ is really an abuse of notation, since even for a given regularity structure there isn't one single space $\mathcal{D}^\gamma$, but a whole collection of them, one for each model $(\Pi, \Gamma) \in \mathscr{M}$. More formally, one should really consider the space $\mathscr{M} \ltimes \mathcal{D}^\gamma$ consisting of pairs $((\Pi, \Gamma), f)$ such that f belongs to the space $\mathcal{D}^\gamma$ based on the model (Π, Γ). The space $\mathscr{M} \ltimes \mathcal{D}^\gamma$ also comes with a natural topology.

In the case where $\mathscr{T}$ is the polynomial regularity structure and (Π, Γ) are the usual Taylor polynomials as above, one can see that this definition coincides with the usual definition of $\mathcal{C}^\gamma$ (except at integer values where $\mathcal{D}^1$ describes Lipschitz continuous functions, etc). In this case, the component $f_0(z) = \langle \mathbf{1}, f(z) \rangle$ of $f(z)$ in T_0 (here we write $\langle \mathbf{1}, \cdot \rangle$ for the basis element of T^* dual to $\mathbf{1}$) is the only reasonable candidate for the function represented by f. Furthermore, $\langle \mathbf{1}, \Gamma_{zz'} f(z') \rangle$ is nothing but the candidate Taylor expansion of f around z', evaluated at z. The bound (20) with $\alpha = 0$ is then just a statement of the fact that f_0 is of class $\mathcal{C}^\gamma$ and that $f(z)$ is its Taylor series of order γ at z. The corresponding bounds for $\alpha > 0$ then follow immediately, since they merely state that the αth derivative of f_0 is of class $\mathcal{C}^{\gamma-\alpha}$.

4.3. The reconstruction operator.

The situation is much less straightforward when the model space T contains components of negative homogeneity. In

this case, the bounds (17) allow the model Π_z to consist of genuine distributions and we do not anymore have an obvious candidate for the distribution represented by f. The following result shows that such a distribution nevertheless always exists and is unique as soon as $\gamma > 0$. This also provides an *a posteriori* justification for our definition of the spaces $\mathcal{D}^\gamma$.

Theorem 4.8. *Consider a regularity structure $\mathscr{T} = (A, T, G)$ and fix $\gamma > r = \inf A$. Then, there exists a continuous map $\mathcal{R}\colon \mathscr{M} \ltimes \mathcal{D}^\gamma \to \mathcal{S}'$ (the "reconstruction map") with the property that*

$$\big|\big(\mathcal{R}(\Pi,\Gamma,f) - \Pi_z f(z)\big)(\varphi_z^\lambda)\big| \lesssim \lambda^\gamma \,, \tag{21}$$

uniformly over $\lambda \in (0,1]$ and $\varphi \in \mathcal{B}_r$, and locally uniformly over $z \in \mathbf{R}^{d+1}$. Furthermore, for any given model (Π,Γ), the map $f \mapsto \mathcal{R}(\Pi,\Gamma,f)$ is linear. If $\gamma > 0$, the map $\mathcal{R}$ is uniquely specified by the requirement (21).

Remark 4.9. In the sequel, we will always consider (Π,Γ) as fixed and view $\mathcal{R}$ as a linear map, writing $\mathcal{R}f$ instead of $\mathcal{R}(\Pi,\Gamma,f)$. The above notation does however make it plain that the full map $\mathcal{R}$ is *not* a linear map.

Remark 4.10. An important special case is given by situations where $\Pi_z\tau$ happens to be a continuous function for every $\tau \in T$ and every z. Then, it turns out that $\mathcal{R}f$ is also a continuous function and one simply has

$$\big(\mathcal{R}f\big)(z) = \big(\Pi_z f(z)\big)(z) \,. \tag{22}$$

In the general case, this formula makes of course no sense since $\Pi_z f(z)$ is a distribution and cannot be evaluated at z.

Remark 4.11. We made a slight abuse of notation here since there is really a family of operators $\mathcal{R}^\gamma$, one for each regularity. However, this abuse is justified by the following consistency relation. Given $f \in \mathcal{D}^\gamma$ and $\tilde\gamma < \gamma$, one can always construct $\tilde f$ by projecting $f(z)$ onto $T_{<\tilde\gamma}$ for every z. It turns out that one then necessarily has $\tilde f \in \mathcal{D}^{\tilde\gamma}$ and $\mathcal{R}\tilde f = \mathcal{R}f$, provided that $\tilde\gamma > 0$. This is also consistent with (22) since, if $\Pi_z\tau$ is a continuous function and the homogeneity of τ is strictly positive, then $\big(\Pi_z\tau\big)(z) = 0$.

We refer to [23, Thm 3.10] for a full proof of Theorem 4.8 and to [22] for a simplified proof that only gives continuity in each "fiber" $\mathcal{D}^\gamma$. The main idea is to use a basis of compactly supported wavelets to construct approximations $\mathcal{R}^n$ in such a way that our definitions can be exploited in a natural way to compare $\mathcal{R}^{n+1}$ with $\mathcal{R}^n$ and show that the sequence of approximations is Cauchy in a suitable space of distributions $\mathcal{C}^\alpha$. In the most important case when $\gamma > 0$, it turns out that while the existence of a map $\mathcal{R}$ with the required properties is highly non-trivial, its uniqueness is actually quite easy to see. If $\gamma \le 0$ on the other hand, it is clear that $\mathcal{R}$ cannot be uniquely determined by (21), since this bound remains unchanged if we add to $\mathcal{R}$ any distribution in $\mathcal{C}^\gamma$. The existence of $\mathcal{R}$ in the case $\gamma < 0$ is however still a non-trivial result since in general one has $\mathcal{R}f \notin \mathcal{C}^\gamma$!

5. Regularity structures for SPDEs

We now return to the problem of providing a robust well-posedness theory for stochastic PDEs of the type (2), (4), (3), or even just (11). Our aim is to build a suitable regularity structure for which we can reformulate our SPDE as a fixed point problem in $\mathcal{D}^\gamma$ for a suitable value of γ.

Remark 5.1. Actually, it turns out that since we are interested in Cauchy problems, there will always be some singularity at $t = 0$. This introduces additional technical complications which we do not wish to dwell upon.

5.1. General construction of the model space. Our first task is to construct the model space T. Since we certainly want to be able to represent arbitrary smooth functions (for example in order to be able to take into account the contribution of the initial condition), we want T to contain the space $\bar{T}$ of abstract polynomials in $d+1$ indeterminates endowed with the parabolic grading described in Section 4.1. Since the noise ξ cannot be adequately represented by polynomials, we furthermore add a basis vector Ξ to T, which we postulate to have some homogeneity $\alpha < 0$ such that $\xi \in \mathcal{C}^\alpha$. In the case of space-time white noise, we would choose $\alpha = -\frac{d}{2} - 1 - \kappa$ for some (typically very small) exponent $\kappa > 0$.

At this stage, the discussion following (12) suggests that if our structure T contains a basis vector τ of homogeneity β representing some distribution η involved in the description of the right hand side of our equation, then it should also contain a basis vector of homogeneity $\beta + 2$ (the "2" here comes from the fact that convolution with the heat kernel yields a gain of 2 in regularity) representing the distribution $K \star \eta$ involved in the description of the solution to the equation. Let us denote this new basis vector by $\mathcal{I}(\tau)$, where $\mathcal{I}$ stands for "integration". In the special case where $\tau \in \bar{T}$, so that it represents an actual polynomial, we do not need any new symbol since K convolved with a polynomial yields a smooth function. One way of formalising this is to simply postulate that $\mathcal{I}(X^k) = 0$ for every multiindex k.

Remark 5.2. For consistency, we will also always assume that $\int K(z)Q(z)\,dz = 0$ for all polynomials Q of some fixed, but sufficiently high, degree. Since K is an essentially arbitrary truncation of the heat kernel, we can do this without loss of generality.

If the right hand side of our equation involves the spatial derivatives of the solution, then, for each basis vector τ of homogeneity β representing some distribution η appearing in the description of the solution, we should also have a basis vector $\mathscr{D}_i\tau$ of homogeneity $\beta - 1$ representing $\partial_i\eta$ and appearing in the description of the derivative of the solution in the direction x_i.

Finally, if the right hand side of our equation involves a product between two terms F and $\bar{F}$, and if basis vectors τ and $\bar{\tau}$ respectively are involved in their description, then we should also have a basis vector $\tau\bar{\tau}$ which would be involved in the description of the product. If τ and $\bar{\tau}$ represent the distributions η and $\bar{\eta}$ respectively, then this new basis vector represents the distribution $\eta\bar{\eta}$, what-

ever this actually means. Regarding its homogeneity, by analogy with the case of polynomials, it is natural to impose that the homogeneity of $\tau\bar{\tau}$ is the sum of the homogeneities of its two factors.

This suggests that we should build T by taking as its basis vectors some formal expressions built from the symbols X and Ξ, together with the operations $\mathcal{I}(\cdot)$, $\mathscr{D}_i$, and multiplication. Furthermore, the natural way of computing the homogeneity of a formal expression in view of the above is to associate homogeneity 2 to X_0, 1 to X_i for $i \neq 0$, α to Ξ, 2 to $\mathcal{I}(\cdot)$, and -1 to $\mathscr{D}_i$, and to simply add the homogeneities of all symbols appearing in any given expression. Denote by $\mathcal{F}$ the collection of all formal expressions that can be constructed in this way and denote by $|\tau|$ the homogeneity of $\tau \in \mathcal{F}$, so we have for example

$$\big|X_i\Xi\big| = \alpha + 1\,, \qquad \big|\mathcal{I}(\Xi)^2\mathcal{I}(X_i\mathscr{D}_j\mathcal{I}(\Xi))\big| = 3\alpha + 8\,, \qquad \text{etc.}$$

We note however that if we simply took for T the space of linear combinations of *all* elements in $\mathcal{F}$ then, since $\alpha < 0$, there would be basis vectors of arbitrarily negative homogeneity, which would go against Definition 4.1. What saves us is that most formal expressions are not needed in order to formulate our equations as fixed point problems. For example, the expression Ξ^2 is useless since we would never try to square the driving noise. Similarly, if we consider (4a), then $\mathcal{I}(\Xi)$ is needed for the description of the solution, which implies that $\mathcal{I}(\Xi)^2$ and $\mathcal{I}(\Xi)^3$ are needed to describe the right hand side, but we do not need $\mathcal{I}(\Xi)^4$ for example.

5.2. Specific model spaces. This suggests that we should take T as the linear combinations of only those formal expressions $\tau \in \mathcal{F}$ that are actually expected to appear in the description of the solution to our equation or its right hand side. Instead of trying to formulate a general construction (see [23, Sec. 8.1] for such an attempt), let us illustrate this by a few examples. We first focus on the case of the KPZ equation (3) and we construct subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{F}$ that are used in the description of the solution and the right hand side of the equation respectively. These are defined as the smallest subsets of $\mathcal{F}$ with the following properties:

$$\mathcal{T} \subset \mathcal{U} \cap \mathcal{V}\,, \quad \{\mathcal{I}(\tau) : \tau \in \mathcal{V} \setminus \mathcal{T}\} \subset \mathcal{U}\,, \quad \{\Xi\} \cup \{\mathscr{D}\tau_1 \cdot \mathscr{D}\tau_2 : \tau_i \in \mathcal{U}\} \subset \mathcal{V}\,. \tag{23}$$

where we used the notation $\mathcal{T} = \{X^k\}$ with k running over all multiindices, so that the space of Taylor polynomials $\bar{T}$ is the linear span of $\mathcal{T}$. We then define T as the space of all linear combinations of elements of $\mathcal{U} \cup \mathcal{V}$. We also denote by $T_{\mathcal{U}}$ the subspace of T spanned by $\mathcal{U}$. This construction is such that if we have any function $H\colon \mathbf{R}^{d+1} \to T_{\mathcal{U}}$, then we can define in a natural way a function $\Xi - (\mathscr{D}H)^2\colon \mathbf{R}^{d+1} \to T$ by the last property. Furthermore, by the second property, one has again $\mathcal{I}(\Xi - (\mathscr{D}H)^2)\colon \mathbf{R}^{d+1} \to T_{\mathcal{U}}$, which suggests that T is indeed sufficiently rich to formulate a fixed point problem mimicking the mild formulation of (3). Furthermore, one has

Lemma 5.3. *If $\mathcal{U}$ and $\mathcal{V}$ are the smallest subsets of $\mathcal{F}$ satisfying* (23) *and one has $|\Xi| > -2$ then, for every $\gamma > 0$, the set $\{\tau \in \mathcal{U} \cup \mathcal{V} : |\tau| < \gamma\}$ is finite.*

The condition $\alpha > -2$ corresponds to the restriction $d < 2$, which makes sense since 2 is the critical dimension for the KPZ equation [32]. The other example we would like to consider is the class of SPDEs (11). In this case, the right hand side is not polynomial. However, we can apply the same methodology as above as if the nonlinear functions f and g were simply polynomials of arbitrary degree. We thus impose $\mathcal{T} \subset \mathcal{U} \cap \mathcal{V}$ and $\{\mathcal{I}(\tau) : \tau \in \mathcal{V} \setminus \mathcal{T}\}$ as before, and then further impose that

$$\Big\{\Xi \prod_{i=1}^m \tau_i : m \geq 1 \,\&\, \tau_i \in \mathcal{U}\Big\} \cup \Big\{\prod_{i=1}^m \tau_i : m \geq 1 \,\&\, \tau_i \in \mathcal{U}\Big\} \subset \mathcal{V} \,.$$

Again, we have $\mathcal{U} \subset \mathcal{V}$ and we define T as before. Furthermore, it is straightforward to verify that the analogue to Lemma 5.3 holds, provided that $|\Xi| > -2$.

5.3. Construction of the structure group. Now that we have some idea on how to construct T for the problems that are of interest to us (with a slightly different construction for each class of models but a clear common thread), we would like to build a corresponding structure group G. In order to give a motivation for the definition of G, it is very instructive to simultaneously think about the structure of the corresponding models. Let us first consider some smooth driving noise, which we call ξ_ε to distinguish it from the limiting noise ξ. At this stage however, this should be thought of as simply a fixed smooth function. In view of the discussion of Section 5.1, for each of the model spaces built in Section 5.2, we can associate to ξ_ε a linear map $\mathbf{\Pi} \colon T \to \mathcal{C}^\infty(\mathbf{R}^{d+1})$ in the following way. We set

$$\big(\mathbf{\Pi} X_i\big)(z) = z_i \,, \qquad \big(\mathbf{\Pi}\Xi\big)(z) = \xi_\varepsilon(z) \,, \tag{24a}$$

and we then define $\mathbf{\Pi}$ recursively by

$$\mathbf{\Pi}\mathcal{I}(\tau) = K \star \mathbf{\Pi}\tau \,, \qquad \mathbf{\Pi}\mathscr{D}_i\tau = \partial_i \mathbf{\Pi}\tau \,, \qquad \mathbf{\Pi}(\tau\bar\tau) = \big(\mathbf{\Pi}\tau\big) \cdot \big(\mathbf{\Pi}\bar\tau\big) \,, \tag{24b}$$

where $\cdot$ simply denotes the pointwise product between smooth functions. At this stage, it is however not clear how one would build an actual model in the sense of Definition 4.2 associated to ξ_ε. It is natural that one would set

$$\big(\Pi_z X_i\big)(z') = z'_i - z_i \,, \qquad \big(\Pi_z \Xi\big)(z') = \xi_\varepsilon(z') \,, \tag{25a}$$

and then

$$\Pi_z \mathscr{D}_i\tau = \partial_i \Pi_z\tau \,, \qquad \Pi_z(\tau\bar\tau) = \big(\Pi_z\tau\big) \cdot \big(\Pi_z\bar\tau\big) \,. \tag{25b}$$

It is less clear *a priori* how to define $\Pi_z\mathcal{I}(\tau)$. The problem is that if we simply set $\Pi_z\mathcal{I}(\tau) = K \star \Pi_z\tau$, then the bound (17) would typically no longer be compatible with the requirement that $|\mathcal{I}(\tau)| = |\tau| + 2$. One way to circumvent this problem is to simply subtract the Taylor expansion of $K \star \Pi_z\tau$ around z up to the required order. We therefore set

$$\big(\Pi_z\mathcal{I}(\tau)\big)(z') = \big(K \star \Pi_z\tau\big)(z') - \sum_{|k|<|\tau|+2} \frac{(z'-z)^k}{k!} \big(D^{(k)} K \star \Pi_z\tau\big)(z) \,. \tag{25c}$$

It can easily be verified (simply proceed recursively) that if we define Π_z in this way and $\mathbf{\Pi}$ as in (24) then, for every z, one can find a linear map $F_z\colon T \to T$ such that $\Pi_z = \mathbf{\Pi} F_z$. In particular, one has $\Pi_{z'} = \Pi_z F_z^{-1} F_{z'}$. Furthermore, F_z is "upper triangular" with the identity on the diagonal in the sense of (13). It is also easily seen by induction that the matrix elements of F_z are all given by some polynomials in z and in the quantities $\big(D^{(k)}K \star \Pi_z \tau\big)(z)$.

This suggests that we should take for G the set of all linear maps that can appear in this fashion. It is however not clear in principle how to describe G more explicitly and it is also not clear that it even forms a group. In order to describe G, it is natural to introduce a space T_+ which is given by all possible polynomials in $d+1$ commuting variables $\{Z_i\}_{i=0}^d$ as well as countably many additional commuting variables $\{\mathcal{J}_k(\tau) : \tau \in (\mathcal{U} \cup \mathcal{V}) \setminus \mathcal{T} \;\&\; |k| < |\tau| + 2\}$. One should think of Z_i as representing z_i and $\mathcal{J}_k(\tau)$ as representing $\big(D^{(k)}K \star \Pi_z \tau\big)(z)$, so that the matrix elements of F_z are represented by elements of T_+. There are no relations between these coefficients, which suggests that elements of G are described by an arbitrary morphism $f\colon T_+ \to \mathbf{R}$, i.e. an arbitrary linear map which furthermore satisfies $f(\sigma\bar\sigma) = f(\sigma)\, f(\bar\sigma)$, so that it is uniquely determined by $f(Z_i)$ and $f(\mathcal{J}_k(\tau))$.

Given any linear map $\Delta\colon T \to T \otimes T_+$ and a morphism f as above, one can then define a linear map $\hat\Gamma_f\colon T \to T$ by

$$\hat\Gamma_f \tau = \big(I \otimes f\big)\Delta\tau\;.$$

(Here we identify T with $T \otimes \mathbf{R}$ in the obvious way.) The discussion given above then suggests that it is possible to construct Δ in such a way that if we define f_z by

$$f_z(Z_i) = z_i\;, \qquad f_z(\mathcal{J}_k(\tau)) = \big(D^{(k)}K \star \Pi_z\tau\big)(z)\;, \tag{26}$$

then one has $\hat\Gamma_{f_z} = F_z$. The precise definition of Δ is irrelevant for our discussion, but a recursive description of it can easily be recovered simply by comparing (25) to (24). In particular, it is possible to show that $\Delta\tau$ is of the form

$$\Delta\tau = \tau \otimes \mathbf{1} + \sum_i c_i^\tau \tau_i \otimes \sigma_i\;, \tag{27}$$

for some expressions $\tau_i \in T$ with $|\tau_i| < |\tau|$ and for some non-empty monomials $\sigma_i \in T_+$ such that $|\sigma_i| + |\tau|_i = |\tau|$. Here, we associate a homogeneity to elements in T_+ by setting $|Z_0| = 2$, $|Z_i| = 1$ for $i \neq 0$, and $|\mathcal{J}_k(\tau)| = |\tau| + 2 - |k|$.

In particular, we see that if we let $e\colon T_+ \to \mathbf{R}$ be the trivial morphism for which $e(Z_i) = e(\mathcal{J}_k(\tau)) = 0$, so that one only has $e(\mathbf{1}) = 1$ where $\mathbf{1}$ is the empty product, then $\hat\Gamma_e \tau = \tau$. The important fact for our purpose is the following, a proof of which can be found in [23, Sec. 8]. Here, we denote by $\mathcal{M}\colon T_+ \otimes T_+ \to T_+$ the multiplication operator $\mathcal{M}(\sigma \otimes \bar\sigma) = \sigma\bar\sigma$ and by I the identity.

Theorem 5.4. *There exists a map $\Delta^+\colon T_+ \to T_+ \otimes T_+$ such that the following identities hold:*

$$\begin{aligned} \Delta^+(\sigma\bar\sigma) &= \big(\Delta^+\sigma\big)\cdot\big(\Delta^+\bar\sigma\big)\;, \qquad (\Delta\otimes I)\Delta = (I\otimes\Delta^+)\Delta\;,\\ (e\otimes I)\Delta^+ &= (I\otimes e)\Delta^+ = I\;, \quad (\Delta^+\otimes I)\Delta^+ = (I\otimes\Delta^+)\Delta^+\;. \end{aligned} \tag{28}$$

Furthermore, there exists a map $\mathcal{A}\colon T_+ \to T_+$ *which is multiplicative in the sense that* $\mathcal{A}(\sigma\bar\sigma) = (\mathcal{A}\sigma)\cdot(\mathcal{A}\bar\sigma)$*, and which is such that* $\mathcal{M}(I\otimes\mathcal{A})\Delta^+ = \mathcal{M}(\mathcal{A}\otimes I)\Delta^+ = e$*, with* $e\colon T_+ \to \mathbf{R}$ *as above.*

Remark 5.5. In technical lingo, this lemma states that $(T_+, \cdot, \Delta^+)$ is a Hopf algebra with antipode $\mathcal{A}$, and that T is a comodule over T_+.

The importance of this result is that it shows that G is indeed a group. For any two morphisms f and g, we can define a linear map $f\circ g\colon T_+ \to \mathbf{R}$ by $\big(f\circ g\big)(\sigma) = \big(f\otimes g\big)\Delta^+\sigma$. As a consequence of the first identity in (28), $f\circ g$ is again a morphism on T_+. As a consequence of the second identity, one has $\hat\Gamma_{f\circ g} = \Gamma_f\,\Gamma_g$. The last identity shows that $(f_1\circ f_2)\circ f_3 = f_1\circ(f_2\circ f_3)$, while the properties of $\mathcal{A}$ ensure that if we set $f^{-1}(\sigma) = f(\mathcal{A}\sigma)$, then $f\circ f^{-1} = f^{-1}\circ f = e$. Finally, the third identity in (28) shows that e is indeed the identity element, thus turning the set of all morphisms of T_+ into a group under $\circ$, acting on T via $\hat\Gamma$.

Let us now turn back to our models. Given a smooth function ξ_ε, we define Π_z as in (25) and f_z by (26). We then also define linear maps $\Gamma_{zz'}$ by $\Gamma_{zz'} = \hat\Gamma_{\gamma_{zz'}}$ with $\gamma_{zz'} = f_z^{-1}\circ f_{z'}$. We then have

Lemma 5.6. *For every smooth function* ξ_ε*, the pair* (Π,Γ) *defined above is a model.*

Proof. The algebraic constraints (16) are satisfied essentially by definition. The first bound of (17) can easily be verified recursively by (25). The only non-trivial fact is that the matrix elements of $\Gamma_{zz'}$ satisfy the right bound. If one can show that $|\gamma_{zz'}(\sigma)| \lesssim |z-z'|^{|\sigma|}$, this in turn follows from (27). This bound is non-trivial and was obtained in [23, Prop. 8.27]. □

5.4. Admissible models.

Thanks to Lemma 5.6, we now have a large class of models for the regularity structures built in the previous two subsections. However, we do not want to restrict ourselves to this class (or even its closure). The reason is that if we define products in the "naïve" way given by the second identity in (25b), then there will typically be some situations where the result diverges as we let $\varepsilon\to 0$ in ξ_ε. Therefore, we do not impose this relation in general but rather view it as the *definition* of the product, i.e. we interpret it as

$$\big(\Pi_z\tau\big)\cdot\big(\Pi_z\bar\tau\big) := \Pi_z(\tau\bar\tau)\;.$$

However, the remainder of the structure described in (25) is required for X_i, $\mathscr{D}_i$ and $\mathcal{I}$ to have the correct interpretation. This motivates the following definition.

Definition 5.7. Given a regularity structure $\mathscr{T}$ constructed as in Sections 5.2 and 5.3, we say that a model (Π,Γ) is *admissible* if it satisfies $\big(\Pi_z X_i\big)(z') = z_i' - z_i$, $\Pi_z\mathscr{D}_i\tau = \partial_i\Pi_z\tau$, as well as (25c) and if furthermore $\Gamma_{zz'} = \hat\Gamma_{f_z}^{-1}\hat\Gamma_{f_{z'}}$ with f_z given by (26). We will denote the space of all admissible models by $\mathscr{M}_0\subset\mathscr{M}$.

Remark 5.8. In the particular case of admissible models for a regularity structure of the type considered here, the data of the single linear map $\mathbf{\Pi}$ as above is sufficient to reconstruct the full model (Π,Γ).

Note that at this stage, it is not clear whether this concept is even well-defined: in general, $D^{(k)}K \star \Pi_z \tau$ will be a distribution and cannot be evaluated at fixed points, so (26) might be meaningless for a general model. It turns out that the definition actually always makes sense, provided that the second identity in (26) is interpreted as

$$f_z(\mathcal{J}_k(\tau)) = \sum_{n \geq 0} \big(D^{(k)} K_n \star \Pi_z \tau\big)(z) \;,$$

where $K = \sum_{n \geq 0} K_n$ as in (6). This is because the bound (6), combined with the bound (17) and the fact that K_n is supported in the ball of radius 2^{-n} imply that

$$\big|\big(D^{(k)} K_n \star \Pi_z \tau\big)(z)\big| \lesssim 2^{(|k|-|\tau|-2)n} \;.$$

The condition $|k| < |\tau| + 2$ appearing in (25c) is then precisely what is required to guarantee that this is always summable.

5.5. Abstract fixed point problem. We now show how to reformulate a stochastic PDE as a fixed point problem in some space $\mathcal{D}^\gamma$ based on an admissible model for the regularity structure associated to the SPDE by the construction of Section 5.2. For definiteness, we focus on the example of the KPZ equation (3), but all other examples mentioned in the introduction can be treated in virtually the same way. Writing P for the heat kernel, the mild formulation of (3) is given by

$$h = P \star \mathbf{1}_{t>0}\big((\partial_x h)^2 + \xi\big) + P h_0 \;, \tag{29}$$

where we write Ph_0 for the harmonic extension of h_0. (This is just the solution to the heat equation with initial condition h_0.) In order to formulate this as a fixed point problem in $\mathcal{D}^\gamma$ for a suitable value of $\gamma > 0$, we will make use of the following far-reaching extension of Schauder's theorem.

Theorem 5.9. *Fix one of the regularity structures built in the previous section and fix an admissible model. Then, for all but a discrete set of values of $\gamma > 0$, there exists a continuous operator $\mathcal{P} \colon \mathcal{D}^\gamma \to \mathcal{D}^{\gamma+2}$ such that the identity*

$$\mathcal{R}\mathcal{P} f = P \star \mathcal{R} f \;, \tag{30}$$

holds for every $f \in \mathcal{D}^\gamma$. Furthermore, one has $\big(\mathcal{P} f\big)(z) - \mathcal{I} f(z) \in \bar{T}$.

Remark 5.10. Recall that $\bar{T} \subset T$ denotes the linear span of the X^k, which represent the usual Taylor polynomials. Again, while $\mathcal{P}$ is a linear map when we consider the underlying model as fixed, it can (and should) also be viewed as a continuous nonlinear map from $\mathscr{M}_0 \ltimes \mathcal{D}^\gamma$ into $\mathscr{M}_0 \ltimes \mathcal{D}^{\gamma+2}$. The reason why some values of γ need to be excluded is essentially the same as for the usual Schauder theorem.

For a proof of Theorem 5.9 and a precise description of the operator $\mathcal{P}$, see [23, Sec. 5]. With the help of the operator $\mathcal{P}$, it is then possible to reformulate (29)

as the following fixed point problem in $\mathcal{D}^\gamma$, provided that we have an admissible model at our disposal:

$$H = \mathcal{P}\mathbf{1}_{t>0}\big((\mathscr{D}H)^2 + \Xi\big) + Ph_0 \;. \tag{31}$$

Here, the smooth function Ph_0 is interpreted as an element in $\mathcal{D}^\gamma$ with values in $\bar{T}$ via its Taylor expansion of order γ. Note that in the context of the regularity structure associated to the KPZ equation in Section 5.2, the right hand side of this equation makes sense for every $H \in \mathcal{D}^\gamma$, provided that H takes values in $T_{\mathcal{U}}$. This is an immediate consequence of the property (23).

Remark 5.11. As already mentioned earlier, we cheat here in the sense that $\mathcal{D}^\gamma$ should really be replaced by a space $\mathcal{D}^{\gamma,\eta}$ allowing for a suitable singular behaviour on the hyperplane $t = 0$.

It is also possible to show (see [23, Thm 4.7]) that if we set $|\Xi| = -\frac{3}{2} - \kappa$ for some sufficiently small $\kappa > 0$, then one has $(\mathscr{D}H)^2 \in \mathcal{D}^{\gamma-\frac{3}{2}-\kappa}$ for $H \in \mathcal{D}^\gamma$. As a consequence, we expect to be able to find local solutions to the fixed point problem (31), provided that we formulate it in $\mathcal{D}^\gamma$ for $\gamma > \frac{3}{2} + \kappa$. This is indeed the case, and a more general instance of this fact can be found in [23, Thm 7.8]. Furthermore, the local solution is locally Lipschitz continuous as a function of both the initial condition h_0 and the underlying admissible model $(\Pi, \Gamma) \in \mathscr{M}_0$.

Now that we have a local solution $H \in \mathcal{D}^\gamma$ for (31), we would like to know how this solution relates to the original problem (3). This is given by the following simple fact:

Proposition 5.12. *If the underlying model* (Π, Γ) *is built from a smooth function* ξ_ε *as in* (25) *and if* H *solves* (31)*, then* $\mathcal{R}H$ *solves* (29)*.*

Proof. As a consequence of (30), we see that $\mathcal{R}H$ solves

$$\mathcal{R}H = P \star \mathbf{1}_{t>0}\big(\mathcal{R}\big((\mathscr{D}H)^2\big) + \xi_\varepsilon\big) + Ph_0 \;.$$

Combining (25b) with (22), it is not difficult to see that in this particular case, one has $\mathcal{R}\big((\mathscr{D}H)^2\big) = (\partial_x \mathcal{R}H)^2$, so that the claim follows. □

The results of the previous subsection yield a robust solution theory for (31) which projects down (via $\mathcal{R}$) to the usual solution theory for (3) for smooth driving noise ξ_ε. If it were the case that the sequence of models $(\Pi^{(\varepsilon)}, \Gamma^{(\varepsilon)})$ associated to the regularised noise ξ_ε via (25) converges to a limit in $\mathscr{M}_0$, then this would essentially conclude our analysis of (3).

Unfortunately, this is *not* the case. Indeed, in all of the examples mentioned in the introduction except for (2), the sequence of models $(\Pi^{(\varepsilon)}, \Gamma^{(\varepsilon)})$ does not converge as $\varepsilon \to 0$. In order to remedy to this situation, the idea is to look for a sequence of "renormalised" models $(\hat{\Pi}^{(\varepsilon)}, \hat{\Gamma}^{(\varepsilon)})$ which are also admissible and also satisfy $\hat{\Pi}_z^{(\varepsilon)}\Xi = \xi_\varepsilon$, but do converge to a limit as $\varepsilon \to 0$. The last section of this article shows how these renormalised models can be constructed.

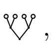
5.6. Renormalisation. In order to renormalise our model, we will build a very natural group of continuous transformations of $\mathscr{M}_0$ that build a new admissible model from an old one. The renormalised model will then be the image of the "canonical" model $(\Pi^{(\varepsilon)}, \Gamma^{(\varepsilon)})$ under a (diverging) sequence of such transformations. Since we want the new model to also be admissible, the only defining property that we are allowed to modify in (25) is the definition of the product. In order to describe the renormalised model, it turns out to be more convenient to consider again its representation by a single linear map $\hat{\mathbf{\Pi}}^{(\varepsilon)} : T \to \mathcal{S}'$ as in (5.3), which is something we can do by Remark 5.8.

At this stage, we do not appear to have much choice: the only "reasonable" way of building $\hat{\mathbf{\Pi}}^{(\varepsilon)}$ from $\mathbf{\Pi}^{(\varepsilon)}$ is to compose it to the right with some fixed linear map $M_\varepsilon \colon T \to T$:

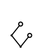
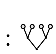
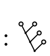
$$\hat{\mathbf{\Pi}}^{(\varepsilon)} = \mathbf{\Pi}^{(\varepsilon)} M_\varepsilon \;. \tag{32}$$

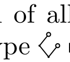
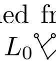
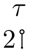
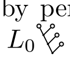
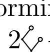
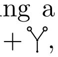
If we do this for an arbitrary map M_ε, we will of course immediately lose the algebraic and analytical properties that allow to associate an admissible model $(\hat{\Pi}^{(\varepsilon)}, \hat{\Gamma}^{(\varepsilon)})$ to the map $\hat{\mathbf{\Pi}}^{(\varepsilon)}$. As a matter of fact, it is completely unclear *a priori* whether there exists *any* non-trivial map M_ε that preserves these properties. Fortunately, these maps do exists and a somewhat indirect characterisation of them can be found in [23, Sec. 8]. Even better, there are sufficiently many of them so that the divergencies of $\mathbf{\Pi}^{(\varepsilon)}$ can be compensated by a judicious choice of M_ε.

Let us just illustrate how this plays out in the case of the KPZ equation already studied in the last subsection. In order to simplify notations, we now use the following shorthand graphical notation for elements of $\mathcal{U} \cup \mathcal{V}$. For Ξ, we draw a small circle. The integration map $\mathcal{I}$ is then represented by a downfacing wavy line and $\mathscr{D}\mathcal{I}$ is represented by a downfacing plain line. The multiplication of symbols is obtained by joining them at the root. For example, we have

$(\mathscr{D}\mathcal{I}(\Xi))^2 =$  $,\qquad (\mathscr{D}\mathcal{I}(\mathscr{D}\mathcal{I}(\Xi)^2))^2 =$  $,\qquad \mathcal{I}(\mathscr{D}\mathcal{I}(\Xi)^2) =$  $.$

In the case of the KPZ equation, it turns out that one can exhibit an explicit four-parameter group of matrices M which preserve admissible models when used in (32). These matrices are of the form $M = \exp(-\sum_{i=0}^3 C_i L_i)$, where the generators L_i are determined by the following contraction rules:

$L_0 \colon$  $\mapsto \mathbf{1}\;,\qquad L_1 \colon$  $\mapsto \mathbf{1}\;,\qquad L_2 \colon$  $\mapsto \mathbf{1}\qquad L_3 \colon$  $\mapsto \mathbf{1}\;.$ (33)

This should be understood in the sense that if τ is an arbitrary formal expression, then $L_0\tau$ is the sum of all formal expressions obtained from τ by performing a substitution of the type  $\mapsto \mathbf{1}$. For example, one has L_0  $= 2$  , L_0  $= 2$  $+$  , etc. The extension of the other operators L_i to all of T is given by $L_i\tau = 0$ for $i \neq 0$ and every τ for which L_i wasn't already defined in (33). We then have the following result, which is a consequence of [23, Sec. 8] and [28] and was implicit in [21]:

Theorem 5.13. *Let M_ε be given as above, let $\mathbf{\Pi}^{(\varepsilon)}$ be constructed from ξ_ε as in (24), and let $\hat{\mathbf{\Pi}}^{(\varepsilon)} = \mathbf{\Pi}^{(\varepsilon)} M_\varepsilon$. Then, there exists a unique admissible model*

$(\hat\Pi^{(\varepsilon)},\hat\Gamma^{(\varepsilon)})$ *such that* $\hat\Pi_z^{(\varepsilon)} = \hat{\mathbf{\Pi}}^{(\varepsilon)}\hat F_z^{(\varepsilon)}$, *where* $\hat F_z^{(\varepsilon)}$ *relates to* $\hat\Pi_z^{(\varepsilon)}$ *as in* (26). *Furthermore, one has the identity*

$$\big(\hat\Pi_z^{(\varepsilon)}\tau\big)(z) = \big(\Pi_z^{(\varepsilon)} M_\varepsilon \tau\big)(z)\;. \tag{34}$$

Finally, there is a choice of M_ε *such that* $(\hat\Pi^{(\varepsilon)},\hat\Gamma^{(\varepsilon)})$ *converges to a limit* $(\hat\Pi,\hat\Gamma)$ *which is universal in that it does not depend on the details of the regularisation procedure.*

Remark 5.14. Despite (34), it is *not* true in general that $\hat\Pi_z^{(\varepsilon)} = \Pi_z^{(\varepsilon)} M_\varepsilon$. The point is that (34) only holds at the point z and not at $z' \neq z$.

In order to complete our survey of Theorem 1.1, it remains to identify the solution to (31) with respect to the renormalised model $(\hat\Pi^{(\varepsilon)},\hat\Gamma^{(\varepsilon)})$ with the classical solution to some modified partial differential equation. The continuity of the abstract solution map then immediately implies that the solutions to the modified PDE converge to a limit. The fact that the limiting model $(\hat\Pi,\hat\Gamma)$ is universal also implies that this limit is universal.

Theorem 5.15. *Let* $M_\varepsilon = \exp(-\sum_{i=0}^3 C_i^{(\varepsilon)} L_i)$ *be as above and let* $(\hat\Pi^{(\varepsilon)},\hat\Gamma^{(\varepsilon)})$ *be the corresponding renormalised model. Let furthermore* H *be the solution to* (31) *with respect to this model. Then, the function* $h(t,x) = \big(\mathcal{R}H\big)(t,x)$ *solves the equation*

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 - 4C_0^{(\varepsilon)}\,\partial_x h + \xi_\varepsilon - (C_1^{(\varepsilon)} + C_2^{(\varepsilon)} + 4C_3^{(\varepsilon)})\;. \tag{35}$$

Remark 5.16. In order to obtain a limit $(\hat\Pi,\hat\Gamma)$, the renormalisation constants $C_i^{(\varepsilon)}$ should be chosen in the following way:

$$C_0^{(\varepsilon)} = 0\;,\qquad C_1^{(\varepsilon)} = \frac{c_1}{\varepsilon}\;,\qquad C_2^{(\varepsilon)} = 4c_2\log\varepsilon + c_3\;,\qquad C_3^{(\varepsilon)} = -c_2\log\varepsilon + c_4\;.$$

Here, the c_i are constants of order 1 that depend on the details of the regularisation procedure for ξ_ε. The fact that $C_0^{(\varepsilon)} = 0$ explains why the corresponding term does not appear in (3). The fact that the diverging parts of $C_2^{(\varepsilon)}$ and $C_3^{(\varepsilon)}$ cancel in (35) explains why this logarithmic sub-divergence was not observed in [4] for example.

Proof. We first note that, as a consequence of Theorem 5.9 and of (31), one can write for $t > 0$

$$H = \mathcal{I}\big((\mathscr{D}H)^2 + \Xi\big) + (...)\;, \tag{36}$$

where $(...)$ denotes some terms belonging to $\bar T \subset T$.

By repeatedly using this identity, we conclude that any solution $H \in \mathcal{D}^\gamma$ to (31) for γ greater than (but close enough to) $3/2$ is necessarily of the form

$$H = h\,\mathbf{1} + \mathcal{I}(\xi) + \mathcal{I}(\mathcal{I}(\xi)^2) + h'\,X_1 + 2\mathcal{I}(\mathcal{I}(\xi)\,X_1) + 2h'\,\mathcal{I}(X_1^2)\;, \tag{37}$$

for some real-valued functions h and h'. Note that h' is treated as an independent function here, we certainly do not mean to suggest that the function h is differentiable! Our notation is only by analogy with the classical Taylor expansion. As an immediate consequence, $\mathscr{D}H$ is given by

$$\mathscr{D}H = \text{Ɩ} + \text{Y} + h'\,\mathbf{1} + 2\,\text{Y} + 2h'\,\text{<}\,, \tag{38}$$

as an element of $\mathcal{D}^\gamma$ for γ close to $1/2$. The right hand side of the equation is then given up to order 0 by

$$(\mathscr{D}H)^2 + \Xi = \Xi + \text{V} + 2\,\text{V} + 2h'\,\text{Ɩ} + \text{V} + 4\,\text{V} + 2h'\,\text{Y} + 4h'\,\text{<} + (h')^2\,\mathbf{1}\,. \tag{39}$$

Using the definition of M_ε, we conclude that

$$M_\varepsilon \mathscr{D}H = \mathscr{D}H - 4C_0^{(\varepsilon)}\,\text{<}\,,$$

so that, as an element of $\mathcal{D}^\gamma$ with very small (but positive) γ, one has the identity

$$(M_\varepsilon \mathscr{D}H)^2 = (\mathscr{D}H)^2 - 8C_0^{(\varepsilon)}\,\text{<}\,.$$

As a consequence, after neglecting again all terms of strictly positive homogeneity, one has the identity

$$M_\varepsilon\big((\mathscr{D}H)^2 + \Xi\big) = (M_\varepsilon \mathscr{D}H)^2 + \Xi - 4C_0^{(\varepsilon)}\, M_\varepsilon \mathscr{D}H - (C_1^{(\varepsilon)} + C_2^{(\varepsilon)} + 4C_3^{(\varepsilon)})\,.$$

Combining this with (34) and (22), we conclude that

$$\mathcal{R}\big((\mathscr{D}H)^2 + \Xi\big) = (\partial_x \mathcal{R}H)^2 + \xi_\varepsilon - 4C_0^{(\varepsilon)}\, \partial_x \mathcal{R}H - (C_1^{(\varepsilon)} + C_2^{(\varepsilon)} + 4C_3^{(\varepsilon)})\,,$$

from which the claim then follows in the same way as for Proposition 5.12. □

Remark 5.17. Ultimately, the reason why the theory mentioned in Section 1.1 (or indeed the theory of controlled rough paths, as originally exploited in [21]) can also be applied in this case is that in (37), only *one* basis vector besides those in $\mathcal{T}$ (i.e. besides $\mathbf{1}$ and X_1) comes with a non-constant coefficient, namely the basis vector <. The methodology explained in Section 3.1 on the other hand can be applied whenever *no* basis vector besides those in $\mathcal{T}$ comes with a non-constant coefficient.

References

[1] S. Albeverio and M. Röckner. Stochastic differential equations in infinite dimensions: solutions via Dirichlet forms. *Probab. Theory Related Fields*, 89(3):347–386, 1991.

[2] H. Bahouri, J.-Y. Chemin, and R. Danchin. *Fourier analysis and nonlinear partial differential equations*, volume 343 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Heidelberg, 2011.

[3] Á. Bényi, D. Maldonado, and V. Naibo. What is ... a paraproduct? *Notices Amer. Math. Soc.*, 57(7):858–860, 2010.

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

[5] L. Bertini, E. Presutti, B. Rüdiger, and E. Saada. Dynamical fluctuations at the critical point: convergence to a nonlinear stochastic PDE. *Teor. Veroyatnost. i Primenen.*, 38(4):689–741, 1993.

[6] J.-M. Bismut. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete*, 56(4):469–505, 1981.

[7] J.-M. Bony. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires. *Ann. Sci. École Norm. Sup. (4)*, 14(2):209–246, 1981.

[8] R. A. Carmona and S. A. Molchanov. Parabolic Anderson problem and intermittency. *Mem. Amer. Math. Soc.*, 108(518):viii+125, 1994.

[9] R. Catellier and K. Chouk. Paracontrolled distributions and the 3-dimensional stochastic quantization equation. *ArXiv e-prints*, Oct. 2013.

[10] G. Da Prato and A. Debussche. Two-dimensional Navier-Stokes equations driven by a space-time white noise. *J. Funct. Anal.*, 196(1):180–210, 2002.

[11] G. Da Prato and A. Debussche. Strong solutions to the stochastic quantization equations. *Ann. Probab.*, 31(4):1900–1916, 2003.

[12] G. Da Prato, A. Debussche, and L. Tubaro. A modified Kardar-Parisi-Zhang model. *Electron. Comm. Probab.*, 12:442–453 (electronic), 2007.

[13] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1992.

[14] R. L. Dobrushin. Gaussian and their subordinated self-similar random generalized fields. *Ann. Probab.*, 7(1):1–28, 1979.

[15] P. K. Friz and M. Hairer. *A course on rough paths*. Universitext. Springer, 2014. To appear.

[16] P. K. Friz and N. B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[17] P. Goncalves and M. Jara. Universality of KPZ equation. *ArXiv e-prints*, Mar. 2010.

[18] M. Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

[19] M. Gubinelli, P. Imkeller, and N. Perkowski. Paraproducts, rough paths and controlled distributions. *ArXiv e-prints*, Oct. 2012.

[20] M. Hairer. Rough stochastic PDEs. *Comm. Pure Appl. Math.*, 64(11):1547–1585, 2011.

[21] M. Hairer. Solving the KPZ equation. *Ann. of Math. (2)*, 178(2):559–664, 2013.

[22] M. Hairer. Introduction to regularity structures. *ArXiv e-prints*, Jan. 2014. Braz. J. Prob. Stat., to appear.

[23] M. Hairer. A theory of regularity structures. *Invent. Math.*, Mar. 2014.

[24] M. Hairer and J. Maas. A spatial version of the Itô-Stratonovich correction. *Ann. Probab.*, 40(4):1675–1714, 2012.

[25] M. Hairer, J. Maas, and H. Weber. Approximating rough stochastic PDEs. *Comm. Pure Appl. Math.*, 67(5):776–870, 2014.

[26] M. Hairer, É. Pardoux, and A. Piatnitksy. A Wong-Zakai theorem for stochastic PDEs. Work in progress, 2014.

[27] M. Hairer and N. S. Pillai. Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths. *Ann. Probab.*, 41(4):2544–2598, 2013.

[28] M. Hairer and J. Quastel. Continuous interface models rescale to KPZ. Work in progress, 2014.

[29] M. Hairer and H. Weber. Rough Burgers-like equations with multiplicative noise. *Probab. Theory Related Fields*, 155(1-2):71–126, 2013.

[30] K. Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

[31] G. Jona-Lasinio and P. K. Mitter. On the stochastic quantization of field theory. *Comm. Math. Phys.*, 101(3):409–436, 1985.

[32] M. Kardar, G. Parisi, and Y.-C. Zhang. Dynamic scaling of growing interfaces. *Phys. Rev. Lett.*, 56(9):889–892, Mar 1986.

[33] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, volume 32 of *North-Holland Math. Library*, pages 271–306. North-Holland, Amsterdam, 1984.

[34] T. J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[35] T. J. Lyons, M. Caruana, and T. Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour.

[36] T. J. Lyons and Z. Qian. *System control and rough paths*. Oxford Mathematical Monographs. Oxford University Press, Oxford, 2002. Oxford Science Publications.

[37] P. Malliavin. *Stochastic analysis*, volume 313 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1997.

[38] J. Norris. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, volume 1204 of *Lecture Notes in Math.*, pages 101–130. Springer, Berlin, 1986.

[39] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[40] G. Parisi and Y. S. Wu. Perturbation theory without gauge fixing. *Sci. Sinica*, 24(4):483–496, 1981.

[41] J. Schauder. Über lineare elliptische Differentialgleichungen zweiter Ordnung. *Math. Z.*, 38(1):257–282, 1934.

[42] L. Simon. Schauder estimates by scaling. *Calc. Var. Partial Differential Equations*, 5(5):391–407, 1997.

[43] R. L. Stratonovič. A new form of representing stochastic integrals and equations. *Vestnik Moskov. Univ. Ser. I Mat. Meh.*, 1964(1):3–12, 1964.

[44] E. Wong and M. Zakai. On the relation between ordinary and stochastic differential equations. *Internat. J. Engrg. Sci.*, 3:213–229, 1965.

Mathematics Institute, The University of Warwick, U.K.
E-mail: M.Hairer@Warwick.ac.uk