Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 2, 175–210
DOI: 10.1214/14-BJPS241
© Brazilian Statistical Association, 2015

Introduction to regularity structures

Martin Hairer
The University of Warwick

Abstract. These are short notes from a series of lectures given at the University of Rennes in June 2013, at the University of Bonn in July 2013, at the XVIIth Brazilian School of Probability in Mambucaba in August 2013, and at ETH Zurich in September 2013. They give a concise overview of the theory of regularity structures as exposed in the article [*Invent. Math.* DOI: 10.1007/s00222-014-0505-4]. In order to allow to focus on the conceptual aspects of the theory, many proofs are omitted and statements are simplified. We focus on applying the theory to the problem of giving a solution theory to the stochastic quantisation equations for the Euclidean Φ^4_3 quantum field theory.

1 Introduction

Very recently, a new theory of "regularity structures" was introduced (Hairer, 2014), unifying various flavours of the theory of (controlled) rough paths (including Gubinelli's theory of controlled rough paths (Gubinelli, 2004), as well as his branched rough paths (Gubinelli, 2010)), as well as the usual Taylor expansions. While it has its roots in the theory of rough paths (Lyons, 1998), the main advantage of this new theory is that it is no longer tied to the one-dimensionality of the time parameter, which makes it also suitable for the description of solutions to stochastic *partial* differential equations, rather than just stochastic ordinary differential equations. The aim of this article is to give a concise survey of the theory while focusing on the construction of the dynamical Φ^4_3 model. While the exposition aims to be reasonably self-contained (in particular no prior knowledge of the theory of rough paths is assumed), most of the proofs will only be sketched.

The main achievement of the theory of regularity structures is that it allows to give a (pathwise!) meaning to ill-posed stochastic PDEs that arise naturally when trying to describe the macroscopic behaviour of models from statistical mechanics near criticality. One example of such an equation is the KPZ equation arising as a natural model for one-dimensional interface motion (Kardar et al., 1986; Bertini and Giacomin, 1997; Hairer, 2013):

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 + \xi - C.$$

Key words and phrases. Stochastic PDEs, regularity structures, renormalisation.
Received December 2013; accepted March 2014.

Another example is the dynamical Φ^4_3 model arising, for example, in the stochastic quantisation of Euclidean quantum field theory (Parisi and Wu, 1981; Jona-Lasinio and Mitter, 1985; Albeverio and Röckner, 1991; Da Prato and Debussche, 2003; Hairer, 2014), as well as a universal model for phase coexistence near the critical point (Giacomin et al., 1999):

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

In both of these examples, ξ formally denotes space–time white noise, C is an arbitrary constant (which will actually turn out to be infinite in some sense), and we consider a bounded square spatial domain with periodic boundary conditions. In the case of the dynamical Φ^4_3 model, the spatial variable has dimension 3, while it has dimension 1 in the case of the KPZ equation. While a full exposition of the theory is well beyond the scope of this short introduction, we aim to give a concise overview to most of its concepts. In most cases, we will only state results in a rather informal way and give some ideas as to how the proofs work, focusing on conceptual rather than technical issues. The only exception is the "reconstruction theorem," Theorem 2.10 below, which is the linchpin of the whole theory. Since its proof (or rather a slightly simplified version of it) is relatively concise, we provide a fully self-contained version. For precise statements and complete proofs of most of the results exposed here, we refer to the original article (Hairer, 2014).

Loosely speaking, the type of well-posedness results that can be proven with the help of the theory of regularity structures can be formulated as follows.

Theorem 1.1. *Let $\xi_\varepsilon = \delta_\varepsilon * \xi$ denote the regularisation of space–time white noise with a compactly supported smooth mollifier δ_ε that is scaled by ε in the spatial direction(s) and by ε^2 in the time direction. Denote by h_ε and Φ_ε the solutions to*

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + (\partial_x h_\varepsilon)^2 - C_\varepsilon + \xi_\varepsilon,$$
$$\partial_t \Phi_\varepsilon = \Delta \Phi_\varepsilon + \tilde{C}_\varepsilon \Phi_\varepsilon - \Phi_\varepsilon^3 + \xi_\varepsilon.$$

Then there exist choices of constants C_ε and $\tilde{C}_\varepsilon$ diverging as $\varepsilon \to 0$, as well as processes h and Φ such that $h_\varepsilon \to h$ and $\Phi_\varepsilon \to \Phi$ in probability. Furthermore, while the constants C_ε and $\tilde{C}_\varepsilon$ do *depend crucially on the choice of mollifiers δ_ε, the limits h and Φ do* not *depend on them.*

Remark 1.2. We made a severe abuse of notation here since the space–time white noise appearing in the equation for h_ε is on $\mathbf{R} \times \mathbf{T}^1$, while the one appearing in the equation for Φ_ε is on $\mathbf{R} \times \mathbf{T}^3$. (Here, we denote by $\mathbf{T}^n$ the n-dimensional torus.)

Remark 1.3. We have not described the topology in which the convergence takes place in these examples. In the case of the KPZ equation, one actually obtains convergence in some space of space–time Hölder continuous functions. In the case of the dynamical Φ^4_3 model, convergence takes place in some space of space–time

distributions. One caveat that also has to be dealt with in the latter case is that the limiting process Φ may in principle explode in finite time for some instances of the driving noise.

From a "philosophical" perspective, the theory of regularity structures is inspired by the theory of controlled rough paths (Lyons, 1998; Gubinelli, 2004; Lyons et al., 2007), so let us rapidly survey the main ideas of that theory. The setting of the theory of controlled rough paths is the following. Let us say that we want to solve a controlled differential equation of the type

$$dY = f(Y)\,dX(t), \tag{1.1}$$

where $X \in \mathcal{C}^\alpha$ is a rather rough function (say a typical sample path for an m-dimensional Brownian motion). It is a classical result by Young (1936) that the Riemann–Stieltjes integral $(X, Y) \mapsto \int_0^\cdot Y\,dX$ makes sense as a continuous map from $\mathcal{C}^\alpha \times \mathcal{C}^\alpha$ into $\mathcal{C}^\alpha$ if and only if $\alpha > \frac{1}{2}$. As a consequence, "naïve" approaches to a pathwise solution to (1.1) are bound to fail if X has the regularity of Brownian motion.

The main idea is to exploit the a priori "guess" that solutions to (1.1) should "look like X at small scales." More precisely, one would naturally expect the solution Y to satisfy

$$Y_t = Y_s + Y'_s X_{s,t} + \mathcal{O}\big(|t-s|^{2\alpha}\big), \tag{1.2}$$

where we wrote $X_{s,t}$ as a shorthand for the increment $X_t - X_s$. As a matter of fact, one would expect to have such an expansion with $Y' = f(Y)$. Denote by $\mathcal{C}^\alpha_X$ the space of pairs (Y, Y') satisfying (1.2) for a given "model path" X. It is then possible to simply "postulate" the values of the integrals

$$\mathbb{X}_{s,t} =: \int_s^t X_{s,r} \otimes dX_r, \tag{1.3}$$

satisfying "Chen's relations"

$$\mathbb{X}_{s,t} - \mathbb{X}_{s,u} - \mathbb{X}_{u,t} = X_{s,u} \otimes X_{u,t}, \tag{1.4}$$

as well as the analytic bound $|\mathbb{X}_{s,t}| \lesssim |t-s|^{2\alpha}$, and to exploit this additional data to give a coherent definition of expressions of the type $\int Y\,dX$, provided that the path X is "enhanced" with its iterated integrals $\mathbb{X}$ and Y is a "controlled path" of the type (1.2). See, for example, Gubinelli (2004) for more information or Hairer (2011) for a concise exposition of this theory.

Compare (1.2) to the fact that a function $f : \mathbf{R} \to \mathbf{R}$ is of class $\mathcal{C}^\gamma$ with $\gamma \in (k, k+1)$ if for every $s \in \mathbf{R}$ there exist coefficients $f_s^{(1)}, \dots, f_s^{(k)}$ such that

$$f_t = f_s + \sum_{\ell=1}^{k} f_s^{(\ell)} (t-s)^\ell + \mathcal{O}\big(|t-s|^\gamma\big). \tag{1.5}$$

Of course, $f_s^{(\ell)}$ is nothing but the ℓth derivative of f at the point s, divided by $\ell!$. In this sense, one should really think of a controlled rough path $(Y, Y') \in \mathcal{C}_X^\alpha$ as a 2α-Hölder continuous function, but with respect to a "model" determined by the function X, rather than by the usual Taylor polynomials. This formal analogy between controlled rough paths and Taylor expansions suggests that it might be fruitful to systematically investigate what are the "right" objects that could possibly take the place of Taylor polynomials, while still retaining many of their nice properties.

2 Definitions and the reconstruction operator

The first step in such an endeavour is to set up an algebraic structure reflecting the properties of Taylor expansions. First of all, such a structure should contain a vector space T that will contain the coefficients of our expansion. It is natural to assume that T has a graded structure: $T = \bigoplus_{\alpha \in A} T_\alpha$, for some set A of possible "homogeneities." For example, in the case of the usual Taylor expansion (1.5), it is natural to take for A the set of natural numbers and to have T_ℓ contain the coefficients corresponding to the derivatives of order ℓ. In the case of controlled rough paths, however, it is natural to take $A = \{0, \alpha\}$, to have again T_0 contain the value of the function Y at any time s, and to have T_α contain the Gubinelli derivative Y'_s. This reflects the fact that the "monomial" $t \mapsto X_{s,t}$ only vanishes at order α near $t = s$, while the usual monomials $t \mapsto (t - s)^\ell$ vanish at integer order ℓ.

This, however, is not the full algebraic structure describing Taylor-like expansions. Indeed, one of the characteristics of Taylor expansions is that an expansion around some point x_0 can be re-expanded around any other point x_1 by writing

$$(x - x_0)^m = \sum_{k+\ell=m} \frac{m!}{k!\ell!} (x_1 - x_0)^k \cdot (x - x_1)^\ell. \tag{2.1}$$

(In the case when $x \in \mathbf{R}^d$, k, ℓ and m denote multi-indices and $k! = k_1! \cdots k_d!$.) Somewhat similarly, in the case of controlled rough paths, we have the (rather trivial) identity

$$X_{s_0,t} = X_{s_0,s_1} \cdot 1 + 1 \cdot X_{s_1,t}. \tag{2.2}$$

What is a natural abstraction of this fact? In terms of the coefficients of a "Taylor expansion," the operation of re-expanding around a different point is ultimately just a linear operation from $\Gamma : T \to T$, where the precise value of the map Γ depends on the starting point x_0, the endpoint x_1, and possibly also on the details of the particular "model" that we are considering. In view of the above examples, it is natural to impose furthermore that Γ has the property that if $\tau \in T_\alpha$, then $\Gamma\tau - \tau \in \bigoplus_{\beta<\alpha} T_\beta$. In other words, when re-expanding a homogeneous monomial

around a different point, the leading order coefficient remains the same, but lower order monomials may appear.

These heuristic considerations can be summarised in the following definition of an abstract object we call a *regularity structure*:

Definition 2.1. Let $A \subset \mathbf{R}$ be bounded from below and without accumulation point, and let $T = \bigoplus_{\alpha \in A} T_\alpha$ be a vector space graded by A such that each T_α is a Banach space. Let furthermore G be a group of continuous operators on T such that, for every $\alpha \in A$, every $\Gamma \in G$, and every $\tau \in T_\alpha$, one has $\Gamma\tau - \tau \in \bigoplus_{\beta<\alpha} T_\beta$. The triple $\mathscr{T} = (A, T, G)$ is called a *regularity structure* with *model space* T and *structure group* G.

Remark 2.2. Given $\tau \in T$, we will write $\|\tau\|_\alpha$ for the norm of its component in T_α.

Remark 2.3. In Hairer (2014), it is furthermore assumed that $0 \in A$, $T_0 \approx \mathbf{R}$, and T_0 is invariant under G. This is a very natural assumption which ensures that our regularity structure is at least sufficiently rich to represent constant functions.

Remark 2.4. In principle, the set A can be infinite. By analogy with the polynomials, it is then natural to consider T as the set of all formal series of the form $\sum_{\alpha \in A} \tau_\alpha$, where only finitely many of the τ_α's are non-zero. This also dovetails nicely with the particular form of elements in G. In practice, however, we will only ever work with finite subsets of A so that the precise topology on T does not matter.

At this stage, a regularity structure is a completely abstract object. It only becomes useful when endowed with a *model*, which is a concrete way of associating to any $\tau \in T$ and $x_0 \in \mathbf{R}^d$, the actual "Taylor polynomial based at x_0" represented by τ. Furthermore, we want elements $\tau \in T_\alpha$ to represent functions (or possibly distributions) that "vanish at order α" around the given point x_0.

Since we would like to allow A to contain negative values and, therefore, allow elements in T to represent actual distributions, we need a suitable notion of "vanishing at order α." We achieve this by considering the size of our distributions, when tested against test functions that are localised around the given point x_0. Given a test function φ on $\mathbf{R}^d$, we write φ_x^λ as a shorthand for

$$\varphi_x^\lambda(y) = \lambda^{-d}\varphi\big(\lambda^{-1}(y - x)\big).$$

Given $r > 0$, we also denote by $\mathcal{B}_r$ the set of all functions $\varphi : \mathbf{R}^d \to \mathbf{R}$ such that $\varphi \in \mathcal{C}^r$ with $\|\varphi\|_{\mathcal{C}^r} \le 1$ that are furthermore supported in the unit ball around the origin. With this notation, our definition of a model for a given regularity structure $\mathscr{T}$ is as follows.

Definition 2.5. Given a regularity structure $\mathscr{T}$ and an integer $d \geq 1$, a *model* for $\mathscr{T}$ on $\mathbf{R}^d$ consists of maps

$$\Pi : \mathbf{R}^d \to \mathcal{L}\big(T, \mathcal{S}'(\mathbf{R}^d)\big), \qquad \Gamma : \mathbf{R}^d \times \mathbf{R}^d \to G,$$
$$x \mapsto \Pi_x, \qquad (x, y) \mapsto \Gamma_{xy}$$

such that $\Gamma_{xy}\Gamma_{yz} = \Gamma_{xz}$ and $\Pi_x \Gamma_{xy} = \Pi_y$. Furthermore, given $r > |\inf A|$, for any compact set $\mathfrak{K} \subset \mathbf{R}^d$ and constant $\gamma > 0$, there exists a constant C such that the bounds

$$\big|(\Pi_x \tau)(\varphi_x^\lambda)\big| \leq C\lambda^{|\tau|}\|\tau\|_\alpha, \qquad \|\Gamma_{xy}\tau\|_\beta \leq C|x-y|^{\alpha-\beta}\|\tau\|_\alpha, \tag{2.3}$$

hold uniformly over $\varphi \in \mathcal{B}_r$, $(x, y) \in \mathfrak{K}$, $\lambda \in (0, 1]$, $\tau \in T_\alpha$ with $\alpha \leq \gamma$, and $\beta < \alpha$.

Remark 2.6. In principle, test functions appearing in (2.3) should be smooth. It turns out that if these bounds hold for smooth elements of $\mathcal{B}_r$, then $\Pi_x\tau$ can be extended canonically to allow any $\mathcal{C}^r$ test function with compact support.

Remark 2.7. The identity $\Pi_x\Gamma_{xy} = \Pi_y$ reflects the fact that Γ_{xy} is the linear map that takes an expansion around y and turns it into an expansion around x. The first bound in (2.3) states what we mean precisely when we say that $\tau \in T_\alpha$ represents a term that vanishes at order α. (Note that α can be negative, so that this may actually not vanish at all.) The second bound in (2.3) is very natural in view of both (2.1) and (2.2). It states that when expanding a monomial of order α around a new point at distance h from the old one, the coefficient appearing in front of lower-order monomials of order β is of order at most $h^{\alpha-\beta}$.

Remark 2.8. In many cases of interest, it is natural to scale the different directions of $\mathbf{R}^d$ in a different way. This is the case, for example, when using the theory of regularity structures to build solution theories for parabolic stochastic PDEs, in which case the time direction "counts as" two space directions. To deal with such a situation, one can introduce a scaling $\mathfrak{s}$ of $\mathbf{R}^d$, which is just a collection of d mutually prime strictly positive integers and to define φ_x^λ in such a way that the ith direction is scaled by $\lambda^{\mathfrak{s}_i}$. In this case, the Euclidean distance between two points should be replaced everywhere by the corresponding scaled distance $|x|_{\mathfrak{s}} = \sum_i |x_i|^{1/\mathfrak{s}_i}$. See also Hairer (2014) for more details.

With these definitions at hand, it is then natural to define an equivalent in this context of the space of γ-Hölder continuous functions in the following way.

Definition 2.9. Given a regularity structure $\mathscr{T}$ equipped with a model (Π, Γ) over $\mathbf{R}^d$, the space $\mathcal{D}^\gamma = \mathcal{D}^\gamma(\mathscr{T}, \Gamma)$ is given by the set of functions $f : \mathbf{R}^d \to \bigoplus_{\alpha<\gamma} T_\alpha$ such that, for every compact set $\mathfrak{K}$ and every $\alpha < \gamma$, the exists a constant C with

$$\big\|f(x) - \Gamma_{xy} f(y)\big\|_\alpha \leq C|x-y|^{\gamma-\alpha} \tag{2.4}$$

uniformly over $x, y \in \mathfrak{K}$.

The most fundamental result in the theory of regularity structures then states that given $f \in \mathcal{D}^\gamma$ with $\gamma > 0$, there exists a *unique* Schwartz distribution $\mathcal{R}f$ on $\mathbf{R}^d$ such that, for every $x \in \mathbf{R}^d$, $\mathcal{R}f$ "looks like $\Pi_x f(x)$ near x." More precisely, one has the following theorem.

Theorem 2.10. *Let $\mathscr{T}$ be a regularity structure as above and let (Π, Γ) a model for $\mathscr{T}$ on $\mathbf{R}^d$. Then, there exists a unique linear map $\mathcal{R}: \mathcal{D}^\gamma \to \mathcal{S}'(\mathbf{R}^d)$ such that*

$$\big|(\mathcal{R}f - \Pi_x f(x))(\varphi_x^\lambda)\big| \lesssim \lambda^\gamma, \tag{2.5}$$

uniformly over $\varphi \in \mathcal{B}_r$ and λ as before, and locally uniformly in x.

Proof. The proof of the theorem relies on the following fact. Given any $r > 0$ (but finite), there exists a function $\varphi: \mathbf{R}^d \to \mathbf{R}$ with the following properties:

(1) The function φ is of class $\mathcal{C}^r$ and has compact support.
(2) For every polynomial P of degree r, there exists a polynomial $\hat{P}$ of degree r such that, for every $x \in \mathbf{R}^d$, one has $\sum_{y \in \mathbf{Z}^d} \hat{P}(y)\varphi(x - y) = P(x)$.
(3) One has $\int \varphi(x)\varphi(x - y)\,dx = \delta_{y,0}$ for every $y \in \mathbf{Z}^d$.
(4) There exist coefficients $\{a_k\}_{k \in \mathbf{Z}^d}$ such that $2^{-d/2}\varphi(x/2) = \sum_{k \in \mathbf{Z}^d} a_k \varphi(x - k)$.

The existence of such a function φ is highly non-trivial. This is actually equivalent to the existence of a wavelet basis consisting of $\mathcal{C}^r$ functions with compact support, a proof of which was first obtained by Daubechies in her seminal article (Daubechies, 1988). From now on, we take the existence of such a function φ as a given for some $r > |\inf A|$. We also set $\Lambda^n = 2^{-n}\mathbf{Z}^d$ and, for $y \in \Lambda^n$, we set $\varphi_y^n(x) = 2^{nd/2}\varphi(2^n(x - y))$. Here, the normalisation is chosen in such a way that the set $\{\varphi_y^n\}_{y \in \Lambda^n}$ is again orthonormal in L^2. We then denote by $V_n \subset \mathcal{C}^r$ the linear span of $\{\varphi_y^n\}_{y \in \Lambda^n}$, so that, by the property (4) above, one has $V_0 \subset V_1 \subset V_2 \subset \cdots$. We furthermore denote by $\hat{V}_n$ the L^2-orthogonal complement of V_{n-1} in V_n, so that $V_n = V_0 \oplus \hat{V}_1 \oplus \cdots \oplus \hat{V}_n$. In order to keep notations compact, it will also be convenient to define the coefficients a_k^n with $k \in \Lambda^n$ by $a_k^n = a_{2^n k}$.

With these notation at hand, we then define a sequence of linear operators $\mathcal{R}^n: \mathcal{D}^\gamma \to \mathcal{C}^r$ by

$$(\mathcal{R}^n f)(y) = \sum_{x \in \Lambda^n} \big(\Pi_x f(x)\big)(\varphi_x^n)\varphi_x^n(y).$$

We claim that there then exists a Schwartz distribution $\mathcal{R}f$ such that, for every compactly supported test function ψ of class $\mathcal{C}^r$, one has $\langle \mathcal{R}^n f, \psi\rangle \to (\mathcal{R}f)(\psi)$, and that $\mathcal{R}f$ furthermore satisfies the properties stated in the theorem.

Let us first consider the size of the components of $\mathcal{R}^{n+1} f - \mathcal{R}^n f$ in V_n. Given $x \in \Lambda^n$, we make use of properties (3) and (4), so that

$$\begin{aligned}\langle \mathcal{R}^{n+1} f - \mathcal{R}^n f, \varphi_x^n \rangle &= \sum_{k \in \Lambda^{n+1}} a_k^n \langle \mathcal{R}^{n+1} f, \varphi_{x+k}^{n+1} \rangle - \big(\Pi_x f(x)\big)(\varphi_x^n) \\ &= \sum_{k \in \Lambda^{n+1}} a_k^n \big(\Pi_{x+k} f(x+k)\big)(\varphi_{x+k}^{n+1}) - \big(\Pi_x f(x)\big)(\varphi_x^n) \\ &= \sum_{k \in \Lambda^{n+1}} a_k^n \big(\big(\Pi_{x+k} f(x+k)\big)(\varphi_{x+k}^{n+1}) - \big(\Pi_x f(x)\big)(\varphi_{x+k}^{n+1})\big) \\ &= \sum_{k \in \Lambda^{n+1}} a_k^n \big(\Pi_{x+k}\big(f(x+k) - \Gamma_{x+k,x} f(x)\big)\big)(\varphi_{x+k}^{n+1}),\end{aligned}$$

where we used the algebraic relations between Π_x and Γ_{xy} to obtain the last identity. Since only finitely many of the coefficients a_k are non-zero, it follows from the definition of $\mathcal{D}^\gamma$ that for the non-vanishing terms in this sum we have the bound

$$\big\| f(x+k) - \Gamma_{x+k,x} f(x) \big\|_\alpha \lesssim 2^{-n(\gamma-\alpha)},$$

uniformly over $n \geq 0$ and x in any compact set. Furthermore, for any $\tau \in T_\alpha$, it follows from the definition of a model that one has the bound

$$\big|(\Pi_x \tau)(\varphi_x^n)\big| \lesssim 2^{-\alpha n - nd/2},$$

again uniformly over $n \geq 0$ and x in any compact set. Here, the additional factor $2^{-nd/2}$ comes from the fact that the functions φ_x^n are normalised in L^2 rather than L^1. Combining these two bounds, we immediately obtain that

$$\big|\langle \mathcal{R}^{n+1} f - \mathcal{R}^n f, \varphi_x^n \rangle\big| \lesssim 2^{-\gamma n - nd/2}, \tag{2.6}$$

uniformly over $n \geq 0$ and x in compact sets. Take now a test function $\psi \in \mathcal{C}^r$ with compact support and let us try to estimate $\langle \mathcal{R}^{n+1} f - \mathcal{R}^n f, \psi \rangle$. Since $\mathcal{R}^{n+1} f - \mathcal{R}^n f \in V_{n+1}$, we can decompose it into a part $\delta \mathcal{R}^n f \in V_n$ and a part $\hat{\delta} \mathcal{R}^n f \in \hat{V}_{n+1}$ and estimate both parts separately. Regarding the part in V_n, we have

$$\big|\langle \delta \mathcal{R}^n f, \psi \rangle\big| = \Big| \sum_{x \in \Lambda^{n+1}} \langle \delta \mathcal{R}^n f, \varphi_x^n \rangle \langle \varphi_x^n, \psi \rangle \Big| \lesssim 2^{-\gamma n - nd/2} \sum_{x \in \Lambda^{n+1}} \big|\langle \varphi_x^n, \psi \rangle\big|, \tag{2.7}$$

where we made use of the bound (2.6). At this stage we use the fact that, due to the boundedness of ψ, we have $|\langle \varphi_x^n, \psi \rangle| \lesssim 2^{-nd/2}$. Furthermore, thanks to the boundedness of the support of ψ, the number of non-vanishing terms appearing in this sum is bounded by 2^{nd}, so that we eventually obtain the bound

$$\big|\langle \delta \mathcal{R}^n f, \psi \rangle\big| \lesssim 2^{-\gamma n}. \tag{2.8}$$

Regarding the second term, we use the standard fact coming from wavelet analysis (Meyer, 1992) that a basis of $\hat{V}_{n+1}$ can be obtained in the same way as the basis of V_n, but replacing the function φ by functions $\hat{\varphi}$ from some finite set Φ. In other

words, $\hat{V}_{n+1}$ is the linear span of $\{\hat{\varphi}_x^n\}_{x\in\Lambda^n;\hat{\varphi}\in\Phi}$. Furthermore, as a consequence of property (2), the functions $\hat{\varphi}\in\Phi$ all have the property that

$$\int \hat{\varphi}(x)P(x)\,dx = 0, \tag{2.9}$$

for any polynomial P of degree less or equal to r. In particular, this shows that one has the bound

$$\bigl|\langle\hat{\varphi}_x^n,\psi\rangle\bigr| \lesssim 2^{-nd/2-nr}.$$

As a consequence, one has

$$\bigl|\langle\hat{\delta}\mathcal{R}^n f,\psi\rangle\bigr| = \Bigl|\sum_{\substack{x\in\Lambda^n\\ \hat{\varphi}\in\Phi}} \langle\mathcal{R}^{n+1}f,\hat{\varphi}_x^n\rangle\langle\hat{\varphi}_x^n,\psi\rangle\Bigr| \lesssim 2^{-nd/2-nr}\Bigl|\sum_{\substack{x\in\Lambda^n\\ \hat{\varphi}\in\Phi}} \langle\mathcal{R}^{n+1}f,\hat{\varphi}_x^n\rangle\Bigr|.$$

At this stage, we note that, thanks to the definition of $\mathcal{R}^{n+1}$ and the bounds on the model (Π,Γ), we have $|\langle\mathcal{R}^{n+1}f,\hat{\varphi}_x^n\rangle| \lesssim 2^{-nd/2-\alpha_0 n}$, where $\alpha_0 = \inf A$, so that $|\langle\hat{\delta}\mathcal{R}^n f,\psi\rangle| \lesssim 2^{-nr-\alpha_0 n}$. Combining this with (2.8), we see that one has indeed $\mathcal{R}^n f \to \mathcal{R}f$ for some Schwartz distribution $\mathcal{R}f$.

It remains to show that the bound (2.5) holds. For this, given a distribution $\eta\in\mathcal{C}^\alpha$ for some $\alpha > -r$, we first introduce the notation

$$\mathcal{P}_n\eta = \sum_{x\in\Lambda^n}\eta(\varphi_x^n)\varphi_x^n, \qquad \hat{\mathcal{P}}_n\eta = \sum_{\hat{\varphi}\in\Phi}\sum_{x\in\Lambda^n}\eta(\hat{\varphi}_x^n)\hat{\varphi}_x^n.$$

We also choose an integer value $n\geq 0$ such that $2^{-n}\sim\lambda$ and we write

$$\begin{aligned}
&\mathcal{R}f - \Pi_x f(x)\\
&\quad= \mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x) + \sum_{m\geq n}\bigl(\mathcal{R}^{m+1}f - \mathcal{R}^m f - \hat{\mathcal{P}}_m\Pi_x f(x)\bigr) \qquad (2.10)\\
&\quad= \mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x) + \sum_{m\geq n}\bigl(\hat{\delta}\mathcal{R}^m f - \hat{\mathcal{P}}_m\Pi_x f(x)\bigr) + \sum_{m\geq n}\delta\mathcal{R}^m f.
\end{aligned}$$

We then test these terms against ψ_x^λ and we estimate the resulting terms separately. For the first term, we have the identity

$$\bigl(\mathcal{R}^n f - \mathcal{P}_n\Pi_x f(x)\bigr)\bigl(\psi_x^\lambda\bigr) = \sum_{y\in\Lambda^n}\bigl(\Pi_y f(y) - \Pi_x f(x)\bigr)\bigl(\varphi_y^n\bigr)\langle\varphi_y^n,\psi_x^\lambda\rangle. \tag{2.11}$$

We have the bound $|\langle\varphi_y^n,\psi_x^\lambda\rangle| \lesssim \lambda^{-d}2^{-dn/2} \sim 2^{dn/2}$. Since one furthermore has $|y-x|\lesssim\lambda$ for all non-vanishing terms in the sum, one also has similarly to before

$$\bigl|\bigl(\Pi_y f(y) - \Pi_x f(x)\bigr)\bigl(\varphi_y^n\bigr)\bigr| \lesssim \sum_{\alpha<\gamma}\lambda^{\gamma-\alpha}2^{-dn/2-\alpha n} \sim 2^{-dn/2-\gamma n}. \tag{2.12}$$

Since only finitely many (independently of n) terms contribute to the sum in (2.11), it is indeed bounded by a constant proportional to $2^{-\gamma n}\sim\lambda^\gamma$ as required.

We now turn to the second term in (2.10), where we consider some fixed value $m \geq n$. We rewrite this term very similarly to before as

$$\begin{aligned}&\big(\hat{\delta}\mathcal{R}^m f - \hat{\mathcal{P}}_m \Pi_x f(x)\big)\big(\psi_x^\lambda\big)\\ &\quad= \sum_{\hat{\varphi}\in\Phi}\sum_{y,z}\big(\Pi_y f(y) - \Pi_x f(x)\big)\big(\varphi_y^{m+1}\big)\langle\varphi_y^{m+1}, \hat{\varphi}_z^m\rangle\langle\hat{\varphi}_z^m, \psi_x^\lambda\rangle,\end{aligned}$$

where the sum runs over $y \in \Lambda^{m+1}$ and $z \in \Lambda^m$. This time, we use the fact that by the property (2.9) of the wavelets $\hat{\varphi}$, one has the bound

$$\big|\langle\hat{\varphi}_z^m, \psi_x^\lambda\rangle\big| \lesssim \lambda^{-d-r}2^{-rm-md/2}, \tag{2.13}$$

and the L^2-scaling implies that $|\langle\varphi_y^{m+1}, \hat{\varphi}_z^m\rangle| \lesssim 1$. Furthermore, for each $z \in \Lambda^m$, only finitely many elements $y \in \Lambda^{m+1}$ contribute to the sum, and these elements all satisfy $|y - z| \lesssim 2^{-m}$. Bounding the first factor as in (2.12) and using the fact that there are of the order of $\lambda^d 2^{md}$ terms contributing for every fixed m, we thus see that the contribution of the second term in (2.10) is bounded by

$$\begin{aligned}&\sum_{m\geq n}\lambda^d 2^{md}\sum_{\alpha<\gamma}\lambda^{\gamma-\alpha-d-r}2^{-dm-\alpha m-rm}\\ &\quad\sim \sum_{\alpha<\gamma}\lambda^{\gamma-\alpha-r}\sum_{m\geq n}2^{-\alpha m-rm} \sim \lambda^\gamma.\end{aligned}$$

For the last term in (2.10), we combine (2.7) with the bound $|\langle\varphi_y^m, \psi_x^\lambda\rangle| \lesssim \lambda^{-d}2^{-dm/2}$ and the fact that there are of the order of $\lambda^d 2^{-md}$ terms appearing in the sum (2.7) to conclude that the mth summand is bounded by a constant proportional to $2^{-\gamma m}$. Summing over m yields again the desired bound and completes the proof. □

Remark 2.11. Note that the space $\mathcal{D}^\gamma$ depends crucially on the choice of model (Π, Γ). As a consequence, the reconstruction operator $\mathcal{R}$ itself also depends on that choice. However, the map $(\Pi, \Gamma, f) \mapsto \mathcal{R}f$ turns out to be locally Lipschitz continuous provided that the distance between (Π, Γ, f) and $(\bar{\Pi}, \bar{\Gamma}, \bar{f})$ is given by the smallest constant ϱ such that

$$\begin{aligned}\big\|f(x) - \bar{f}(x) - \Gamma_{xy}f(y) + \bar{\Gamma}_{xy}\bar{f}(y)\big\|_\alpha &\leq \varrho|x-y|^{\gamma-\alpha},\\ \big|(\Pi_x\tau - \bar{\Pi}_x\tau)\big(\varphi_x^\lambda\big)\big| &\leq \varrho\lambda^\alpha\|\tau\|,\\ \|\Gamma_{xy}\tau - \bar{\Gamma}_{xy}\tau\|_\beta &\leq \varrho|x-y|^{\alpha-\beta}\|\tau\|.\end{aligned}$$

Here, in order to obtain bounds on $(\mathcal{R}f - \bar{\mathcal{R}}\bar{f})(\psi)$ for some smooth compactly supported test function ψ, the above bounds should hold uniformly for x and y in a neighbourhood of the support of ψ. The proof that this stronger continuity property also holds is actually crucial when showing that sequences of solutions to mollified equations all converge to the same limiting object. However, its proof is somewhat more involved which is why we chose not to give it here.

Remark 2.12. In the particular case where $\Pi_x\tau$ happens to be a continuous function for every $\tau \in T$ (and every $x \in \mathbf{R}^d$), $\mathcal{R}f$ is also a continuous function and one has the identity

$$(\mathcal{R}f)(x) = \big(\Pi_x f(x)\big)(x). \tag{2.14}$$

This can be seen from the fact that

$$(\mathcal{R}f)(y) = \lim_{n\to\infty} \big(\mathcal{R}^n f\big)(y) = \lim_{n\to\infty} \sum_{x\in\Lambda^n} \big(\Pi_x f(x)\big)\big(\varphi_x^n\big)\varphi_x^n(y).$$

Indeed, our assumptions imply that the function $(x, z) \mapsto (\Pi_x f(x))(z)$ is jointly continuous and since the non-vanishing terms in the above sum satisfy $|x - y| \lesssim 2^{-n}$, one has $2^{dn/2}(\Pi_x f(x))(\varphi_x^n) \approx (\Pi_y f(y))(y)$ for large n. Since furthermore $\sum_{x\in\Lambda^n} \varphi_x^n(y) = 2^{dn/2}$, the claim follows.

3 Examples of regularity structures

3.1 The polynomial structure

It should by now be clear how the structure given by the usual Taylor polynomials fits into this framework. A natural way of setting it up is to take for T the space of all abstract polynomials in d commuting variables, denoted by $X_1, \dots, X_d$, and to postulate that T_k consists of the linear span of monomials of degree k. As an abstract group, the structure group G is then given by $\mathbf{R}^d$ endowed with addition as its group operation, which acts onto T via $\Gamma_h X^k = (X - h)^k$, where $h \in \mathbf{R}^d$ and we use the notation X^k as a shorthand for $X_1^{k_1} \cdots X_d^{k_d}$ for any multiindex k.

The canonical polynomial model is then given by

$$\big(\Pi_x X^k\big)(y) = (y - x)^k, \qquad \Gamma_{xy} = \Gamma_{y-x}.$$

We leave it as an exercise to the reader to verify that this does indeed satisfy the bounds and relations of Definition 2.5.

In the particular case of the canonical polynomial model and for $\gamma \notin \mathbf{N}$, the spaces $\mathcal{D}^\gamma$ then coincide precisely with the usual Hölder spaces $\mathcal{C}^\gamma$. In the case of integer values, this should be interpreted as bounded functions for $\gamma = 0$, Lipschitz continuous functions for $\gamma = 1$, etc.

3.2 Controlled rough paths

Let us see now how the theory of controlled rough paths can be reinterpreted in the light of this theory. For given $\alpha \in (\frac{1}{3}, \frac{1}{2})$ and $n \geq 1$, we can define a regularity structure $\mathscr{T}$ by setting $A = \{\alpha - 1, 2\alpha - 1, 0, \alpha\}$. We furthermore take for T_0 a copy of $\mathbf{R}$ with unit vector $\mathbf{1}$, for T_α and $T_{\alpha-1}$ a copy of $\mathbf{R}^n$ with respective unit

vectors $\mathbb{W}_j$ and Ξ_j, and for $T_{2\alpha-1}$ a copy of $\mathbf{R}^{n\times n}$ with unit vectors $\mathbb{W}_j\Xi_i$. The structure group G is taken to be isomorphic to $\mathbf{R}^n$ and, for $x\in\mathbf{R}^n$, it acts on T via

$$\Gamma_x\mathbf{1}=\mathbf{1},\qquad \Gamma_x\Xi_i=\Xi_i,$$
$$\Gamma_x\mathbb{W}_i=\mathbb{W}_i-x^i\mathbf{1},\qquad \Gamma_x(\mathbb{W}_j\Xi_i)=\mathbb{W}_j\Xi_i-x^j\Xi_i.$$

Let now $\mathbf{X}=(X,\mathbb{X})$ be an α-Hölder continuous rough path with values in $\mathbf{R}^n$. In other words, the functions X and $\mathbb{X}$ are as in the Introduction, satisfying the relation (1.4) and the analytic bounds $|X_t-X_s|\lesssim|t-s|^\alpha$, $|\mathbb{X}_{s,t}|\lesssim|t-s|^{2\alpha}$. It turns out that this defines a model for $\mathscr{T}$ in the following way (recall that $X_{s,t}$ is a shorthand for X_t-X_s):

Lemma 3.1. *Given an α-Hölder continuous rough path $\mathbf{X}$, one can define a model for $\mathscr{T}$ on $\mathbf{R}$ by setting $\Gamma_{su}=\Gamma_{X_{s,u}}$ and*

$$(\Pi_s\mathbf{1})(t)=1,\qquad (\Pi_s\mathbb{W}_j)(t)=X^j_{s,t},$$
$$(\Pi_s\Xi_j)(\psi)=\int\psi(t)\,dX^j_t,\qquad (\Pi_s\mathbb{W}_j\Xi_i)(\psi)=\int\psi(t)\,d\mathbb{X}^{i,j}_{s,t}.$$

Here, both integrals are perfectly well-defined Riemann integrals, with the differential in the second case taken with respect to the variable t. Given a controlled rough path $(Y,Y')\in\mathcal{C}^\alpha_X$ as in (1.2), this then defines an element $\hat{Y}\in\mathcal{D}^{2\alpha}$ by setting

$$\hat{Y}(s)=Y(s)\mathbf{1}+Y'_i(s)\mathbb{W}_i,$$

with summation over i implied.

Proof. We first check that the algebraic properties of Definition 2.5 are satisfied. It is clear that $\Gamma_{su}\Gamma_{ut}=\Gamma_{st}$ and that $\Pi_s\Gamma_{su}\tau=\Pi_u\tau$ for $\tau\in\{\mathbf{1},\mathbb{W}_j,\Xi_j\}$. Regarding $\mathbb{W}_j\Xi_i$, we differentiate Chen's relations (1.4) which yields the identity

$$d\mathbb{X}^{i,j}_{s,t}=d\mathbb{X}^{i,j}_{u,t}+X^i_{s,u}\,dX^j_t.$$

The last missing algebraic relation then follows at once. The required analytic bounds follow immediately from the definition of the rough path space $\mathcal{D}^\alpha$.

Regarding the function $\hat{Y}$ defined in the statement, we have

$$\|\hat{Y}(s)-\Gamma_{su}\hat{Y}(u)\|_0=|Y(s)-Y(u)+Y'_i(u)X^i_{s,u}|,$$
$$\|\hat{Y}(s)-\Gamma_{su}\hat{Y}(u)\|_\alpha=|Y'(s)-Y'(u)|,$$

so that the condition (2.4) with $\gamma=2\alpha$ does indeed coincide with the definition of a controlled rough path given in the Introduction. □

In this context, the reconstruction theorem allows us to define an integration operator with respect to W. We can formulate this as follows where one should really think of Z as providing a consistent definition of what one means by $\int Y\,dX^j$.

Lemma 3.2. *In the same context as above, let $\alpha \in (\frac{1}{3}, \frac{1}{2})$, and consider $\hat{Y} \in \mathcal{D}^{2\alpha}$ built as above from a controlled rough path. Then the map $\hat{Y}\,\Xi_i$ given by*

$$(\hat{Y}\,\Xi_j)(s) = Y(s)\,\Xi_j + Y_i'(s)\mathbb{W}_i\,\Xi_j$$

belongs to $\mathcal{D}^{3\alpha-1}$. Furthermore, there exists a function Z such that, for every smooth test function ψ, one has

$$(\mathcal{R}\hat{Y}\,\Xi_j)(\psi) = \int \psi(t)\,dZ(t),$$

and such that $Z_{s,t} = Y(s)X^j_{s,t} + Y_i'(s)\mathbb{X}^{i,j}_{s,t} + \mathcal{O}(|t-s|^{3\alpha})$.

Proof. The fact that $\hat{Y}\,\Xi_i \in \mathcal{D}^{3\alpha-1}$ is an immediate consequence of the definitions. Since $\alpha > \frac{1}{3}$ by assumption, we can apply the reconstruction theorem to it, from which it follows that there exists a unique distribution η such that, if ψ is a smooth compactly supported test function, one has

$$\eta(\psi^\lambda_s) = \int \psi^\lambda_s(t)Y(s)\,dX^j_t + \int \psi^\lambda_s(t)Y_i'(s)\,d\mathbb{X}^{i,j}_{s,t} + \mathcal{O}(\lambda^{3\alpha-1}).$$

By a simple approximation argument, it turns out that one can take for ψ the indicator function of the interval $[0, 1]$, so that

$$\eta(\mathbf{1}_{[s,t]}) = Y(s)X^j_{s,t} + Y_i'(s)\mathbb{X}^{i,j}_{s,t} + \mathcal{O}(|t-s|^{3\alpha}).$$

Here, the reason why one obtains an exponent 3α rather than $3\alpha - 1$ is that it is really $|t-s|^{-1}\mathbf{1}_{[s,t]}$ that scales like an approximate δ-distribution as $t \to s$. □

Remark 3.3. Using the formula (2.14), it is straightforward to verify that if X happens to be a smooth function and $\mathbb{X}$ is defined from X via (1.3), but this time viewing it as a definition for the left-hand side, with the right-hand side given by a usual Riemann integral, then the function Z constructed in Lemma 3.2 coincides with the usual Riemann integral of Y against X^j.

3.3 A classical result from harmonic analysis

The considerations above suggest that a very natural space of distributions is obtained in the following way. For some $\alpha > 0$, we denote by $\mathcal{C}^{-\alpha}$ the space of all Schwartz distributions η such that η belongs to the dual of $\mathcal{C}^r$ with $r > \alpha$ some integer and such that

$$|\eta(\varphi^\lambda_x)| \lesssim \lambda^{-\alpha},$$

uniformly over all $\varphi \in \mathcal{B}_r$ and $\lambda \in (0, 1]$, and locally uniformly in x. Given any compact set $\mathfrak{K}$, the best possible constant such that the above bound holds uniformly over $x \in \mathfrak{K}$ yields a seminorm. The collection of these seminorms endows $\mathcal{C}^{-\alpha}$ with a Fréchet space structure.

Remark 3.4. It turns out that the space $\mathcal{C}^{-\alpha}$ is independent of the choice of r in the definition given above, which justifies the notation. Different values of r give raise to equivalent seminorms.

Remark 3.5. In terms of the scale of classical Besov spaces, the space $\mathcal{C}^{-\alpha}$ is a local version of $\mathcal{B}^{-\alpha}_{\infty,\infty}$. It is in some sense the largest space of distributions that is invariant under the scaling $\varphi(\cdot) \mapsto \lambda^{-\alpha}\varphi(\lambda^{-1}\cdot)$; see, for example, Bourgain and Pavlović (2008).

It is then a classical result in the "folklore" of harmonic analysis that the product extends naturally to $\mathcal{C}^{-\alpha} \times \mathcal{C}^{\beta}$ into $\mathcal{S}'(\mathbf{R}^d)$ if and only if $\beta > \alpha$. The reconstruction theorem yields a straightforward proof of the "if" part of this result.

Theorem 3.6. *There is a continuous bilinear map $B : \mathcal{C}^{-\alpha} \times \mathcal{C}^{\beta} \to \mathcal{S}'(\mathbf{R}^d)$ such that $B(f, g) = fg$ for any two continuous functions f and g.*

Proof. Assume from now on that $\xi \in \mathcal{C}^{-\alpha}$ for some $\alpha > 0$ and that $f \in \mathcal{C}^{\beta}$ for some $\beta > \alpha$. We then build a regularity structure $\mathscr{T}$ in the following way. For the set A, we take $A = \mathbf{N} \cup (\mathbf{N} - \alpha)$ and for T, we set $T = V \oplus W$, where each one of the spaces V and W is a copy of the polynomial model in d commuting variables constructed in Section 3.1. We also choose Γ as in the canonical model, acting simultaneously on each of the two instances.

As before, we denote by X^k the canonical basis vectors in V. We also use the suggestive notation "ΞX^k" for the corresponding basis vector in W, but we postulate that $\Xi X^k \in T_{\alpha+|k|}$ rather than $\Xi X^k \in T_{|k|}$. Given any distribution $\xi \in \mathcal{C}^{-\alpha}$, we then define a model (Π^{ξ}, Γ), where Γ is as in the canonical model, while Π^{ξ} acts as

$$\bigl(\Pi^{\xi}_x X^k\bigr)(y) = (y - x)^k, \qquad \bigl(\Pi^{\xi}_x \Xi X^k\bigr)(y) = (y - x)^k \xi(y),$$

with the obvious abuse of notation in the second expression. It is then straightforward to verify that $\Pi_y = \Pi_x \circ \Gamma_{xy}$ and that the relevant analytical bounds are satisfied, so that this is indeed a model.

Denote now by $\mathcal{R}^{\xi}$ the reconstruction map associated to the model (Π^{ξ}, Γ) and, for $f \in \mathcal{C}^{\beta}$, denote by F the element in $\mathcal{D}^{\beta}$ given by the local Taylor expansion of f of order β at each point. Note that even though the space $\mathcal{D}^{\beta}$ does in principle depend on the choice of model, in our situation $F \in \mathcal{D}^{\beta}$ for any choice of ξ. It follows immediately from the definitions that the map $x \mapsto \Xi F(x)$ belongs to $\mathcal{D}^{\beta-\alpha}$ so that, provided that $\beta > \alpha$, one can apply the reconstruction operator to it. This suggests that the multiplication operator we are looking for can be defined as

$$B(f, \xi) = \mathcal{R}^{\xi}(\Xi F).$$

By Theorem 2.10, this is a jointly continuous map from $\mathcal{C}^{\beta} \times \mathcal{C}^{-\alpha}$ into $\mathcal{S}'(\mathbf{R}^d)$, provided that $\beta > \alpha$. If ξ happens to be a smooth function, then it follows immediately from Remark 2.12 that $B(f, \xi) = f(x)\xi(x)$, so that B is indeed the requested continuous extension of the usual product. □

Remark 3.7. As a consequence of (2.5), it is actually easy to show that $B : \mathcal{C}^{-\alpha} \times \mathcal{C}^{\beta} \to \mathcal{C}^{-\alpha}$.

4 Products and composition by smooth functions

One of the main purposes of the theory presented here is to give a robust way to multiply distributions (or functions with distributions) that goes beyond the barrier illustrated by Theorem 3.6. Provided that our functions/distributions are represented as elements in $\mathcal{D}^{\gamma}$ for some model and regularity structure, we can multiply their "Taylor expansions" pointwise, provided that we give ourselves a table of multiplication on T.

It is natural to consider products with the following properties. Here, given a regularity structure, we say that a subspace $V \subset T$ is a *sector* if it is invariant under the action of the structure group G and if it can furthermore be written as $V = \bigoplus_{\alpha \in A} V_\alpha$ with $V_\alpha \subset T_\alpha$.

Definition 4.1. Given a regularity structure (T, A, G) and two sectors $V, \bar{V} \subset T$, a *product* on $(V, \bar{V})$ is a bilinear map $\star : V \times \bar{V} \to T$ such that, for any $\tau \in V_\alpha$ and $\bar{\tau} \in \bar{V}_\beta$, one has $\tau \star \bar{\tau} \in T_{\alpha+\beta}$ and such that, for any element $\Gamma \in G$, one has $\Gamma(\tau \star \bar{\tau}) = \Gamma\tau \star \Gamma\bar{\tau}$.

Remark 4.2. The condition that homogeneities add up under multiplication is very natural bearing in mind the case of the polynomial regularity structure. The second condition is also very natural since it merely states that if one re-expands the product of two "polynomials" around a different point, one should obtain the same result as if one reexpands each factor first and then multiplies them together.

Given such a product, we can ask ourselves when the pointwise product of an element $\mathcal{D}^{\gamma_1}$ with an element in $\mathcal{D}^{\gamma_2}$ again belongs to some $\mathcal{D}^{\gamma}$. In order to answer this question, we introduce the notation $\mathcal{D}^{\gamma}_{\alpha}$ to denote those elements $f \in \mathcal{D}^{\gamma}$ such that furthermore

$$f(x) \in T_\alpha^+ \equiv \bigoplus_{\beta \geq \alpha} T_\beta,$$

for every x. With this notation at hand, it is not too difficult to verify that one has the following result.

Theorem 4.3. *Let $f_1 \in \mathcal{D}^{\gamma_1}_{\alpha_1}(V)$, $f_2 \in \mathcal{D}^{\gamma_2}_{\alpha_2}(\bar{V})$, and let $\star$ be a product on $(V, \bar{V})$. Then, the function f given by $f(x) = f_1(x) \star f_2(x)$ belongs to $\mathcal{D}^{\gamma}_{\alpha}$ with*

$$\alpha = \alpha_1 + \alpha_2, \qquad \gamma = (\gamma_1 + \alpha_2) \wedge (\gamma_2 + \alpha_1). \tag{4.1}$$

Proof. It is clear that $f(x) \in \bigoplus_{\beta>\alpha} T_\beta$, so it remains to show that it belongs to $\mathcal{D}^\gamma$. Furthermore, since we are only interested in showing that $f_1 \star f_2 \in \mathcal{D}^\gamma$, we discard all of the components in T_β for $\beta \geq \gamma$.

By the properties of the product $\star$, it remains to obtain a bound of the type

$$\|\Gamma_{xy} f_1(y) \star \Gamma_{xy} f_2(y) - f_1(x) \star f_2(x)\|_\beta \lesssim |x-y|^{\gamma-\beta}.$$

By adding and subtracting suitable terms, we obtain

$$\begin{aligned}\|\Gamma_{xy} f(y) - f(x)\|_\beta \leq{}& \|(\Gamma_{xy} f_1(y) - f_1(x)) \star (\Gamma_{xy} f_2(y) - f_2(x))\|_\beta \\ &+ \|(\Gamma_{xy} f_1(y) - f_1(x)) \star f_2(x)\|_\beta \\ &+ \|f_1(x) \star (\Gamma_{xy} f_2(y) - f_2(x))\|_\beta.\end{aligned} \tag{4.2}$$

It follows from the properties of the product $\star$ that the first term in (4.2) is bounded by a constant times

$$\begin{aligned}&\sum_{\beta_1+\beta_2=\beta} \|\Gamma_{xy} f_1(y) - f_1(x)\|_{\beta_1} \|\Gamma_{xy} f_2(y) - f_2(x)\|_{\beta_2} \\ &\quad\lesssim \sum_{\beta_1+\beta_2=\beta} \|x-y\|^{\gamma_1-\beta_1} \|x-y\|^{\gamma_2-\beta_2} \lesssim \|x-y\|^{\gamma_1+\gamma_2-\beta}.\end{aligned}$$

Since $\gamma_1 + \gamma_2 \geq \gamma$, this bound is as required. The second term is bounded by a constant times

$$\begin{aligned}&\sum_{\beta_1+\beta_2=\beta} \|\Gamma_{xy} f_1(y) - f_1(x)\|_{\beta_1} \|f_2(x)\|_{\beta_2} \\ &\quad\lesssim \|x-y\|^{\gamma_1-\beta_1} \mathbf{1}_{\beta_2 \geq \alpha_2} \lesssim \|x-y\|^{\gamma_1+\alpha_2-\beta},\end{aligned}$$

where the second inequality uses the identity $\beta_1 + \beta_2 = \beta$. Since $\gamma_1 + \alpha_2 \geq \gamma$, this bound is again of the required type. The last term is bounded similarly by reversing the roles played by f_1 and f_2. □

Remark 4.4. It is clear that the formula (4.1) for γ is optimal in general as can be seen from the following two "reality checks." First, consider the case of the polynomial model and take $f_i \in \mathcal{C}^{\gamma_i}$. In this case, the truncated Taylor series F_i for f_i belong to $\mathcal{D}_0^{\gamma_i}$. It is clear that in this case, the product cannot be expected to have better regularity than $\gamma_1 \wedge \gamma_2$ in general, which is indeed what (4.1) states. The second reality check comes from the example of Section 3.3. In this case, one has $F \in \mathcal{D}_0^\beta$, while the constant function $x \mapsto \Xi$ belongs to $\mathcal{D}_{-\alpha}^\infty$ so that, according to (4.1), one expects their product to belong to $\mathcal{D}_{-\alpha}^{\beta-\alpha}$, which is indeed the case.

It turns out that if we have a product on a regularity structure, then in many cases this also naturally yields a notion of composition with smooth functions. Of course, one could in general not expect to be able to compose a smooth function

with a distribution of negative order. As a matter of fact, we will only define the composition of smooth functions with elements in some $\mathcal{D}^\gamma$ for which it is guaranteed that the reconstruction operator yields a continuous function. One might think at this case that this would yield a triviality, since we know of course how to compose arbitrary continuous function. The subtlety is that we would like to design our composition operator in such a way that the result is again an element of $\mathcal{D}^\gamma$.

For this purpose, we say that a given sector $V \subset T$ is *function-like* if $\alpha < 0 \Rightarrow V_\alpha = 0$ and if V_0 is one-dimensional. (Denote the unit vector of V_0 by $\mathbf{1}$.) We will furthermore always assume that our models are *normal* in the sense that $(\Pi_x \mathbf{1})(y) = 1$. I this case, it turns out that if $f \in \mathcal{D}^\gamma(V)$, then $\mathcal{R}f$ is a continuous function and one has the identity $(\mathcal{R}f)(x) = \langle \mathbf{1}, f(x)\rangle$, where we denote by $\langle \mathbf{1}, \cdot\rangle$ the element in the dual of V which picks out the pre-factor of $\mathbf{1}$.

Assume now that we are given a regularity structure with a function-like sector V and a product $\star : V \times V \to V$. For any smooth function $G : \mathbf{R} \to \mathbf{R}$ and any $f \in \mathcal{D}^\gamma(V)$ with $\gamma > 0$, we can then *define* $G(f)$ to be the V-valued function given by

$$(G \circ f)(x) = \sum_{k \geq 0} \frac{G^{(k)}(\bar{f}(x))}{k!} \tilde{f}(x)^{\star k},$$

where we have set

$$\bar{f}(x) = \langle \mathbf{1}, f(x)\rangle, \qquad \tilde{f}(x) = f(x) - \bar{f}(x)\mathbf{1}.$$

Here, $G^{(k)}$ denotes the kth derivative of G and $\tau^{\star k}$ denotes the k-fold product $\tau \star \cdots \star \tau$. We also used the usual conventions $G^{(0)} = G$ and $\tau^{\star 0} = \mathbf{1}$.

Note that as long as G is $\mathcal{C}^\infty$, this expression is well-defined. Indeed, by assumption, there exists some $\alpha_0 > 0$ such that $\tilde{f}(x) \in T^+_{\alpha_0}$. By the properties of the product, this implies that one has $\tilde{f}(x)^{\star k} \in T^+_{k\alpha_0}$. As a consequence, when considering the component of $G \circ f$ in T_β for $\beta < \gamma$, the only terms that give a contribution are those with $k < \gamma/\alpha_0$. Since we cannot possibly hope in general that $G \circ f \in \mathcal{D}^{\gamma'}$ for some $\gamma' > \gamma$, this is all we really need.

It turns out that if G is sufficiently regular, then the map $f \mapsto G \circ f$ enjoys similarly nice continuity properties to what we are used to from classical Hölder spaces. The following result is the analogue in this context to the well-known fact that the composition of a $\mathcal{C}^\gamma$ function with a sufficiently smooth function G is again of class $\mathcal{C}^\gamma$.

Proposition 4.5. *In the same setting as above, provided that G is of class $\mathcal{C}^k$ with $k > \gamma/\alpha_0$, the map $f \mapsto G \circ f$ is continuous from $\mathcal{D}^\gamma(V)$ into itself. If $k > \gamma/\alpha_0 + 1$, then it is locally Lipschitz continuous.*

The proof of this result can be found in Hairer (2014). It is somewhat lengthy, but ultimately rather straightforward.

4.1 A simple example

A very important remark is that even if both $\mathcal{R}f_1$ and $\mathcal{R}f_2$ happens to be continuous functions, this does *not* in general imply that $\mathcal{R}(f_1 \star f_2)(x) = (\mathcal{R}f_1)(x)(\mathcal{R}f_2)(x)$. For example, fix $\kappa < 0$ and consider the regularity structure given by $A = (-2\kappa, -\kappa, 0)$, with each T_α being a copy of $\mathbf{R}$ given by $T_{-n\kappa} = \langle \Xi^n \rangle$. We furthermore take for G the trivial group. This regularity structure comes with an obvious product by setting $\Xi^m \star \Xi^n = \Xi^{m+n}$ provided that $m + n \le 2$.

Then, we could for example take as a model for $\mathscr{T} = (T, A, G)$:

$$\big(\Pi_x \Xi^0\big)(y) = 1, \qquad (\Pi_x \Xi)(y) = 0, \qquad \big(\Pi_x \Xi^2\big)(y) = c, \tag{4.3}$$

where c is an arbitrary constant. Let furthermore

$$F_1(x) = f_1(x)\Xi^0 + f_1'(x)\Xi, \qquad F_2(x) = f_2(x)\Xi^0 + f_2'(x)\Xi.$$

Since our group G is trivial, one has $F_i \in \mathcal{D}^\gamma$ provided that each of the f_i belongs to $\mathcal{D}^\gamma$ and each of the f_i' belongs to $\mathcal{D}^{\gamma+\kappa}$. (And one has $\gamma + \kappa < 1$.) One furthermore has the identity $(\mathcal{R}F_i)(x) = f_i(x)$.

However, the pointwise product is given by

$$(F_1 \star F_2)(x) = f_1(x)f_2(x)\Xi^0 + \big(f_1'(x)f_2(x) + f_2'(x)f_1(x)\big)\Xi + f_1'(x)f_2'(x)\Xi^2,$$

which by Theorem 4.3 belongs to $\mathcal{D}^{\gamma-\kappa}$. Provided that $\gamma > \kappa$, one can then apply the reconstruction operator to this product and one obtains

$$\mathcal{R}(F_1 \star F_2)(x) = f_1(x)f_2(x) + cf_1'(x)f_2'(x),$$

which is obviously different from the pointwise product $\mathcal{R}F_1 \cdot \mathcal{R}F_2$.

How should this be interpreted? For $n > 0$, we could have defined a model $\Pi^{(n)}$ by

$$\big(\Pi_x \Xi^0\big)(y) = 1, \qquad (\Pi_x \Xi)(y) = \sqrt{2c}\sin(nx), \qquad \big(\Pi_x \Xi^2\big)(y) = 2c\sin^2(nx).$$

Denoting by $\mathcal{R}^{(n)}$ the corresponding reconstruction operator, we have the identity

$$\big(\mathcal{R}^{(n)}F_i\big)(x) = f_i(x) + \sqrt{2c}\,f_i'(x)\sin(nx),$$

as well as $\mathcal{R}^{(n)}(F_1 \star F_2) = \mathcal{R}^{(n)}F_1 \cdot \mathcal{R}^{(n)}F_2$. As a model, the model $\Pi^{(n)}$ actually converges to the limiting model Π defined in (4.3). As a consequence of the continuity of the reconstruction operator, this implies that

$$\mathcal{R}^{(n)}F_1 \cdot \mathcal{R}^{(n)}F_2 = \mathcal{R}^{(n)}(F_1 \star F_2) \to \mathcal{R}(F_1 \star F_2) \neq \mathcal{R}F_1 \cdot \mathcal{R}F_2,$$

which is of course also easy to see "by hand." This shows that in some cases, the "non-standard" models as in (4.3) can be interpreted as limits of "standard" models for which the usual rules of calculus hold. Even this is, however, not always the case.

5 Schauder estimates and admissible models

One of the reasons why the theory of regularity structures is very successful at providing detailed descriptions of the small-scale features of solutions to semilinear (S)PDEs is that it comes with very sharp Schauder estimates. Recall that the classical Schauder estimates state that if $K : \mathbf{R}^d \to \mathbf{R}$ is a kernel that is smooth everywhere, except for a singularity at the origin that is (approximately) homogeneous of degree $\beta - d$ for some $\beta > 0$, then the operator $f \mapsto K * f$ maps $\mathcal{C}^\alpha$ into $\mathcal{C}^{\alpha+\beta}$ for every $\alpha \in \mathbf{R}$, except for those values for which $\alpha + \beta \in \mathbf{N}$. (See, e.g., Simon (1997).)

It turns out that similar Schauder estimates hold in the context of general regularity structures in the sense that it is in general possible to build an operator $\mathcal{K} : \mathcal{D}^\gamma \to \mathcal{D}^{\gamma+\beta}$ with the property that $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$. Of course, such a statement can only be true if our regularity structure contains not only the objects necessary to describe $\mathcal{R}f$ up to order γ, but also those required to describe $K * \mathcal{R}f$ up to order $\gamma + \beta$. What are these objects? At this stage, it might be useful to reflect on the effect of the convolution of a singular function (or distribution) with K.

Let us assume for a moment that f is also smooth everywhere, except at some point x_0. It is then straightforward to convince ourselves that $K * f$ is also smooth everywhere, except at x_0. Indeed, for any $\delta > 0$, we can write $K = K_\delta + K_\delta^c$, where K_δ is supported in a ball of radius δ around 0 and K_δ^c is a smooth function. Similarly, we can decompose f as $f = f_\delta + f_\delta^c$, where f_δ is supported in a δ-ball around x_0 and f_δ^c is smooth. Since the convolution of a smooth function with an arbitrary distribution is smooth, it follows that the only non-smooth component of $K * f$ is given by $K_\delta * f_\delta$, which is supported in a ball of radius 2δ around x_0. Since δ was arbitrary, the statement follows. By linearity, this strongly suggests that the local structure of the singularities of $K * f$ can be described completely by only using knowledge on the local structure of the singularities of f. It also suggests that the "singular part" of the operator $\mathcal{K}$ should be local, with the non-local parts of $\mathcal{K}$ only contributing to the "regular part."

This discussion suggests that we certainly need the following ingredients to build an operator $\mathcal{K}$ with the desired properties:

- The canonical polynomial structure should be part of our regularity structure in order to be able to describe the "regular parts."
- We should be given an "abstract integration operator" $\mathcal{I}$ on T which describes how the "singular parts" of $\mathcal{R}f$ transform under convolution by K.
- We should restrict ourselves to models which are "compatible" with the action of $\mathcal{I}$ in the sense that the behaviour of $\Pi_x \mathcal{I}\tau$ should relate in a suitable way to the behaviour of $K * \Pi_x \tau$ near x.

One way to implement these ingredients is to assume first that our model space T contains abstract polynomials in the following sense.

Assumption 5.1. *There exists a sector $\bar{T} \subset T$ isomorphic to the space of abstract polynomials in d commuting variables. In other words, $\bar{T}_\alpha \neq 0$ if and only if $\alpha \in \mathbf{N}$, and one can find basis vectors X^k of $T_{|k|}$ such that every element $\Gamma \in G$ acts on $\bar{T}$ by $\Gamma X^k = (X - h)^k$ for some $h \in \mathbf{R}^d$.*

Furthermore, we assume that there exists an abstract integration operator $\mathcal{I}$ with the following properties.

Assumption 5.2. *There exists a linear map $\mathcal{I} : T \to T$ such that $\mathcal{I}T_\alpha \subset T_{\alpha+\beta}$, such that $\mathcal{I}\bar{T} = 0$, and such that, for every $\Gamma \in G$ and $\tau \in T$, one has*

$$\Gamma \mathcal{I}\tau - \mathcal{I}\Gamma\tau \in \bar{T}. \tag{5.1}$$

Finally, we want to consider models that are compatible with this structure for a given kernel K. For this, we first make precise what we mean exactly when we said that K is approximately homogeneous of degree $\beta - d$.

Assumption 5.3. *One can write $K = \sum_{n \geq 0} K_n$ where each of the kernels $K_n : \mathbf{R}^d \to \mathbf{R}$ is smooth and compactly supported in a ball of radius 2^{-n} around the origin. Furthermore, we assume that for every multi-index k, one has a constant C such that the bound*

$$\sup_x |D^k K_n(x)| \leq C 2^{n(d-\beta+|k|)}, \tag{5.2}$$

holds uniformly in n. Finally, we assume that $\int K_n(x) P(x)\, dx = 0$ for every polynomial P of degree at most N, for some sufficiently large value of N.

Remark 5.4. It turns out that in order to define the operator $\mathcal{K}$ on $\mathcal{D}^\gamma$, we will need K to annihilate polynomials of degree N for some $N \geq \gamma + \beta$.

Remark 5.5. The last assumption may appear to be extremely stringent at first sight. In practice, this turns out not to be a problem at all. Say, for example, that we want to define an operator that represents convolution with $\mathcal{G}$, the Green's function of the Laplacian. Then $\mathcal{G}$ can be decomposed into a sum of terms satisfying the bound (5.2) with $\beta = 2$, but it does of course not annihilate generic polynomials and it is not supported in the ball of radius 1.

However, for any fixed value of $N > 0$, it is straightforward to decompose $\mathcal{G}$ as $\mathcal{G} = K + R$, where the kernel K is compactly supported and satisfies all of the properties mentioned above, and the kernel R is smooth. Lifting the convolution with R to an operator from $\mathcal{D}^\gamma \to \mathcal{D}^{\gamma+\beta}$ (actually to $\mathcal{D}^{\bar{\gamma}}$ for any $\bar{\gamma} > 0$) is straightforward, so that we have reduced our problem to that of constructing an operator describing the convolution by K.

Given such a kernel K, we can now make precise what we meant earlier when we said that the models under consideration should be compatible with the kernel K.

Definition 5.6. Given a kernel K as in Assumption 5.3 and a regularity structure $\mathscr{T}$ satisfying Assumptions 5.1 and 5.2, we say that a model (Π, Γ) is *admissible* if the identities

$$\big(\Pi_x X^k\big)(y) = (y - x)^k, \qquad \Pi_x \mathcal{I}\tau = K * \Pi_x \tau - \Pi_x \mathcal{J}(x)\tau, \tag{5.3}$$

holds for every $\tau \in T$ with $|\tau| \le N$. Here, $\mathcal{J}(x) : T \to \bar{T}$ is the linear map given on homogeneous elements by

$$\mathcal{J}(x)\tau = \sum_{|k| < |\tau| + \beta} \frac{X^k}{k!} \int D^{(k)} K(x - y)(\Pi_x \tau)(dy). \tag{5.4}$$

Remark 5.7. Note first that if $\tau \in \bar{T}$, then the definition given above is coherent as long as $|\tau| < N$. Indeed, since $\mathcal{I}\tau = 0$, one necessarily has $\Pi_x \mathcal{I}\tau = 0$. On the other hand, the properties of K ensure that in this case one also has $K * \Pi_x \tau = 0$, as well as $\mathcal{J}(x)\tau = 0$.

Remark 5.8. While $K * \xi$ is well defined for any distribution ξ, it is not so clear *a priori* whether the operator $\mathcal{J}(x)$ given in (5.4) is also well defined. It turns out that the axioms of a model do ensure that this is the case. The correct way of interpreting (5.4) is by

$$\mathcal{J}(x)\tau = \sum_{|k| < |\tau| + \beta} \sum_{n \ge 0} \frac{X^k}{k!} (\Pi_x \tau)\big(D^{(k)} K_n(x - \cdot)\big).$$

Note now that the scaling properties of the K_n ensure that $2^{(\beta - |k|)n} D^{(k)} K_n(x - \cdot)$ is a test function that is localised around x at scale 2^{-n}. As a consequence, one has

$$\big|(\Pi_x \tau)\big(D^{(k)} K_n(x - \cdot)\big)\big| \lesssim 2^{(|k| - \beta - |\tau|)n},$$

so that this expression is indeed summable as long as $|k| < |\tau| + \beta$.

Remark 5.9. As a matter of fact, it turns out that the above definition of an admissible model dovetails very nicely with our axioms defining a general model. Indeed, starting from *any* regularity structure $\mathscr{T}$, *any* model (Π, Γ) for $\mathscr{T}$, and a kernel K satisfying Assumption 5.3, it is usually possible to build a larger regularity structure $\hat{\mathscr{T}}$ containing $\mathscr{T}$ (in the "obvious" sense that $T \subset \hat{T}$ and the action of $\hat{G}$ on T is compatible with that of G) and endowed with an abstract integration map $\mathcal{I}$, as well as an admissible model $(\hat{\Pi}, \hat{\Gamma})$ on $\hat{\mathscr{T}}$ which reduces to (Π, Γ) when restricted to T. See Hairer (2014) for more details.

The only exception to this rule arises when the original structure T contains some homogeneous element τ which does not represent a polynomial and which is such that $|\tau| + \beta \in \mathbf{N}$. Since the bounds appearing both in the definition of a model and in Assumption 5.3 are only upper bounds, it is in practice easy to exclude such a situation by slightly tweaking the definition of either the exponent β or of the original regularity structure $\mathscr{T}$.

With all of these definitions in place, we can finally build the operator $\mathcal{K} : \mathcal{D}^\gamma \to \mathcal{D}^{\gamma+\beta}$ announced at the beginning of this section. Recalling the definition of $\mathcal{J}$ from (5.4), we set

$$(\mathcal{K}f)(x) = \mathcal{I}f(x) + \mathcal{J}(x)f(x) + (\mathcal{N}f)(x), \tag{5.5}$$

where the operator $\mathcal{N}$ is given by

$$(\mathcal{N}f)(x) = \sum_{|k|<\gamma+\beta} \frac{X^k}{k!} \int D^{(k)}K(x-y)\big(\mathcal{R}f - \Pi_x f(x)\big)(dy). \tag{5.6}$$

Note first that thanks to the reconstruction theorem, it is possible to verify that the right-hand side of (5.6) does indeed make sense for every $f \in \mathcal{D}^\gamma$ in virtually the same way as in Remark 5.8. One has the following.

Theorem 5.10. *Let K be a kernel satisfying Assumption 5.3, let $\mathscr{T} = (A, T, G)$ be a regularity structure satisfying Assumptions 5.1 and 5.2, and let (Π, Γ) be an admissible model for $\mathscr{T}$. Then, for every $f \in \mathcal{D}^\gamma$ with $\gamma \in (0, N-\beta)$ and $\gamma + \beta \notin \mathbf{N}$, the function $\mathcal{K}f$ defined in (5.5) belongs to $\mathcal{D}^{\gamma+\beta}$ and satisfies $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$.*

Proof. The complete proof of this result can be found in Hairer (2014) and will not be given here. Let us simply show that one has indeed $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$ in the particular case when our model consists of continuous functions so that Remark 2.12 applies. In this case, one has

$$(\mathcal{R}\mathcal{K}f)(x) = \big(\Pi_x\big(\mathcal{I}f(x) + \mathcal{J}(x)f(x)\big)\big)(x) + \big(\Pi_x(\mathcal{N}f)(x)\big)(x).$$

As a consequence of (5.3), the first term appearing in the right-hand side of this expression is given by

$$\big(\Pi_x\big(\mathcal{I}f(x) + \mathcal{J}(x)f(x)\big)\big)(x) = \big(K * \Pi_x f(x)\big)(x).$$

On the other hand, the only term contributing to the second term is the one with $k = 0$ (which is always present since $\gamma > 0$ by assumption) which then yields

$$\big(\Pi_x(\mathcal{N}f)(x)\big)(x) = \int K(x-y)\big(\mathcal{R}f - \Pi_x f(x)\big)(dy).$$

Adding both of these terms, we see that the expression $(K * \Pi_x f(x))(x)$ cancels, leaving us with the desired result. □

6 Application of the theory to semilinear SPDEs

Let us now briefly explain how this theory can be used to make sense of solutions to very singular semilinear stochastic PDEs. We will keep the discussion in this section at a very informal level without attempting to make mathematically precise statements. The interested reader may find more details in Hairer (2014).

For definiteness, we focus on the case of the dynamical Φ^4_3 model, which is formally given by

$$\partial_t \Phi = \Delta \Phi - \Phi^3 + \xi, \tag{6.1}$$

where ξ denotes space–time white noise and the spatial variable takes values in the three-dimensional torus. The problem with such an equation is that even the solution to the linear part of the equation, namely

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

only admits solutions in some spaces of Schwartz distributions. As a matter of fact, one has $\Psi(t,\cdot) \in \mathcal{C}^{-\alpha}$ if and only if $\alpha > 1/2$. As a consequence, it turns out that the only way of giving meaning to (6.1) is to "renormalise" the equation by adding an "infinite" linear term "$\infty\Phi$" which counteracts the strong dissipativity of the term $-\Phi^3$. To be slightly more precise, one can prove a statement of the following kind:

Theorem 6.1. *Consider the sequence of equations*

$$\partial_t \Phi_\varepsilon = \Delta \Phi_\varepsilon + C_\varepsilon \Phi_\varepsilon - \Phi_\varepsilon^3 + \xi_\varepsilon, \tag{6.2}$$

where $\xi_\varepsilon = \delta_\varepsilon * \xi$ with $\delta_\varepsilon(t,x) = \varepsilon^{-5}\varrho(\varepsilon^{-2}t, \varepsilon^{-1}x)$, for some smooth and compactly supported function ϱ, and ξ denotes space–time white noise. Then there exists a choice of constants C_ε such that the sequence Φ_ε converges in probability to a limiting (distributional) process Φ. Furthermore, the limiting process Φ does not depend on the choice of mollifier ϱ.

Remark 6.2. It turns out that in order to obtain a limit that is independent of the choice of mollifier ϱ, one should take C_ε of the form

$$C_\varepsilon = \frac{c_1}{\varepsilon} + \tilde{c} \log \varepsilon + c_3,$$

where $\tilde{c}$ is universal, but c_1 and c_3 depend on the choice of ϱ.

Remark 6.3. The limiting solution Φ is only local in time, so that the precise statement has to be slightly tweaked to allow for finite-time blow-ups. Regarding the initial condition, one can take $\Phi_0 \in \mathcal{C}^{-\beta}$ for any $\beta < 2/3$. This is expected to be optimal, even for the deterministic equation.

The aim of this section is to sketch how the theory of regularity structures can be used to obtain this kind of convergence results. First of all, we note that while our solution Φ will be a space–time distribution (or rather an element of $\mathcal{D}^\gamma$ for some regularity structure with a model over $\mathbf{R}^4$), the "time" direction has a different scaling behaviour from the three "space" directions. As a consequence, it turns out to be effective to slightly change our definition of "localised test functions" by setting

$$\varphi^\lambda_{(s,x)}(t,y) = \lambda^{-5}\varphi\big(\lambda^{-2}(t-s), \lambda^{-1}(y-x)\big).$$

Accordingly, the "effective dimension" of our space–time is actually 5, rather than 4. The theory presented above extends *mutatis mutandis* to this setting. (Note in particular that when considering the degree of a regular monomial, powers of the time variable should now be counted double.) Note also that with this way of measuring regularity, space–time white noise belongs to $\mathcal{C}^{-\alpha}$ for every $\alpha > \frac{5}{2}$. This is because of the bound

$$\big(\mathbf{E}\langle \xi, \varphi^\lambda_x\rangle^2\big)^{1/2} = \|\varphi^\lambda_x\|_{L^2} \approx \lambda^{-5/2},$$

combined with an argument similar to the proof of Kolmogorov's continuity lemma.

6.1 Construction of the associated regularity structure

Our first step is to build a regularity structure that is sufficiently large to allow to reformulate (6.1) as a fixed point in $\mathcal{D}^\gamma$ for some $\gamma > 0$. Denoting by $\mathcal{G}$ the heat kernel (i.e., the Green's function of the operator $\partial_t - \Delta$), we can write the solution to (6.1) with initial condition Φ_0 as

$$\Phi = \mathcal{G} * \big(\xi - \Phi^3\big) + \mathcal{G}\Phi_0,$$

where $*$ denotes space–time convolution and where we denote by $\mathcal{G}\Phi_0$ the harmonic extension of Φ_0. In order to have a chance of fitting this into the framework described above, we first decompose the heat kernel $\mathcal{G}$ as

$$\mathcal{G} = K + \hat{K},$$

where the kernel K satisfies all of the assumptions of Section 5 (with $\beta = 2$) and the remainder $\hat{K}$ is smooth. If we consider any regularity structure containing the usual Taylor polynomials and equipped with an admissible model, is straightforward to associate to $\hat{K}$ an operator $\hat{\mathcal{K}} : \mathcal{D}^\gamma \to \mathcal{D}^\infty$ via

$$(\hat{\mathcal{K}} f)(z) = \sum_k \frac{X^k}{k!}\big(D^{(k)}\hat{K} * \mathcal{R} f\big)(z),$$

where z denotes a space–time point and k runs over all possible four-dimensional multi-indices. Similarly, the harmonic extension of Φ_0 can be lifted to an element

in $\mathcal{D}^\infty$ which we denote again by $\mathcal{G}\Phi_0$ by considering its Taylor expansion around every space–time point. At this stage, we note that we actually cheated a little: while $\mathcal{G}\Phi_0$ is smooth in $\{(t,x): t>0, x\in\mathbf{T}^3\}$ and vanishes when $t<0$, it is of course singular on the time-0 hyperplane $\{(0,x): x\in\mathbf{T}^3\}$. This problem can be cured by introducing weighted versions of the spaces $\mathcal{D}^\gamma$ allowing for singularities on a given hyperplane. A precise definition of these spaces and their behaviour under multiplication and the action of the integral operator $\mathcal{K}$ can be found in Hairer (2014). For the purpose of the informal discussion given here, we will simply ignore this problem.

This suggests that the "abstract" formulation of (6.1) should be given by

$$\Phi = \mathcal{K}\big(\Xi - \Phi^3\big) + \hat{\mathcal{K}}\big(\Xi - \Phi^3\big) + \mathcal{G}\Phi_0. \tag{6.3}$$

In view of (5.5), this equation is of the type

$$\Phi = \mathcal{I}\big(\Xi - \Phi^3\big) + (\cdots), \tag{6.4}$$

where the terms $(\cdots)$ consist of functions that take values in the subspace $\bar{T}$ of T spanned by regular Taylor polynomials. In order to build a regularity structure in which (6.4) can be formulated, it is natural to start with the structure given by abstract polynomials (again with the parabolic scaling which causes the abstract "time" variable to have homogeneity 2 rather than 1), and to add a symbol Ξ to it which we postulate to have homogeneity $-\frac{5}{2}^-$, where we denote by α^- an exponent strictly smaller than, but arbitrarily close to, the value α.

We then simply add to T all of the formal expressions that an application of the right-hand side of (6.4) can generate for the description of Φ, Φ^2, and Φ^3. The homogeneity of a given expression is completely determined by the rules $|\mathcal{I}\tau| = |\tau| + 2$ and $|\tau\bar{\tau}| = |\tau| + |\tau|$. More precisely, we consider a collection $\mathcal{U}$ of formal expressions which is the smallest collection containing $\mathbf{1}$, X, and $\mathcal{I}(\Xi)$, and such that

$$\tau_1, \tau_2, \tau_3 \in \mathcal{U} \quad \Rightarrow \quad \mathcal{I}(\tau_1\tau_2\tau_3) \in \mathcal{U},$$

where it is understood that $\mathcal{I}(X^k) = 0$ for every multiindex k. We then set

$$\mathcal{W} = \{\Xi\} \cup \{\tau_1\tau_2\tau_3 : \tau_i \in \mathcal{U}\},$$

and we define our space T as the set of all linear combinations of elements in $\mathcal{W}$. (Note that since $\mathbf{1}\in\mathcal{U}$, one does in particular have $\mathcal{U}\subset\mathcal{W}$.) Naturally, $\mathcal{T}_\alpha$ consists of those linear combinations that only involve elements in $\mathcal{W}$ that are of homogeneity α. It is not too difficult to convince oneself that, for every $\alpha\in\mathbf{R}$, $\mathcal{W}$ contains only finitely many elements of homogeneity less than α, so that each $\mathcal{T}_\alpha$ is finite-dimensional.

In order to simplify expressions later, we will use the following shorthand graphical notation for elements of $\mathcal{W}$. For Ξ, we simply draw a dot. The integration map is then represented by a down-facing line and the multiplication of symbols is obtained by joining them at the root. For example, we have

$$\mathcal{I}(\Xi) = \,\uparrow, \qquad \mathcal{I}(\Xi)^3 = \,\Vee, \qquad \mathcal{I}(\Xi)\mathcal{I}\big(\mathcal{I}(\Xi)^3\big) = \,\Vee\!\!\uparrow.$$

Symbols containing factors of X have no particular graphical representation, so we will for example write $X_i\mathcal{I}(\Xi)^2 = X_i\vee$. With this notation, the space T is given by

$$T = \langle \Xi, \varv, \vee, \varv\!\!\vee, \uparrow, \vee\!\!\vee, \varv\!\!\vee\!\!\vee, X_i\vee, \mathbf{1}, \vee\!\!\vee\!\!\vee, \vee\!\!\vee\!\!\vee\!\!\vee, \ldots\rangle,$$

where we ordered symbols in increasing order of homogeneity and used $\langle\cdot\rangle$ to denote the linear span. Given any sufficiently regular function ξ (say a continuous space–time function), there is then a canonical way of lifting ξ to a model $\iota\xi = (\Pi, \Gamma)$ for T by setting

$$(\Pi_x \Xi)(y) = \xi(y), \qquad \bigl(\Pi_x X^k\bigr)(y) = (y - x)^k,$$

and then recursively by

$$(\Pi_x \tau\bar{\tau})(y) = (\Pi_x \tau)(y) \cdot (\Pi_x \bar{\tau})(y), \tag{6.5}$$

as well as (5.3). (Note that here we used x and y as notation for generic space–time points in order not to overload the notations.)

It turns out furthermore that there is a canonical way of building a structure group G for T and to also lift ξ to a family of operators Γ_{xy}, in such a way that all of the algebraic and analytic properties of an admissible model are satisfied. With such a model $\iota\xi$ at hand, it follows from (6.5) and the admissibility of $\iota\xi$ that the associated reconstruction operator satisfies the properties

$$\mathcal{R}\mathcal{K}f = K * \mathcal{R}f, \qquad \mathcal{R}(fg) = \mathcal{R}f \cdot \mathcal{R}g,$$

as long as all the functions to which $\mathcal{R}$ is applied belong to $\mathcal{D}^\gamma$ for some $\gamma > 0$. As a consequence, applying the reconstruction operator $\mathcal{R}$ to both sides of (6.3), we see that if Φ solves (6.3) then, provided that the model $(\Pi, \Gamma) = \iota\xi$ was built as above starting from any *continuous* realisation ξ of the driving noise, $\mathcal{R}\Phi$ solves the equation (6.1).

At this stage, the situation is as follows. For any *continuous* realisation ξ of the driving noise, we have factored the solution map $(\Phi_0, \xi) \to \Phi$ associated to (6.1) into maps

$$(\Phi_0, \xi) \to (\Phi_0, \iota\xi) \to \Phi \to \mathcal{R}\Phi,$$

where the middle arrow corresponds to the solution to (6.3) in some weighted $\mathcal{D}^\gamma$-space. The advantage of such a factorisation is that the last two arrows yield *continuous* maps, even in topologies sufficiently weak to be able to describe driving noise having the lack of regularity of space–time white noise. The only arrow that is not continuous in such a weak topology is the first one. At this stage, it should be believable that a similar construction can be performed for a very large class of semi-linear stochastic PDEs. In particular, the KPZ equation can also be analysed in this framework.

Given this construction, one is lead naturally to the following question: given a sequence ξ_ε of "natural" regularisations of space–time white noise, for example,

as in (6.2), do the lifts $\iota\xi_\varepsilon$ converge in probably in a suitable space of admissible models? Unfortunately, unlike in the case of the theory of rough paths where this is very often the case, the answer to this question in the context of SPDEs is often an emphatic **no**. Indeed, if it were the case for the dynamical Φ^4_3 model, then one could have chosen the constant C_ε to be independent of ε in (6.2), which is certainly not the case.

7 Renormalisation of the dynamical Φ^4_3 model

One way of circumventing the fact that $\iota\xi_\varepsilon$ does not converge to a limiting model as $\varepsilon \to 0$ is to consider instead a sequence of *renormalised* models. The main idea is to exploit the fact that our abstract definitions of a model do not impose the identity (6.5), even in situations where ξ itself happens to be a continuous function. One question that then imposes itself is: what are the natural ways of "deforming" the usual product which still lead to lifts to an admissible model? It turns out that the regularity structure whose construction was sketched above comes equipped with a natural *finite-dimensional* group of continuous transformations $\mathfrak{R}$ on its space of admissible models (henceforth called the "renormalisation group"), which essentially amounts to the space of all natural deformations of the product. It then turns out that even though $\iota\xi_\varepsilon$ does not converge, it is possible to find a sequence M_ε of elements in $\mathfrak{R}$ such that the sequence $M_\varepsilon\iota\xi_\varepsilon$ converges to a limiting model $(\hat\Pi, \hat\Gamma)$. Unfortunately, the elements M_ε no *not* preserve the image of ι in the space of admissible models. As a consequence, when solving the fixed-point map (6.3) with respect to the model $M_\varepsilon\iota\xi_\varepsilon$ and inserting the solution into the reconstruction operator, it is not clear *a priori* that the resulting function (or distribution) can again be interpreted as the solution to some modified PDE. It turns out that in our case, at least for a certain two-parameter subgroup of $\mathfrak{R}$, this is again the case and the modified equation is precisely given by (6.2), where C_ε is some linear combination of the two constants appearing in the description of M_ε.

There are now three questions that remain to be answered:

1. How does one construct the renormalisation group $\mathfrak{R}$?
2. How does one derive the new equation obtained when renormalising a model?
3. What is the right choice of M_ε ensuring that the renormalised models converge?

7.1 The renormalisation group

In order to construct $\mathfrak{R}$, it is essential to first have some additional knowledge of the structure group G for the type of regularity structures considered above. Recall that the purpose of the group G is to provide a class of linear maps $\Gamma : T \to T$ arising as possible candidates for the action of "re-expanding" a "Taylor series" around a different point. In our case, in view of (5.3), the coefficients of these re-expansions will naturally be some polynomials in x and in the expressions appearing in (5.4).

This suggests that we should define a space T^+ whose basis vectors consist of formal expressions of the type

$$X^k \prod_{i=1}^{N} \mathcal{J}_{\ell_i} \tau_i, \tag{7.1}$$

where N is an arbitrary but finite number, the τ_i are basis elements of T, and the ℓ_i are d-dimensional multi-indices satisfying $|\ell_i| < |\tau_i| + 2$. (The last bound is a reflection of the restriction of the summands in (5.4) with $\beta = 2$.) The space T^+ also has a natural graded structure $T^+ = \bigoplus T_\alpha^+$ by setting

$$|\mathcal{J}_\ell \tau| = |\tau| + 2 - |\ell|, \qquad |X^k| = |k|,$$

and by postulating that the degree of a product is the sum of the degrees. Unlike in the case of T however, elements of T^+ all have strictly positive homogeneity, except for the empty product $\mathbf{1}$ which we postulate to have degree 0.

To any given admissible model (Π, Γ), it is then natural to associate linear maps $f_x : T^+ \to \mathbf{R}$ by $f_x(X^k) = x^k$, $f_x(\sigma\bar{\sigma}) = f_x(\sigma) f_x(\bar{\sigma})$, and

$$f_x(\mathcal{J}_{\ell_i} \tau_i) = \int D^{(\ell_i)} K(x - y)(\Pi_x \tau_i)(dy). \tag{7.2}$$

It then turns out that it is possible to build a linear map $\Delta : T \to T \otimes T^+$ such that if we define $F_x : T \to T$ by

$$F_x \tau = (I \otimes f_x)\Delta\tau, \tag{7.3}$$

where I denotes the identity operator on T, then these maps are invertible and $\Pi_x F_x^{-1}$ is independent of x. Furthermore, there exists a map $\Delta^+ : T^+ \to T^+ \otimes T^+$ such that

$$(\Delta \otimes I)\Delta = \big(I \otimes \Delta^+\big)\Delta, \qquad \Delta^+(\sigma\bar{\sigma}) = \Delta^+\sigma \cdot \Delta^+\bar{\sigma}. \tag{7.4}$$

With this map at hand, we can define a product $\circ$ on the space of linear functionals $f : T^+ \to \mathbf{R}$ by

$$(f \circ g)(\sigma) = (f \otimes g)\Delta^+\sigma.$$

If we furthermore denote by Γ_f the operator T associated to any such linear functional as in (7.3), the first identity of (7.4) yields the identity $\Gamma_f \Gamma_g = \Gamma_{f \circ g}$. The second identity of (7.4) furthermore ensures that if f and g are both multiplicative in the sense that $f(\sigma\bar{\sigma}) = f(\sigma) f(\bar{\sigma})$, then $f \circ g$ is again multiplicative. It also turns out that every multiplicative linear functional f admits a unique inverse f^{-1} such that $f^{-1} \circ f = f \circ f^{-1} = e$, where $e : T^+ \to \mathbf{R}$ maps every basis vector of the form (7.1) to zero, except for $e(\mathbf{1}) = 1$. The element e is neutral in the sense that Γ_e is the identity operator.

It is now natural to define the structure group G associated to T as the set of all multiplicative linear functionals on T^+, acting on T via (7.3). Furthermore, for any admissible model, one has the identity

$$\Gamma_{xy} = F_x^{-1} F_y = \Gamma_{\gamma_{xy}}, \qquad \gamma_{xy} = f_x^{-1} \circ f_y.$$

How does all this help with the identification of a natural class of deformations for the usual product? First, it turns out that for every continuous function ξ, if we denote again by (Π, Γ) the model $\iota\xi$, then the linear map $\mathbf{\Pi} : T \to \mathcal{C}$ given by

$$\mathbf{\Pi} = \Pi_y F_y^{-1},$$

which is independent of the choice of y by the above discussion, is given by

$$(\mathbf{\Pi}\Xi)(x) = \xi(x), \qquad \big(\mathbf{\Pi} X^k\big)(x) = x^k, \tag{7.5}$$

and then recursively by

$$\mathbf{\Pi}\tau\bar{\tau} = \mathbf{\Pi}\tau \cdot \mathbf{\Pi}\bar{\tau}, \qquad \mathbf{\Pi}\mathcal{I}\tau = K * \mathbf{\Pi}\tau.$$

Note that this is very similar to the definition of $\iota\xi$, with the notable exception that (5.3) is replaced by the more "natural" identity $\mathbf{\Pi}\mathcal{I}\tau = K * \mathbf{\Pi}\tau$. It turns out that the knowledge of $\mathbf{\Pi}$ and the knowledge of (Π, Γ) are equivalent since one has $\Pi_x = \Pi F_x$ and the map F_x can be recovered from Π_x by (7.2). (This argument appears circular but it is possible to put a suitable recursive structure on T and T^+ ensuring that this actually works.) Furthermore, the translation $(\Pi, \Gamma) \leftrightarrow \mathbf{\Pi}$ actually works for *any* admissible model and does not at all rely on the fact that it was built by lifting a continuous function. However, in the general case, the first identity in (7.5) does not of course not make any sense anymore and might fail even if the coordinates of $\mathbf{\Pi}$ consist of continuous functions.

At this stage, we note that if ξ happens to be a stationary stochastic process and $\mathbf{\Pi}$ is built from ξ by following the above procedure, then $\mathbf{\Pi}\tau$ is a stationary stochastic process for every $\tau \in T$. In order to define $\mathfrak{R}$, it is natural to consider only transformations of the space of admissible models that preserve this property. Since we are not in general allowed to multiply components of $\mathbf{\Pi}$, the only remaining operation is to form linear combinations. It is therefore natural to describe elements of $\mathfrak{R}$ by linear maps $M : T \to T$ and to postulate their action on admissible models by $\mathbf{\Pi} \mapsto \mathbf{\Pi}^M$ with

$$\mathbf{\Pi}^M \tau = \mathbf{\Pi} M \tau.$$

It is not clear a priori whether given such a map M and an admissible model (Π, Γ) there is a coherent way of building a new model (Π^M, Γ^M) such that $\mathbf{\Pi}^M$ is the map associated to (Π^M, Γ^M) as above. It turns out that one has the following statement.

Proposition 7.1. *In the above context, for every linear map $M : T \to T$ commuting with $\mathcal{I}$ and multiplication by X^k, there exist unique linear maps $\Delta^M : T \to T \otimes T^+$ and $\hat{\Delta}^M : T^+ \to T^+ \otimes T^+$ such that if we set*

$$\Pi_x^M \tau = (\Pi_x \otimes f_x)\Delta^M \tau, \qquad \gamma_{xy}^M(\sigma) = (\gamma_{xy} \otimes f_x)\hat{\Delta}^M \sigma,$$

then Π_x^M satisfies again (5.3) and the identity $\Pi_x^M \Gamma_{xy}^M = \Pi_y^M$.

At this stage, it may look like *any* linear map $M : T \to T$ commuting with $\mathcal{I}$ and multiplication by X^k yields a transformation on the space of admissible models by Proposition 7.1. This, however, is not true since we have completely disregarded the *analytical* bounds that every model has to satisfy. It is clear from Definition 2.5 that these are satisfied if and only if $\Pi_x^M \tau$ is a linear combination of the $\Pi_x \tau_j$ with $|\tau_j| \geq |\tau|$. This suggests the following definition.

Definition 7.2. The renormalisation group $\mathfrak{R}$ consists of the set of linear maps $M : T \to T$ commuting with $\mathcal{I}$ and with multiplication by X^k, such that for $\tau \in T_\alpha$ and $\sigma \in T_\alpha^+$, one has

$$\Delta^M \tau - \tau \otimes \mathbf{1} \in \bigoplus_{\beta > \alpha} T_\alpha \otimes T^+, \qquad \hat{\Delta}^M \sigma - \sigma \otimes \mathbf{1} \in \bigoplus_{\beta > \alpha} T_\alpha^+ \otimes T^+.$$

Its action on the space of admissible models is given by Proposition 7.1.

7.2 The renormalised equations

In the case of the dynamical Φ^4 model considered in this article, it turns out that we only need a two-parameter subgroup of $\mathfrak{R}$ to renormalise the equations. More precisely, we consider elements $M \in \mathfrak{R}$ of the form $M = \exp(-C_1 L_1 - C_2 L_2)$, where the two generators L_1 and L_2 are determined by the substitution rules

$$L_1 : \Xi \mapsto \mathbf{1}, \qquad L_2 : \Xi\!\bullet \mapsto \mathbf{1}.$$

This should be understood in the sense that if τ is an arbitrary formal expression, then $L_1 \tau$ is the sum of all formal expressions obtained from τ by performing a substitution of the type $\Xi \mapsto 1$, and similarly for L_2. For example, one has

$$L_1 \Xi\!\bullet = 3\uparrow, \qquad L_1 \bullet\!\bullet = \bullet, \qquad L_2 \Xi\!\bullet = 3\uparrow.$$

One then has the following result.

Proposition 7.3. *The linear maps M of the type just described belong to $\mathfrak{R}$. Furthermore, if (Π, Γ) is an admissible model such that $\Pi_x \tau$ is a continuous function for every $\tau \in T$, then one has the identity*

$$\big(\Pi_x^M \tau\big)(x) = (\Pi_x M \tau)(x). \tag{7.6}$$

Remark 7.4. Note that it is the same value x that appears twice on each side of (7.6). It is in fact *not* the case that one has $\Pi_x^M \tau = \Pi_x M\tau$. However, the identity (7.6) is all we need to derive the renormalised equations.

It is now rather straightforward to show the following.

Proposition 7.5. *Let $M = \exp(-C_1 L_1 - C_2 L_2)$ as above and let $(\Pi^M, \Gamma^M) = M\iota\xi$ for some smooth function ξ. Let furthermore Φ be the solution to (6.3) with respect to the model (Π^M, Γ^M). Then the function $u(t,x) = (\mathcal{R}^M\Phi)(t,x)$ solves the equation*

$$\partial_t u = \Delta u - u^3 + (3C_1 - 9C_2)u + \xi.$$

Proof. By Theorem 4.3, it turns out that (6.3) can be solved in $\mathcal{D}^\gamma$ as soon as γ is a little bit greater than 1. Therefore, we only need to keep track of its solution Φ up to terms of homogeneity 1. By repeatedly applying the identity (6.4), we see that the solution Φ is necessarily of the form

$$\Phi = \varUparrow + \varphi\mathbf{1} - \varPsi - 3\varphi\varUpsilon + \langle \nabla\varphi, X\rangle, \tag{7.7}$$

for some real-valued function φ and some $\mathbf{R}^3$-valued function $\nabla\varphi$. (Note that $\nabla\varphi$ is treated as an independent function here, we certainly do not suggest that the function φ is differentiable. Our notation is only by analogy with the classical Taylor expansion....) Similarly, the right-hand side of the equation is given up to order 0 by

$$\Xi - \Phi^3 = \Xi - \varPsi - 3\varphi\varV + 3\varXi - 3\varphi^2\varUparrow + 6\varphi\varPsi + 9\varphi\varXi - 3\langle\nabla\varphi, \varV X\rangle - \varphi^3\mathbf{1}. \tag{7.8}$$

Combining this with the definition of M, it is straightforward to see that, modulo terms of strictly positive homogeneity, one has

$$\begin{aligned} M(\Xi - \Phi^3) &= \Xi - (M\Phi)^3 + 3C_1\varUparrow + 3C_1\varphi\mathbf{1} - 9C_2\varUparrow - 9C_2\varphi\mathbf{1} \\ &= \Xi - (M\Phi)^3 + (3C_1 - 9C_2)M\Phi. \end{aligned}$$

Combining this with (7.6), the claim now follows at once. □

7.3 Convergence of the renormalised models

It remains to argue why one expects to be able to find constants C_1^ε and C_2^ε such that the sequence of renormalised models $M^\varepsilon \iota\xi_\varepsilon$ converges to a limiting model. Instead of considering the actual sequence of models, we only consider the sequence of stationary processes $\hat{\mathbf{\Pi}}^\varepsilon\tau := \mathbf{\Pi}^\varepsilon M^\varepsilon\tau$, where $\mathbf{\Pi}^\varepsilon$ is associated to $(\Pi^\varepsilon, \Gamma^\varepsilon) = \iota\xi_\varepsilon$ as before. Since there are general arguments available to deal with all the expressions τ of positive homogeneity, we restrict ourselves to those of negative homogeneity which, leaving out Ξ which is easy to treat, are given by

$$\varPsi, \varV, \varXi, \varUparrow, \varPsi, \varXi, X_i\varV.$$

For this section, some elementary notions from the theory of Wiener chaos expansions are required, but we will try to hide this as much as possible. At a formal level, one has the identity

$$\boldsymbol{\Pi}^\varepsilon \tau = K * \xi_\varepsilon = K_\varepsilon * \xi,$$

where the kernel K_ε is given by $K_\varepsilon = K * \delta_\varepsilon$. This shows that, at least formally, one has

$$\big(\boldsymbol{\Pi}^\varepsilon \mathcal{V}\big)(z) = (K * \xi_\varepsilon)(z)^2 = \int\int K_\varepsilon(z - z_1) K_\varepsilon(z - z_2) \xi(z_1) \xi(z_2)\, dz_1\, dz_2.$$

Similar but more complicated expressions can be found for any formal expression τ. This naturally leads to the study of random variables of the type

$$I_k(f) = \int \cdots \int f(z_1, \ldots, z_k) \xi(z_1) \cdots \xi(z_k)\, dz_1 \cdots dz_k. \tag{7.9}$$

Ideally, one would hope to have an Itô isometry of the type $\mathbf{E} I_k(f) I_k(g) = \langle f^{\mathrm{sym}}, g^{\mathrm{sym}} \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the L^2-scalar product and f^{sym} denotes the symmetrisation of f. This is unfortunately *not* the case. Instead, one should replace the products in (7.9) by *Wick products*, which are formally generated by all possible *contractions* of the type

$$\xi(z_i)\xi(z_j) \mapsto \xi(z_i) \diamond \xi(z_j) + \delta(z_i - z_j).$$

If we then set

$$\hat{I}_k(f) = \int \cdots \int f(z_1, \ldots, z_k) \xi(z_1) \diamond \cdots \diamond \xi(z_k)\, dz_1 \cdots dz_k,$$

one has indeed

$$\mathbf{E} \hat{I}_k(f) \hat{I}_k(g) = \langle f^{\mathrm{sym}}, g^{\mathrm{sym}} \rangle.$$

See Nualart (1995) for a more thorough description of this construction, which also goes under the name of *Wiener chaos*. It turns out that one has equivalence of moments in the sense that, for every $k > 0$ and $p > 0$ there exists a constant $C_{k,p}$ such that

$$\mathbf{E} |\hat{I}_k(f)|^p \le C_{k,p} \|f^{\mathrm{sym}}\|^p \le C_{k,p} \|f\|^p,$$

where the second bound comes from the fact that symmetrisation is a contraction in L^2. Finally, one has $\mathbf{E} \hat{I}_k(f) \hat{I}_\ell(g) = 0$ if $k \neq \ell$. Random variables of the form $\hat{I}_k(f)$ for some $k \ge 0$ and some square integrable function f are said to belong to the *kth homogeneous Wiener chaos*.

Returning to our problem, we first argue that it should be possible to choose M in such a way that $\hat{\boldsymbol{\Pi}}^\varepsilon \mathcal{V}$ converges to a limit as $\varepsilon \to 0$. The above considerations suggest that one should rewrite $\boldsymbol{\Pi}^\varepsilon$ as

$$\begin{aligned} \big(\boldsymbol{\Pi}^\varepsilon \mathcal{V}\big)(z) &= (K * \xi_\varepsilon)(z)^2 \\ &= \int\int K_\varepsilon(z - z_1) K_\varepsilon(z - z_2) \xi(z_1) \diamond \xi(z_2)\, dz_1\, dz_2 + C_\varepsilon, \end{aligned} \tag{7.10}$$

where the constant C_ε is given by

$$C_\varepsilon = \int K_\varepsilon^2(z_1)\,dz_1 = \int K_\varepsilon^2(z - z_1)\,dz_1.$$

Note now that K_ε is an ε-approximation of the kernel K which has the same singular behaviour as the heat kernel. In terms of the parabolic distance, the singularity of the heat kernel scales like $K(z) \sim |z|^{-3}$ for $z \to 0$. (Recall that we consider the parabolic distance $|(t,x)| = \sqrt{|t|} + |x|$, so that this is consistent with the fact that the heat kernel is bounded by $t^{-3/2}$.) This suggests that one has $K_\varepsilon^2(z) \sim |z|^{-6}$ for $|z| \gg \varepsilon$. Since parabolic space–time has scaling dimension 5 (time counts double), this is a non-integrable singularity. As a matter of fact, there is a whole power of z missing to make it borderline integrable, which suggests that one has

$$C_\varepsilon \sim \frac{1}{\varepsilon}.$$

This already shows that one should not expect $\mathbf{\Pi}^\varepsilon \mathcal{V}$ to converge to a limit as $\varepsilon \to 0$. However, it turns out that the first term in (7.10) converges to a distribution-valued stationary space–time process, so that one would like to somehow get rid of this diverging constant C_ε. This is exactly where the renormalisation map M (in particular the factor $\exp(-C_1 L_1)$) enters into play. Following the above definitions, we see that one has

$$(\hat{\mathbf{\Pi}}^\varepsilon \mathcal{V})(z) = (\mathbf{\Pi}^\varepsilon M \mathcal{V})(z) = (\mathbf{\Pi}^\varepsilon \mathcal{V})(z) - C_1.$$

This suggests that if we make the choice $C_1 = C_\varepsilon$, then $\hat{\mathbf{\Pi}}^\varepsilon \mathcal{V}$ does indeed converge to a non-trivial limit as $\varepsilon \to 0$. This limit is a distribution given by

$$(\mathbf{\Pi}^\varepsilon \mathcal{V})(\psi) = \int\int \psi(z) K(z - z_1) K(z - z_2)\,dz\,\xi(z_1) \diamond \xi(z_2)\,dz_1\,dz_2.$$

Using again the scaling properties of the kernel K, it is not too difficult to show that this yields indeed a random variable belonging to the second homogeneous Wiener chaos for every choice of smooth test function ψ. Once we know that $\hat{\mathbf{\Pi}}^\varepsilon \mathcal{V}$ converges, it is immediate that $\hat{\mathbf{\Pi}}^\varepsilon X \mathcal{V}$ converges as well, since this amounts to just multiplying a distribution by a smooth function.

A similar argument to what we did for $\mathcal{V}$ allows to take care of $\tau = \mathcal{V}$ since one then has

$$\begin{aligned}(\mathbf{\Pi}^\varepsilon \mathcal{V})(z) = \int\int & K_\varepsilon(z - z_1) K_\varepsilon(z - z_2) K_\varepsilon(z - z_3) \\ & \times \xi(z_1) \diamond \xi(z_2) \diamond \xi(z_3)\,dz_1\,dz_2\,dz_3 \\ + 3 \int\int & K_\varepsilon(z - z_1) K_\varepsilon(z - z_2) K_\varepsilon(z - z_3) \\ & \times \delta(z_1 - z_2)\xi(z_3)\,dz_1\,dz_2\,dz_3.\end{aligned}$$

Noting that the second term in this expression is nothing but

$$3C_\varepsilon \int K_\varepsilon(z - z_1)\xi(z_1)\,dz_1 = 3C_\varepsilon\big(\mathbf{\Pi}^\varepsilon \varUparrow\big)(z),$$

we see that in this case, provided again that $C_1 = C_\varepsilon$, $\hat{\mathbf{\Pi}}^\varepsilon \Vee$ is given by only the first term in the expression above, which turns out to converge to a non-degenerate limiting random distribution in a similar way to what happened for $\Vee$.

Turning to our list of terms of negative homogeneity, it remains to consider $\Varpsi$, $\Varpsi$, and $\Varpsi$. It turns out that the latter two are the more difficult ones, so we only discuss these. Let us first argue why we expect to be able to choose the constants C_1 and C_2 in such a way that $\hat{\mathbf{\Pi}}^\varepsilon \Varpsi$ converges to a limit. In this case, the "bad" terms comes from the part of $(\mathbf{\Pi}^\varepsilon \Varpsi)(z)$ belonging to the homogeneous chaos of order 0. This is simply a constant, which turns out to be given by

$$\hat{C}_\varepsilon = 2\int K(z)\,Q_\varepsilon^2(z)\,dz, \tag{7.11}$$

where the kernel Q_ε is given by

$$Q_\varepsilon(z) = \int K_\varepsilon(\bar{z})K_\varepsilon(\bar{z} - z)\,d\bar{z}.$$

Since K_ε is an ε-mollification of a kernel with a singularity of order -3 and the scaling dimension of the underlying space is 5, we see that Q_ε behaves like an ε-mollification of a kernel with a singularity of order $-3-3+5=-1$ at the origin. As a consequence, the singularity of the integrand in (7.11) is of order -5, which gives rise to a logarithmic divergence as $\varepsilon \to 0$. This suggests that one should choose $C_2 = \hat{C}_\varepsilon$ in order to cancel out this diverging term and obtain a non-trivial limit for $\hat{\mathbf{\Pi}}^\varepsilon \Varpsi$ as $\varepsilon \to 0$. This is indeed the case.

We finally turn to the symbol $\Varpsi$. In this case, the "bad" terms appearing in the Wiener chaos decomposition of $\mathbf{\Pi}^\varepsilon \Varpsi$ are the terms in the first homogeneous Wiener chaos, which are of the form

$$\begin{aligned} &3\int \hat{Q}_\varepsilon(z - z_1)K_\varepsilon(z_1 - z_2)\xi(z_2)\,dz_1\,dz_2 \\ &\quad = 3\int (\hat{Q}_\varepsilon * K_\varepsilon)(z - z_2)\xi(z_2)\,dz_2, \end{aligned} \tag{7.12}$$

where $\hat{Q}_\varepsilon$ is the kernel given by

$$\hat{Q}_\varepsilon(z) = 2K(z)\,Q_\varepsilon^2(z).$$

As already mentioned above, the problem here is that as $\varepsilon \to 0$, $\hat{Q}_\varepsilon$ converges to a kernel $\hat{Q} = 2KQ^2$, which has a non-integrable singularity at the origin. In particular, the action of integrating a test function against $\hat{Q}_\varepsilon$ does not converge to a limiting distribution as $\varepsilon \to 0$.

This is akin to the problem of making sense of integration against a one-dimensional kernel with a singularity of type $1/|x|$ at the origin. For the sake of the argument, let us consider a function $W : \mathbf{R} \to \mathbf{R}$ which is compactly supported and smooth everywhere except at the origin, where it diverges like $W(x) \sim 1/|x|$. It is then natural to associate to W a "renormalised" distribution $\mathscr{R}W$ given by

$$(\mathscr{R}W)(\varphi) = \int W(x)\big(\varphi(x) - \varphi(0)\big)\,dx.$$

Note that $\mathscr{R}W$ has the property that if $\varphi(0) = 0$, then it simply corresponds to integration against W, which is the standard way of associating a distribution to a function. In a way, the extra term can be interpreted as subtracting a Dirac distribution with an "infinite mass" located at the origin, thus cancelling out the divergence of the non-integrable singularity. It is also straightforward to verify that if W_ε is a sequence of smooth approximations to W (say one has $W_\varepsilon(x) = W(x)$ for $|x| > \varepsilon$ and $W_\varepsilon \sim 1/\varepsilon$ otherwise), then $\mathscr{R}W^\varepsilon \to \mathscr{R}W$ in a distributional sense, and (using the usual correspondence between functions and distributions) one has

$$\mathscr{R}W^\varepsilon = W^\varepsilon - \hat{C}_\varepsilon \delta_0, \qquad \hat{C}_\varepsilon = \int W^\varepsilon(x)\,dx.$$

The cure to the problem we are facing for showing the convergence of $\mathbf{\Pi}^\varepsilon \Varv$ is virtually identical. Indeed,by choosing $C_2 = \hat{C}_\varepsilon$ as in (7.11), the term in the first homogeneous Wiener chaos for $\hat{\mathbf{\Pi}}^\varepsilon$ corresponding to (7.12) is precisely given by

$$\begin{aligned} 3\int \hat{Q}_\varepsilon(z - z_1) K_\varepsilon(z_1 - z_2)\xi(z_2)\,dz_1\,dz_2 - 3C_2 \int K_\varepsilon(z - z_2)\xi(z_2)\,dz_2 \\ = 3\int (\mathscr{R}\hat{Q}_\varepsilon * K_\varepsilon)(z - z_2)\xi(z_2)\,dz_2. \end{aligned}$$

It turns out that the convergence of $\mathscr{R}\hat{Q}_\varepsilon$ to a limiting distribution $\mathscr{R}\hat{Q}$ takes place in a sufficiently strong topology to allow to conclude that $\hat{\mathbf{\Pi}}^\varepsilon \Varv$ does indeed converge to a non-trivial limiting random distribution.

It should be clear from this whole discussion that while the precise values of the constants C_1 and C_2 depend on the details of the mollifier δ_ε, the limiting (random) model $(\hat{\Pi}, \hat{\Gamma})$ obtained in this way is independent of it. Combining this with the continuity of the solution to the fixed point map (6.3) and of the reconstruction operator $\mathcal{R}$ with respect to the underlying model, we see that the statement of Theorem 6.1 follows almost immediately.

Acknowledgment

Financial support from the Leverhulme trust through a leadership award is gratefully acknowledged.

References

Albeverio, S. and Röckner, M. (1991). Stochastic differential equations in infinite dimensions: Solutions via Dirichlet forms. *Probab. Theory Related Fields* **89**, 347–386. MR1113223

Bertini, L. and Giacomin, G. (1997). Stochastic Burgers and KPZ equations from particle systems. *Comm. Math. Phys.* **183**, 571–607. MR1462228

Bourgain, J. and Pavlović, N. (2008). Ill-posedness of the Navier–Stokes equations in a critical space in 3D. *J. Funct. Anal.* **255**, 2233–2247. MR2473255

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41**, 909–996. MR0951745

Da Prato, G. and Debussche, A. (2003). Strong solutions to the stochastic quantization equations. *Ann. Probab.* **31**, 1900–1916. MR2016604

Giacomin, G., Lebowitz, J. L. and Presutti, E. (1999). Deterministic and stochastic hydrodynamic equations arising from simple microscopic model systems. In *Stochastic Partial Differential Equations: Six Perspectives*. *Math. Surveys Monogr.* **64**, 107–152. Providence, RI: Amer. Math. Soc. MR1661764

Gubinelli, M. (2004). Controlling rough paths. *J. Funct. Anal.* **216**, 86–140. MR2091358

Gubinelli, M. (2010). Ramification of rough paths. *J. Differential Equations* **248**, 693–721. MR2578445

Hairer, M. (2011). Rough stochastic PDEs. *Comm. Pure Appl. Math.* **64**, 1547–1585. MR2832168

Hairer, M. (2013). Solving the KPZ equation. *Ann. of Math. (2)* **178**, 559–664. MR3071506

Hairer, M. (2014). A theory of regularity structures. *Invent. Math.* DOI: 10.1007/s00222-014-0505-4.

Jona-Lasinio, G. and Mitter, P. K. (1985). On the stochastic quantization of field theory. *Comm. Math. Phys.* **101**, 409–436. MR0815192

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

Lyons, T. J., Caruana, M. and Lévy, T. (2007). *Differential Equations Driven by Rough Paths*. *Lecture Notes in Mathematics* **1908**. Berlin: Springer. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004. With an introduction concerning the Summer School by Jean Picard. MR2314753

Lyons, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14**, 215–310. MR1654527

Meyer, Y. (1992). *Wavelets and Operators*. *Cambridge Studies in Advanced Mathematics* **37**. Cambridge: Cambridge Univ. Press. Translated from the 1990 French original by D. H. Salinger. MR1228209

Nualart, D. (1995). *The Malliavin Calculus and Related Topics*. *Probability and Its Applications (New York)*. New York: Springer-Verlag. MR1344217

Parisi, G. and Wu, Y. S. (1981). Perturbation theory without gauge fixing. *Sci. Sinica* **24**, 483–496. MR0626795

Simon, L. (1997). Schauder estimates by scaling. *Calc. Var. Partial Differential Equations* **5**, 391–407. MR1459795

Young, L. C. (1936). An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67**, 251–282. MR1555421

The University of Warwick
Coventry, Warwickshire CV4 7AL
United Kingdom
E-mail: M.Hairer@Warwick.ac.uk