Chapter 14
Operations on modelled distributions

The original motivation for the development of the theory of regularity structures was to provide robust solution theories for singular stochastic PDEs like the KPZ equation or the dynamical Φ^4_3 model. The idea is to reformulate them as fixed point problems in some space $\mathscr{D}^\gamma$ (or rather a slightly modified version that takes into account possible singular behaviour near time 0) based on a suitable random model in a regularity structure purpose-built for the problem at hand. In order to achieve this this chapter provides a systematic way of formulating the standard operations arising in the construction of the corresponding fixed point problem (differentiation, multiplication, composition by a regular function, convolution with the heat kernel) as operations on the spaces $\mathscr{D}^\gamma$.

14.1 Differentiation

Being a local operation, differentiating a modelled distribution is straightforward, provided that the model one works with is sufficiently rich. Denote by $\mathcal{L}$ some (formal) differential operator with constant coefficients that is homogeneous of degree m, i.e. it is of the form

$$\mathcal{L} = \sum_{|k|=m} a_k D^k \;,$$

where k is a d-dimensional multi-index, $a_k \in \mathbf{R}$, and D^k denotes the kth mixed derivative in the distributional sense.

Given a regularity structure (T, G), it is convenient to define "abstract" differentiation only on suitable substructures. The appropriate notion of sector was already introduced in Definition 13.1. We have

Definition 14.1. Consider a sector $V \subset T$. A linear operator $\partial \colon V \to T$ is said to *realise* $\mathcal{L}$ (of degree m) for the model (Π, Γ) if

P. K. Friz, M. Hairer, *A Course on Rough Paths*, Universitext,
https://doi.org/10.1007/978-3-030-41556-3_14

- one has $\partial\tau \in T_{\alpha-m}$ for every $\tau \in V_\alpha$,
- one has $\Gamma\partial\tau = \partial\Gamma\tau$ for every $\tau \in V$ and every $\Gamma \in G$.
- one has $\Pi_x\partial\tau = \mathcal{L}\Pi_x\tau$ for every $\tau \in V$ and every $x \in \mathbf{R}^d$.

Writing $\mathscr{D}^\gamma(V)$ for those elements in $\mathscr{D}^\gamma$ taking values in the sector V, it then turns out that one has the following fact:

Proposition 14.2. *Assume that ∂ realises $\mathcal{L}$ for the model (Π,Γ) and let $f \in \mathscr{D}^\gamma(V)$ for some $\gamma > m$. Then, $\partial f \in \mathscr{D}^{\gamma-m}$ and the identity $\mathcal{R}\partial f = \mathcal{L}\mathcal{R}f$ holds.*

Proof. The fact that $\partial f \in \mathscr{D}^{\gamma-m}$ is an immediate consequence of the definitions, so we only need to show that $\mathcal{R}\partial f = \mathcal{L}\mathcal{R}f$.

By the "uniqueness" part of the reconstruction theorem, this on the other hand follows immediately if we can show that, for every fixed test function ψ and every $x \in \mathbf{R}^d$, one has

$$\big(\Pi_x\partial f(x) - \mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) \lesssim \lambda^\delta\;,$$

for some $\delta > 0$. Here, we defined ψ_x^λ as before. By the assumption on the model Π, we have the identity

$$\big(\Pi_x\partial f(x)-\mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) = \big(\mathcal{L}\Pi_x f(x)-\mathcal{L}\mathcal{R}f\big)(\psi_x^\lambda) = -\big(\Pi_x f(x)-\mathcal{R}f\big)(\mathcal{L}^*\psi_x^\lambda)\;,$$

where $\mathcal{L}^*$ is the formal adjoint of $\mathcal{L}$. Since, as a consequence of the homogeneity of $\mathcal{L}$, one has the identity $\mathcal{L}^*\psi_x^\lambda = \lambda^{-m}\big(\mathcal{L}^*\psi\big)_x^\lambda$, it then follows immediately from the reconstruction theorem that the right-hand side of this expression is of order $\lambda^{\gamma-m}$, as required. □

14.2 Products and composition by regular 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 13.18. Provided that our functions / distributions are represented as elements in $\mathscr{D}^\gamma$ 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.

Definition 14.3. Given a regularity structure (T,G) and two sectors $V,\bar V \subset T$, a *product* on $(V,\bar V)$ is a bilinear map $\star\colon V\times\bar V \to T$ 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 14.4. The condition that degrees 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 reexpands 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 $\mathscr{D}^{\gamma_1}$ with an element in $\mathscr{D}^{\gamma_2}$ again belongs to some $\mathscr{D}^{\gamma}$. In order to answer this question, we introduce the notation $\mathscr{D}^{\gamma}_{\alpha}$ to denote those elements $f \in \mathscr{D}^{\gamma}$ such that furthermore

$$f(x) \in T_{\geq\alpha} = \bigoplus_{\beta\geq\alpha} T_\beta \;,$$

for every x. With this notation at hand, it is not hard to show:

Theorem 14.5. *Let $f_1 \in \mathscr{D}^{\gamma_1}_{\alpha_1}(V)$, $f_2 \in \mathscr{D}^{\gamma_2}_{\alpha_2}(\bar{V})$, 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 $\mathscr{D}^{\gamma}_{\alpha}$ with*

$$\alpha = \alpha_1 + \alpha_2 \;, \qquad \gamma = (\gamma_1 + \alpha_2) \wedge (\gamma_2 + \alpha_1) \;. \tag{14.1}$$

Proof. It is clear that $f(x) \in T_{\geq\alpha}$, so it remains to show that it belongs to $\mathscr{D}^{\gamma}$. Furthermore, since we are only interested in showing that $f_1 \star f_2 \in \mathscr{D}^{\gamma}$, 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

$$\begin{aligned} \|\Gamma_{xy} f(y) - f(x)\|_\beta \leq \|&\big(\Gamma_{xy} f_1(y) - f_1(x)\big) \star \big(\Gamma_{xy} f_2(y) - f_2(x)\big)\|_\beta \\ &+ \|\big(\Gamma_{xy} f_1(y) - f_1(x)\big) \star f_2(x)\|_\beta \\ &+ \|f_1(x) \star \big(\Gamma_{xy} f_2(y) - f_2(x)\big)\|_\beta \;. \end{aligned} \tag{14.2}$$

It follows from the properties of the product $\star$ that the first term in (14.2) is bounded by a constant times

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

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

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

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 14.6. Strictly speaking, it is the projection of $f(x) = f_1(x) \star f_2(x)$ to $T_{<\gamma}$ that belongs to $\mathscr{D}_\alpha^\gamma$, see Exercise 13.4.

Remark 14.7. It is clear that the formula (14.1) 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 (abstract) truncated Taylor series f_i for f_i belong to $\mathscr{D}_0^{\gamma_i}$. 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 (14.1) states. The second reality check comes from (the proof of) Theorem 13.18. In this case, with $\beta > \alpha \geq 0$, one has $f \in \mathscr{D}_0^\beta$, while the constant function $x \mapsto \Xi$ belongs to $\mathscr{D}_{-\alpha}^\infty$ so that, according to (14.1), one expects their product to belong to $\mathscr{D}_{-\alpha}^{\beta-\alpha}$, 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 regular functions. Of course, one could in general not expect to be able to compose a regular function with a distribution of negative order. As a matter of fact, we will only define the composition of regular functions with elements in some $\mathscr{D}^\gamma$ 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 $\mathscr{D}^\gamma$.

For this purpose, we say that a given sector $V \subset T$ is *function-like* if $\alpha < 0 \implies V_\alpha = 0$ 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 $\big(\Pi_x \mathbf{1}\big)(y) = 1$. In this case, it turns out that if $f \in \mathscr{D}^\gamma(V)$ for a function-like sector V, then $\mathcal{R}f$ is a continuous function and one has the identity $\big(\mathcal{R}f\big)(x) = \langle \mathbf{1}, f(x)\rangle$, where we denote by $\langle \mathbf{1}, \bullet\rangle$ the element in the dual of V which picks out the prefactor of $\mathbf{1}$.

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

$$\big(G \circ f\big)(x) = \sum_{k \geq 0} \frac{G^{(k)}(\bar f(x))}{k!} \mathcal{Q}_{<\gamma} \tilde f(x)^{\star k} ,$$

where we have set

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

and weher $\mathcal{Q}_{<\gamma}\colon T \to T_{<\gamma}$ is the natural projection. 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 $\tilde f(x) \in T_{\geq \alpha_0}$. By the properties of the product, this implies that one has $\tilde{f}(x)^{\star k} \in T_{\geq k\alpha_0}$. 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 $G \circ f \in \mathscr{D}^{\gamma'}$ 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 Lemma 7.3:

Proposition 14.8. *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 $\mathscr{D}^\gamma(V)$ into itself. If $k > \gamma/\alpha_0 + 1$, then it is locally Lipschitz continuous.*

The proof of the first statement can be found in [Hai14b], while the second statement was shown in [HP15]. It is a somewhat lengthy, but ultimately rather straightforward calculation.

14.3 Classical Schauder estimates

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. A full proof of the Schauder estimates for regularity structures is beyond the scope of this book, but we want to convey the flavour of the proof. The aim of this section is therefore to give a self-contained proof of the classical Schauder estimates which state that for any (compactly supported) kernel K that is approximately homogeneous of degree $\beta - d$, the convolution map $\zeta \mapsto K * \zeta$ is continuous from $\mathcal{C}^\alpha$ to $\mathcal{C}^{\alpha+\beta}$, provided that $\alpha + \beta$ is not a positive integer. We first make precise our assumptions on the kernel K.

Definition 14.9. Given $\beta > 0$, a kernel $K \colon \mathbf{R}^d \setminus \{0\} \to \mathbf{R}$, smooth except for a singularity at the origin, is said to be *β-regularising* if it is supported in the unit ball around the origin and, for every $k \in \mathbf{N}^d$, there exists a constant C such that $|D^k K(x)| \leq C|x|^{\beta-d-|k|}$.

Immediate examples are (smooth truncations of) the Newton potential in dimension $d \geq 3$, proportional to $1/|x|^{d-2}$ and hence 2-regularising, the fractional Volterra kernel $(x^{H-1/2}1_{x>0})$ with $d = 1$ and $\beta = H + 1/2$. The heat kernel on space-time $\mathbf{R}^{n+1}$, proportional to $(t, x) \mapsto t^{-n/2} \exp(-\frac{|x|^2}{4t})1_{t>0}$, also fits in this setting (and is 2-regularising), provided one works with "parabolic" scaling (cf. Remark 13.9).

As in Section 13.3, and for any $r \in \mathbf{N}$, we work with $\mathcal{B}_r \subset \mathcal{D}$, the set of smooth test functions with $\mathcal{C}^r$-norm bounded by 1 and supported in the unit ball. It will be convenient for the purpose of this section to write $\mathcal{B}_{r,x}^\lambda$ for the set of all test functions of the form φ_x^λ with $\varphi \in \mathcal{B}_r$. Such $\psi \in \mathcal{B}_{r,x}^\lambda$ are characterised by having support in the ball of radius λ centred at x and derivatives bounds $|D^k\psi| \leq \lambda^{-d-|k|}$ for $|k| \leq r$. We also note that, for any real $s \in [0, r]$, the estimate $\|\psi\|_{\mathcal{C}^s} \lesssim \lambda^{-d-s}$ holds true.

Lemma 14.10. *Given a β-regularising kernel K and $r \geq 0$, one can write $K = \sum_{n\geq -1} K_n$ in such a way that $2^{\beta n} K_n \in C\mathcal{B}^{2^{-n}}_{r,0}$ for some $C > 0$.*

Proof. As is common in the construction of Paley–Littlewood blocks, we work with a dyadic partitions of unity, based on a smooth "cutoff" function" $\varphi : \mathbf{R}_+ \to [0,1]$, supported in $[2^{-1}, 2^1]$, such that $\sum_{n\geq 0} \varphi_n \equiv 1$ on $(0,1]$, where $\varphi_n := \varphi(2^n \bullet)$ is supported in $[2^{-n-1}, 2^{-n+1}]$. Since K is supported in $\{x : |x| \leq 1\}$, the stated decomposition clearly holds with (smooth) $K_n(x) := \varphi_{n+1}(|x|)K(x)$, supported in the ball of radius 2^{-n} centred at the origin. To see that $2^{\beta n} K_n \in C\mathcal{B}^{2^{-n}}_{r,0}$, for given $r \geq 0$, it remains to see that $|D^j K_n| \lesssim (2^{-n})^{\beta - d - |j|}$ for $|j| \leq r$. This estimate holds, with K_n replaced by K, by the defining property of a β-regularising kernel, restricted to $x \asymp 2^{-n}$. On the other hand, $|D^i \varphi_n| = |(2^n)^{|i|} D^i \varphi| \lesssim (2^n)^{|i|}$, and we conclude with Leibnitz' product rule. □

The following simple proposition is the first crucial ingredient in our approach. Loosely speaking, it states that the convolution of two test functions localised at two distinct scales is localised at the sum (or equivalently maximum) of the two scales and that one gains in amplitude if the tighter of the two test functions annihilates polynomials of a certain degree.

Proposition 14.11. *There exists $C > 0$ such that, for all $\varphi \in \mathcal{B}^\lambda_{r,x}$ and $\psi \in \mathcal{B}^\mu_{r,y}$, one has $\psi * \varphi \in C\mathcal{B}^{\lambda+\mu}_{r,x+y}$. If furthermore $\lambda \leq \mu$ and $\int P(z)\varphi(z)\,dz = 0$ for every polynomial P with $\deg P < \gamma \leq r$, some $\gamma \in \mathbf{R}_+$, then $\psi * \varphi \in C(\lambda/\mu)^\gamma \mathcal{B}^{2\mu}_{\lfloor r-\gamma \rfloor, x+y}$.*

Proof. Clearly, $\psi * \varphi$ is supported in the ball of radius $\lambda + \mu$ centred at $x + y$. For the first claim, by swapping the roles of φ and ψ if necessary, we may assume $\lambda \leq \mu$. To see that the convolution yields an element in $\mathcal{B}^{\lambda+\mu}_{r,x+y}$, in view of the characterisation of such spaces, it suffices to estimate, for $|k| \leq r$, $D^k(\psi * \varphi) = (D^k\psi) * \varphi$ using $|(D^k\psi)| \lesssim \mu^{-d-|k|} \asymp (\lambda + \mu)^{-d-|k|}$ and $\int |\varphi(z)|\,dz \leq C$ (independent of λ). Regarding the second claim, we write

$$
\begin{aligned}
D^k(\psi * \varphi)(\bullet) &= \int \psi^{(k)}(\bullet - z)\,\varphi(z)\,dz \\
&= \int \big(\psi^{(k)}(\bullet - z) - P^{\gamma;(k)}_{\bullet}(\bullet - z)\big)\,\varphi(z)\,dz\;,
\end{aligned}
$$

for $0 \leq |k| \leq r - \gamma$, where $P^{\gamma;(k)}_{\bullet}$ denotes the Taylor expansion (at the dotted base-point) of $\psi^{(k)} \equiv D^k\psi$ of integer degree $\gamma - \{\gamma\} < \gamma$ (annihilated by φ). It remains to be seen that, for all such k,

$$
|D^k(\varphi * \psi)(\bullet)| \lesssim (\lambda/\mu)^\gamma \mu^{-d-|k|}\,.
$$

To this end, using that $\gamma + |k| \leq r$, one has the estimate

$$
|\psi^{(k)}(\bullet - z) - P^{k,\gamma}_{\bullet}(\bullet - z)| \lesssim \|\psi\|_{\mathcal{C}^{\gamma+|k|}} |z|^\gamma \lesssim \mu^{-d-\gamma-|k|} |z|^\gamma\;.
$$

We only need to consider z in the support of φ, and in fact can assume without loss of generality that $x = 0$ (otherwise subtract another annihilated Taylor polynomial...), so that $\int |z|^\gamma |\varphi(z)|\,dz \le \lambda^\gamma \int |\varphi(z)|\,dz \lesssim \lambda^\gamma$. The desired estimate now follows. □

Our second crucial ingredient is a characterisation of Hölder spaces that is well adapted to our approach. For this, we define the following scale of spaces of distributions.

Definition 14.12. For $\alpha \in \mathbf{R}$, write $r = r_o(\alpha)$ for the smallest non-negative integer such that $r + \alpha > 0$. We then define $\mathcal{Z}^\alpha$ as the space of distributions on $\mathbf{R}^d$ such that for every compact set $\mathfrak{K} \subset \mathbf{R}^d$ there exists a constant C such that the bound

$$|\zeta(\varphi)| \le C\lambda^\alpha\;,$$

holds uniformly $\lambda \in (0,1], x \in \mathfrak{K}$ and all $\varphi \in \mathcal{B}_{r,x}^\lambda$ such that $\int \varphi(z)P(z)\,dz = 0$ for all polynomials P with $\deg P \le \alpha$. For 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 $\mathcal{Z}^\alpha$ with a Fréchet space structure.

The precise choice of r in Definition 14.12 is not very important, as one could have taken any other choice $r \ge r_o(\alpha)$. More precisely, one has the following result.

Lemma 14.13. *For $r \ge r_o(\alpha)$, write $\mathcal{Z}_r^\alpha$ for $\mathcal{Z}^\alpha$ as defined above, but with $r_o(\alpha)$ replaced by r. Then $\mathcal{Z}_r^\alpha = \mathcal{Z}^\alpha$.*

Proof. We fix a partition of unity $\{\chi_y\}_{y\in\Lambda}$ for $\mathbf{R}^d$ such that all the χ_y are translates of χ_0 by $y \in \mathbf{R}^d$ and $\Lambda \subset \mathbf{R}^d$ is a lattice. In particular, we make sure that $\chi_y \in \mathcal{B}_{r,y}^\lambda$. Given any $\lambda > 0$, we write $\chi_{y,\lambda}(x) = \chi_{y/\lambda}(x/\lambda)$ and we set $\Lambda_\lambda = \Lambda/\lambda$. We also fix a function $\psi \in \mathcal{C}^\infty$ with support in the centred unit ball and such that

$$\int_{\mathbf{R}^d} x^k \psi(x)\,dx = \delta_{k,0}\;, \qquad \forall k\,:\, |k| \le r\;. \tag{14.3}$$

(Such functions exist by Exercise 13.8.) We then write $\tilde\psi(x) = 2^d\psi(2x) - \psi(x)$ and note that by (14.3) one has $\int_{\mathbf{R}^d} x^k \tilde\psi(x)\,dx = 0$ for $|k| \le r$.

Let now $\alpha < 0$ and take $\zeta \in \mathcal{Z}_r^\alpha$, we want to show that $\zeta \in \mathcal{Z}^\alpha$. Given $\varphi \in \mathcal{B}_{r_o,x}^\lambda$ and setting $\lambda_n = 2^{-n}\lambda$, we write

$$\varphi = \varphi * \psi^\lambda + \sum_{n\ge 0}\sum_{y\in\Lambda_{\lambda_n}} \varphi_{n,y}\;, \qquad \varphi_{n,y} = \bigl(\varphi * \tilde\psi^{\lambda_n}\bigr)\cdot \chi_{y,\lambda_n}\;. \tag{14.4}$$

As a simple consequence of the Taylor remainder theorem, one has the bound

$$\bigl\|\varphi * D^k\tilde\psi^{\lambda_n}\bigr\|_\infty \lesssim \lambda^{-d}2^{-r_o n}\lambda_n^{-|k|} = 2^{-(d+r_o)n}\lambda_n^{-d-|k|}\;,$$

so that there exists a constant C independent of φ such that $\varphi_{n,y} \in C2^{-(d+r_o)n}\mathcal{B}_{r,y}^{\lambda_n}$, which in particular implies that

$$|\zeta(\varphi_{n,y})| \lesssim \lambda^\alpha 2^{-(d+r_o+\alpha)n} . \tag{14.5}$$

Since the number of terms in Λ_{λ_n} such that $\varphi_{n,y}$ is non-zero is of order 2^{nd}, we conclude that

$$|\zeta(\varphi)| \lesssim \lambda^\alpha + \sum_{n\geq 0} \lambda^\alpha 2^{-(r_o+\alpha)n} \lesssim \lambda^\alpha ,$$

where we used the fact that $r_o + \alpha > 0$ by definition.

Note that the assumption $\alpha < 0$ was used in order to obtain the bound (14.5) since there is no reason for $\varphi_{n,y}$ to annihilate polynomials even if φ does. The case $\alpha > 0$ is easier, noting that the definition of $\mathcal{Z}_r^\alpha$ implies that $\zeta * \tilde{\psi}^{\lambda_n}$ is a continuous function bounded by $\mathcal{O}(\lambda_n^\alpha)$. We then use the fact that

$$\zeta(\varphi) = \zeta(\varphi * \psi^\lambda) + \sum_{n\geq 0} \langle \zeta * \tilde{\psi}^{\lambda_n}, \varphi\rangle ,$$

with $\langle\cdot,\cdot\rangle$ denoting the L^2 scalar product, combined with the fact that φ integrates to $\mathcal{O}(1)$, to conclude that $|\zeta(\varphi)| \lesssim \lambda^\alpha(1 + \sum_{n\geq 0} 2^{-\alpha n}) \lesssim \lambda^\alpha$ as required.

The case $\alpha = 0$ is a bit more delicate and we leave it as Exercise 14.3. □

Remark 14.14. Validity of the stated bounds implies that distributions in $\mathcal{Z}^\alpha \subset \mathcal{D}'$ can be extended canonically to test functions in $\mathcal{C}_c^r$ (elements in $\mathcal{C}^r$ with compact support). In this sense, $\mathcal{Z}^\alpha$ is contained in the topological dual of $\mathcal{C}_c^r$. (The situation is similar in the definition of models, cf. Remark 13.7.)

For $\alpha < 0$, the polynomial-annihilation condition is void and there is no additional condition on φ besides $\varphi \in \mathcal{B}_{r,x}^\lambda$. In this case $\mathcal{Z}^\alpha$ is precisely the negative Hölder space $\mathcal{C}^\alpha$ introduced in Section 13.3.1. The following proposition shows that to some extent this is also true in case of positive Hölder spaces, as previously encountered in Section 13.3.1.

Proposition 14.15. *For $\alpha \notin \mathbf{N}$, one has $\mathcal{Z}^\alpha = \mathcal{C}^\alpha$.*

Proof. There is nothing to prove for $\alpha < 0$, so let $\alpha > 0$. We first show that $\mathcal{C}^\alpha \subset \mathcal{Z}^\alpha$, this inclusion also being valid for integer values of α. In fact, it suffices to note that, given $f \in \mathcal{C}^\alpha$ and $\varphi \in \mathcal{B}_{r,x}^\lambda$ as in Definition 14.12, one has

$$\int f(y)\varphi(y)\,dy = \int \bigl(f(y) - P_x^\alpha(y-x)\bigr)\varphi(y)\,dy \lesssim \lambda^\alpha ,$$

where the identity follows from the fact that φ annihilates P_x^α, the Taylor expansion at order α of f, based at x, and the bound is as in the proof of Proposition 14.11.

For the converse inclusion, we first consider the case $\alpha \in (0,1)$ and let $\zeta \in \mathcal{Z}^\alpha$. Let $\varrho\colon \mathbf{R}^d \to \mathbf{R}$ be a smooth function that is compactly supported in the unit ball around the origin and such that $\int \varrho(z)\,dz = 1$. Note first that, for any $x \in \mathbf{R}^d$ and $\lambda \in (0,1]$, it follows from the definition of $\mathcal{Z}^\alpha$ that one has the bound

$$|\zeta(\varrho_x^{2^{-n}\lambda}) - \zeta(\varrho_x^{2^{-n-1}\lambda})| = |\zeta(\varrho_x^{2^{-n}\lambda} - \varrho_x^{2^{-n-1}\lambda})| \leq C\lambda^\alpha 2^{-\alpha n} .$$

It follows that $f(x) = \lim_{n\to\infty} \zeta(\varrho_x^{2^{-n}\lambda})$ is well-defined and that

$$|f(x) - \zeta(\varrho_x^\lambda)| \lesssim \lambda^\alpha \;.$$

As a consequence, one has

$$|f(x) - f(y)| \lesssim \lambda^\alpha + \big|\zeta(\varrho_x^\lambda - \varrho_y^\lambda)\big| \;.$$

Choosing $\lambda = |x-y|$, it follows that $f \in \mathcal{C}^\alpha$. The fact that $f = \zeta$ in the sense that $\zeta(\varphi) = \int f(z)\,\varphi(z)\,dz$ follows immediately from the fact that

$$\zeta(\varphi) = \lim_{\lambda\to 0} \zeta(\varphi * \varrho^\lambda) = \lim_{\lambda\to 0} \int \zeta(\varrho_x^\lambda)\,\varphi(x)\,dx \;.$$

The claim for general non-integer α can then be seen from the fact that $\zeta \in \mathcal{Z}^\alpha$ implies $D^k\zeta \in \mathcal{Z}^{\alpha-|k|}$ (interpreted as distributional derivatives) for every multi-index k. Details are left to the reader. □

Remark 14.16. For $n \in \mathbf{N}$, the spaces $\mathcal{Z}^n$ are usually called *Hölder–Zygmund spaces* in the literature (thus our choice of symbol $\mathcal{Z}$). They are distinct from the usual Hölder spaces since one can check that $x \mapsto x^n \log x$ belongs to $\mathcal{Z}^n$, but not to $\mathcal{C}^n$.

With all of these preliminaries in place, we can give a very simple proof of Schauder's theorem. (See for example [Sim97] for an alternative proof of a very similar statement.)

Theorem 14.17. *For any β-regularising kernel K, the map $\zeta \mapsto K * \zeta$ is continuous from $\mathcal{Z}^\alpha$ to $\mathcal{Z}^{\alpha+\beta}$ for every $\alpha \in \mathbf{R}$.*

Proof. Let $\zeta \in \mathcal{Z}^\alpha$ and let $\varphi \in \mathcal{B}_{r,x}^\lambda$ where we will (and can by Lemma 14.13) work with suitable $r \geq r_o(\alpha+\beta)$, chosen below, such that $\int \varphi(z)P(z)\,dz = 0$ for every P with $\deg P \leq \alpha+\beta$. Lemma 14.10 yields a decomposition $(K_n : n \geq -1)$ for $\check K(x) = K(-x)$, so that

$$(K * \zeta)(\varphi) = \zeta(\check K * \varphi). = \sum_n \zeta(K_n * \varphi) = \sum_n 2^{-\beta n}\zeta(2^{\beta n}K_n * \varphi) \;, \quad (14.6)$$

with $2^{\beta n}K_n \in C\mathcal{B}_{r,0}^{2^{-n}}$ for some $C > 0$. It then follows from Proposition 14.11 (applied with $\mu = 2^{-n}$, noting that $K_n * \varphi$ also annihilates polynomials of degree up to $\alpha+\beta$) and the definition of $\mathcal{Z}^\alpha$ that

$$|\zeta(2^{\beta n}K_n * \varphi)| \lesssim \begin{cases} \lambda^\alpha & \text{if } 2^{-n} \leq \lambda, \\ (2^n\lambda)^\gamma 2^{-\alpha n} & \text{otherwise,} \end{cases}$$

provided $\lfloor r-\gamma\rfloor \geq r_o(\alpha+\beta)$. We will also need $\gamma > \alpha+\beta$, so that for instance $r := 2(|\alpha|+\beta)+2$ is a safe choice. Inserting this bound into (14.6), and using $\beta > 0, \gamma > \alpha+\beta$ to estimate the geometric sums, one has the bounds

$$\sum_{\substack{n\geq 0 \\ 2^{-n}\leq\lambda}} 2^{-\beta n}\lambda^{\alpha} \lesssim \lambda^{\alpha+\beta}, \qquad \sum_{\substack{n\geq 0 \\ 2^{-n}\geq\lambda}} 2^{(\gamma-\alpha-\beta)n}\lambda^{\gamma} \lesssim \lambda^{\alpha+\beta},$$

it follows that $|(K * \zeta)(\varphi)| \lesssim \lambda^{\alpha+\beta}$, whence the claim follows. □

Remark 14.18. The proof is (much) simpler in the "negative" case, with Hölder exponents $\alpha < \alpha + \beta < 0$. In essence, this is due to the absence of polynomial vanishing conditions. More specifically, one can take $r = r_o(\alpha + \beta)$ in the above proof, and then $\gamma = 0$ later on, so that only the easy (first) part of Proposition 14.15 is used. A reduction of the general to the negative case, in dimension $d = 1$, is discussed in Exercise 14.2.

Remark 14.19. One can verify that the proof never made explicit use of the Euclidean scaling and can be adapted mutatis mutandis to the case of arbitrary scalings as mentioned in Remark 13.9, provided that the notion of "β-regularising kernel" is adjusted accordingly (replace the exponent $\beta - d - |k|$ by $\beta - |\mathfrak{s}| - |k|_{\mathfrak{s}}$).

14.4 Multilevel Schauder estimates and admissible models

As we saw in the previous section, the classical Schauder estimates state that if $K\colon \mathbf{R}^d \to \mathbf{R}$ is a kernel that is smooth everywhere, except for a singularity at the origin that is approximately homogeneous of degree $\beta - d$ for some fixed $\beta > 0$ (i.e. it is β-regularising in the sense of Definition 14.9), 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 $\alpha + \beta \in \mathbf{N}$.

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 $\mathcal{K}\colon \mathscr{D}^\gamma \to \mathscr{D}^{\gamma+\beta}$ with the property that $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$. We call such a statement a "multi-level Schauder estimate" since it is a form of Schauder estimate for all the components of f in T_α for all $\alpha < \gamma$. Of course, such a statement can only be expected to hold 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 a given real-valued function f is 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 need the following ingredients to build an operator $\mathcal{K}$ with the desired properties:

- The 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" $\mathcal{I}$ (of order β) 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 regularity structure contains abstract polynomials in the following sense.

Assumption 14.20 *There exists a sector $\bar{T} \subset T$ isomorphic to the polynomial regularity structure. 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 $\Gamma X^k = (X + h\mathbf{1})^k$ for some $h \in \mathbf{R}^d$.*

Furthermore, we assume that there exists an abstract integration operator $\mathcal{I}$, of fixed order $\beta > 0$, with the following properties.

Assumption 14.21 *There exists a linear map $\mathcal{I}\colon V \to T$ for some sector $V \subset T$ such that $\mathcal{I}V_\alpha \subset T_{\alpha+\beta}$ and, for every $\Gamma \in G$ and $\tau \in T$,*

$$\Gamma \mathcal{I}\tau - \mathcal{I}\Gamma\tau \in \bar{T} \,. \tag{14.7}$$

Remark 14.22. We do not want to assume $\Gamma\mathcal{I} = \mathcal{I}\Gamma$. This is already seen in case of the rough path structure given by Definition 13.4. The map $\mathcal{I} : \dot{W}^i \mapsto W^i$, $1 \leq i \leq e$, constitutes an abstract integration operator (defined on the sector $T_{\alpha-1}$). Since a generic $\Gamma_h \in G$ maps W^i to $W^i + h^i\mathbf{1}$, we see that $\Gamma\mathcal{I} - \mathcal{I}\Gamma \neq 0$ (for $h \neq 0$) and takes values in $T_0 = \langle \mathbf{1} \rangle$.

Finally, we want to restrict our attention to models that are compatible with this structure for a given kernel K in the following sense.

Definition 14.23. Given a β-regularising kernel K and a regularity structure $\mathscr{T}$ satisfying Assumptions 14.20 and 14.21, we say that a model (Π, Γ) is *admissible* if the identities

$$\big(\Pi_x X^k\big)(y) = (y-x)^k \,, \qquad \Pi_x \mathcal{I}\tau = K * \Pi_x \tau - \Pi_x \mathcal{J}_x \tau \,, \tag{14.8}$$

hold for every $\tau \in V$. Here, $\mathcal{J}_x\colon V \to \bar{T}$ is the linear map given on homogeneous elements by

$$\mathcal{J}_x \tau = \sum_{|k| < \deg \tau + \beta} \frac{X^k}{k!} \int D^k K(x-y)\,\big(\Pi_x \tau\big)(dy) \,. \tag{14.9}$$

Remark 14.24. In some cases, it will be convenient to introduce a whole family $\mathcal{I}_k$ of integration operators of order $\beta - |k|$. The notion of admissibility is then defined similarly, with $\mathcal{I}$ replaced by $\mathcal{I}_k$ and K replaced by $D^k K$, to the extent that these symbols are included in the structure space.

Remark 14.25. If ξ is smooth and we furthermore impose that Π_x is multiplicative (which is not enforced in general!), this yields a recursion to define the *canonical model* associated to ξ provided one manages to construct Γ_{xy} at the same time. The correct recursion to do this is

$$\Gamma_{xy}(\mathcal{I} + \mathcal{J}_y)\tau = (\mathcal{I} + \mathcal{J}_x)\Gamma_{xy}\tau \;, \tag{14.10}$$

which is clearly consistent with the constraint (14.7) and which one can show guarantees that $\Pi_x \Gamma_{xy} \mathcal{I}\tau = \Pi_y \mathcal{I}\tau$. See also Exercise 14.6.

Remark 14.26. Recall that if P is a polynomial and K is a compactly supported function, then $K * P$ is again a polynomial of the same degree as P. Since, for $\Pi_x\tau$ smooth enough, the term $\Pi_x \mathcal{J}_x \tau$ appearing in (14.8) is nothing but the Taylor expansion of $K * \Pi_x \tau$ around x, it follows that one has $\Pi_x \mathcal{I} X^k = 0$ for any multi-index k and any admissible model, which would suggest that one could have imposed the identity $\mathcal{I} X^k = 0$ already at the algebraic level. This would however create inconsistencies later on when incorporating renormalisation, unless we assume that $\int K(x)P(x)\,dx = 0$ for every polynomial P of degree N, for some sufficiently large value of N. Here, we chose to simply add instead $\mathcal{I}X^k$ as separate symbols to our regularity structure and to then set $\mathcal{I}X^k = \mathcal{I}X^k$.

Remark 14.27. 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 (14.9) 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 (14.9) is by

$$\mathcal{J}_x \tau = \sum_{|k| < \deg \tau + \beta} \sum_{n \geq 0} \frac{X^k}{k!} \big(\Pi_x \tau\big)\big(D^k K_n(x - \bullet)\big) \;,$$

with K_n as in Lemma 14.10. The scaling properties of the K_n ensure that the function $2^{(\beta - |k|)n} D^k K_n(x - \bullet)$ is a test function that is localised around x at scale 2^{-n}. As a consequence, one has

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

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

Remark 14.28. 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 β-regularising kernel K, 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 [Hai14b] 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 $\deg \tau + \beta \in \mathbf{N}$. Since the bounds appearing both in the definition of a model and in that of a β-regularising kernel 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 $\mathcal{K}\colon \mathscr{D}^\gamma \to \mathscr{D}^{\gamma+\beta}$ announced at the beginning of this section. Recalling the definition of $\mathcal{J}$ from (14.9), we set

$$\big(\mathcal{K}f\big)(x) = \mathcal{I}f(x) + \mathcal{J}_x f(x) + \big(\mathcal{N}f\big)(x)\;, \tag{14.11}$$

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

$$\big(\mathcal{N}f\big)(x) = \sum_{|k|<\gamma+\beta} \frac{X^k}{k!} \int D^k K(x-y)\,\big(\mathcal{R}f - \Pi_x f(x)\big)(dy)\;. \tag{14.12}$$

Note first that thanks to the reconstruction theorem, it is possible to verify that the right-hand side of (14.12) does indeed make sense for every $f \in \mathscr{D}^\gamma$ in virtually the same way as in Remark 14.27. One has:

Theorem 14.29. *Let K be a β-regularising kernel, let $\mathscr{T} = (T, G)$ be a regularity structure satisfying Assumptions 14.20 and 14.21, and let (Π, Γ) be an admissible model for $\mathscr{T}$. Then, for every $f \in \mathscr{D}^\gamma$ with $\gamma \in (0, N-\beta)$ and $\gamma + \beta \notin \mathbf{N}$, the function $\mathcal{K}f$ defined in (14.11) belongs to $\mathscr{D}^{\gamma+\beta}$ and satisfies $\mathcal{R}\mathcal{K}f = K * \mathcal{R}f$.*

Proof. The complete proof of this result can be found in [Hai14b] and will not be given here. Since it is rather straightforward, we will however give a proof of Schauder's estimate in the classical case (i.e. that of the polynomial regularity structure) in Section 14.3 below.

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 13.27 applies. In this case, one has

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

As a consequence of (14.8), the first term appearing in the right-hand side of this expression is given by

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

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 $\big(K * \Pi_x f(x)\big)(x)$ cancels, leaving us with the desired result. □

We are now in principle in possession of all of the ingredients required to formulate fixed point problems for a large number of semilinear stochastic PDEs: multiplication, composition by regular functions, differentiation, and integration against the Green's function of the linearised equation. Before we show how this can be leveraged in practice in order to build a robust solution theory for the KPZ equation, we briefly explore some of main concepts in setting of (very) rough paths.

14.5 Rough volatility and robust Itô integration revisited

Recent applications from mathematical finance, where $\sigma(t,\omega) = \sigma(\widehat{W}_t)$ models *rough stochastic volatility*, involve (standard) Itô integrals of the form

$$\int_0^T \sigma(\widehat{W}_t)d(W_t,\bar{W}_t) \equiv \int_0^T f(\widehat{W}_t)dW_t + \int_0^T \bar{f}(\widehat{W}_t)d\bar{W}_t\;, \tag{14.13}$$

where $\sigma = (f,\bar{f}) : \mathbf{R} \to \mathbf{R}^2$ is a sufficiently smooth map, $(W,\bar{W})$ is a 2-dimensional standard Brownian motion, and $\widehat{W}_t$ given by

$$\int K^H(t-s)\,dW_s\;, \tag{14.14}$$

with Riemann–Liouville kernel $K^H(x) = x^{H-1/2}1_{x>0}$. Since $K^H \in L^2_{\text{loc}}(\mathbf{R})$ but not in $L^2(\mathbf{R})$, we replace it in the sequel by a compactly supported K, smooth away from zero and equal to K^H in some neighbourhood of zero. We then require W to be a two-sided Brownian motion, so that $\xi := \dot{W}$ defines Gaussian white noise on $\mathbf{R}$, and

$$\widehat{W} = K * \xi\;. \tag{14.15}$$

Alternatively, as done in [BFG+19], see also [BFG20], one can restrict integration in (14.14) to $[0,t]$ with the benefit of exactly recovering Brownian motion $\widehat{W} = W$ for $H = 1/2$ in which case the integral (14.13) fits squarely into rough integration theory (namely Theorem 4.4, applied with the Itô Brownian rough path from Proposition 3.4). However, for $H \in (0,1/2)$ rough integration must fail. Indeed, K is $(1/2+H)$-regularising so that it follows from Schauder's Theorem 14.17 that $\widehat{W}$ and then $\sigma(\widehat{W})$ have generically H^--Hölder regularity and hence cannot be expected to be controlled by $W \in \mathcal{C}^{1/2^-}$. We can make (minor) progress by noting that $(\widehat{W},\bar{W})$ is a 2-dimensional Gaussian process with *independent* components. At least for $H > 1/3$, the results of Section 10.3 for Gaussian rough paths apply essentially directly to the final integral $\int \bar{f}(\widehat{W})d\bar{W}$ above and Exercise 14.8 allows to deal with arbitrary $H > 0$.

The remainder of this section will focus on the other, seemingly harmless, one-dimensional Itô integral, with $\widehat{W}$ as given in (14.15),

$$\int_0^T f(\widehat{W})dW \,. \tag{14.16}$$

We are interested in a robust form of this Itô stochastic integral. In case of $\widehat{W} = W$ we can in fact express (14.16) via Itô's formula, which immediately gives a version of this integral which is continuous in W, even in uniform topology. Certainly, this trick fails when $\widehat{W} \neq W$.

In this section we set up a regularity structure that provides a full solution to this problem. Needless to say, this structure is much simpler than what is needed for the KPZ equation in the next chapter. Yet, it showcases a number of features omnipresent for singular SPDEs, but without some of the added complexity coming from PDE theory.

Recall that the Hölder exponent of $\widehat{W}$ is $H - \kappa$ for any $\kappa > 0$. As a result, we have $|\widehat{W}^m_{s,t}| \lesssim |t-s|^{m(H-\kappa)}$ and the building blocks for a robust representation of (14.16) are

$$\mathbb{W}^m_{s,t} = \int_s^t (\widehat{W}_{s,r})^m \, dW_r \quad , \tag{14.17}$$

with $m = 0, 1, 2, \ldots, M$ where M is the smallest integer such that $(M+1)H + 1/2 > 1$, which reflects the analytic redundancy of $\mathbb{W}^{M+1}$ in the sense of

$$|\mathbb{W}^{M+1}(s,t)| \lesssim |t-s|^{(M+1)(H-\kappa)+1/2} = o(t-s) \,,$$

for small enough $\kappa > 0$. For definiteness, let us focus on the case

$$H > \frac{1}{8}, \quad M = 3 \,.$$

We first define symbols (these will be the basis vectors of our regularity structure) to represent $(\widehat{W}_{s,t})^m, 0 \leq m \leq 3$. If $\Xi \equiv \circ$ is the symbol for white noise $\xi \equiv \dot{W}$, we can write the required symbols indifferently as

$$\{\mathbf{1}, \mathcal{I}(\Xi), \mathcal{I}(\Xi)^2, \mathcal{I}(\Xi)^3\} \;\equiv\; \{\mathbf{1}, ⫯, \Psi, \Psi\!\!\!|\}.$$

The map $\mathcal{I} : \Xi \mapsto \mathcal{I}(\Xi)$ represents convolution with K and is graphically represented by a downfacing plain line; multiplication (which we postulate to be commutative and associative) is depicted by joining trees at their roots. For instance, $⫯ \star \Psi = \Psi\!\!\!|$ (we will omit $\star$ in the sequel). Similarly, the symbols denoting $(\widehat{W}_{s,t})^m \dot{W}_t$, defined as the generalised derivative $\partial \mathbb{W}^m_{s,\bullet}$, are given in the same pictorial representation as $\{\circ, ⫯_\circ, \Psi_\circ, \Psi\!\!\!|_\circ\}$ (with for example $\Psi_\circ = \mathcal{I}(\Xi)^2\Xi$). We then define the structure space of our regularity structure as the free vector space generated by these symbols, namely

$$T = \langle \Xi, \Xi\mathcal{I}(\Xi), \Xi\mathcal{I}(\Xi)^2, \Xi\mathcal{I}(\Xi)^3, \mathbf{1}, \mathcal{I}(\Xi), \mathcal{I}(\Xi)^2, \mathcal{I}(\Xi)^3 \rangle \; . \tag{14.18}$$

The partial product defined on T (for example $\Xi \cdot \mathcal{I}(\Xi)^2 = \Xi\mathcal{I}(\Xi)^2$) does not extend to all of T.[1] It is natural to postulate that Ξ has degree $\deg \Xi = -\frac{1}{2}^-$ (the presence of the exponent '$-$' reflects the fact that in order for the bound (13.13) to be satisfied when $\Pi_t \Xi$ is given by white noise, we need to make sure that $\deg \Xi$ is strictly smaller than $-\frac{1}{2}$, but by how much exactly is irrelevant as long as it is a small enough quantity), that $\mathcal{I}$ increases degree by $H + \frac{1}{2}$, and that the degree is additive under multiplication. Since it is natural to take $\deg \mathbf{1} = 0$ to retain consistency with the polynomial regularity structure, this uniquely determines the degree of each of the basis vectors of T, for instance

$$\deg \Xi\mathcal{I}(\Xi)^3 = \deg \Xi + 3 \deg \mathcal{I}(\Xi) = (3H - \tfrac{1}{2})^- \; .$$

To understand the structure group, we shift from a base point s to a new base point t. Basic additivity properties of the integral in (14.17) show that

$$\mathbb{W}^3_{s,\cdot} = \mathbb{W}^3_{t,\cdot} + 3\mathbb{W}^2_{t,\cdot}\widehat{W}_{s,t} + 3\mathbb{W}^1_{t,\cdot}\widehat{W}^2_{s,t} + \mathbb{W}^0_{t,\cdot}\widehat{W}^3_{s,t} + \mathbb{W}^3_{s,t} \; .$$

Considering the (generalised) derivative in the free variable, we have

$$\partial\mathbb{W}^3_{s,\cdot} = \partial\mathbb{W}^3_{t,\cdot} + 3(\partial\mathbb{W}^2_{t,\cdot})\widehat{W}_{s,t} + 3(\partial\mathbb{W}^1_{t,\cdot})\widehat{W}^2_{s,t} + (\partial\mathbb{W}^0_{t,\cdot})\widehat{W}^3_{s,t} \; . \tag{14.19}$$

This suggests to "break up" the symbol $\Xi\mathcal{I}(\Xi)^3$ (for $\partial\mathbb{W}^3_{*,\cdot}$) in the form

$$\Delta^+(\Xi\mathcal{I}(\Xi)^3) := \Xi\mathcal{I}(\Xi)^3 \otimes \mathbf{1} + 3\Xi\mathcal{I}(\Xi)^2 \otimes \mathcal{J}(\Xi) + 3\Xi\mathcal{I}(\Xi) \otimes \mathcal{J}(\Xi)^2 + \Xi \otimes \mathcal{J}(\Xi)^3 \in T \otimes T^+ \; ,$$

where the introduction of a new space T^+ is justified by the fact that elements in T^+ represent functions of two variables (s and t here), while elements of T represent functions of one variable (the base point s resp. t) that are distributions in the remaining free variable. In particular, it is rather natural that T^+ (unlike T) contains no symbols of negative degree and that elements of T^+ can be multiplied freely. In other words, it is natural in this context to define T^+ as the free commutative algebra generated by the single element $\mathcal{J}(\Xi) \stackrel{\text{def}}{=} \mathcal{J}(\circ)$. The difference between T^+ and T is emphasised in our notation by drawing basis vectors of T^+ in black.

The action of the linear map $\Delta^+ : T \to T \otimes T^+$ has the appealing graphical interpretation of *cutting off positive branches*: for instance, the summand $3\Xi\mathcal{I}(\Xi)^2 \otimes \mathcal{J}(\Xi) = \Xi\mathcal{I}(\Xi)^2 \otimes 3\mathcal{J}(\Xi)$ in $\Delta^+(\Xi\mathcal{I}(\Xi)^3)$ is explained as follows: there are three ways to "cut off" a "lollipop" $\mathcal{I}(\Xi)$ from $\Xi\mathcal{I}(\Xi)^3$, which are then painted black and put as $3\mathcal{J}(\Xi) \in T^+$ to the right-hand side; the remaining "pruned" tree $\Xi\mathcal{I}(\Xi)^2 \in T$ goes to the left. Similarly, there are three ways to cut off two lollipops from $\Xi\mathcal{I}(\Xi)^3$, which then appear as $3\mathcal{J}(\Xi)^2 \in T^+$ on the right-hand side, while the pruned remainder $\Xi\mathcal{I}(\Xi) \in T$ appears on the left.

[1] For instance, we do *not* want our regularity structure to contain a symbol Ξ^2 denoting the square of white-noise. We also have no need for trees with ≥ 4 branches so that products like $\mathcal{I}(\Xi)\,\Xi\mathcal{I}(\Xi)^3$, $\mathcal{I}(\Xi)^2\,\Xi\mathcal{I}(\Xi)^2$ etc. remain deliberately undefined within T.

A concise recursive algebraic description of Δ^+ starts with

$$\Delta^+\mathbf{1} = \mathbf{1}\otimes\mathbf{1}\,, \quad \Delta^+\Xi = \Xi\otimes\mathbf{1}\,,$$

followed by an extension to all of T by imposing the identities[2]

$$\begin{aligned}\Delta^+(\tau\bar\tau) &= \Delta^+\tau\cdot\Delta^+\bar\tau\,,\\ \Delta^+\mathcal{I}(\tau) &= (\mathcal{I}\otimes \mathrm{Id})\Delta^+\tau + \mathbf{1}\otimes\mathcal{J}(\tau)\,.\end{aligned}$$

Here, $\mathcal{J}(\tau)$ is the element in T^+ obtained from a (then painted black) symbol τ. In our pictorial representation $\mathcal{J}$ is visualised by a (black) downfacing line. The tree associated to $\mathcal{J}(\tau)$ has exactly one line emerging from the root (such trees are called *planted*). In the present example, $\tau = \circ$ is the only symbol in T, as given in (14.18), with image under $\mathcal{I}$ in T, so that the second relation above can only produce ⫯ $= \mathcal{J}(\circ)\in T^+$; whereas the first relation leads to powers thereof (in T^+).

Let now G_+ denote the set of characters on T^+, i.e. all linear maps $g\colon T^+\to\mathbf{R}$ with the property that $g(\sigma\bar\sigma) = g(\sigma)g(\bar\sigma)$ for any two elements σ and $\bar\sigma$ in T^+. There is not much choice here, since $c = g($⫯$)\in\mathbf{R}$ fully determines any such map. In order to get back to (14.19), we introduce $\Gamma_g\colon T\to T$ by

$$\Gamma_g\tau = (\mathrm{Id}\otimes g)\Delta^+\tau\,, \tag{14.20}$$

so that, for instance, $\Gamma_g(\Psi) = \Psi + 3c\,Y + 3c^2\,\wr + c^3\circ\in T$, and with $c = g($⫯$) = \widehat{W}_{s,t}$ this precisely captures (14.19) as an abstract shift map $\Gamma_{st} = \Gamma_{g_{s,t}}$ with $g_{s,t}($⫯$) = \widehat{W}_{s,t}$. In principle, (14.20) makes sense for every $g\in(T^+)^*$, but it turns out that the set of those maps Γ_g with $g\in G_+$ forms a group, which is precisely our structure group:

$$G := \{\Gamma_g : g\in G_+\}. \tag{14.21}$$

Written in matrix form, with respect to the ordered basis of T consisting of 4 negative and 4 non-negative symbols, each Γ_g is block-diagonal with two (4×4)-blocks of the form

$$\begin{pmatrix} 1 & c & c^2 & c^3\\ 0 & 1 & 2c & 3c^2\\ 0 & 0 & 1 & 3c\\ 0 & 0 & 0 & 1\end{pmatrix} =: N_c$$

One can check that $N_cN_{\bar c} = N_{c+\bar c}$ with $c,\bar c\in\mathbf{R}$ so that, as a group, G is isomorphic to $(\mathbf{R},+)$. This completes the construction of the regularity structure (T,G). We leave it to the reader to identify pairs of sectors on which (the usually omitted) $\star$ defines a product in the sense of Section 14.2 and to show that $\mathcal{I}$ is indeed an abstract integration operator[3] in the sense of Definition 14.21.

[2] The multiplicative property is understood for all symbols $\tau,\bar\tau\in T$ which can be multiplied in T.

[3] In the present setting there is no need to include higher order abstract polynomials $X, X^2,\dots$ as part of T.

As already hinted at, the natural *Itô model* $\mathrm{M}^{\text{Itô}} := (\Pi, \Gamma)$ in this context is defined by setting

$$\Pi_s \mathbf{1} = 1\,, \quad \Pi_s \Xi = \dot{W}\,, \quad \Pi_s(\mathcal{I}(\Xi)^m) = \widehat{W}^m_{s,\bullet}\,, \quad \Pi_s(\Xi\mathcal{I}(\Xi)^m) = \partial \mathbb{W}\,,$$

as well as $\Gamma_{st} = \Gamma_{g_{s,t}}$ with $g_{s,t}(\rotatebox{90}{$\circ$}) = \widehat{W}_{s,t}$. We leave it to the reader to check that $\mathrm{M}^{\text{Itô}}$ satisfies the required bounds (13.13) and therefore really defines a random model for the regularity structure (T, G). We also note that the model is *admissible* in the sense of Definition 14.23: in essence, this is seen from the identity

$$\Pi_s \mathcal{I}\Xi = K * \Pi_s \Xi - \Pi_s \mathcal{J}(s)\Xi = K * \dot{W} - (K * \dot{W})(s) = \widehat{W}_{s,\cdot} \qquad (14.22)$$

where we used that only $k = 0$ figures in the sum of (14.9), so that

$$\mathcal{J}_s \Xi = \mathbf{1} \int K(s-t)\,\big(\Pi_t \Xi\big)(dt) = (K * \dot{W})(s)\,\mathbf{1}\,.$$

On the other hand, we can replace white noise $\dot{W} = \dot{W}(\omega)$ by a mollification $\dot{W}^\varepsilon := \delta^\varepsilon * \dot{W}$ with $\delta^\varepsilon(t) = \varepsilon^{-1}\varrho(\varepsilon^{-1}t)$, for some $\varrho \in \mathcal{C}^\infty_c$ with $\int \varrho = 1$, or indeed any smooth function ξ, and define the associated *canonical model* $\mathscr{L}(\xi) = (\Pi, \Gamma)$ by prescribing

$$\Pi_s \Xi = \xi, \quad \Pi_s(\mathcal{I}(\Xi)^m) = (K * \xi)^m_{s,\cdot}, \quad \Pi_s(\Xi\mathcal{I}(\Xi)^m) = \xi(\cdot)(K * \xi)^m_{s,\bullet}\,,$$

as well as $g_{s,t}(\rotatebox{90}{$\circ$}) = (K * \xi)_{s,t}$. We again leave it to the reader to check that $\mathscr{L}(\xi)$ is indeed an admissible model for our regularity structure.

It is interesting to consider the canonical model $\mathscr{L}(\dot{W}^\varepsilon)$ as $\varepsilon \to 0$. Formally, one would expect convergence to a "Stratonovich model", but this does not exist because of an *infinite Itô–Stratonovich correction*. To wit, assume the approximate bracket

$$[W, \widehat{W}]^\pi := \sum_{[s,t]\in\pi} W_{s,t}\widehat{W}_{s,t}$$

converges, say in L^1, upon refinement $|\pi| \to 0$. Then the mean would have to convergence, which is contradicted by the computation, using Itô isometry,

$$\begin{aligned}\mathbf{E} W_{s,t}\widehat{W}_{s,t} &= \int_s^t K(t-r)dr = \int_0^{t-s} K(r)dr \\ &\sim \int_0^{t-s} K^H(r)dr = c_H (t-s)^{H+\frac{1}{2}}\,,\end{aligned}$$

and the standing assumption that $H < 1/2$. As a consequence, the canonical model $\mathscr{L}(\dot{W}^\varepsilon)$ will not converge as $\varepsilon \to 0$, although the previous discussion suggests to "cure" this by subtracting a diverging term, namely to consider[4]

[4] This is an instance of *Wick renormalisation* where one replaces the product of two scalar Gaussian random variables X, Y by $X \diamond Y := XY - \mathbf{E}[XY]$.

$$\int \widehat{W}^{\varepsilon} dW^{\varepsilon} - \mathbf{E}\Big(\int \widehat{W}^{\varepsilon} dW^{\varepsilon} \Big) , \tag{14.23}$$

with integration understood over $[s,t]$ with re-centred integrand $\widehat{W}_{s,\cdot}$. However, such Wick renormalisation at the level of generalised increments may destroy the algebraic Chen relations. (Indeed, they only hold when the expectation is proportional to $[s,t]$, which has no reason to be the case in general.)

In fact, our admissible model (Π, Γ) here can be described in terms of a single "base-point free" realisation map $\boldsymbol{\Pi} : T \to \mathcal{D}'$ which enjoys somewhat more natural relations, such as

$$\boldsymbol{\Pi} \mathcal{I} \Xi = K * \boldsymbol{\Pi} \Xi = K * \dot{W} = K * \xi$$

instead of (14.22) in the Itô-model case, and similarly for $\boldsymbol{\Pi}^{\varepsilon}$ with $\dot{W}$ replaced by $\dot{W}^{\varepsilon} = \xi^{\varepsilon}$. The full specification reads[5]

$$\begin{aligned} \boldsymbol{\Pi}^{\varepsilon} \mathbf{1} &= 1, & \boldsymbol{\Pi}^{\varepsilon} \Xi &= \xi^{\varepsilon}, \\ \boldsymbol{\Pi}^{\varepsilon} (\mathcal{I}(\Xi)^m) &= (K * \xi^{\varepsilon})^m, & \boldsymbol{\Pi}^{\varepsilon} (\Xi \mathcal{I}(\Xi)^m) &= \xi^{\varepsilon} (K * \xi^{\varepsilon})^m . \end{aligned} \tag{14.24}$$

Remark 14.30. Define a character f_t on T^+ by specifying (in the Itô model[6])

$$f_t(\mathord{\uparrow}) = f_t(\mathcal{J}(\Xi)) := \int K(t-s) \, \big(\Pi_t \Xi\big)(s) = (K * \xi)_t , \tag{14.25}$$

and also a linear map $F_t \colon T \to T$ by $F_t \tau = (\mathrm{Id} \otimes f_t) \Delta^+ \tau$. One checks without difficulty that F_t is an invertible map, $\Gamma_{ts} = F_t^{-1} \circ F_s$ and

$$\boldsymbol{\Pi} = \Pi_s F_s^{-1} = \Pi_t F_t^{-1} \implies \Pi_s = \Pi_t F_t^{-1} \circ F_s = \boldsymbol{\Pi} \circ F_s .$$

At the level of the canonical model $\boldsymbol{\Pi}^{\varepsilon}$, switching to $\Pi_t^{\varepsilon} = \boldsymbol{\Pi}^{\varepsilon} F_t$, this construction merely replaces $K * \xi^{\varepsilon}$ with the "base-pointed" expression $(K * \xi^{\varepsilon})_{t,\cdot}$ and tracks the induced changes to the higher levels.

The Wick renormalisation in (14.23) points us to the (divergent) quantity[7]

$$\begin{aligned} \mathord{\rceil} \stackrel{\text{def}}{=} \mathbf{E}(\boldsymbol{\Pi}^{\varepsilon}(\mathord{\updownarrow})) &= \mathbf{E}[(K * \delta^{\varepsilon} * \xi)(t)(\delta^{\varepsilon} * \xi)(t)] \\ &= \int_{\mathbf{R}} (K * \delta^{\varepsilon})(t-s) \delta^{\varepsilon}(t-s) ds = (K * \bar{\delta}^{\varepsilon})(0) . \end{aligned}$$

where we recall $\delta^{\varepsilon} = \varepsilon^{-1} \varrho(\varepsilon^{-1} \bullet)$; and similarly for $\bar{\delta}^{\varepsilon}$ with $\bar{\varrho} = \varrho(-(\bullet)) * \varrho$. Since $K(x) = x^{H-1/2} 1_{x>0}$ in a neighbourhood of zero, there is no loss of generality in assuming that this includes the support of $\bar{\varrho}$. For $\varepsilon \in (0,1]$, it follows that[8]

[5] One defines $\boldsymbol{\Pi}(\Xi \mathcal{I}(\Xi)^m)$ as the distributional derivative of an Itô integral.

[6] ...and similarly in the canonical one, with $(K * \xi)_t$ replaced by $(K * \xi^{\varepsilon})_t$...

[7] Thanks to stationarity, this quantity is independent of t. In particular, one could immediately take $t = 0$.

[8] In the case of $H = 1/2$, so that $K^H \equiv 1$, noting that $\varrho(\bullet)$, and hence $\bar{\varrho} = \varrho(-(\bullet)) * \varrho$, has unit mass, the constant equals $1/2$, which is the same $1/2$ appearing in the Itô–Stratonovich correction.

$$\mathfrak{d} = (K * \bar{\delta}^\varepsilon)(0) = \int_0^\infty K^H(s)\frac{1}{\varepsilon}\bar{\varrho}\Big(\frac{s}{\varepsilon}\Big)ds = \varepsilon^{H-1/2}\int_0^\infty K^H(s)\; \bar{\varrho}(s)ds\;.$$

We can now replace the informal (14.23) by defining a "renormalised" (admissible) model

$$\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}(\Xi\mathcal{I}(\Xi)) := \boldsymbol{\Pi}^\varepsilon(\Xi\mathcal{I}(\Xi) + c_1^\varepsilon\mathbf{1})\;,$$

with diverging constant

$$c_1^\varepsilon = -\mathbf{E}(\boldsymbol{\Pi}^\varepsilon(\mathfrak{t})) = -\mathfrak{d}\;.$$

In essence, we can leave it to the algebra to handle the correct shifting to different base points (in other words: to recover $(\Pi^{\varepsilon;\mathrm{ren}}, \Gamma^{\varepsilon;\mathrm{ren}})$ from knowledge of $\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}$) in the same spirit as Chen's relation allows to work out increments $\mathbf{X}_{s,t}$ of a given rough path $t \mapsto \mathbf{X}_t$.) On the analytic side, we note that the right-hand side still has controlled blow up of order deg $\Xi\mathcal{I}(\Xi) = (-1/2 + H)^- < 0$. This further suggests that the renormalisation procedure can be described by suitable (linear) maps, say $M : T \to T$, which are (only) allowed to produces additional terms (of higher degrees) as, for instance, $M_{c_1} : \Xi\mathcal{I}(\Xi) \mapsto \Xi\mathcal{I}(\Xi) + c_1\mathbf{1}$ in our present example.

At this stage we could proceed "by hand" and try to work out the correct fixes for all $\Pi_s^\varepsilon(\Xi\mathcal{I}(\Xi)^m),\ m = 1,2,3$, but care is necessary since "curing" level $m = 1$, as done above, will spill over to the higher levels. This is already seen in the instructive case when $m = 0$, i.e. for $\Pi_s(\Xi) = \dot{W}$. Indeed, if one "renormalises" $\dot{W} \implies \dot{W} + c_0$, then writing $V(t) := t$, this leads to[9]

$$\mathbb{W}^m_{s,t} = \int_s^t (\widehat{W}_{s,r})^m\, dW_r \mapsto \int_s^t (\widehat{W}_{s,r} + c_0\widehat{V}_{s,r})^m\,(dW_r + c_0 dV_r)\;.$$

and hence affects all higher levels ($m = 1,2,\ldots$). While $\dot{V} = 1$ naturally has $\mathbf{1}$ as associated symbol, $\widehat{V}$ leads to a new symbol, indifferently written as $\mathcal{I}\mathbf{1} \equiv \mathcal{I}()$ or $|$, in agreement with out earlier convention to represent action of $\mathcal{I}$ as single downfacing line.

$$\Xi(\mathcal{I}\Xi)^m \mapsto (\Xi + c_0\mathbf{1})(\mathcal{I}\Xi + c_0\mathcal{I}\mathbf{1})^m\;.$$

Provided we manage to define all these "fixes" (for $m = 0,1,2,3$) consistently, we can expect a family of linear maps $M = M_c$ indexed by $c = (c_0, c_1, c_2, c_3) \in \mathbf{R}^4$ which furthermore constitutes a group in the sense that of (the matrix identity) $M_c M_{\bar{c}} = M_{c+\bar{c}}$ with $c, \bar{c} \in \mathbf{R}^4$. This is the *renormalisation group*, here isomorphic to $(\mathbf{R}^4, +)$. There was a cheat here, in that our initial collection of symbols (with linear span T) was not rich enough to define M_c as linear map from T into itself. In this sense T was incomplete, and one should work on a space $\tilde{T} \supset T$ which contains required symbols such as $|$ or $\vee$. (The notion of *complete rule* put forward in [BHZ19] formalises this.) However, in the present example this was really a consequence of the (analytically unnecessary!) level-0 renormalisation. In fact, $c_0 = 0$ is the only possible choice that respects the symmetry of the noise, in the sense that $\dot{W}$ and $-\dot{W}$

[9] This is nothing but a variation of the concept of *translation of rough paths*.

have identical law. This reduces the renormalisation group to $(\mathbf{R}^3, +)$ and reflects a general principle: symmetries help to reduce the dimension of the renormalisation group. See [BGHZ19] for an example where this principle takes centre stage in a striking manner.

In general one proceeds as follows. Define T^- as the free commutative algebra generated by all negative symbols in T; that is,

$$T^- := \mathrm{Alg}(\{\circ, \mathsf{I}, \mathsf{V}, \Psi\}) \;. \tag{14.26}$$

(Similarly to before, we colour basis elements of T^- differently to distinguish them from those of T and / or T^+.) Elements in T^- are naturally represented as linear combination of (unordered) *forests*; for instance

$$-\tfrac{1}{2}\mathbf{1} - 3\mathsf{I} + \circ\,\circ + \tfrac{4}{3}\circ\Psi\Psi \in T^- \;,$$

where $\mathbf{1}$ denotes the empty forest. As before, it is useful to introduce a linear map $\Delta^- : T \to T^- \otimes T$ which iterates over all possible ways of extracting possibly empty collections of subtrees of negative degree, putting them as a forest on the left-hand side, and leaving the remaining tree (where all "extracted" subtrees have now been contracted to a point) on the T-valued right-hand side. For instance,

$$\Delta^-(\Psi) = \mathbf{1} \otimes \Psi + \ldots + 3\mathsf{I} \otimes \mathsf{V} + \ldots + 3\circ\,\circ \otimes \mathsf{V} + \ldots + \Psi \otimes \mathbf{1} \;.$$

The resulting renormalisation maps $M : T \to T$ are then parametrised by characters on T^-, similar to the construction of the structure group. Consider for instance the case of a character $g = g^\varepsilon$ defined by $g(\mathsf{I}) = c_1^\varepsilon$, $g(\Psi) = c_3^\varepsilon$, and set to vanish on the remaining two generators $\circ$ and V. Then, the map M_g given by

$$M_g = (g \otimes \mathrm{Id})\Delta^-$$

acts as the identity on all symbols of T other than

$$M_g\mathsf{I} = \mathsf{I} + c_1^\varepsilon\mathbf{1}, \quad M_g\mathsf{V} = \mathsf{V} + 2c_1^\varepsilon\,\mathsf{i}, \quad M_g\Psi = \Psi + 3c_1^\varepsilon\mathsf{V} + c_3^\varepsilon\mathbf{1} \;. \tag{14.27}$$

The resulting renormalised model $\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}} \equiv \boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}$ realises, for instance, the symbol Ψ as

$$\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}\Psi = \boldsymbol{\Pi}^\varepsilon M_g\Psi = \xi^\varepsilon\,(K * \xi^\varepsilon)^3 + 3c_1^\varepsilon(K * \xi^\varepsilon)^2 + c_3^\varepsilon \;.$$

It is a non-trivial but nevertheless fairly general fact that it is possible to choose the character g^ε in such a way that the model $\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}$ converges to a limiting model. This is the case if we choose g^ε as the *BPHZ character* (see [BHZ19, Thm 6.18]) associated to $\boldsymbol{\Pi}^\varepsilon$. This is defined in general as the *unique* character g^ε of T^- such that the renormalised model $\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}$ satisfies $\mathbf{E}\boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}}\tau = 0$ for every symbol τ of strictly negative degree. With our earlier choice

$$c_1^\varepsilon = -\mathbf{E}(\boldsymbol{\Pi}^\varepsilon\mathsf{I})(0) = -\mathsf{D}$$

it is immediate from (14.27) that one has indeed $\mathbf{E}(\boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}\mathbf{I}) = 0$. Furthermore, since first and third moments of centred Gaussians vanish, we also have $\mathbf{E}(\boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}\mathbf{Y}) = \mathbf{E}(\boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}\mathbf{Y}) = 0$ as a consequence of the fact that we set $g(\circ) = g(\mathbf{Y}) = 0$. Finally, it follows from Wick's formula that

$$\begin{aligned}\mathbf{E}\boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}\Psi &= \mathbf{E}[\xi^\varepsilon(K*\xi^\varepsilon)^3] + 3c_1^\varepsilon\mathbf{E}(K*\xi^\varepsilon)^2 + c_3^\varepsilon \\ &= 3\Big(\mathbf{E}[\xi^\varepsilon(K*\xi^\varepsilon)] + c_1^\varepsilon\Big)\mathbf{E}(K*\xi^\varepsilon)^2 + c_3^\varepsilon \\ &= 3(\,\mathbb{I} + c_1^\varepsilon)\,\nabla + c_3^\varepsilon = c_3^\varepsilon\,,\end{aligned}$$

so that $\boldsymbol{\Pi}^\varepsilon M_{g^\varepsilon}\Psi$ has vanishing mean if and only if we also choose $c_3^\varepsilon = 0$.

We have made it plausible that

$$\mathrm{M}^{\varepsilon;\mathrm{ren}} := (\Pi^{\varepsilon;\mathrm{ren}}, \Gamma^{\varepsilon;\mathrm{ren}}) \leftrightarrow \boldsymbol{\Pi}^{\varepsilon;\mathrm{ren}},$$

indeed gives rise to an (admissible) model, with all analytic bounds and algebraic constraints intact, and such that in the sense of model convergence,

$$\mathrm{M}^{\varepsilon;\mathrm{ren}} \to \mathrm{M}^{\mathrm{BPHZ}} = \mathrm{M}^{\mathrm{Itô}}\,. \tag{14.28}$$

The main result of [CH16] is that the convergence $\mathrm{M}^{\varepsilon;\mathrm{ren}} \to \mathrm{M}^{\mathrm{BPHZ}}$ remains true in vastly greater generality and that the limiting model is independent of the specific choice of M^ε for a large class of stationary approximations ξ^ε to the noise ξ.

At last, we leave it to the reader to adapt the material of Section 13.3.2 to define the modelled distribution that allows to reconstruct the Itô integral $\int_0^t f(\widehat{W}_s)dW_s$ and further deduce from (14.28) the following (renormalised) *Wong–Zakai result*,

$$\int_0^t f(\widehat{W}_s^\varepsilon)dW_s^\varepsilon - c_1^\varepsilon\int_0^t f'(\widehat{W}_s^\varepsilon)ds \to \int_0^t f(\widehat{W}_s)dW_s \tag{14.29}$$

where we recall that $c_1^\varepsilon = \varepsilon^{H-1/2}\int_0^\infty K^H(s)\,\bar\varrho(s)ds$. Noting that $\bar\varrho = \varrho(-(\bullet))*\varrho$ is even and has unit mass, we see that $c_1^\varepsilon = \frac{1}{2}$ when $H = 1/2$. We can then pass to the limit for each term on the right-hand side of (14.29) separately. This allows us to recover the identity

$$\int_0^t f(W_s)\circ dW_s - \frac{1}{2}\int_0^t f'(W_s)ds = \int_0^t f(W_s)dW_s\,,$$

in agreement with the usual Itô–Stratonovich correction familiar from stochastic calculus.

14.6 Exercises

Exercise 14.1 *a) Construct an example of a regularity structure with trivial group G, as well as a model and modelled distributions f_i such that both $\mathcal{R}f_1$ and $\mathcal{R}f_2$ are continuous functions but the identity*

$$\mathcal{R}(f_1 \star f_2)(x) = (\mathcal{R}f_1)(x)\,(\mathcal{R}f_2)(x)$$

fails.

b) Transfer Exercise 2.10 to the present context.

Solution. (We only address the first part.) Consider for instance the regularity structure given by $A = (-2\kappa, -\kappa, 0)$ for fixed $\kappa > 0$ 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 $m + n \leq 2$.

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

$$\big(\Pi_x \Xi^0\big)(y) = 1\;, \quad \big(\Pi_x \Xi\big)(y) = 0\;, \quad \big(\Pi_x \Xi^2\big)(y) = c\;, \tag{14.30}$$

where c is an arbitrary constant. Let furthermore

$$f_1(x) = f_1(x)\Xi^0 + \tilde f_1(x)\Xi\;, \qquad f_2(x) = f_2(x)\Xi^0 + \tilde f_2(x)\Xi\;.$$

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

However, the pointwise product is given by

$$\big(f_1 \star f_2\big)(x) = f_1(x)f_2(x)\Xi^0 + \big(\tilde f_1(x)f_2(x) + \tilde f_2(x)f_1(x)\big)\Xi + \tilde f_1(x)\tilde f_2(x)\Xi^2\;,$$

which by Theorem 14.5 belongs to $\mathscr{D}^{\gamma-\kappa}$. Provided that $\gamma > \kappa$, one can then apply the reconstruction operator to this product and one obtains

$$\mathcal{R}\big(f_1 \star f_2\big)(x) = f_1(x)f_2(x) + c\tilde f_1(x)\tilde f_2(x)\;,$$

which is obviously quite different from the pointwise product $(\mathcal{R}f_1)(x)\cdot(\mathcal{R}f_2)(x)$.

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

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

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

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

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 (14.30). 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-canonical" models as in (14.30) can be interpreted as limits of "canonical" models for which the usual rules of calculus hold. Even this is however not always the case (think of the Itô Brownian rough path).

Exercise 14.2 *Consider $\mathcal{Z}^\alpha = \mathcal{Z}^\alpha(\mathbf{R}^d)$.*

a) Show that distributional derivatives satisfy $D^k \mathcal{Z}^\alpha \subset \mathcal{Z}^{\alpha-|k|}$ for any multi-index k. Show that for $d = 1$ equality holds. That is, any $g \in \mathcal{Z}^{\alpha-k}$, with $k \in \mathbf{N}$, is the kth distributional derivative of some $f \in \mathcal{Z}^\alpha$.

b) The proof of Schauder's theorem in Section 14.3 was more involved in the "positive" case, when $0 \le \alpha + \beta \in [n-1, n)$, some $n \in \mathbf{N}$. Give an easier proof in the case $d = 1$ by reducing the positive to the negative case.

$\ast\ast$ **Exercise 14.3** *Provide a proof of the case $\alpha = 0$ in Lemma 14.13.*

Solution. As in Lemma 14.13, we aim to bound $|\zeta(\varphi)|$ for $\varphi \in \mathcal{B}^\lambda_{r_o,x}$ and $\zeta \in \mathcal{Z}^\alpha_r$ for some $r \ge r_o$. One strategy is to consider a compactly supported wavelet basis of regularity r and to separately bound the terms in the wavelet expansion of φ.

If we wish to rely purely on elementary arguments, one strategy goes as follows.

a) Show first that $\zeta \in \mathcal{Z}^\alpha_r$ if and only if $\zeta\chi \in \mathcal{Z}^\alpha_r$ for every smooth compactly supported function χ. This allows us to reduce ourselves to the case when ζ itself is compactly supported and we assume this from now on.

b) Show that if $\zeta \in \mathcal{Z}^0_r$ is supported in a ball of radius 1 and if ψ is such that $\int \psi(x)\,dx = 0$ and such that $|D^k\psi(x)| \le (1+|x|)^{-\beta-|k|}$ for $|k| \le r$ and some large enough exponent k, then $|\zeta(\psi^\lambda_x)| \lesssim 1$, uniformly over such ψ and over $x \in \mathbf{R}^d$ and $\lambda \in (0,1]$.

c) Choose a function ψ with the property that its Fourier transform is smooth, identically 1 in the ball of radius 1, and identically 0 outside of the ball of radius 2 and define $\tilde\psi$ as in the proof of Lemma 14.13. Write

$$\varphi = \varphi * \psi^\lambda + \sum_{n \ge 0} \varphi * \tilde\psi^{\lambda_n}$$

as in the proof of Lemma 14.13.

d) Choose χ such that its Fourier transform is smooth, identically equal to 1 on the annulus of radii in $[1,4]$ and vanishes outside the annulus of radii in $[1/2,5]$. Note that this implies that $\tilde\psi^{\lambda_n} = \tilde\psi^{\lambda_n} * \chi^{\lambda_n}$ and conclude that

$$\zeta(\varphi * \tilde\psi^{\lambda_n}) = \langle \zeta * \tilde\psi^{\lambda_n}, \varphi * \chi^{\lambda_n} \rangle \;.$$

e) Use the fact that $\varphi \in \mathcal{C}^1$ and χ integrates to 0 to conclude that $|\varphi * \chi^{\lambda_n}| \lesssim 2^{-n}\lambda^{-d}$ and therefore that $|\zeta(\varphi * \tilde{\psi}^{\lambda_n})| \lesssim 2^{-n}$, which is summable as required.

$*$ **Exercise 14.4** *Show that, for g smooth enough, one has $K * (g\eta) - g(K * \eta) \in \mathcal{C}^{\alpha+\beta+1}$ for every β-regularising kernel K and $\eta \in \mathcal{C}^\alpha$ with $\alpha < 0$. How smooth is smooth enough? Compare the following two strategies.*

Strategy 1: *Go through the proof of the Schauder estimate in Section 14.3 and estimate the difference $\langle K_n * (g\eta) - g(K_n * \eta), \psi_\lambda \rangle$.*

Strategy 2: *Consider the regularity structure T spanned by the Taylor polynomials and an additional symbol Ξ of degree α, with the structure group acting trivially on Ξ. We extend this by adding an integration operator of order β and all products with Taylor polynomials. We also consider on it the natural model mapping Ξ to η. Writing $g \in \mathscr{D}^\gamma$ for the Taylor lift of g as in Proposition 13.16, verify that $g\Xi \in \mathscr{D}^{\gamma+\alpha}$. The multilevel Schauder estimate then shows that, provided that $\gamma + \alpha > 0$, one has $\mathcal{K}(g\Xi) \in \mathscr{D}^{\gamma+\alpha+\beta}$ and $g\mathcal{K}(\Xi) \in \mathscr{D}^{\gamma+\min\{0,\alpha+\beta\}}$, so in particular*

$$F \stackrel{\text{def}}{=} \mathcal{K}(g\Xi) - g\mathcal{K}(\Xi) \in \mathscr{D}^{1+\alpha+\beta} ,$$

provided that $\gamma > \max\{1, -\alpha, 1 + \alpha + \beta\}$. Furthermore, the explicit expression for $\mathcal{K}$ shows that

$$\mathcal{K}(g\Xi) = g\mathcal{I}(\Xi) + g'\mathcal{I}(X\Xi) + (\dots) , \qquad g\mathcal{K}(\Xi) = g\mathcal{I}(\Xi) + (\dots) ,$$

where $(\dots)$ denotes terms that either belong to the polynomial part of the regularity structure or are of degree strictly greater than $\alpha + \beta + 1$ (which is the degree of $\mathcal{I}(X\Xi)$). In particular, the truncation of F at level $\alpha + \beta + 1$ belongs to $\mathscr{D}_{\mathrm{P}}^{\alpha+\beta+1}$, and we conclude by the second part of Proposition 13.16.

Exercise 14.5 *Consider space-time $\mathbf{R}^d$ with one temporal and $(d-1)$ spatial dimensions, under the parabolic scaling $(2, 1, \dots, 1)$, as introduced in Remark 13.9. Denote by $\mathcal{G}$ the heat kernel (i.e. the Green's function of the operator $\partial_t - \partial_x^2$). Show that one has the decomposition*

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

where the kernel K satisfies all of the assumptions of Section 14.4 (with $\beta = 2$) and the remainder $\hat{K}$ is smooth and bounded.

Exercise 14.6 (From [Bru18]) *In the context of Remark 14.25, establish the recursion*

$$\Gamma_{xy}\mathcal{I}\tau = \mathcal{I}(\Gamma_{xy}\tau) - \Gamma_{xy}\mathcal{J}_{xy}\tau , \qquad (14.31)$$

with

$$\mathcal{J}_{xy}\tau := \sum_{|k| < \deg \tau + \beta} \frac{X^k}{k!} \Pi_x(\mathcal{I}_k(\Gamma_{xy}\tau))(y) .$$

Exercise 14.7 *Show that if one defines $\Gamma_{xy}\mathcal{I}\tau$ in such a way that (14.10) holds, then it guarantees that $\Pi_x \Gamma_{xy}\mathcal{I}\tau = \Pi_y \mathcal{I}\tau$.*

Exercise 14.8 *Adapt the material in Section 14.5 and construct a suitable regularity structure and model so that the two-dimensional Itô integral (14.13) is obtained as reconstruction of a suitable modelled distribution.*

14.7 Comments

The material on differentiation, products and admissible models follows essentially [Hai14b], although the conditions on the kernel K – previously assumed to annihilate certain polynomials – are now more flexible. In particular, we do not enforce the identity $\mathcal{I}(X^k) = 0$ and instead allow for the possibility of simply including symbols $\mathcal{I}(X^k)$ as basis vectors of our regularity structure. It is the case that any admissible model will necessarily satisfy $\Pi_x \mathcal{I}(X^k) = 0$, but in general $\Gamma_{xy}\mathcal{I}(X^k) \neq 0$. The material of Section 14.5 is essentially taken from [BFG$^+$19], with a viewpoint similar to [BCFP19].