©2015 The Mathematical Society of Japan
J. Math. Soc. Japan
Vol. 67, No. 4 (2015) pp. 1551–1604
doi: 10.2969/jmsj/06741551

A Wong-Zakai theorem for stochastic PDEs

By Martin Hairer and Étienne Pardoux

Dedicated to the memory of Kiyosi Itô on the occasion of the 100th anniversary of his birth

(Received Sep. 11, 2014)
(Revised Oct. 28, 2014)

Abstract. We prove a version of the Wong-Zakai theorem for one-dimensional parabolic nonlinear stochastic PDEs driven by space-time white noise. As a corollary, we obtain a detailed local description of solutions.

1. Introduction.

A series of classical results pioneered by Wong and Zakai [**WZ65a**], [**WZ65b**] for one-dimensional SDEs and by Stroock and Varadhan [**SV72**] in the multidimensional case, states that if B_ε denotes some "natural" smooth ε-approximation to a d-dimensional Brownian motion B (for example piecewise linear approximation or convolution with a mollifier), g and h are smooth functions, and x_ε denotes the solution to the ODE

$$\dot{x}_\varepsilon = h(x_\varepsilon) + g(x_\varepsilon)\dot{B}_\varepsilon, \tag{1.1}$$

then x_ε converges in probability, as $\varepsilon \to 0$, to the solution to the SDE

$$dx = h(x)\,dt + g(x) \circ dB,$$

where $\circ dB$ denotes Stratonovich integration against B.

It has been an open problem for some time to obtain an analogous result in the case of one dimensional stochastic PDEs driven by space-time white noise of the type

$$du = \partial_x^2 u\,dt + H(u)\,dt + G(u)\,dW(t). \tag{1.2}$$

The problem is that there is no Stratonovich formulation for such an equation since the Itô-Stratonovich correction term would be infinite. It is formally given by $(1/2)G'(u)G(u)\,\mathrm{tr}\,Q$ where tQ is the covariance operator of $W(t)$. In the case of space-time white noise, Q is the identity operator on L^2, which is of course not trace class. On the other hand, we know that if one subtracts a suitable correction term from the random ODE (1.1), then it is possible to ensure that solutions converge to the Itô solution. More precisely, if one considers the sequence of equations given by

2010 *Mathematics Subject Classification.* Primary 60H15; Secondary 60H05.
Key Words and Phrases. stochastic PDEs, approximation, pathwise solutions.

$$\dot{x}_\varepsilon = h(x_\varepsilon) - \frac{1}{2} Dg(x_\varepsilon)\, g(x_\varepsilon) + g(x_\varepsilon)\dot{B}_\varepsilon, \tag{1.3}$$

then x_ε converges, as $\varepsilon \to 0$, to the Itô solution to $dx = h(x)\,dt + g(x)\,dB$. Since the Itô solution is the only "natural" notion of solution available for (1.2), this suggests that if one considers approximations of the type

$$\partial_t u_\varepsilon = \partial_x^2 u_\varepsilon + H(u_\varepsilon) - C_\varepsilon G'(u_\varepsilon) G(u_\varepsilon) + G(u_\varepsilon)\,\xi_\varepsilon, \tag{1.4}$$

where ξ_ε denotes an ε-approximation to space-time white noise and C_ε is a suitable constant which diverges as $\varepsilon \to 0$, then one might expect u_ε to converge to the solution to (1.2), interpreted in the Itô sense.

The main result of this article demonstrates that this is *almost* the case, at least for a large class of Gaussian approximations to space-time white noise, and on bounded domains. While it is true that the solutions u_ε to (1.4) converge to a limit if C_ε is suitably chosen, it is *not* true in general that this limit is given by the Itô solution to (1.2). Instead, the limit solves the same equation, but with a modified drift term H. Alternatively, this can be formulated as stating that, in order to obtain the limiting equation (1.2), the "correct" approximations are of the form (1.4), but with H replaced by some different nonlinearity $\bar{H}$. It furthermore turns out that the constants describing $\bar{H}$ are not universal, but depend on the details of the regularisation ξ_ε.

In order to formulate our results more precisely, we fix an even, smooth, compactly supported function $\varrho\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ with $\int \varrho = 1$ and $\varrho(t,x) = \varrho(t,-x)$, and we set

$$\varrho_\varepsilon(t,x) = \varepsilon^{-3}\varrho(\varepsilon^{-2}t, \varepsilon^{-1}x), \qquad c_\varrho = \int P(z)(\varrho \star \varrho)(z)\,dz, \tag{1.5}$$

where P denotes the heat kernel on $\boldsymbol{R}$ and $\star$ denotes space-time convolution. Here and everywhere in the paper, we use the *parabolic scaling*, which means that t is rescaled as if it were a 2–dimensional variable. The distance associated to this scaling is defined by $|(t,x)| = \sqrt{|t|} + |x|$. We also define our regularised noise ξ_ε by

$$\xi_\varepsilon(t,x) = \int_{-\infty}^{\infty} \big\langle \varrho_\varepsilon(t-s, x-\cdot), dW(s) \big\rangle, \tag{1.6}$$

where W is the cylindrical Wiener process driving (1.2). Finally, we will assume from now on that the spatial variable x takes values in the one-dimensional torus S^1, so that $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L^2(S^1)$. With these notations at hand, the main result of this article is as follows.

THEOREM 1.1. *Let H and G be of classes $\mathcal{C}^2$ and $\mathcal{C}^5$ respectively, both with bounded first derivatives, let u denote the solution to* (1.2)*, and let u_ε denote the classical solution to the random PDE* (1.4) *with $C_\varepsilon = \varepsilon^{-1} c_\varrho$ and H replaced by*

$$\bar{H}(u) = H(u) - c_\varrho^{(1)} G'(u)^3 G(u) - c_\varrho^{(2)} G''(u) G'(u) G^2(u), \tag{1.7}$$

for some constants $c_\varrho^{(i)}$ possibly depending on ϱ but not on ε. Both solutions are started with the same initial condition $u_\varepsilon(0,\cdot) = u(0,\cdot) \in \mathcal{C}(S^1)$.

Then, there exists a choice of $c_\varrho^{(i)}$ such that, for any $T > 0$, one has

$$\lim_{\varepsilon\to 0} \sup_{(t,x)\in[0,T]\times S^1} |u(t,x) - u_\varepsilon(t,x)| = 0,$$

in probability. Moreover, for any $\alpha \in (0,1/2)$ and $t > 0$, the restriction of u_ε to $[t,T]\times S^1$ converges to u in probability for the topology of $\mathcal{C}^{\alpha/2,\alpha}$.

Before we proceed, let us remark on several possible straightforward (and not so straightforward) extensions to Theorem 1.1, as well as a few reality-checks.

REMARK 1.2. There are explicit formulae for the constants $c_\varrho^{(i)}$, see Section 2 below. One important remark is that if we set $\varrho(t,x) = \delta^{-1}\tilde\varrho(\delta^{-1}t,x)$ for some fixed $\tilde\varrho$ and let $\delta \to 0$, then we have $c_\varrho^{(i)} \to 0$. This is not too surprising since it intuitively corresponds to first removing the temporal regularisation at fixed spatial regularisation (there, traditional generalisations of the Wong-Zakai theorem apply and only an Itô-Stratonovich correction term appears) and then only removing the spatial regularisation, under which we know that (1.2) is stable [**DPZ92**].

REMARK 1.3. As will be clear, our proof automatically yields a situation analogous to that arising when building a solution theory to finite-dimensional SDEs using the theory of rough paths [**LCL07**] (see also [**GT10**] for an earlier attempt in this direction in the context of evolution equations): one obtains a natural notion of solution to (1.2) which is pathwise, i.e. defined for some set of measure 1 independent of the initial condition and even of the choice of nonlinearities G and H. Note however that even if G and H are bounded with bounded derivatives of all orders, our construction does *not* yield the existence of a stochastic flow, but only of local solutions. The reason in a nutshell is that in order to obtain our convergence result we reformulate (1.4) as a fixed point problem in a space of functions of high "regularity" (where regularity is to be understood in an unconventional sense though). In this space, just like in the classical spaces $\mathcal{C}^\alpha$ for $\alpha > 1$, the map $u \mapsto G(u)$ is not globally Lipschitz continuous, even if all derivatives of G are bounded. As a consequence, we cannot rule out that exceptional realisations of the driving noise and exceptional initial conditions lead to a finite-time blow-up.

REMARK 1.4. Using the same methodology, one can also treat systems of equations of the type

$$du_i = \partial_x^2 u_i\,dt + H_i(u)\,dt + G_{ij}(u)\,dW_j(t),$$

with summation over repeated indices implied and W_j a finite collection of i.i.d. cylindrical Wiener processes. The only difficulty that arises is notational, so we will stick to the case of one single equation.

In the multi-dimensional case, it is however not obvious *a priori* how indices are paired in the correction terms. The "Itô-Stratonovich" correction multiplying C_ε has the

same pairing of indices as in the case of classical SDEs, namely $\partial_k G_{ij}(u)\, G_{kj}(u)$, with summation over k and j implied. Keeping track of the indices in (4.10) below shows that the additional order 1 correction term is given by

$$\begin{aligned}\bar{H}_i(u) = H_i(u) &- c_\varrho^{(1,1)}\, \partial_k G_{ij}(u)\, \partial_\ell G_{km}(u)\, \partial_n G_{\ell j}(u)\, G_{nm}(u) \\ &- c_\varrho^{(1,2)}\, \partial_k G_{ij}(u)\, \partial_\ell G_{km}(u)\, \partial_n G_{\ell m}(u)\, G_{nj}(u) \\ &- c_\varrho^{(2,1)}\, \partial^2_{k\ell} G_{ij}(u)\, \partial_m G_{kn}(u)\, G_{mj}(u)\, G_{\ell n}(u) \\ &- c_\varrho^{(2,2)}\, \partial^2_{k\ell} G_{ij}(u)\, \partial_m G_{kj}(u)\, G_{mn}(u)\, G_{\ell n}(u), \end{aligned} \tag{1.8}$$

with implied summation over repeated indices. The constants $c_\varrho^{(i,j)}$ appearing in this expression will be given in Section 2 and are such that $c_\varrho^{(1)} = c_\varrho^{(1,1)} + c_\varrho^{(1,2)}$ and similarly for $c_\varrho^{(2)}$.

REMARK 1.5. The appearance of a correction term involving second derivatives of G explains why we need G to be at least of class $\mathcal{C}^3$. The reason why we actually need it to be of class $\mathcal{C}^5$ is twofold. First, at this level of "irregularity", one could in principle also have the appearance of correction terms involving the third derivative of G. (See in particular the right hand side of (4.10) below.) In our case however, the corresponding term vanishes thanks to the specific properties of Gaussian random variables. Second, in order to guarantee that the expression (4.10) below belongs to $\mathcal{D}^{\gamma,\eta}$ for some $\gamma > 0$, we need some control on the next order terms, so that some additional regularity beyond $\mathcal{C}^4$ is required.

REMARK 1.6. If, in the first expression of (2.1), we assume that the integration variable z is real-valued and that P is the Heaviside function (which is nothing but the Green's function for the differential operator ∂_t on $\boldsymbol{R}$), then we obtain $c_\varrho = 1/2$, independently of ϱ, as a consequence of the fact that $\varrho^{(2)}$ is symmetric and integrates to 1. This shows that the dominant correction term is indeed compatible with its interpretation as an Itô-Stratonovich correction.

REMARK 1.7. In law, the noise $\xi_\varepsilon(t,x)$ is the same as $\varepsilon^{-3/2}\xi_1(\varepsilon^{-2}t, \varepsilon^{-1}x)$. It is then natural to ask whether a similar Wong-Zakai theorem still holds if one sets $\xi_\varepsilon(t,x) = \varepsilon^{-3/2}\xi_1(\varepsilon^{-2}t, \varepsilon^{-1}x)$, but with ξ_1 an arbitrary (non-Gaussian) space-time stationary process with sufficiently rapid mixing properties. It is at this stage still unclear whether (1.4) is the correct approximation in this case, as there may be additional correction terms needed. In particular, it seems plausible that correction terms of order $\varepsilon^{-1/2}$ appear in general.

REMARK 1.8. If one sets $\xi_\varepsilon(t,x) = \varepsilon^{-1}\xi_1(\varepsilon^{-2}t, \varepsilon^{-1}x)$, then the sequence of solutions to (1.4) with $C_\varepsilon = 0$ converges to a deterministic limit, given by the solution to the corrected PDE

$$\partial_t u = \partial_x^2 u + H(u) + c_\varrho G'(u) G(u).$$

See [**PP12**] for a treatment of the linear case.

REMARK 1.9. The results presented in this article are somewhat orthogonal to the results of [**HW13**] where the authors considered a more general class of (also possibly vector-valued) equations of the type

$$du = \partial_x^2 u\,dt + H_1(u)\,dt + H_2(u)\partial_x u\,dt + G(u)\,dW,$$

for W an L^2-cylindrical Wiener process. Indeed, the problems tackled in [**HW13**] include a larger class of equations and allows to consider approximations where the linear operator ∂_x^2 is replaced by a quite general ε-approximation to the Laplacian. However, the allowable approximations to W included only *spatial* regularisations (with the integral already treated as an Itô integral for $\varepsilon > 0$), while the whole point of the current article is to allow for *space-time* regularisations. In particular, no additional correction term appears in [**HW13**].

Theorem 1.1 is a consequence of a much stronger convergence result given in Theorem 4.7 below. The statement of this stronger result does however first require the introduction of the algebraic/analytical machinery of regularity structures developed in [**Hai14**], so we refrain from stating it here. As a consequence of this stronger result, one immediately obtains a number of results that have traditionally been considered out of the scope of current techniques of proof. For example, it immediately follows that solution to (1.2) are locally continuous with respect to their initial condition in a pathwise sense, rather than just in probability.

COROLLARY 1.10. *Let H be $\mathcal{C}^2$ and G be $\mathcal{C}^5$, both with bounded first derivatives. Then, for every $u_0 \in \mathcal{C}(S^1)$, every $T > 0$, and every $\delta > 0$, there exists a neighbourhood $\mathcal{U}$ of u_0, an event $\Omega_0 \subset \Omega$ with $\boldsymbol{P}(\Omega_0) > 1 - \delta$ and a map $\Phi\colon \mathcal{U} \times \Omega \to \mathcal{C}([0,T] \times S^1)$ such that $\Phi(\cdot,\omega)$ is continuous for every $\omega \in \Omega_0$ and such that $\Phi(\bar{u}_0,\cdot)$ solves (1.2) with initial condition $\bar{u}_0$ for every $\bar{u}_0 \in \mathcal{U}$.*

One also obtains the following very sharp regularity result (see for example the recent article [**KSXZ13**] for a similar but weaker result):

COROLLARY 1.11. *Let H and G be of class $\mathcal{C}^2$ and $\mathcal{C}^5$ respectively, both with bounded first derivatives, and let u denote the solution to (1.2). Let furthermore v denote the solution to the linear equation*

$$dv = \partial_x^2 v\,dt + dW, \qquad v(0,x) = 0.$$

Then, for every random space-time point $(t,x) \in \boldsymbol{R}_+ \times S^1$ and every $\kappa > 0$, there exists a random constant C such that the bound

$$\big|u(s,y) - u(t,x) - G(u(t,x))\big(v(s,y) - v(t,x)\big)\big| \le C(\sqrt{|t-s|} + |x-y|)^{1-\kappa},$$

holds uniformly over $y \in S^1$ and $|t-s| < 1 \wedge |t|/2$.

REMARK 1.12. Note here that the time t does not need to be a stopping time! This is a consequence of the pathwise nature of our analysis.

As a final application of our results, we can show that if the solutions to two stochastic PDEs driven by the same realisation of the noise and with the same diffusion coefficient G cross, then they are tangent to even higher order at their intersection.

COROLLARY 1.13. *Let H, $\bar{H}$ and G be $\mathcal{C}^2$, $\mathcal{C}^2$ and $\mathcal{C}^5$ respectively with bounded first derivatives and let u and $\bar{u}$ denote the solutions to* (1.2)*, as well as the same equation with H replaced by $\bar{H}$. Let $(t,x) \in \boldsymbol{R}_+ \times S^1$ be a random space-time point such that $u(t,x) = \bar{u}(t,x)$ and let $\kappa > 0$.*

Then, there exist random constants C and D such that

$$\bigl|u(s,y) - \bar{u}(s,y) - D(y-x)\bigr| \le C(|t-s| + |x-y|^2)^{(3/4)-\kappa},$$

holds uniformly for (s,y) in a compact subset of $\boldsymbol{R}_+ \times S^1$. (With C depending on the set.) If one has $H = \bar{H}$ and one knows furthermore that $u(s,y) \ge \bar{u}(s,y)$ in a region of the type $\{(s,y) : |x-y| \le \delta \;\&\; s \in (t-\delta, t]\}$ for some $\delta > 0$, then this bound improves to

$$\bigl|u(s,y) - \bar{u}(s,y) - D_1(t-s) - D_2(y-x)^2\bigr| \le C(|t-s| + |x-y|^2)^{(5/4)-\kappa},$$

and furthermore $D_1, D_2 \ge 0$.

REMARK 1.14. The assumption that H and G have bounded first derivatives ensures the existence of global solutions to (1.2), which makes the statement easier to formulate. It is straightforward to localise the statements to cover situations with possible blow-ups.

REMARK 1.15. The main result of this paper, in the particular case $H \equiv 0$ and $G(u) = u$, is equivalent to the convergence to its Hopf–Cole solution of the regularised KPZ equation. Indeed, consider equation (1.4) with $H(u) = -c_\varrho^{(1)}\, u$ and $G(u) = u$, and with an initial condition which satisfies $u_\varepsilon(0,x) > 0$, for all $x \in S^1$, which implies that the same is true for $u_\varepsilon(t,x)$ for all $t > 0$, $x \in S^1$. Let now $Z_\varepsilon(t,x) = \log[u_\varepsilon(t,x)]$. Dividing (1.4) by u_ϵ, we obtain that

$$\partial_t Z_\varepsilon = \partial_x^2 Z_\varepsilon + (\partial_x Z_\varepsilon)^2 - c_\varrho^{(1)} - C_\varepsilon + \xi_\varepsilon.$$

A consequence of Theorem 1.1 is that $Z_\epsilon(t,x) \to Z(t,x) := \log u(t,x)$, where u solves the SPDE

$$du = \partial_x^2 u\, dt + u\, dW(t)$$

in the Itô sense, i.e. Z is the Hopf–Cole solution of the KPZ equation. This result was implicit in [**Hai13**], [**Hai14**] but not formulated there in detail. It is also used in [**HQ15**] to identify the limit for a class of weakly asymmetric interface fluctuation models as being the Hopf–Cole solution of the KPZ equation.

ACKNOWLEDGEMENTS. MH gratefully acknowledges financial support from the Philip Leverhulme Trust, the Royal Society, and the ERC.

2. Values of constants.

In this section, we give explicit expressions for $c_\varrho^{(1)}$ and $c_\varrho^{(2)}$ and we discuss the limiting behaviours of these constants under scalings of the type $\varrho = \delta^{-1}\tilde{\varrho}(\delta^{-1}t, x)$ and $\varrho = \delta^{-1}\tilde{\varrho}(t, \delta^{-1}x)$. This allows us to also consider different scaling behaviours for the noise.

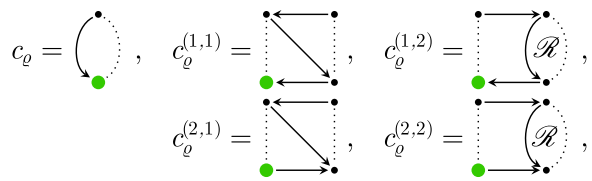
Denoting by P the heat kernel (with the convention that $P(t,x) = 0$ for $t \le 0$), by δ the Dirac at the origin, and using the shorthand $\varrho^{(2)} = \varrho \star \varrho$, the constants $c_\varrho^{(i)}$ are given by the expression

$$c_\varrho^{(i)} = c_\varrho^{(i,1)} + c_\varrho^{(i,2)},$$

with

$$\begin{aligned}
c_\varrho^{(1,1)} &= \int P(z_1)P(z_2)P(z_3)\,\varrho^{(2)}(z_1+z_2)\varrho^{(2)}(z_2+z_3)\prod_{i=1}^3 dz_i,\\
c_\varrho^{(1,2)} &= \int P(z_1)P(z_2)\big(P(z_3)\,\varrho^{(2)}(z_3) - c_\varrho\delta(z_3)\big)\varrho^{(2)}(z_1+z_2+z_3)\prod_{i=1}^3 dz_i,\\
c_\varrho^{(2,1)} &= \int P(z_1)P(z_2)P(z_3)\,\varrho^{(2)}(z_1+z_2)\varrho^{(2)}(z_2-z_3)\prod_{i=1}^3 dz_i,\\
c_\varrho^{(2,2)} &= \int P(z_1)P(z_2)\big(P(z_3)\,\varrho^{(2)}(z_3) - c_\varrho\delta(z_3)\big)\varrho^{(2)}(z_1-z_2+z_3)\prod_{i=1}^3 dz_i.
\end{aligned} \tag{2.1}$$

Here, the integration variables z_i are space-time variables and are integrated over all of $\boldsymbol{R}^2$. More concisely, if we draw a node • for each integration variable, a special node ● for the origin, an arrow for the heat kernel (evaluated at the difference between the two variables that it connects), and a dotted line for $\varrho^{(2)}$ (also evaluated at the difference between the two variables that it connects), then (2.1) can be rewritten much more concisely and suggestively as



(2.2)

where the symbol $\mathscr{R}$ inside a loop indicates that it was "renormalised" by subtracting a delta-function with weight identical to the integral of the kernel represented by the loop. In this case, the kernel in question is $P\varrho^{(2)}$ and its integral is precisely c_ϱ. This kind of

notation will be used extensively in Section 5 below, so we urge the reader to familiarise themselves with the link between (2.2) and (2.1). Note also that in this particular case, the location on the graph of the distinguished variable ● representing the origin does not matter.

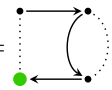
The fact that the integrals depicted by (2.2) converge (and the precise meaning of the ones involving "renormalisation") is not completely obvious, so we will discuss this now.

2.1. Convergence of the renormalisation constants.

The convergence of the integral defining c_ϱ is very easy to verify, so we focus on the remaining four terms. There are three issues with the definitions (2.2): integrability of the expressions at short scales, integrability at large scales, and the meaning of the "renormalised" integrals. Regarding integrability at small scales, it is very easy to convince oneself that this is not a problem, thanks to the relatively mild singularity of the heat kernel and the boundedness of the function $\varrho^{(2)}$.

Regarding the integrability at large scales, we first consider the terms $c_\varrho^{(i,1)}$. For these terms, we see that thanks to the fact that $P(t,x) = 0$ for $t < 0$ and $\varrho^{(2)}$ is compactly supported, the time variables take values in a compact domain. Pictorially, if two nodes are connected by $\varrho^{(2)}$ like this •⋯⋯•, then the corresponding time variables can be separated by at most a fixed finite distance. Furthermore, for any configuration of the type •⟶•⟶• (here, the direction of the arrows does matter), the time coordinate of the middle variables has to lie between the time coordinates of the other two. Since, for any fixed time, the heat kernel decays exponentially fast in the space variable, this shows that the expressions for the constants $c_\varrho^{(i,1)}$ are well-defined.

Regarding $c_\varrho^{(1,2)}$, we can simply rewrite it as

$$c_\varrho^{(1,2)} =$$  $$- \; c_\varrho$$  .

Using the same considerations as those used to bound $c_\varrho^{(i,1)}$, we conclude that each of these two terms converges separately.

The constant $c_\varrho^{(2,2)}$ is a little bit more delicate to bound. Indeed, if we decompose it as we did for $c_\varrho^{(1,2)}$, then we see that each of these terms taken individually diverges. This can be easily seen for the second term, since this is given by $\int \varrho^{(2)}(z_1 - z_2)P(z_1)P(z_2)\,dz_1\,dz_2$ and $|P(z)|^2$ decays like $|z|^{-2}$ which is far from being integrable at large scales. (Recall that we endow $\boldsymbol{R}^2$ with the parabolic distance which, for $z = (t,x)$, is given by $|z| = |x| + \sqrt{|t|}$. When endowed with the parabolic scaling, the scaling dimension of $\boldsymbol{R}^2$ is equal to 3, so we are one whole power short of being integrable, and not just borderline.) Denote now by Q the distribution formally given by $Q(z) = P(z)\varrho^{(2)}(z) - c_\varrho\,\delta_0(z)$, then $c_\varrho^{(2,2)}$ should be interpreted as

$$c_\varrho^{(2,2)} = \int \varrho^{(2)}(z_1 - z_2)(P \star Q)(z_1)\,P(z_2)\,dz_1\,dz_2.$$

A finite bound on this expression then follows immediately from the following very simple result.

LEMMA 2.1. *There exists a constant C such that the function $P \star Q$ is bounded by $C\big(|z|^{-1} \wedge |z|^{-3}\big)$, uniformly in z.*

PROOF. We use the the fact that Q has compact support (since ϱ, hence also $\varrho^{(2)}$, has this property), integrates to 0 and is even in the spatial variable. As a consequence, it annihilates every function of the form $(t,x) \mapsto a + bx$ with $a, b \in \boldsymbol{R}$. The claim then follows immediately from [**Hai14**, Proposition A.1] (applying it with $A = \{(0,0),(0,1)\}$), combined with the scaling properties of P. □

2.2. Non-parabolic scalings: spatial regularisation.

The case of purely spatial regularisation formally corresponds to $\varrho(t,x) = \delta_0(t)\bar{\varrho}(x)$ (with δ_0 the Dirac distribution centred at the origin). In general, one could consider a regularisation of the noise of the type $\xi_\varepsilon = \varepsilon^{-(1+\beta)/2}\xi_1(\varepsilon^{-\beta}t, \varepsilon^{-1}x)$ for some $\beta > 2$, which corresponds to the case where we consider approximations that are "more regular" in space than in time. We can recast this in the previous framework, but now ϱ (or rather $\varrho^{(2)}$ since this is all that ever matters for the calculation of our constants) is replaced by

$$\varrho_\delta^{(2)}(t,x) = \delta^{-1}\varrho^{(2)}(\delta^{-1}t, x), \qquad \delta = \varepsilon^{\beta-2}. \tag{2.3}$$

With such a choice, we then obtain

$$c_\varrho = c_\varrho(\delta) = \int P(z)\,\varrho_\delta^{(2)}(z)\,dz \to c_\varrho(0) = \int_0^\infty \varrho^{(2)}(t,0)\,dt,$$

which is consistent with the results obtained in [**PP12**], [**HPP13**].

We claim that the constants $c_\varrho^{(i,j)}$ converge to 0 as $\delta \to 0$ if the regularisation $\varrho^{(2)}$ is of the form (2.3). Regarding the constants $c_\varrho^{(i,2)}$, it is straightforward to see that, weakly, one has $P\varrho_\delta^{(2)} \to c_\varrho(0)\,\delta_0$, so that these two constants do indeed converge to 0 as $\delta \to 0$. Regarding the constants $c_\varrho^{(i,1)}$, we obtain an upper bound on their absolute values by replacing $\varrho_\delta^{(2)}(t,x)$ by the function $\delta^{-1}\mathbf{1}_{|t|<\delta}(t,x)$. For any fixed values of the three time variables, the integration over the spatial variables is then bounded by δ^{-2} as a simple consequence of the fact that $\int P(t,x)\,dx = 1$ for every t. (The factor δ^{-2} comes from the two factors $\varrho_\delta^{(2)}$.) Since the time variables are bounded by some multiple of δ and there are three time integrals, we conclude that $|c_\varrho^{(i,1)})| \le C\delta$ for some C, so that they also converge to 0. This is also consistent with [**BMS95**, Theorem 1.11], where the authors obtain a convergence result similar in spirit to our Theorem 1.1, but with ξ_ε given by a regularisation of space-time white noise at scale ε in space and scale $\exp(-c/\varepsilon)$ (which is of course *much* smaller than ε^2) in time.

2.3. Non-parabolic scalings: temporal regularisation.

The case of temporal regularisation is much less obvious, and this was already the "difficult case" in [**PP12**], [**HPP13**]. This time, we consider a regularisation of the noise

of the type $\xi_\varepsilon = \varepsilon^{-1+(\beta/2)}\xi_1(\varepsilon^{-2}t, \varepsilon^{-\beta}x)$ for some $\beta > 1$. Again, we can recast this into our framework, with $\varrho^{(2)}$ replaced by

$$\varrho^{(2)}_\delta(t,x) = \delta^{-1}\varrho^{(2)}(t,\delta^{-1}x), \qquad \delta = \varepsilon^{\beta-1}.$$

With this choice it is then straightforward to verify that one has

$$c_\varrho = \int P(z)\,\varrho^{(2)}_\delta(z)\,dz \to \int_0^\infty \frac{\bar\varrho(t)}{\sqrt{4\pi t}}\,dt,$$

where we used the shorthand notation $\bar\varrho(t) = \int_{\boldsymbol{R}} \varrho^{(2)}(t,x)\,dx$. This is due to the fact that $\varrho^{(2)}_\delta$ degenerates to a Dirac delta-function in the spatial variable and $P(t,0) = 1/\sqrt{4\pi t}$. Inspection of (2.1), combined with the fact that the heat kernel P satisfies the identities

$$\int_{\boldsymbol{R}} P_s(x)P_t(x)\,dx = \frac{1}{\sqrt{4\pi(s+t)}},$$
$$\int_{\boldsymbol{R}} P_s(x)P_t(x)P_u(x)\,dx = \frac{1}{4\pi\sqrt{st+tu+su}},$$

yields furthermore the expressions

$$\begin{aligned} c^{(1)}_\varrho &= \frac{1}{4\pi}\int_{\boldsymbol{R}^3_+} \frac{\bar\varrho(t+s)\bar\varrho(t+u)}{\sqrt{st+tu+su}}\,ds\,dt\,du \\ &\quad + \frac{1}{4\pi}\int_{\boldsymbol{R}^3_+} \frac{\bar\varrho(t+u+s)}{\sqrt{t+s}}\left(\frac{\bar\varrho(u)}{\sqrt{u}} - c_\varrho\,\delta(u)\right)du\,ds\,dt, \\ c^{(2)}_\varrho &= \frac{1}{4\pi}\int_{\boldsymbol{R}^3_+} \frac{\bar\varrho(t-s)\bar\varrho(t+u)}{\sqrt{st+tu+su}}\,ds\,dt\,du \\ &\quad + \frac{1}{4\pi}\int_{\boldsymbol{R}^3_+} \frac{\bar\varrho(t+u-s)}{\sqrt{t+s}}\left(\frac{\bar\varrho(u)}{\sqrt{u}} - c_\varrho\,\delta(u)\right)du\,ds\,dt, \end{aligned}$$

for the constants governing the remaining two correction terms.

3. Regularity structures.

In order to prove the type of convergence result mentioned in the introduction, we make use of the theory of regularity structures developed in [**Hai14**]. A complete self-contained exposition of the theory is of course beyond the scope of this article, so we will content ourselves with a short summary of the theory's main concepts and results, when specialised to the specific example of the class (1.2). For a more concise exposition of the general theory, see also the lecture notes [**Hai15**], as well as [**FH14**, Chapter 15].

The main ingredient of the theory is that of a *regularity structure*. This consists first of a graded vector space $\mathcal{T} = \bigoplus_{\alpha\in A}\mathcal{T}_\alpha$ where A denotes a set of real-valued indices

(called homogeneities) that is locally finite and bounded from below. In our specific situation, each of the spaces $\mathcal{T}_\alpha$ is finite-dimensional and comes with a distinguished canonical basis. The space $\mathcal{T}$ also comes endowed with a group $\mathcal{G}$ of continuous linear transformations of $\mathcal{T}$ with the property that, for every $\Gamma \in \mathcal{G}$, every $\alpha \in A$, and every $\tau \in \mathcal{T}_\alpha$ one has

$$\Gamma\tau - \tau \in \bigoplus_{\beta<\alpha} \mathcal{T}_\beta. \tag{3.1}$$

The canonical example to keep in mind is the space $\bar{\mathcal{T}} = \bigoplus_{n\in\boldsymbol{N}} \bar{\mathcal{T}}_n$ of abstract polynomials in finitely many indeterminates, with $A = \boldsymbol{N}$ and $\bar{\mathcal{T}}_n$ denoting the space of monomials that are homogeneous of degree n. In this case, a natural group of transformations $\mathcal{G}$ acting on $\bar{\mathcal{T}}$ is given by the group of translations, which does indeed satisfy (3.1).

3.1. Specific regularity structure.

The regularity structure that is relevant for the analysis of (1.2) is built in the following way. First, we start with the regularity structure $\bar{\mathcal{T}}$ given by all polynomials in two indeterminates, let us call them X_1 and X_0, which denote the space and time directions respectively. We do however endow these with the usual parabolic space-time scaling so that each factor of the "time" variable X_0 increases the homogeneity by 2. In particular, one has $\mathbf{1} \in \bar{\mathcal{T}}_0$, $X_1 \in \bar{\mathcal{T}}_1$, $X_0 \in \bar{\mathcal{T}}_2$, $X_1^2 \in \bar{\mathcal{T}}_2$, etc.

We then introduce two additional symbols, Ξ and $\mathcal{I}$, which will be interpreted as an abstract representation of the driving noise ξ and of the operation of convolution with the heat kernel respectively. Fixing some (sufficiently small in the sequel) exponent $\kappa > 0$, we then define $\mathcal{T}_{-(3/2)-\kappa}$ as the copy of $\boldsymbol{R}$ with unit vector denoted by Ξ and we postulate that if τ is some formal expression with homogeneity $|\tau| = \alpha$, then $\mathcal{I}(\tau)$ is a new formal expression with homogeneity $|\mathcal{I}(\tau)| = \alpha + 2$. Furthermore, we also postulate that $\mathcal{I}(X^k) = 0$ for every multiindex k. (Here, for $k = (k_0, k_1)$, we have used the shorthand $X^k = X_0^{k_0} X_1^{k_1}$.) See Section 3.3 below, and in particular (3.8), for an interpretation of this fact. Furthermore, if $\tau, \bar{\tau}$ are formal expressions with respective homogeneities $\alpha, \bar{\alpha}$, then $\tau\bar{\tau} = \bar{\tau}\tau$ is postulated to be a new formal expression with homogeneity $\alpha + \bar{\alpha}$.

A few examples of formal expression with their respective homogeneities that can in principle be built in this way are given by

$$|X_0\Xi| = \frac{1}{2} - \kappa, \quad |\Xi\mathcal{I}(\Xi)| = -1 - 2\kappa, \quad |\Xi^2\mathcal{I}(\Xi)^2| = -2 - 4\kappa. \tag{3.2}$$

In order to define our regularity structure $\mathcal{T}$, we do not keep all of these formal expressions, but only those that are actually useful for the abstract reformulation of (1.2). More precisely, we consider a collection $\mathcal{U}$ of formal expressions which is the smallest collection containing $\mathbf{1}$, X_0, and X_1, and such that

$$\begin{aligned} \tau \in \mathcal{U} \quad &\Rightarrow \quad \mathcal{I}(\tau) \in \mathcal{U}, \\ \tau \in \mathcal{U} \quad &\Rightarrow \quad \mathcal{I}(\Xi\tau) \in \mathcal{U}, \\ \tau, \bar{\tau} \in \mathcal{U} \quad &\Rightarrow \quad \tau\bar{\tau} \in \mathcal{U}. \end{aligned} \tag{3.3}$$

We then set

$$\mathcal{W} = \mathcal{U} \cup \{\Xi\tau : \tau \in \mathcal{U}\},$$

and we define $\mathcal{T}$ as the set of all linear combinations of elements in $\mathcal{W}$. Naturally, $\mathcal{T}_\alpha$ consists of those linear combinations that only involve elements in $\mathcal{W}$ that are of homogeneity α. Furthermore, $\mathcal{W}$ is the previously announced set of canonical basis elements of $\mathcal{T}$. In particular, $\mathcal{T}$ contains the first two expressions of (3.2), but not the last one. It follows furthermore from [**Hai14**, Lemma 8.10] that, for every $\alpha \in \boldsymbol{R}$, $\mathcal{W}$ contains only finitely many elements of homogeneity less than α, so that each $\mathcal{T}_\alpha$ is finite-dimensional and $A \cap (-\infty, \alpha]$ is finite.

We also decompose $\mathcal{T}$ into a direct sum as

$$\mathcal{T} = \mathcal{T}_\Xi \oplus \mathcal{T}_\mathcal{U}, \tag{3.4}$$

where $\mathcal{T}_\mathcal{U}$ is the linear span of $\mathcal{U}$ and $\mathcal{T}_\Xi$ is the linear span of $\Xi\mathcal{U}$, which are all those symbols in $\mathcal{W}$ containing a factor Ξ.

3.2. Structure group.

Let us now describe the structure group $\mathcal{G}$ associated to the space $\mathcal{T}$. For this, we first introduce $\mathcal{T}_+$, the free commutative algebra generated by the formal expressions

$$\mathcal{W}_+ \stackrel{\text{def}}{=} \{X_0, X_1\} \cup \{\mathcal{J}_k(\tau) : \tau \in \mathcal{W} \setminus \bar{\mathcal{T}},\ |k| < |\tau| + 2\}, \tag{3.5}$$

where k is an arbitrary 2-dimensional multiindex and $|k|$ denotes its "parabolic length", i.e.

$$|k| = 2k_0 + k_1.$$

In other words, $\mathcal{T}_+$ consists of all linear combinations of products of formal expressions in $\mathcal{W}_+$. We will view $\mathcal{J}_k$ as a map from $\mathcal{T}$ into $\mathcal{T}_+$ by postulating that it acts linearly on $\mathcal{T}$ and that $\mathcal{J}_k(\tau) = 0$ for those formal expressions τ for which $|\tau| + 2 \le |k|$ or $\tau \in \bar{\mathcal{T}}$. Note that for the moment, elements of $\mathcal{T}_+$ are formal objects. They will be used later on to index matrix elements for some useful linear transformations on $\mathcal{T}$. We will give explicit formulae for the relations between these formal expressions and the numbers they represent in Section 3.3.

With this definition at hand, we construct a linear map $\Delta\colon \mathcal{T} \to \mathcal{T} \otimes \mathcal{T}_+$ in a recursive way. In order to streamline notations, we shall write $\tau^{(1)} \otimes \tau^{(2)}$ as a shorthand for $\Delta\tau$. (This is an abuse of notation, following Sweedler, since in general $\Delta\tau$ is a linear combination of such terms.) We then define Δ via the identities

$$\Delta\mathbf{1} = \mathbf{1} \otimes \mathbf{1}, \qquad \Delta\Xi = \Xi \otimes \mathbf{1}, \qquad \Delta X_i = X_i \otimes \mathbf{1} + \mathbf{1} \otimes X_i, \tag{3.6}$$

and then recursively by the following relations:

$$\Delta\tau\overline{\tau} = \tau^{(1)}\overline{\tau}^{(1)} \otimes \tau^{(2)}\overline{\tau}^{(2)},$$

$$\Delta\mathcal{I}(\tau) = \mathcal{I}(\tau^{(1)}) \otimes \tau^{(2)} + \sum_{\ell,k} \frac{X^\ell}{\ell!} \otimes \frac{X^k}{k!} \mathcal{J}_{\ell+k}(\tau). \tag{3.7}$$

For any linear functional $f : \mathcal{T}_+ \to \boldsymbol{R}$, we can now define in a natural way a map $\Gamma_f : \mathcal{T} \to \mathcal{T}$ by

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

Let now $\mathcal{G}_+$ denote the set of all such linear functionals f which are multiplicative in the sense that $f(\tau\bar{\tau}) = f(\tau)f(\bar{\tau})$ for any two elements $\tau, \bar{\tau} \in \mathcal{T}_+$. With this definition at hand, we set

$$\mathcal{G} = \{\Gamma_f : f \in \mathcal{G}_+\}.$$

It is not difficult to see that these operators satisfy the property (3.1), but it is a highly non-trivial fact that the set $\mathcal{G}$ of these linear operators does indeed form a group under composition, see [**Hai14**, Section 8.1].

Remark 3.1. As a matter of fact, we will never need to consider the full space $\mathcal{T}$ as defined above, but it will be sufficient to consider the subspace generated by all elements of homogeneity less than some large enough number ζ. In practice, it actually turns out to be sufficient to choose any $\zeta > (3/2) + \kappa$, except when proving Corollary 1.13 for which we require $\zeta = 5/2$.

Remark 3.2. As a consequence of (3.6), (3.7), and the fact that linear functionals in $\mathcal{G}_+$ are multiplicative, one has $\Gamma\Xi = \Xi$ and $\Gamma(\tau\bar{\tau}) = (\Gamma\tau)(\Gamma\bar{\tau})$ for every $\Gamma \in \mathcal{G}$ and every $\tau, \bar{\tau} \in \mathcal{T}$ such that the product $\tau\bar{\tau}$ also belongs to $\mathcal{T}$.

3.3. Models.

Now that we have fixed our algebraic regularity structure $(\mathcal{T}, \mathcal{G})$, we introduce a family of analytical objects associated to it that will play the role of Taylor polynomials in our theory in order to allow us to describe solutions to (1.2) locally, up to arbitrarily high order, despite the fact that they are not smooth in the conventional sense.

From now on, we also fix a value $\zeta \geq 2$ as in Remark 3.1 and we set $\mathcal{T} = \bigoplus_{\alpha \in A : \alpha \leq \zeta} \mathcal{T}_\alpha$. This also has the advantage that $\mathcal{T}$ itself is finite-dimensional so we do not need to worry about topologies. In order to describe our "polynomial-like" objects, we first fix a kernel $K : \boldsymbol{R}^2 \to \boldsymbol{R}$ with the following properties:

1. The kernel K is compactly supported in $\{x^2 + |t| \leq 1\}$.
2. One has $K(t, x) = 0$ for $t \leq 0$ and $K(t, -x) = K(t, x)$.
3. For (t, x) with $x^2 + t < 1/2$ and $t > 0$, one has

$$K(t, x) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t},$$

and K is smooth on $\{x^2 + |t| \geq 1/4\}$.

4. For every polynomial $Q\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ of parabolic degree less than ζ, one has

$$\int_{\boldsymbol{R}^2} K(t,x)Q(t,x)\,dx\,dt = 0. \tag{3.8}$$

In other words, K has essentially all the properties of the heat kernel P, except that it is furthermore compactly supported and satisfies (3.8). The constants $1/2$ and $1/4$ appearing in the third point are of course completely arbitrary as long as they are strictly between 0 and 1. The existence of a kernel K satisfying these properties is very easy to show.

We now denote by $\mathcal{S}'$ the space of Schwartz distributions on $\boldsymbol{R}^2$ and by $\mathcal{L}(\mathcal{T},\mathcal{S}')$ the space of (necessarily continuous) linear maps from $\mathcal{T}$ to $\mathcal{S}'$. Furthermore, given a continuous test function $\varphi\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ and a point $z=(t,x)\in\boldsymbol{R}^2$, we set

$$\varphi_z^\lambda(\bar z) = \lambda^{-3}\varphi\big((\lambda^{-2}(\bar t - t), \lambda^{-1}(\bar x - x)\big),$$

where we also used the shorthand $\bar z = (\bar t, \bar x)$. Finally, we write $\mathcal{B}$ for the set of functions $\varphi\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ that are smooth, compactly supported in the ball of radius one, and with their values and both first and second derivatives bounded by 1.

Given a kernel K as above, we then introduce a set $\mathscr{M}$ of *admissible models* which are analytical objects built upon our regularity structure $(\mathcal{T},\mathcal{G})$ that will play a role for our solutions that is similar to that of the usual Taylor polynomials for smooth functions. An admissible model consists of a pair (Π, F) of functions

$$\begin{aligned} \Pi\colon \boldsymbol{R}^2 &\to \mathcal{L}(\mathcal{T},\mathcal{S}'), \qquad & F\colon \boldsymbol{R}^2 &\to \mathcal{G},\\ z &\mapsto \Pi_z, & z &\mapsto F_z, \end{aligned}$$

with the following properties. First, writing $\gamma_{z\bar z}\in\mathcal{G}_+$ for the element such that $F_z^{-1}\circ F_{\bar z} = \Gamma_{\gamma_{z\bar z}}$, we impose that they satisfy the analytical bounds

$$\big|\big(\Pi_z\tau\big)(\varphi_z^\lambda)\big| \lesssim \lambda^{|\tau|}, \qquad \big|\gamma_{z\bar z}(\bar\tau)\big| \lesssim |z-\bar z|^{|\bar\tau|}, \tag{3.9}$$

uniformly over $\varphi\in\mathcal{B}$, $\lambda\in(0,1]$, $\tau\in\mathcal{W}$, and $\bar\tau\in\mathcal{W}_+$. Here, with the same shorthand as before, we set

$$|z-\bar z| = |x-\bar x| + \sqrt{|t-\bar t|}.$$

Also, the proportionality constants implicit in the notation $\lesssim$ are assumed to be bounded uniformly for z and $\bar z$ taking values in any compact set. We furthermore assume that one has the algebraic identity

$$\Pi_z F_z^{-1} = \Pi_{\bar z} F_{\bar z}^{-1}, \tag{3.10}$$

valid for every $z,\bar z$ in $\boldsymbol{R}^2$.

REMARK 3.3. We will write f_z for the element in $\mathcal{G}_+$ such that $F_z = \Gamma_{f_z}$ and we will provide explicit expressions for f_z. We will also use interchangeably the notations (Π, F) and (Π, f) for the model. Note also that there is an explicit formula for the bilinear map giving $\gamma_{z\bar{z}}$ in terms of f_z and $f_{\bar{z}}$, but its expression is somewhat complicated and of no particular use for this article. See [**Hai14**, Section 8.1] for more details.

Finally, and this is why our models are called *admissible*, we assume that, for every multiindex k,

$$\big(\Pi_z X^k\big)(\bar{z}) = (\bar{z} - z)^k, \qquad f_z(X^k) = (-z)^k, \tag{3.11a}$$

and that, for every $\tau \in \mathcal{W}$ with $\mathcal{I}(\tau) \in \mathcal{T}$ (since we truncated $\mathcal{T}$, this is not the case for *all* τ), one has the identities

$$f_z(\mathcal{J}_k\tau) = -\int_{\boldsymbol{R}^2} D^k K(z - \bar{z})\big(\Pi_z\tau\big)(d\bar{z}), \qquad |k| < |\tau| + 2, \tag{3.11b}$$

$$\big(\Pi_z\mathcal{I}\tau\big)(\bar{z}) = \int_{\boldsymbol{R}^2} K(\bar{z} - \bar{\bar{z}})\big(\Pi_z\tau\big)(d\bar{\bar{z}}) + \sum_k \frac{(\bar{z} - z)^k}{k!} f_z(\mathcal{J}_k\tau). \tag{3.11c}$$

Recall that we have set $\mathcal{J}_k\tau = 0$ if $|k| \geq |\tau| + 2$, so that the sum on the second line is finite. It is not clear in principle that these integrals converge, but it turns out that the analytical conditions (3.9) guarantee that this is always the case, see [**Hai14**, Section 5].

REMARK 3.4. Note that since $f_z \in \mathcal{G}_+$, so that it is multiplicative, (3.11a) and (3.11b) do specify f_z (and therefore F_z) completely once we know Π_z. There is therefore quite a lot of rigidity in these definitions, which makes the mere existence of admissible models a highly non-trivial fact.

REMARK 3.5. Building further on Remark 3.4, it actually turns out that if $\Pi\colon \boldsymbol{R}^2 \to \mathcal{L}(\mathcal{T}, \mathcal{S}')$ satisfies the *first* analytical bound in (3.9) and is such that, for F defined from Π via (3.11), one has the identities (3.11) and (3.10), then the second analytical bound in (3.9) is *automatically* satisfied. This is a consequence of [**Hai14**, Theorem 5.14].

Given any smooth space-time function ξ_ε, there is a canonical way of building an admissible model $\Psi(\xi_\varepsilon) = (\Pi^\varepsilon, f^\varepsilon)$ as follows. First, we set $\Pi^\varepsilon_z\Xi = \xi_\varepsilon$, independently of z, and we define it on X^k as in (3.11a). Then, we define Π^ε_z recursively by (3.11c), as well as the identity

$$\big(\Pi^\varepsilon_z\tau\bar{\tau}\big)(\bar{z}) = \big(\Pi^\varepsilon_z\tau\big)(\bar{z})\big(\Pi^\varepsilon_z\bar{\tau}\big)(\bar{z}). \tag{3.12}$$

Note that this is only guaranteed to makes sense if ξ_ε is a function! It was shown in [**Hai14**, Proposition 8.27] that if we furthermore define f^ε via (3.11), then this does indeed define an admissible model for every continuous function ξ_ε. It is however very important to keep in mind that not every admissible model is obtained in this way, or

even as a limit of such models! This will be very important in Section 4 below when we discuss the renormalisation procedure that relates (1.2) to (1.4).

In our case, we would like to define a limiting random model $(\hat\Pi, \hat f)$, naturally called the "Itô model", based on ξ being space-time white noise. One could imagine doing this in a very natural way as follows. As before, one sets $\hat\Pi_z\Xi = \xi$ and $\hat\Pi_z X^k$ as in (3.11a). In other words, we have the identity

$$\big(\hat\Pi_z\Xi\big)(\psi) = \int_{-\infty}^{\infty} \big\langle \psi(t,\cdot), dW(t)\big\rangle,$$

where W is our L^2-cylindrical Wiener process.

We will only consider admissible models, so it is again the case that once we know $\hat\Pi_z\tau$, $\hat\Pi_z\mathcal{I}(\tau)$ is determined by (3.11), so it remains to define $\hat\Pi_z\tau\bar\tau$. To do this, recall the decomposition $\mathcal{T} = \mathcal{T}_\Xi \oplus \mathcal{T}_\mathcal{U}$ as in (3.4). It is straightforward to verify that the structure group $\mathcal{G}_+$ leaves both of these subspaces invariant and that $|\tau| > 0$ for every $\tau \in \mathcal{U}\setminus\{\mathbf{1}\}$. It then follows from [**Hai14**, Proposition 3.28] that if (Π, f) is a model, then $\Pi_z\tau$ is a continuous function for every $\tau \in \mathcal{T}_\mathcal{U}$. As a consequence, if $\tau, \bar\tau \in \mathcal{U}$, we can again define $\hat\Pi_z\tau\bar\tau$ by

$$\big(\hat\Pi_z\tau\bar\tau\big)(\bar z) = \big(\hat\Pi_z\tau\big)(\bar z)\,\big(\hat\Pi_z\bar\tau\big)(\bar z).$$

It remains to define $\hat\Pi_z$ on elements of the form $\Xi\tau$ with $\tau \in \mathcal{T}_\mathcal{U}$. In this case, it would seem natural to postulate that $\hat\Pi_z\Xi\tau$ is a random distribution which acts on test functions ψ by

$$\big(\hat\Pi_z\Xi\tau\big)(\psi) = \int_{-\infty}^{\infty} \big\langle \psi(s,\cdot)\,\big(\hat\Pi_{(t,x)}\tau\big)(s,\cdot), dW(s)\big\rangle. \tag{3.13}$$

Unfortunately, things are not quite that easy. Indeed, for this to make sense as an Itô integral, we need the integrand to be adapted. This is unfortunately the case only if the support of ψ is included in $(t,\infty)\times S^1$. Indeed, it can easily be verified recursively that, as a consequence of the non-anticipativity of the kernel K and the definition (3.11) for an admissible model, the random variables $\big(\hat\Pi_{(t,x)}\tau\big)(s,y)$ are $\mathcal{F}_{s\vee t}$-measurable, where $\mathcal{F}$ is the filtration generated by the increments of W, but they are in general *not* $\mathcal{F}_s$-measurable for $s < t$. As a consequence, (3.13) does not make sense as an Itô integral in general. Since we know that as far as the limiting equation (1.2) is concerned Itô integration coincides with Skorokhod integration, one may think that it suffices to interpret (3.13) in the Skorokhod sense, which is always meaningful if the integrand is sufficiently "nice". Unfortunately, this is not the case either. As a matter of fact, replacing Itô integration by Skorokhod integration would not even allow us to satisfy the consistency equations for the "model" defined in this way.

Our main result will be a consequence of the convergence of a suitable sequence of renormalised models toward an "Itô model" $(\hat\Pi, \hat f)$ which does indeed satisfy the property (3.13) for test functions that are supported "in the future". Here, renormalisation is crucial: if ξ_ε denotes an ε-mollification of our space-time white noise, then one does *not*

expect the sequence of models $\Psi(\xi_\varepsilon)$ to converge to $(\hat\Pi, \hat f)$. Indeed, even in the analogous case of finite-dimensional SDEs, the sequence of solutions to random ODEs obtained from natural regularisations of the noise converges to the Stratonovich solution and not the Itô solution.

In the case of SPDEs of the type (1.2), we expect to have to subtract an asymptotically infinite correction term in order to obtain a finite limit. This subtraction will be made at the level of the *model* rather than at the level of the equation and the ability to do this is one of the main strengths of the theory used in this article. We will describe below in Section 4 the precise renormalisation procedure required to achieve this convergence.

An admissible model (Π, F) really defines an extension of the usual Taylor polynomials, which are given by (3.11a). It is then natural to define spaces $\mathcal{D}^{\gamma,\eta}$ which mimic a weighted version of the Hölder spaces $\mathcal{C}^\gamma$ in the following way. In order to state our definition, given a compact space-time domain D, we denote by $D^{(2)}$ the set of pairs of points $(z,\bar z) \in D^2$ such that furthermore $|z-\bar z| \le 1 \wedge (1/2)\sqrt{|t| \wedge |\bar t|}$, where we used t and $\bar t$ as before as a shorthand for the time components of z and $\bar z$.

DEFINITION 3.6. A function $U\colon \boldsymbol{R}^2 \to \bigoplus_{\alpha<\gamma} \mathcal{T}_\alpha$ belongs to $\mathcal{D}^{\gamma,\eta}$ if, for every compact domain D, one has

$$\|U\|_{\gamma,\eta} \stackrel{\text{def}}{=} \sup_{z\in D}\sup_{\alpha<\gamma} \frac{\|U(z)\|_\alpha}{|t|^{((\eta-\alpha)/2)\wedge 0}} + \sup_{(z,\bar z)\in D^{(2)}} \sup_{\alpha<\gamma} \frac{\|U(z) - \Gamma_{z\bar z}U(\bar z)\|_\alpha}{(|t|\wedge|\bar t|)^{(\eta-\gamma)/2}|z-\bar z|^{\gamma-\alpha}} < \infty. \quad (3.14)$$

Here, we wrote $\|\tau\|_\alpha$ for the norm of the component of τ in $\mathcal{T}_\alpha$ and we used the notation $\Gamma_{z\bar z}$ as a shorthand for $\Gamma_{\gamma_{z\bar z}} = F_z^{-1} \circ F_{\bar z}$ as above.

REMARK 3.7. The powers of t appearing in this definition allow elements of $\mathcal{D}^{\gamma,\eta}$ to exhibit a singularity on the line $\{(t,x) : t = 0\}$. This is essential in order to be able to deal with solutions to (1.2) with "rough" initial conditions.

REMARK 3.8. In order to streamline notations, we suppressed the dependence on the domain D in this norm. This is because in practice, we will only ever use this on some fixed space-time domain.

Note that the space $\mathcal{D}^{\gamma,\eta}$ does depend in a crucial way on the underlying model (Π, F). Therefore, it is not obvious *a priori* how to compare elements belonging to $\mathcal{D}^{\gamma,\eta}$, but based on two different models. This is however crucial when investigating the convergence of solutions to (1.4) as $\varepsilon \to 0$ since these will be obtained from a fixed point problem in $\mathcal{D}^{\gamma,\eta}$, but where the underlying model depends on ε. Given two admissible models (Π, F) and $(\bar\Pi, \bar F)$, the bounds (3.9) yield a natural notion of a semi-distance between the two models by considering, for a given compact domain $D \subset \boldsymbol{R}^2$, the quantity

$$\|\Pi;\bar\Pi\| = \sup_{z\in D} \sup_{\substack{\varphi\in\mathcal{B}\\ \lambda\in(0,1]}} \sup_{\tau\in\mathcal{W}} \frac{\big|\big(\Pi_z\tau - \bar\Pi_z\tau\big)(\varphi_z^\lambda)\big|}{\lambda^{|\tau|}} + \sup_{z,\bar z\in D} \sup_{\tau\in\mathcal{W}_+} \frac{\big|\gamma_{z\bar z}\tau - \bar\gamma_{z\bar z}\tau\big|}{|z-\bar z|^{|\tau|}}. \quad (3.15)$$

We write this as a distance between Π and $\bar\Pi$ only since, as already remarked, γ and $\bar\gamma$ are determined uniquely by Π and $\bar\Pi$ via (3.11). Again, we intentionally do not make the domain D explicit in our notation. This is because our spatial domain is bounded and we will only consider some finite time horizon T. As a consequence, we only ever care about our models in some sufficiently large bounded domain anyway.

A natural distance between elements $U \in \mathcal{D}^{\gamma,\eta}$ and $\bar U \in \bar{\mathcal{D}}^{\gamma,\eta}$ (denoting by $\bar{\mathcal{D}}^{\gamma,\eta}$ the space built over the model $(\bar\Pi, \bar F)$), is given by (3.14), with $U(z)$ replaced by $U(z) - \bar U(z)$ in the first term and $U(z) - \Gamma_{z\bar z}U(\bar z)$ replaced by

$$U(z) - \bar U(z) - \Gamma_{z\bar z}U(\bar z) + \bar\Gamma_{z\bar z}\bar U(\bar z) \tag{3.16}$$

in the second term. We call this quantity $\|U;\bar U\|_{\gamma,\eta}$. Note that this distance is *not* a function of $U - \bar U$! It does however define a distance function on the "fibred space" $\mathscr{M} \ltimes \mathcal{D}^{\gamma,\eta}$ which consists of pairs $((\Pi, F), U)$ of models and modelled distribution such that U belongs to the space $\mathcal{D}^{\gamma,\eta}$ associated to the model (Π, F).

The idea now is to reformulate (1.4), but with $C_\varepsilon = 0$ for the moment, as a fixed point problem in $\mathcal{D}^{\gamma,\eta}$ (based on the canonical model $\Psi(\xi_\varepsilon)$ built above) for suitable values of the exponents γ and η. As a matter of fact, we will view it as a fixed point problem in the subspace $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}} \subset \mathcal{D}^{\gamma,\eta}$ consisting of those functions that take values in $\mathcal{T}_{\mathcal{U}}$. Any element $U \in \mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ can be written uniquely as

$$U(z) = u(z)\mathbf{1} + \tilde U(z), \tag{3.17}$$

where $\tilde U(z)$ takes values in $\bigoplus_{\alpha>0}\mathcal{T}_\alpha$. Furthermore, it is a consequence of [**Hai14**, Proposition 3.28] that if $\gamma > (1/2) - \kappa$, then the function u is necessarily Hölder continuous, in the parabolic sense, with Hölder exponent $(1/2) - \kappa$ on $\boldsymbol{R}_+ \times \boldsymbol{R}$. (Its modulus of Hölder continuity might become singular near $t = 0$.) We will denote by $\mathcal{R}\colon \mathcal{D}^{\gamma,\eta}_{\mathcal{U}} \to \mathcal{C}^{(1/2)-\kappa}(\boldsymbol{R}_+ \times \boldsymbol{R})$ the map $U \mapsto u$. Note that with $U \in \mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ based on an admissible model (Π, F), an equivalent way of defining $\mathcal{R}$ is given by

$$\big(\mathcal{R}U\big)(z) = \big(\Pi_z U(z)\big)(z) \tag{3.18}$$

since, as a consequence of (3.9), $\big(\Pi_z\tilde U(z)\big)(z) = 0$.

This definition makes sense since it turns out that $\Pi_z\tau$ is necessarily a function (and not just a distribution) for every $\tau \in \mathcal{U}$. If it so happens that this is so for every $\tau \in \mathcal{W}$ (as it is for example for the models $(\Pi^\varepsilon, F^\varepsilon)$ mentioned above), then (3.18) actually makes sense for every $U \in \mathcal{D}^{\gamma,\eta}$ (and not just for $U \in \mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$). A remarkable fact, and this is the content of [**Hai14**, Theorem 3.10], is that provided that $\gamma > 0$, the map

$$(\Pi, U) \mapsto \mathcal{R}U \tag{3.19}$$

given by (3.18) is jointly (locally) Lipschitz continuous with respect to the metric defined in (3.14) and (3.15), so that the map (3.19) makes sense even in situations where the definition (3.18) is nonsensical! This of course relies very heavily on the fact that we *only* consider admissible models in (3.19) and not arbitrary functions $\Pi\colon \boldsymbol{R}^2 \to \mathcal{L}(\mathcal{T}, \mathcal{S}')$. The

map $\mathcal{R}$ is called the *reconstruction operator* since it reconstructs the (global) distribution $\mathcal{R}U$ from the (local) data U and Π.

REMARK 3.9. It is possible to verify that if $U \in \mathcal{D}^{\gamma,\eta}$ and one defines $\bar{U}$ by $\bar{U}(z) = \mathcal{Q}_{\bar{\gamma}}U(z)$, where $\mathcal{Q}_{\bar{\gamma}}$ is the projection onto the subspace $\bigoplus_{\alpha<\bar{\gamma}} \mathcal{T}_\alpha$, then one has $\bar{U} \in \mathcal{D}^{\bar{\gamma},\eta}$, provided of course that $\bar{\gamma} \le \gamma$. If one still has $\bar{\gamma} > 0$, then it is the case that $\mathcal{R}\bar{U} = \mathcal{R}U$. In view of this, one might think that only symbols with negative homogeneity "matter". This is not the case however, since one can easily loose regularity. In particular, if $U \in \mathcal{D}^{\gamma,\eta}$, then, combining the definition of $\mathcal{D}^{\gamma,\eta}$ with the facts that $|\Xi| = -(3/2) - \kappa$, homogeneities are additive, and by Remark 3.2 $\Gamma\Xi\tau = \Xi\Gamma\tau$ for every $\Gamma \in \mathcal{G}$, one easily verifies that $\Xi U \in \mathcal{D}^{\gamma-(3/2)-\kappa,\eta-(3/2)-\kappa}$. As a consequence, one needs $\gamma > (3/2) + \kappa$ if one wishes the reconstruction operator $\mathcal{R}$ to be uniquely defined on ΞU.

3.4. Abstract fixed point problem.

We now reformulate (1.4) as a fixed point problem in $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ for suitable values of γ and η. Note first that by Duhamel's formula, the unrenormalised version of (1.4) (i.e. the equation with $C_\varepsilon = 0$) is equivalent to the integral equation

$$u = P \star \big((H(u) + G(u)\xi_\varepsilon)\mathbf{1}_{t>0}\big) + Pu_0. \tag{3.20}$$

Here, P denotes the heat kernel, $\star$ denotes space-time convolution, and Pu_0 denotes the solution to the heat equation with initial condition u_0. A local solution is a pair (u, T) with $T > 0$ and such that (3.20) holds on $[0, T] \times \boldsymbol{R}$. (Here we formulated the problem as if it were on $\boldsymbol{R}$, which one can easily reduce oneself to by considering the periodic extension of the solution.) For this, we need to reformulate the operations of composition with H and G, multiplication by ξ_ε, and convolution against P.

Given $U \in \mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ as in (3.17) (which defines $\tilde{U}$) and a smooth function $G\colon \boldsymbol{R} \to \boldsymbol{R}$, we write

$$\big(\hat{G}(U)\big)(z) = G(u(z))\mathbf{1} + \sum_{k\ge 1} \frac{D^k G(u(z))}{k!} \tilde{U}(z)^k, \tag{3.21}$$

with the understanding that the product between any number of terms such that their homogeneity adds up to γ or more vanishes. It was then shown in [**Hai14**, Proposition 6.13] that the map $U \mapsto \hat{G}(U)$ is locally Lipschitz continuous from $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ to itself, provided that $\gamma > 0$ and $\eta \in [0, \gamma]$.

Furthermore, for every $\delta > 0$ such that furthermore $\delta < (1/2) - \kappa - \eta$, and for every $\gamma > 2 - \delta$, it is possible to construct a linear operator $\mathcal{P}\colon \mathcal{D}^{\gamma-2+\delta,\eta-2+\delta} \to \mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ with the following properties:

1. One has the identity $\mathcal{R}\mathcal{P}U = P \star \mathcal{R}U$, so $\mathcal{P}$ represents space-time convolution by the heat kernel.
2. One has $\mathcal{P}U = \mathcal{I}U + \tilde{\mathcal{P}}U$, where $\tilde{\mathcal{P}}U$ only takes values in $\bar{\mathcal{T}}$, the linear span of the Taylor polynomials $\{X^k\}$.
3. There exists $\theta > 0$ such that

$$\|\mathcal{P}\mathbf{1}_{t>0}U\|_{\gamma,\eta} \lesssim T^{\theta}\|U\|_{\gamma-2+\delta,\eta-2+\delta},$$

where the norms are taken over the domain $[0,T]\times\boldsymbol{R}$.

For a proof of these properties, see Equation 5.15, Theorem 5.12, Proposition 6.16, and Theorem 7.1 in [**Hai14**]. Finally, given a $\mathcal{C}^\gamma$ function u, we write $\mathscr{T}_\gamma u$ for its Taylor expansion of (parabolic) order γ, namely

$$\big(\mathscr{T}_\gamma u\big)(z) = \sum_{|k|<\gamma} \frac{X^k}{k!}\big(D^k u\big)(z) \in \bar{\mathcal{T}} \subset \mathcal{T}.$$

With all of these notations at hand, we can lift (1.2) in a very natural way to a fixed point problem in $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$, by looking for solutions U to

$$U = \mathcal{P}\big((\hat{H}(U) + \hat{G}(U)\Xi)\mathbf{1}_{t>0}\big) + \mathscr{T}_\gamma P u_0. \tag{3.22}$$

The main results of [**Hai14**, Section 7] then allow us to obtain the following result.

THEOREM 3.10. *Fix $\gamma \in ((3/2)+\kappa,\zeta)$ and let F, G be smooth. Then, for every initial condition $u_0 \in \mathcal{C}(S^1)$ and every admissible model (Π,F), there exists a time T such that the fixed point map* (3.22) *has a unique solution in $\mathcal{D}^{\gamma,0}([0,T]\times S^1)$. Furthermore, the solution is locally Lipschitz continuous as a function from $\mathcal{C}(S^1)\times\mathscr{M}$ into $\mathscr{M}\ltimes\mathcal{D}^{\gamma,0}$.*

In principle, one would think that the proof of Theorem 3.10 follows immediately from Theorem 7.8 and Proposition 7.11 in [**Hai14**]. The only caveat is that we do not know whether the map $U\mapsto\hat{G}(U)$ (and similarly for $\hat{F}$) is locally Lipschitz continuous in the strong sense on $\mathcal{D}^{\gamma,0}$. Indeed, while [**Hai14**, Proposition 6.13] yields local Lipschitz continuity when considering arguments built on the same model, it does *not* make any claim regarding the comparison of arguments based on different models. There is however a "trick" that allows us to obtain such a result as a corollary of [**Hai14**, Proposition 6.13], thus yielding the following statement.

PROPOSITION 3.11. *Let $\gamma>0$ and let $(\mathcal{T},\mathcal{G})$ be a regularity structure with no elements of negative homogeneity, endowed with a γ-regular*[1] *product and denote by $\chi>0$ the lowest non-zero homogeneity appearing in $\mathcal{T}$. For $G\colon\boldsymbol{R}\to\boldsymbol{R}$ a function of class $\mathcal{C}^{(\gamma/\chi\vee 1)+1}$, let $\hat{G}$ be defined as in* (3.21).

Then, provided that $\eta\in[0,\gamma]$, one has the bound

$$\|\hat{G}(f);\hat{G}(\bar{f})\|_{\gamma,\eta} \lesssim \|f;\bar{f}\|_{\gamma,\eta} + |\!|\!|\Gamma-\bar{\Gamma}|\!|\!|_\gamma\big(\|f\|_{\gamma,\eta}+\|\bar{f}\|_{\gamma,\eta}\big), \tag{3.23}$$

where we used the notation

[1] See [**Hai14**, Definition 4.6] for this terminology. In particular, the product used in this article is γ-regular for every γ by Remark 3.2.

$$|\!|\!|\Gamma|\!|\!|_\gamma = \sup_{\alpha<\gamma} \sup_{\substack{\tau\in\mathcal{T}_\alpha \\ \|\tau\|=1}} \sup_{\beta<\alpha} \sup_{z\neq\bar z} \frac{\big\|\Gamma_{z\bar z}\tau\big\|_\beta}{|z-\bar z|^{\alpha-\beta}}, \tag{3.24}$$

with the innermost supremum running over the same domain as the one on which the norms in (3.23) *are taken. Here,* $f \in \mathcal{D}^{\gamma,\eta}$ *and* $\bar f \in \bar{\mathcal{D}}^{\gamma,\eta}$*, where* $\bar{\mathcal{D}}^{\gamma,\eta}$ *is the space based on* $\bar\Gamma$*.*

REMARK 3.12. The proportionality constant implicit in the above statement is uniform over the set of all f, $\bar f$, Γ and $\bar\Gamma$ with

$$\|f\|_{\gamma,\eta} + \|\bar f\|_{\gamma,\eta} + |\!|\!|\Gamma|\!|\!|_\gamma + |\!|\!|\bar\Gamma|\!|\!|_\gamma \le R,$$

for arbitrary $R > 0$. Since G is a smooth function with arbitrary growth, one cannot expect better in general.

PROOF. The proof relies on the following construction. Given our regularity structure $\mathcal{T}$ with structure group $\mathcal{G}$, we can double the structure in the following way. First, we set

$$\hat{\mathcal{T}} = \mathcal{T} \oplus \mathcal{T}/\sim,$$

where $\sim$ is the equivalence relation such that $(\tau_1,\tau_2) \sim (\bar\tau_1,\bar\tau_2)$ if and only if there exists $c \in \boldsymbol{R}$ such that $\bar\tau_1 = \tau_1 + c\mathbf{1}$ and $\bar\tau_2 = \tau_2 - c\mathbf{1}$. Since $\mathbf{1}$ is always invariant under the structure group $\mathcal{G}$, the group $\hat{\mathcal{G}} = \mathcal{G}\oplus\mathcal{G}$ acting on $\mathcal{T}\oplus\mathcal{T}$ induces a natural action on $\hat{\mathcal{T}}$.

We now introduce a parameter $\delta > 0$ and, for $\alpha > 0$, we equip the spaces $\hat{\mathcal{T}}_\alpha = \mathcal{T}_\alpha \oplus \mathcal{T}_\alpha$ with δ-dependent norms by setting

$$\|(\tau,\bar\tau)\|_\alpha = \|\tau+\bar\tau\|_\alpha + \delta\,\|\tau-\bar\tau\|_\alpha. \tag{3.25}$$

In the special case $\alpha = 0$, we simply set

$$\|(\tau,\bar\tau)\|_0 = \|\tau+\bar\tau\|_0,$$

which is independent of the representative of our equivalence class. We also have natural injection maps $\iota,\bar\iota\colon \mathcal{T} \to \hat{\mathcal{T}}$ given by $\iota\tau = (\tau,0)$ and $\bar\iota\tau = (0,\tau)$. While these are not isometries, one has $\max\{\|\iota\|,\|\bar\iota\|\} \le 2$, uniformly over $\delta \in (0,1]$. Note that every element in $\hat{\mathcal{T}}$ has a canonical representative for which the two components proportional to $\mathbf{1}$ are equal, so that (3.25) holds for every component. Given an element $\hat\Gamma = (\Gamma,\bar\Gamma) \in \hat{\mathcal{G}}$ and $\hat\tau \in \hat{\mathcal{T}}$ with canonical representative $\hat\tau = (\tau,\bar\tau)$, we then have for any exponent α the bound

$$\begin{aligned}
\|\hat\Gamma\hat\tau\|_\alpha &= \|\Gamma\tau + \bar\Gamma\bar\tau\|_\alpha + \delta\|\Gamma\tau - \bar\Gamma\bar\tau\|_\alpha \\
&\le \frac{1}{2}\big(\|(\Gamma+\bar\Gamma)(\tau+\bar\tau)\|_\alpha + \|(\Gamma-\bar\Gamma)(\tau-\bar\tau)\|_\alpha\big) \\
&\quad + \frac{\delta}{2}\big(\|(\Gamma+\bar\Gamma)(\tau-\bar\tau)\|_\alpha + \|(\Gamma-\bar\Gamma)(\tau+\bar\tau)\|_\alpha\big),
\end{aligned} \tag{3.26}$$

uniformly over the choice of δ. In particular, if we set $\hat{\Gamma}_{xy} = (\Gamma_{xy}, \bar{\Gamma}_{xy})$, then it follows from (3.26) and (3.25) that we have

$$|||\hat{\Gamma}|||_\gamma \le \frac{\delta + \delta^{-1}}{2} |||\Gamma - \bar{\Gamma}|||_\gamma + |||\Gamma|||_\gamma + |||\bar{\Gamma}|||_\gamma. \tag{3.27}$$

The idea will be to consider situations where Γ and $\bar{\Gamma}$ are very close to each other, so that even a choice $\delta \ll 1$ leads to order one bounds on the norm of $\hat{\Gamma}$.

Consider now two modelled distributions f and $\bar{f}$, where $f \in \mathcal{D}^{\gamma,\eta}(\Gamma)$ and $\bar{f} \in \mathcal{D}^{\gamma,\eta}(\bar{\Gamma})$, for two models (Π, Γ) and $(\bar{\Pi}, \bar{\Gamma})$. It is very natural to lift these two models to a single model on $\hat{\mathcal{T}}$ by setting $\hat{\Gamma}_{xy} = (\Gamma_{xy}, \bar{\Gamma}_{xy})$ and $\hat{\Pi}_x(\tau, \bar{\tau}) = \Pi_x \tau + \bar{\Pi}_x \bar{\tau}$. With these notations at hand, it follows immediately from the definitions and from the fact that

$$\hat{\Gamma}_{xy} \iota = \iota \Gamma_{xy}, \qquad \hat{\Gamma}_{xy} \bar{\iota} = \bar{\iota} \bar{\Gamma}_{xy},$$

that both ιf and $\bar{\iota} \bar{f}$ belong to $\mathcal{D}^{\gamma,\eta}(\hat{\Gamma})$ with norms that are bounded by at most twice their original norms, provided that $\delta \in (0, 1]$.

More precisely, it follows immediately from (3.25), combined with the definitions of the (semi-)norms $\|\cdot\|_{\gamma,\eta}$, that one has the two-sided bound

$$\|f; \bar{f}\|_{\gamma,\eta} \le \|\iota f - \bar{\iota} \bar{f}\|_{\gamma,\eta} \le \|f; \bar{f}\|_{\gamma,\eta} + \delta\big(\|f\|_{\gamma,\eta} + \|\bar{f}\|_{\gamma,\eta}\big),$$

uniformly over $\delta \in (0, 1]$. We are now at the stage where we can use [**Hai14**, Proposition 6.13], so that

$$\begin{aligned} \|\hat{G}(f); \hat{G}(\bar{f})\|_{\gamma,\eta} &\le \|\iota \hat{G}(f) - \bar{\iota} \hat{G}(\bar{f})\|_{\gamma,\eta} = \|\hat{G}(\iota f) - \hat{G}(\bar{\iota} \bar{f})\|_{\gamma,\eta} \lesssim \|\iota f - \bar{\iota} \bar{f}\|_{\gamma,\eta} \\ &\le \|f; \bar{f}\|_{\gamma,\eta} + \delta\big(\|f\|_{\gamma,\eta} + \|\bar{f}\|_{\gamma,\eta}\big). \end{aligned}$$

This bound is uniform over all $\delta \in (0, 1]$, all pairs of models (Π, Γ) and $(\bar{\Pi}, \bar{\Gamma})$ such that the induced model $(\hat{\Pi}, \hat{\Gamma})$ has "norm" bounded by a fixed constant, and all functions $f \in \mathcal{D}^{\gamma,\eta}(\Gamma)$ and $\bar{f} \in \mathcal{D}^{\gamma,\eta}(\bar{\Gamma})$ with corresponding norms bounded by a fixed constant. At this stage it looks like one could take δ arbitrarily small. The choice of δ is however limited by (3.27), which suggests that an optimal choice is given by

$$\delta = |||\Gamma - \bar{\Gamma}|||_\gamma,$$

With this particular choice of δ, it follows from (3.27) that for every $R > 0$ there exists C such that $|||\hat{\Gamma}|||_\gamma \le C$ for any two models Γ and $\bar{\Gamma}$ such that $|||\Gamma|||_\gamma + |||\bar{\Gamma}|||_\gamma \le R$. The claim now follows at once. □

PROOF OF THEOREM 3.10. This is now an almost immediate corollary of [**Hai14**, Theorem 7.8], noting that Proposition 3.11 guarantees that the nonlinearity is "strongly locally Lipschitz" in the terminology of that article.

The only thing that needs to be verified is that one has indeed $\mathscr{I}_\gamma P u_0 \in \mathcal{D}^{\gamma,0}$. For

this, we first note that the heat kernel P satisfies for every $k, \ell \in \boldsymbol{N}$ the bound

$$\int_{\boldsymbol{R}} |\partial_x^k \partial_t^\ell P(t,x)|\, dx \lesssim t^{-\ell-(k/2)},$$

uniformly over $(t,x) \in [0,T] \times S^1$ for every fixed final time T. It immediately follows from this bound that the first quantity in (3.14) is bounded, for every fixed $\gamma > 0$, by a multiple of $\|u_0\|_{L^\infty}$. The second quantity is then bounded as a consequence of this by using the remainder formula of [**Hai14**, Theorem A.1]. □

REMARK 3.13. Note that, a consequence of the second property of $\mathcal{P}$, any solution U to (3.22) satisfies

$$U(z) - \mathcal{I}\big(\hat{H}(U(z)) + \hat{G}(U(z))\Xi\big) \in \bar{\mathcal{T}}, \tag{3.28}$$

for all points $z = (t,x)$ with $t \in (0,T)$. This fact will be very important in the sequel when we study the effect of our renormalisation procedure on solutions.

REMARK 3.14. When applied to our situation, the exponent χ appearing in the statement of Proposition 3.11 is equal to $\chi = (1/2) - \kappa$. As a consequence, provided that we restrict ourselves to $\gamma < 2 - 4\kappa$, we only need $F \in \mathcal{C}^2$ and $G \in \mathcal{C}^5$ for the conclusion of Theorem 3.10 to hold.

4. Renormalisation procedure and main result.

At this stage, we note that while Theorem 3.10 allows us to identify solutions to (1.4) with $C_\varepsilon = 0$ with solutions to the abstract fixed point problem (3.22) for a suitable model, we announced a convergence result where we simultaneously let $C_\varepsilon \to \infty$ and introduce additional correction terms of order 1. In particular, this (correctly) suggests that one has no hope to prove that the sequence of models $\Psi(\xi_\varepsilon)$ converges to a limit as $\varepsilon \to 0$.

In order to obtain our main result, the strategy is to build a sequence $M_\varepsilon \Psi(\xi_\varepsilon)$ of *renormalised* models, where M_ε denotes a suitable continuous map on the space of admissible models, such that the following properties hold. First, we show that the sequence $M_\varepsilon \Psi(\xi_\varepsilon)$ converges to a limiting model in $\mathscr{M}$. Second, we show that if U is a (local) solution to (3.22) for the model $M_\varepsilon \Psi(\xi_\varepsilon)$, then the function u in the decomposition (3.17) is the classical solution to the PDE (1.4), but with H replaced by $\bar{H}$ as in (1.7). Finally, we show that solutions to (3.22) depend continuously on the model, thus implying that solutions to (1.4) converge to a limit, and we identify this limit as the Itô solution to (1.2).

In order to implement this strategy, we first explain how the one-parameter family of transformations M_ε is built. This requires a better understanding of the algebraic properties of our regularity structure. In order to simplify notations, we first introduce a graphical shorthand notation for the elements in $\mathcal{W}$.

4.1. Shorthand notation.

From now on, we will highlight the canonical basis vectors of $\mathcal{T}$ (i.e. the elements of $\mathcal{W}$) by drawing them in blue, so that it is easy to distinguish them from either real-valued coefficients or elements of $\mathcal{T}_+$. Note that while the formal structure of $\mathcal{T}_+$ is quite similar to that of $\mathcal{T}$, its role in the theory is very different. Elements of $\mathcal{T}$ are there to index the different components of the underlying model Π_z, while elements of $\mathcal{T}_+$ index the matrix elements of the linear maps F_z. While the notation introduced in Section 3.1 is convenient for making general statements about $\mathcal{T}$ or $\mathcal{G}$, it is not very efficient when talking about any one specific formal expression, since it soon becomes rather lengthy.

We therefore introduce the following alternative graphical notation. Instead of Ξ, we just draw a circle, i.e. we have $\Xi = \circ$. Each occurrence of the abstract integration map $\mathcal{I}$ is then denoted by a downward facing straight line and expressions are multiplied by simply joining the trees representing them by their roots. We also use the shorthand $\Xi X_1 = \otimes$. For example, we have

$$\mathcal{I}(\Xi) = \,, \quad \Xi\,\mathcal{I}(\Xi) = \,, \quad \Xi\,\mathcal{I}(X_1\Xi) = \,, \quad \Xi\,\mathcal{I}^2(\Xi) = \,, \quad \dots$$

While this graphical notation does not allow to describe *every* formal expression in $\mathcal{W}$ (we have no notation for $X_0\Xi$ for example), it will be sufficient for our needs. In order to describe elements in $\mathcal{T}_+$, we will also use $\mathcal{J}'$ for $\mathcal{J}_{(0,1)}$, $\mathcal{J}''$ for $\mathcal{J}_{(0,2)}$ and $\dot{\mathcal{J}}$ for $\mathcal{J}_{(1,0)}$. In view of Remark 3.9, an important role will be played by elements in $\mathcal{W}$ of negative homogeneity, so we list all of them here:

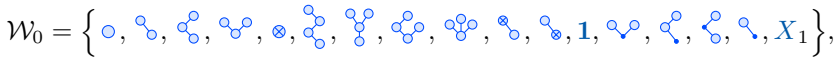
Homogeneity	Symbol(s)
$-(3/2)-\kappa$	
$-1-2\kappa$	
$-(1/2)-3\kappa$	,
$-(1/2)-\kappa$	
-4κ	, , ,
-2κ	,
0	**1**

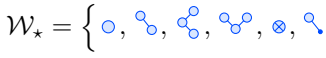
(4.1)

Note that in principle this list may get longer if we take κ too large. One can see by simple inspection that for sufficiently small κ, the elements of smallest positive homogeneity have homogeneity $(1/2)-5\kappa$. Therefore, as long as we assume that $\kappa < 1/10$, the list (4.1) does not change, so we make this a standing assumption.

4.2. General renormalisation group.

It was shown in [**Hai14**, Section 8.3] that one can build a natural family of continuous transformations of $\mathscr{M}$ in the following way. First, we write

$\mathcal{W}_0 =$ 

$\mathcal{W}_\star =$ 

we denote by $\mathcal{T}_0 \subset \mathcal{T}$ the linear span of $\mathcal{W}_0$, and by $\mathcal{T}_0^+ \subset \mathcal{T}_+$ the free algebra generated as in (3.5) by X and $\mathcal{W}_0^+ \stackrel{\text{def}}{=} \{\mathcal{J}_k(\tau) : \tau \in \mathcal{W}_\star,\ |k| < |\tau| + 2\}$. The set $\mathcal{W}_0$ consists of all symbols in $\mathcal{W}$ of negative homogeneity, as well as those symbols generated from them by the renormalisation procedure described in Section 4.3 below. The set $\mathcal{T}_0^+$ generated from $\mathcal{W}_\star$ consists of the smallest collection of symbols required to describe the action of $\mathcal{G}$ on $\mathcal{T}_0$. Consider then an arbitrary linear map $M\colon \mathcal{T}_0 \to \mathcal{T}_0$ which is such that

$$\begin{aligned} M\mathbf{1} &= \mathbf{1}, \qquad & M(X^k\tau) &= X^k M\tau, \\ M\Xi &= \Xi, \qquad & M(\mathcal{I}(\tau)) &= \mathcal{I}(M\tau), \end{aligned} \tag{4.2}$$

where the last identity is assumed to hold for every $\tau \in \mathcal{T}_0$ such that $\mathcal{I}(\tau) \in \mathcal{T}_0$. It was then shown in [**Hai14**, Proposition 8.36] that there exist *unique* maps $\hat{M}\colon \mathcal{T}_0^+ \to \mathcal{T}_0^+$ and $\Delta^M\colon \mathcal{T}_0 \to \mathcal{T}_0 \otimes \mathcal{T}_0^+$ satisfying the identities

$$\begin{aligned} \hat{M}\mathcal{J}_k(\tau) &= \mathcal{M}(\mathcal{J}_k \otimes I)\Delta^M\tau, \\ \hat{M}(\tau\bar{\tau}) &= (\hat{M}\tau)(\hat{M}\bar{\tau}), \\ (I \otimes \mathcal{M})(\Delta \otimes I)\Delta^M\tau &= (M \otimes \hat{M})\Delta\tau, \end{aligned} \tag{4.3}$$

where $\mathcal{M} : \mathcal{T}_0^+ \otimes \mathcal{T}_0^+ \to \mathcal{T}_0^+$ denotes the multiplication map, and such that $\hat{M}$ leaves X^k invariant. Note that the first identity in (4.3) should be checked for all $\tau \in \mathcal{W}_*$, the second for all $\tau, \bar{\tau} \in \mathcal{T}_0^+$, and the third for all $\tau \in \mathcal{W}_0$. With these notations at hand, we have the following definition.

Definition 4.1. The *renormalisation group* $\mathfrak{R}$ associated to our regularity structure is given by the set of all linear maps M as above such that furthermore, for every $\tau \in \mathcal{W}_0$, one has

$$\Delta^M\tau = \tau \otimes \mathbf{1} + \sum \tau^{(1)} \otimes \tau^{(2)},$$

where each of the terms $\tau^{(1)}$ appearing in these sums satisfies $|\tau^{(1)}| > |\tau|$.

Given any $M \in \mathfrak{R}$ and given an admissible model (Π, F), we can define a "renormalised model" $(\hat{\Pi}, \hat{F})$ by setting

$$\hat{\Pi}_x\tau = (\Pi_x \otimes f_x)\Delta^M\tau, \quad \hat{f}_x = f_x\hat{M}. \tag{4.4}$$

Note that in principle $(\hat{\Pi}, \hat{F})$ is only defined on the smaller regularity structure $\mathcal{T}_0$. However, it follows from [**Hai14**, Proposition 4.11] and [**Hai14**, Theorem 5.14] that any

admissible model on $\mathcal{T}_0$ extends uniquely and continuously to a model on all of $\mathcal{T}$. We will implicitly use this extension in the sequel. One then has the following:

THEOREM 4.2. *The map* $(M, \Pi, F) \mapsto (\hat{\Pi}, \hat{F})$ *is continuous from* $\mathfrak{R} \times \mathscr{M}$ *to* $\mathscr{M}$.

PROOF. The fact that $(\hat{\Pi}, \hat{F})$ satisfies the first bound in (3.9) as well as (3.10) and (3.11) was shown in [**Hai14**, Theorem 8.44]. The fact that the second bound in (3.9) also holds does in turn follow automatically from [**Hai14**, Theorem 5.14], using the fact that $\hat{\Pi}$ is again admissible. The continuity with respect to M follows in the same way. □

4.3. Renormalization map.

We will not give a full characterisation of $\mathfrak{R}$, but we will instead describe a three-dimensional subgroup which is sufficient for our needs. We write $M = \exp(-cL - c^{(1)}L^{(1)} - c^{(2)}L^{(2)})$, where $c, c^{(1)}, c^{(2)} \in \boldsymbol{R}$ and L, $L^{(1)}$, $L^{(2)}$ are linear maps on $\mathcal{T}_0$. The map $L^{(1)}$ is simply given by $L^{(1)}$  $= \mathbf{1}$, as well as $L^{(1)}\tau = 0$ for every $\tau \in \mathcal{W}_0 \setminus \{$ $\}$. Similarly, $L^{(2)}$ is given by $L^{(2)}$  $= \mathbf{1}$, and $L^{(2)}\tau = 0$ otherwise.

The map L on the other hand is more complicated to describe. First, one has $L:$  $\mapsto \mathbf{1}$. Furthermore, if τ is a more complicated expression, then L iterates over all occurrences of  as a "subsymbol" of τ and "erases" it in the graphical notation.

More precisely, one has the identities

$$
\begin{aligned}
&L \text{  } = \mathbf{1}, && L \text{ } = \text{ }, && L \text{  } = 2 \text{ },\\
&L \text{  } = 3 \text{  }, && L \text{ } = \text{ } + \text{ }, && L \text{  } = \text{ } + \text{  },\\
&L \text{  } = \text{  } + 2 \text{ }, && L \text{  } = X_1, && L \text{  } = X_1.
\end{aligned}
\tag{4.5}
$$

(Recall that $\mathcal{I}(\mathbf{1}) = 0$, so there is no term  appearing in L  and similarly for the other terms.) We furthermore have

$$
L\mathbf{1} = L \circ = L \otimes = L \text{ } = L \text{ } = L \text{  } = L \text{ } = 0.
$$

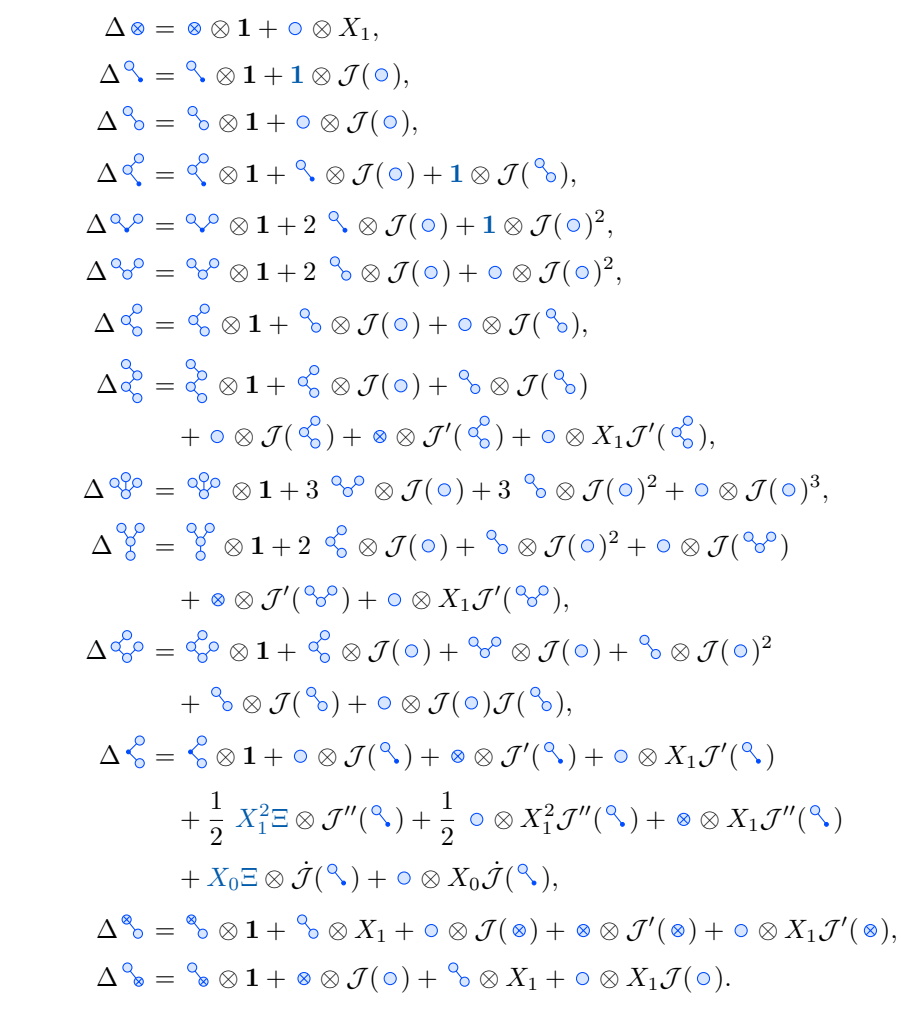
In particular, this shows immediately that $L^2\tau = 0$ for all $\tau \in \mathcal{W}_0$ and L and the $L^{(i)}$ all commute, so that one simply has $M = I - cL - c^{(1)}L^{(1)} - c^{(2)}L^{(2)}$. The main result of this section is the following.

PROPOSITION 4.3. *With* M *defined as above, one has* $M \in \mathfrak{R}$.

PROOF. Since $\mathfrak{R}$ is a group and the operators L, $L^{(1)}$, $L^{(2)}$ commute, it suffices to verify this separately for $\exp(-cL)$ and $\exp(-cL^{(i)})$. We note that if we set $M = \exp(-cL^{(i)})$ for $i \in \{1, 2\}$, then M satisfies the identity $\Delta M\tau = (M \otimes I)\Delta\tau$. As a consequence, it is easy to verify that (4.3) holds, provided that we set $\Delta^M\tau = M\tau \otimes \mathbf{1}$, and $\hat{M}\tau = \tau$. It follows from the "upper triangular" structure of M that Δ^M then satisfy

the properties of Definition 4.1.

Verifying these properties for $M = \exp(-cL)$ is less straightforward. It follows from the recursive definition of Δ that one has the identities

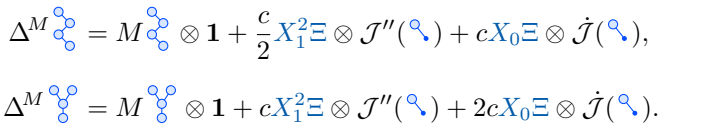


(4.6)

For all above symbols, except for  and  , it turns out that we have

$$\Delta^M \tau = M\tau \otimes \mathbf{1}. \tag{4.7}$$

Moreover, for these two exceptional symbols, one has



Finally $\hat{M} : \mathcal{T}_+ \to \mathcal{T}_+$ is such that $\hat{M}(\sigma\overline{\sigma}) = \hat{M}(\sigma)\hat{M}(\overline{\sigma})$ and for all $\tau \in \mathcal{W}_*$,

$$\hat{M}\mathcal{J}_k(\tau) = \mathcal{J}_k(M\tau).$$

In order to prove that this is the case, it suffices to check that Δ^M and $\hat{M}$ defined in this way do indeed satisfy (4.3). The first two identities are essentially obvious consequences of the definition of $\hat{M}$. Except for the two cases $\tau =$  and  , the third one follows from the identity

$$\Delta M\tau = (M \otimes \hat{M})\Delta\tau,$$

or equivalently

$$\Delta(M\tau - \tau) = \big((M \otimes \hat{M}) - I\big)\Delta\tau.$$

It is not hard to check that identity for all $\tau \in \mathcal{W}_0 \backslash \{$  ,  $\}$. Let us give the details of the computation for the case $\tau =$  . We need to check that

$$\Delta(M - I)$$  $$= \big((M \otimes \hat{M}) - I\big)\Delta$$  .

It is plain that

$$M$$  $$=$$  $$- c$$  $$- c$$  ,

so that

$$\begin{aligned}
\Delta(M - I)\,&\text{  } = -c\,\Delta\,\text{  } - c\,\Delta\,\text{  }\\
&= -c\,\text{  } \otimes \mathbf{1} - c\,\text{  } \otimes \mathcal{J}(\text{  }) - c\,\mathbf{1} \otimes \mathcal{J}(\text{  })\\
&\quad - c\,\text{  } \otimes \mathbf{1} - 2c\,\text{  } \otimes \mathcal{J}(\text{  }) - c\,\mathbf{1} \otimes \mathcal{J}(\text{})^2.
\end{aligned}$$

On the other hand, using the expression for Δ given in (4.6), we obtain

$$\begin{aligned}
\big((M \otimes \hat{M}) - I\big)\Delta\,&\text{  } = -c\,\text{  } \otimes \mathbf{1} - c\,\text{  } \otimes \mathbf{1} - 3c\,\text{} \otimes \mathcal{J}(\text{  })\\
&\quad - c\,\mathbf{1} \otimes \mathcal{J}(\text{})^2 - c\,\mathbf{1} \otimes \mathcal{J}(\text{}),
\end{aligned}$$

thus establishing the required identity.

The two cases $\tau =$  and $\tau =$  are similar, so we only give the detailed computations for the case $\tau =$  . We need to check that

$$(I \otimes \mathcal{M})(\Delta \otimes I)\Delta^M \; = (M \otimes \hat{M})\Delta \; .$$

We first note that

$$\Delta^M \; = \; \otimes \mathbf{1} - c \; \otimes \mathbf{1} - c \; \otimes \mathbf{1} + \frac{c}{2} X_1^2 \Xi \otimes \mathcal{J}''(\;) + c X_0 \Xi \otimes \dot{\mathcal{J}}(\;),$$

from which we deduce that the left hand side reads

$$\begin{aligned}
&(I \otimes \mathcal{M})(\Delta \otimes I)\Delta^M \\
&\quad = \otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \otimes \mathcal{J}(\;) + \circ \otimes \mathcal{J}(\;) + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \\
&\qquad - c \Big(\otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \mathbf{1} \otimes \mathcal{J}(\;) + \otimes \mathbf{1} + \circ \otimes \mathcal{J}(\;) + \otimes \otimes \mathcal{J}'(\;) \\
&\qquad\quad + \circ \otimes X_1 \mathcal{J}'(\;) + \frac{1}{2} X_1^2 \Xi \otimes \mathcal{J}''(\;) + X_1 \Xi \otimes X_1 \mathcal{J}''(\;) \\
&\qquad\quad + \frac{1}{2} \Xi \otimes X_1^2 \mathcal{J}''(\;) + X_0 \Xi \otimes \dot{\mathcal{J}}(\;) + \Xi \otimes X_0 \dot{\mathcal{J}}(\;) \Big) \\
&\qquad + c \Big(\frac{1}{2} X_1^2 \Xi \otimes \mathcal{J}''(\;) + X_1 \Xi \otimes X_1 \mathcal{J}''(\;) + \frac{1}{2} \Xi \otimes X_1^2 \mathcal{J}''(\;) \\
&\qquad\quad + X_0 \Xi \otimes \dot{\mathcal{J}}(\;) + \Xi \otimes X_0 \dot{\mathcal{J}}(\;) \Big) \\
&\quad = \otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \otimes \mathcal{J}(\;) + \circ \otimes \mathcal{J}(\;) + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \\
&\qquad - c \Big(\otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \mathbf{1} \otimes \mathcal{J}(\;) + \otimes \mathbf{1} + \circ \otimes \mathcal{J}(\;) \\
&\qquad\quad + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \Big),
\end{aligned}$$

while the right hand side reads

$$\begin{aligned}
&(M \otimes \hat{M}) \Big(\otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \otimes \mathcal{J}(\;) \\
&\qquad + \circ \otimes \mathcal{J}(\;) + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \Big) \\
&\quad = \otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \otimes \mathcal{J}(\;) + \circ \otimes \mathcal{J}(\;) + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \\
&\qquad - c \Big(\otimes \mathbf{1} + \otimes \mathbf{1} + \otimes \mathcal{J}(\circ) + \mathbf{1} \otimes \mathcal{J}(\;) + \circ \otimes \mathcal{J}(\;) \\
&\qquad\quad + \otimes \otimes \mathcal{J}'(\;) + \circ \otimes X_1 \mathcal{J}'(\;) \Big).
\end{aligned}$$

The required identity is established. By inspection, we then see that Δ^M is indeed of the form required by Definition 4.1, thus concluding the proof. □

4.4. Renormalised solutions.

We now have all the tools required to prove the following result.

PROPOSITION 4.4. *Let $H \in \mathcal{C}^2$, $G \in \mathcal{C}^5$, let $M_\varepsilon = \exp\bigl(-C_\varepsilon L - c^{(1)}L^{(1)} - c^{(2)}L^{(2)}\bigr)$ be as in Section* 4.3*, let $u_0 \in \mathcal{C}(S^1)$, let ξ_ε be a smooth function, and denote by U the local solution to* (3.22) *with model $M_\varepsilon\Psi(\xi_\varepsilon)$ given by Theorem* 3.10*. Then, the function u as in* (3.17) *is the classical solution to the PDE* (1.4)*, with H replaced by $\bar{H}$ as in* (1.7)*.*

If furthermore the PDE (1.4) *has classical solutions that remain bounded up to time $T > 0$, then the fixed point problem* (3.22) *can also be solved uniquely over the same interval $[0, T]$.*

PROOF. Let $U(t,x)$ denote the local solution to (3.22) with model $(\hat{\Pi}^{(\varepsilon)}, \hat{F}^{(\varepsilon)}) \stackrel{\text{def}}{=} M_\varepsilon\Psi(\xi_\varepsilon)$. Applying the reconstruction operator $\mathcal{R}_\varepsilon$ associated to this model to both sides of (3.22), we obtain the identity

$$u = P \star \bigl(\bigl(H(u) + \mathcal{R}_\varepsilon(\hat{G}(U)\Xi)\bigr)\mathbf{1}_{t>0}\bigr) + Pu_0, \tag{4.8}$$

where we used the fact that $\mathcal{R}_\varepsilon\bigl(\hat{H}(U)\bigr) = H(u)$, as well as the defining properties of the operator $\mathcal{P}$. While it is also the case that $\mathcal{R}_\varepsilon\Xi = \xi_\varepsilon$, it is *not* the case in general that $\mathcal{R}_\varepsilon\bigl(\hat{G}(U)\Xi\bigr) = G(u)\xi_\varepsilon$.

It then follows from (3.28) and (3.21) that if we consider it as an element of $\mathcal{D}^{\gamma,\eta}$ with γ greater than, but sufficiently close to, $3/2$, then U is of the form

$$\begin{aligned} U = {} & u\,\mathbf{1} + G(u)\,\mathcal{I}(\Xi) + G'(u)G(u)\,\mathcal{I}(\mathcal{I}(\Xi)^2) + u'\,X_1 \\ & + G'(u)^2G(u)\,\mathcal{I}(\mathcal{I}(\Xi)\mathcal{I}(\Xi)^2) + \frac{1}{2}G''(u)G^2(u)\,\mathcal{I}(\mathcal{I}(\Xi)^3) + G'(u)u'\,\mathcal{I}(X_1), \end{aligned} \tag{4.9}$$

for some functions u and u'. In particular, as an element of $\mathcal{D}^{\gamma'}$ for $\gamma' > 0$ sufficiently close to 0, we have the identity

$$\begin{aligned} \hat{G}(U)\Xi = {} & G(u)\,\Xi + G'(u)G(u)\,\mathcal{I}(\Xi)\Xi + G'(u)^2G(u)\,\mathcal{I}(\mathcal{I}(\Xi)^2)\Xi + G'(u)u'\,\mathcal{I}(X_1)\Xi \\ & + \frac{1}{2}G''(u)G^2(u)\,\mathcal{I}(\mathcal{I}(\Xi)^3)\Xi + \frac{1}{6}G'''(u)G^3(u)\,\mathcal{I}(\mathcal{I}(\Xi)^4)\Xi + G'(u)^3G(u)\,\mathcal{I}(\mathcal{I}(\Xi)^2\mathcal{I}(\Xi)^2)\Xi \\ & + \frac{1}{2}G''(u)G'(u)G^2(u)\,\mathcal{I}(\mathcal{I}(\Xi)^2\mathcal{I}(\Xi)^3)\Xi + G''(u)G'(u)G^2(u)\,\mathcal{I}(\mathcal{I}(\Xi)^3\mathcal{I}(\Xi)^2)\Xi \\ & + G'(u)^2u'\,\mathcal{I}(\mathcal{I}(\Xi)X_1)\Xi + G''(u)G(u)u'\,\mathcal{I}(X_1)\Xi. \end{aligned} \tag{4.10}$$

At this stage, we note that if we write $(\Pi^{(\varepsilon)}, \Gamma^{(\varepsilon)}) = \Psi(\xi_\varepsilon)$, then we have the identity

$$\bigl(\hat{\Pi}_z^{(\varepsilon)}\tau\bigr)(z) = \bigl(\Pi_z^{(\varepsilon)}M_\varepsilon\tau\bigr)(z),$$

with M_ε as in the statement. Combining this with (4.5), we see that this expression is

non-zero for the symbols $\mathbf{1}$, $\circ$, $\circ\!\!\circ$, $\circ\circ\circ\circ$ and $\circ\circ\circ\circ$, where one has

$$\big(\hat{\Pi}_z^{(\varepsilon)}\mathbf{1}\big)(z) = 1, \quad \big(\hat{\Pi}_z^{(\varepsilon)}\circ\big)(z) = \xi_\varepsilon(z), \quad \big(\hat{\Pi}_z^{(\varepsilon)}\circ\!\!\circ\big)(z) = -C_\varepsilon,$$
$$\big(\hat{\Pi}_z^{(\varepsilon)}\circ\circ\circ\circ\big)(z) = -c^{(1)}, \quad \big(\hat{\Pi}_z^{(\varepsilon)}\circ\circ\circ\circ\big)(z) = -c^{(2)}.$$

The third identity follows from the fact that $M_\varepsilon\, \circ\!\!\circ = \circ\!\!\circ - C_\varepsilon \mathbf{1}$ and

$$\Pi_z^{(\varepsilon)}\circ\!\!\circ = \Pi_z^{(\varepsilon)}\circ\!\!\searrow \cdot\, \Pi_z^{(\varepsilon)}\circ,$$

by the definition (3.12) of the canonical lift, noting that the first factor vanishes at the point z because $|\circ\!\!\searrow| > 0$. The last two identities hold for similar reasons.

Combining this with (4.10) and the expression (3.18) for the reconstruction operator, it follows that one has the identity

$$\begin{aligned}\mathcal{R}_\varepsilon\big(\hat{G}(U)\Xi\big)(z) &= G(u(z))\xi_\varepsilon(z) - C_\varepsilon G'(u(z))G(u(z)) \\ &\quad - c^{(1)}G'(u(z))^3 G(u(z)) - c^{(2)}G''(u(z))G'(u(z))G(u(z))^2.\end{aligned}$$

The first claim now follows by combining this with (4.8).

Regarding the possibility to solve (3.22) up to the classical blow-up time of the corresponding PDE, this was shown in [**Hai14**, Proposition 7.11]. □

4.5. Main results.

Given a cylindrical Wiener process W, we define ξ_ε as in (1.6), we set $C_\varepsilon = \varepsilon^{-1}c_\varrho$ with c_ϱ given in (1.5), and we define $c_\varrho^{(1)}$ and $c_\varrho^{(2)}$ as in Section 2. We also define $(\Pi^\varepsilon, F^\varepsilon) = \Psi(\xi_\varepsilon)$, the canonical lift of ξ_ε to the regularity structure $\mathcal{T}$. As before, we write $M_\varepsilon = \exp(-C_\varepsilon L - c_\varrho^{(1)}L^{(1)} - c_\varrho^{(2)}L^{(2)})$ and we denote by $(\hat{\Pi}^\varepsilon, \hat{F}^\varepsilon)$ the renormalised model obtained from $(\Pi^\varepsilon, F^\varepsilon)$ by the action of M_ε given in (4.4). With this notation, our main convergence result at the level of models is the following.

THEOREM 4.5. *Let $(\hat{\Pi}^\varepsilon, \hat{F}^\varepsilon)$ be the renormalised model described above with ξ_ε as in (1.6), $C_\varepsilon = \varepsilon^{-1}c_\varrho$, $c^{(1)} = c_\varrho^{(1)}$, and $c^{(2)} = c_\varrho^{(2)}$ as defined in Section 2. Then, there exists a random model $(\hat{\Pi}, \hat{F})$ and a constant C such that*

$$\boldsymbol{E}\|\hat{\Pi}^\varepsilon; \hat{\Pi}\| \le C\varepsilon^{\kappa/2}, \tag{4.11}$$

for every underlying compact space-time domain.

Furthermore, for every $\tau \in \mathcal{U}$ and every (t,x), the process $s \mapsto \big(\hat{\Pi}_{(t,x)}\tau\big)(s,\cdot)$ is $\mathcal{F}_s$-adapted for $s > t$ and, for every smooth test function φ supported in the future $\{(s,y) : s > t\}$, one has the identity

$$\big(\hat\Pi_{(t,x)}\Xi\tau\big)(\varphi) = \int_t^\infty \big\langle \big(\hat\Pi_{(t,x)}\tau\big)(s,\cdot)\varphi(s,\cdot), dW(s)\big\rangle, \tag{4.12}$$

where the integral on the right is the Itô integral. We call $(\hat\Pi, \hat F)$ *the Itô model.*

REMARK 4.6. In this statement, we did again denote by $\mathcal{F}$ the filtration generated by the underlying cylindrical Wiener process W. Note also that (3.13) does *not* hold in general if φ is not supported in the future. In fact, the statement may not even make any sense in this case since in general $\big(\Pi_{(t,x)}\tau\big)(s,\cdot)$ is not adapted to $\mathcal{F}$ for $s < t$. One may wonder if in this case (3.13) still holds, but with the integral on the right interpreted as a Skorokhod integral in situations where the integrand is anticipative. Again, this is *not* the case in general.

PROOF. The proof is the content of Section 5 below. Instead of (4.11), we will however only show the seemingly weaker statement that

$$\boldsymbol{E}\big|\big(\hat\Pi_0^\varepsilon\tau - \hat\Pi_0^{\bar\varepsilon}\tau\big)(\varphi_\lambda)\big|^2 \le C\varepsilon^\kappa\lambda^{2|\tau|+\kappa}, \qquad \boldsymbol{E}\big|\big(\hat\Pi_0^\varepsilon\tau\big)(\varphi_\lambda)\big|^2 \le C\lambda^{2|\tau|+\kappa}, \tag{4.13}$$

uniformly over $\lambda \in (0,1]$, over smooth test functions φ supported in the ball of radius 1 and with $\mathcal{C}^2$ norm bounded by 1, over all $\tau \in \mathcal{W}$ with $|\tau| < 0$, and over all $0 < \bar\varepsilon \le \varepsilon \le 1$. Here, we wrote φ_λ as a shorthand for the function

$$\varphi_\lambda(t,x) = \lambda^{-3}\varphi(\lambda^{-2}t, \lambda^{-1}x).$$

Taking this bound for granted, the existence of a random model $(\hat\Pi, \hat F)$ satisfying the required bound (4.11) is then an immediate consequence of [**Hai14**, Theorem 10.7]. □

We now have all the tools in place to formulate the main convergence result of this article.

THEOREM 4.7. *Consider the setting of Theorem* 4.5. *Fix* $\gamma \in ((3/2)+\kappa, \zeta]$ *and let* $H \in \mathcal{C}^{\chi_1}$ *and* $G \in \mathcal{C}^{\chi_2}$ *with*

$$\chi_1 > \left(1 + \frac{2\gamma - 4}{1 - 2\kappa}\right) \vee 2, \qquad \chi_2 > \frac{4 \vee 2\gamma}{1 - 2\kappa} \vee 2. \tag{4.14}$$

Assume that the derivative of both H *and* G *is uniformly bounded and let* $u_0 \in \mathcal{C}(S^1)$ *and* $T > 0$. *For any* $\varepsilon > 0$, *denote by* U^ε *the maximal (up to time* T*) solution to the fixed point problem in* $\mathcal{D}^{\gamma,0}$ *constructed in Theorem* 3.10 *with respect to the model* $(\hat\Pi^\varepsilon, \hat F^\varepsilon)$. *Denote by* U *the same solution, but with respect to the Itô model* $(\hat\Pi, \hat F)$.

Then, the maximal existence time for U *is almost surely equal to* T *and* $\mathcal{R}U$ *coincides almost surely with the Itô solution to* (1.2). *Recalling the distance* $\|U; U^\varepsilon\|_{\gamma,\eta}$ *stated just after* (3.16), *one then has for every* $\theta < \kappa/2$ *the estimate*

$$\lim_{\varepsilon\to 0} \boldsymbol{P}\big(\|U; U^\varepsilon\|_{\gamma,0} > \varepsilon^\theta\big) = 0.$$

Here, the distance $\|\cdot;\cdot\|_{\gamma,0}$ is taken over the domain $[0,T]\times S^1$. In particular, it implies that with high probability, the maximal existence time for U^ε is at least T.

PROOF. The proof of the theorem is essentially just a collection of the results of this article. For the Itô model $(\hat\Pi,\hat F)$, solutions are shown in Corollary 6.5 below to coincide with the solutions to (1.2). Since these are known to be global almost surely [**DPZ92**], it follows from [**Hai14**, Proposition 7.11] that the solutions to (3.22) are also global.

The convergence in probability of U^ε to U is then an immediate consequence of Theorem 4.5, combined with the local Lipschitz continuity of the solution map given in Theorem 3.10. In particular, it immediately follows that the existence time for U^ε is at least T with a probability converging to 1 as $\varepsilon\to 0$.

The assumptions (4.14) on the regularity of H and G are precisely what is needed in Proposition 3.11 to ensure that $\hat H$ is locally Lipschitz continuous from $\mathcal{D}^\gamma$ into $\mathcal{D}^{\gamma-2+\kappa}$, and that $\hat G$ is locally Lipschitz continuous from $\mathcal{D}^\gamma$ into $\mathcal{D}^{\gamma-(1/2)+\kappa}$. (Using the fact that the element of lowest non-zero homogeneity appearing in the description of U is Ξ with homogeneity $(1/2)-\kappa$.) □

It is now very easy to prove the various results stated in the introduction.

PROOF OF THEOREM 1.1. This is now an immediate corollary of Theorem 4.7, noting that if we write $u=\mathcal{R}U$ and $u_\varepsilon=\mathcal{R}U^\varepsilon$ then, for $\alpha\le(1/2)-\kappa$ and $t>0$ both the $\mathcal{C}^{\alpha,\alpha/2}$ norm over $(t,T]\times S^1$ and the supremum norm over $[0,T]\times S^1$ are controlled by $\|U;U^\varepsilon\|_{\gamma,0}$. Furthermore, Proposition 4.4 identifies u^ε with the classical solution to (1.4) with modified drift $\bar H$, while Corollary 6.5 identifies u with the Itô solution to (1.2). □

PROOF OF COROLLARY 1.10. This follows immediately from the local Lipschitz continuity of the solution map of Theorem 3.10, combined with the identification of solutions given by Corollary 6.5. □

PROOF OF COROLLARIES 1.11 AND 1.13. These corollaries are both a consequence of the form of the solution. Recall that the solution U to the fixed point problem will necessarily be of the form (4.9). Furthermore, it follows from our definition of the Itô model that the solution v to the linearised equation is given by

$$v(\bar z)-v(z)=\bigl(\Pi_z\Xi\bigr)(\bar z)+R_v(z,\bar z),$$

where the remainder R_v satisfies the bound

$$R_v(z,\bar z)\lesssim\frac{|x-\bar x|+\sqrt{|t-\bar t|}}{(|t|\wedge|\bar t|)^{1/4}},$$

uniformly over $z,\bar z\in[0,T]\times S^1$ for any fixed $T>0$. (Here, we used again the shorthands $z=(t,x)$ and $\bar z=(\bar t,\bar x)$.) In particular, we have

$$G(u(z))\big(v(\bar z)-v(z)\big)=G(u(z))\big(\Pi_z \Xi\big)(\bar z)+\bar R_v(z,\bar z),$$

with $\bar R_v$ satisfying $R_v(z,\bar z)\lesssim |x-\bar x|+\sqrt{|t-\bar t|}$, uniformly away of the line $\{t=0\}$.

On the other hand, it follows from (4.9) and the definition of $\mathcal{D}^{\gamma,0}$ that

$$u(\bar z)-u(z)=G(u(z))\big(\Pi_z \Xi\big)(\bar z)+R_u(z,\bar z),$$

for a remainder R_u satisfying the bound $R_u(z,\bar z)\lesssim (|x-\bar x|+\sqrt{|t-\bar t|})^{1-2\kappa}$. This is thanks to the definition of a model, combined with the fact that the term of lowest homogeneity different from $\mathbf{1}$ and Ξ appearing in the description of U is $\mathcal{I}(\Xi)$, which has homogeneity $1-2\kappa$. Comparing both expressions, Corollary 1.11 follows.

To prove Corollary 1.13, we first note that, as a consequence of (4.9), if z is such that $u(z)=\bar u(z)$, then the coefficients multiplying Ξ and $\mathcal{I}(\Xi)$ in the description of U and $\bar U$ necessarily also coincide at the point z. As a consequence, $\bar u-u$ is differentiable at z (in the spatial direction) and, after subtracting the corresponding linear term, the remainder is given by

$$G'(u(z))\big(u'(z)-\bar u'(z)\big)\big(\Pi_z \mathcal{I}(\Xi)\big)(\bar z)+R,$$

where R is bounded by a multiple of $|\bar t-t|+|\bar x-x|^2$. (Again, locally uniformly away from $\{t=0\}$.) Since $\mathcal{I}(\Xi)$ is of homogeneity $(3/2)-\kappa$, the first claim follows.

In the context of the last statement of Corollary 1.13, the point z furthermore has the property that $u'(z)=\bar u'(z)$. Indeed, if this were not the case then, by the first part, $u-\bar u$ would have a non-zero spatial derivative at z, which would contradict our assumption on the behaviour of the solutions near z. We then combine this with the fact that, to order 2, the expansion for U is the same as that for $\hat G(U)\Xi$ given in (4.10) (except for the terms multiplying the "Taylor polynomials"). In particular, as a consequence of the identities $u(z)=\bar u(z)$ and $u'(z)=\bar u'(z)$, all of these terms agree at the point z. Furthermore, the terms of lowest homogeneity larger than 2 have homogeneity $(5/2)-5\kappa$, from which the bound given in the statement follows at once. The claim concerning the signs of constants D_i is an immediate consequence of the assumption that $u(s,y)\geq \bar u(s,y)$ in $\{(s,y):|x-y|\leq \delta \,\&\, s\in(t-\delta,t]\}$. □

5. Construction of the Itô model.

The aim of this section is to obtain the bound (4.13) as well as the identity (4.12).

5.1. Estimate of the first term.

In order to motivate the technique of proof, we first show in detail as an example how one shows convergence of $\big(\hat\Pi_0^\varepsilon \tau\big)(\varphi_\lambda)$ to a limit.

5.1.1. Graphical notation.

Since the random variables $\big(\hat\Pi_0^\varepsilon\tau\big)(\varphi)$ belong to Wiener chaoses of finite order, they can be described in terms of their Wiener chaos decomposition. Similarly to what was

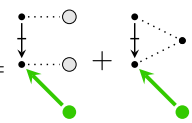
already done in [**Hai13**], [**Hai14**], we use a graphical notation to describe these random variables. A random variable belonging to the kth homogeneous Wiener chaos can be described by a kernel in $L^2(\boldsymbol{R}^2)^{\otimes k}$, i.e. the space of square-integrable functions in k space-time variables via the correspondence $f \mapsto I_k(f)$ given in [**Nua06**, p. 8]. Such a kernel will always be constructed from elementary kernels by multiplication and integration.

By translation invariance, we will only ever need to consider random variables of the type $\big(\Pi_0^\varepsilon \tau\big)(\varphi_\lambda)$, where φ_λ denotes a test function that is localised around the origin. Nodes in our graph will represent variables in $\boldsymbol{R}^2$, with one special node ● representing the origin. The nodes ○ represent the arguments of our kernel, so that a random variable in the kth (homogeneous) Wiener chaos is represented by a graph with exactly k such nodes. The remaining nodes, which we draw as •, represent dummy variables that are to be integrated out.

Each line then represents a kernel, with ⟶ representing the kernel K, ····· representing the kernel ϱ_ε, and ⟹ representing a generic test function φ_λ rescaled to scale λ.

Whenever we draw a barred arrow ⟼ this represents a factor $K(t-s, y-x) - K(-s,-x)$, where (s,x) and (t,y) are the coordinates of the starting and end point respectively. Finally, a double barred arrow ⟾ represents a factor $K(t-s, y-x) - K(-s,-x) - y\,K'(-s,-x)$.

With these graphical notations at hand, it follows for example from the recursive definition of $\Pi_0^{(\varepsilon)}$  combined with the contraction formula for the Wiener chaos decomposition of a product (see [**Nua06**, Proposition 1.1.2]), that one has the identity

$\big(\Pi_0^{(\varepsilon)}$ $\big)(\varphi_\lambda) =$  . (5.1)

In other words, it consists of the sum of one term belonging to the second homogeneous Wiener chaos and one term belonging to the zeroth chaos. (So this is just a real number.) The first term is described by a kernel $(z, \bar z) \mapsto \mathcal{W}^{(2;\varepsilon)}(z, \bar z)$ given by

$$\mathcal{W}^{(2;\varepsilon)}(z, \bar z) = \int \varrho_\varepsilon(z - z_1)\varrho_\varepsilon(\bar z - z_2)\big(K(z_2 - z_1) - K(-z_1)\big)\,\varphi_\lambda(z_2)\,dz_1\,dz_2,$$

while the constant term is given by $\int \mathcal{W}^{(2;\varepsilon)}(z, z)\,dz$.

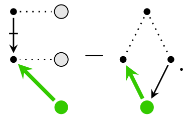
The problem with this is that the second term diverges as $\varepsilon \to 0$, so we would like our renormalisation procedure to cancel this term out. This is why our renormalisation map M was of the form M  = $- C_\varepsilon \mathbf{1}$ and it motivates the definition

$$C_\varepsilon = \int K(t,x)\varrho_\varepsilon^{\star 2}(t,x)\,dt\,dx = \text{  }.$$

For ε small enough, it follows from the scaling properties of the heat kernel that this

constant is indeed equal to $\varepsilon^{-1}c_\varrho$ with c_ϱ as defined in the introduction and already mentioned several times. Since one then has $(\hat\Pi_0^{(\varepsilon)} \circ\!\circ)(\varphi_\lambda) = (\Pi_0^{(\varepsilon)} \circ\!\circ)(\varphi_\lambda) - C_\varepsilon \int \varphi_\lambda$ as a consequence of (4.5), the renormalised model is given by

$$(\hat\Pi_0^{(\varepsilon)} \circ\!\circ)(\varphi_\lambda) = \quad (5.2)$$



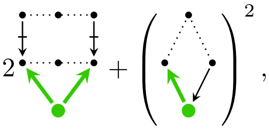
The reason why the second term is present in this expression is that the second term in (5.1) really consists of two terms since, as mentioned above, the arrow represents a factor $K(t-s, y-x) - K(-s,-x)$. The renormalisation term only cancels the first of these two terms.

In order to obtain the second bound in (4.13), we would now like to show that

$$\boldsymbol{E}\big|(\hat\Pi_0^{(\varepsilon)} \circ\!\circ)(\varphi_\lambda)\big|^2 \lesssim \lambda^{2|\circ\!\circ|+\kappa} = \lambda^{-2-3\kappa},$$

where we used the fact that $|\circ\!\circ| = -1-2\kappa$. For this, we use the fact that if X is a random variable belonging to the kth homogeneous Wiener chaos described by a kernel $\mathcal{W}$ with k variables, then one has the bound $\boldsymbol{E}X^2 \le k!\|\mathcal{W}\|^2$, where $\|\cdot\|$ denotes the L^2-norm. The reason why we do not have equality is that in general equality holds only when $\mathcal{W}$ is symmetrised under permutations of its k arguments, which is not the case here. Using furthermore the orthogonality of Wiener chaoses, we therefore deduce from (5.2) the bound

$$\boldsymbol{E}\big|(\hat\Pi_0^{(\varepsilon)} \circ\!\circ)(\varphi_\lambda)\big|^2 \le 2 \quad (5.3)$$



where we used the same graphical notations as before. Note that this is simply a real number (depending of course on the scale λ at which the test function is localised), as can be seen from the absence of free variables $\circ$.

The second term in (5.3) can be bounded with relative ease. Indeed, it suffices to note that it is equal to

$$\Big(\int \varphi_\lambda(z)\big(\varrho_\varepsilon^{\star 2} * K\big)(z)\,dz\Big)^2 .$$

This in turn is bounded by a multiple of λ^{-2} (uniformly in ε) as a consequence of [**Hai14**, Lemma 10.17], which is precisely the desired result.

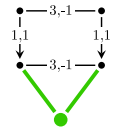
To bound the first term in (5.3), one realises that it is not really necessary to keep track of the precise form of the kernels represented by each edge. The only relevant information is their singular behaviour near the origin. There are however two additional

pieces of information that we need to keep track of. First, there is the fact that the two arrows really correspond to a difference between two kernels where one of them is evaluated at the root. Then, there is the fact that, although the dashed lines correspond to a kernel that has homogeneity -3 (if one wants a bound uniform in ε, then the best one can do is to bound the function $\varrho_\varepsilon(z)$ by $|z|^{-3}$ in parabolic space-time), we know for a fact that it is integrable even though such a homogeneity could in principle lead to a logarithmic divergence.

5.1.2. Bounds on labelled graphs.

One might then rewrite the bound (5.3) as follows:

$$\boldsymbol{E}\big|\big(\hat{\Pi}_0^{(\varepsilon)}\,\circ\!-\!\circ\big)(\varphi_\lambda)\big|^2 \lesssim$$



$$+\,\lambda^{-2}. \tag{5.4}$$

Here, each vertex v represents an integration variable x_v as before, except for the larger root which represents the origin. Edges $e = (v, \bar{v})$ are oriented and decorated with a label $(m_e, r_e) \in \boldsymbol{R} \times \boldsymbol{Z}$. The orientation of an edge matters only if $r_e > 0$.

If $r_e = 0$, then the corresponding edge represents a factor $\hat{J}_e(x_v, x_{\bar{v}}) = J_e(x_{\bar{v}} - x_v)$, where J_e is a smooth compactly supported function with a singularity of order m at the origin. See [**HQ15**] for a precise definition of what we mean by a singularity of order m. It suffices to know at this stage that ϱ_ε and $\varrho_\varepsilon^{\star 2}$ have a singularity of order 3 and K has a singularity of order 1.

If $r_e > 0$, then the corresponding edge represents a factor

$$\hat{J}_e(x_v, x_{\bar{v}}) = J_e(x_{\bar{v}} - x_v) - \sum_{|k|_{\mathfrak{s}} < r_e} \frac{x_{\bar{v}}^k}{k!} D^k J_e(-x_v).$$

Here we see why the orientation matters in this case. Previously, changing the orientation of the edge yields exactly the same factor, provided that we simultaneously change the function J_e into the function $x \mapsto J_e(-x)$. When $r > 0$, this is no longer the case.

The description of the edges with $r_e < 0$ is slightly more delicate. As before, we give ourselves a kernel J_e with a singularity of order m at the origin. This time however, J_e is not necessarily integrable, so we build from it the distribution

$$\big(\mathscr{R} J_e\big)(\varphi) = \int J_e(x) \Big(\varphi(x) - \sum_{|k|_{\mathfrak{s}} < |r_e|} \frac{x^k}{k!} D^k \varphi(0) \Big)\, dx + \sum_{|k|_{\mathfrak{s}} < |r_e|} \frac{I_e^{(k)}}{k!} D^k \varphi(0),$$

where the $I_e^{(k)}$ are some finite numbers that we also need to specify for such an edge. Provided that $m_e + r_e < 3$ (here 3 is the scaling dimension of parabolic space-time), this yields a well-posed distribution thanks to the fact that J_e is integrated against a function that vanishes to sufficiently high order at the origin. In our particular case, we only need to subtract the value of the test function φ at the origin. Furthermore, since

the kernel $\varrho_\varepsilon^{(2)}$ integrates to 1, we actually have that, in the particular case described by (5.4), $\mathscr{R}J_e = J_e$ if we choose $I_0 = 1$. With such a "renormalised kernel" at hand, we then set somewhat informally $\hat{J}_e(x_v, x_{\bar{v}}) = \big(\mathscr{R}J_e\big)(x_{\bar{v}} - x_v)$ in cases where $r < 0$.

There are furthermore two distinguished edges (represented in boldface) that necessarily connect to the origin, so they are always of the type $e = (v, 0)$, and that represent a factor $\hat{J}_e(x_v, x_0) = \varphi_\lambda(x_v - x_0)$. One should think of these edges as being decorated with the label $(0, 0)$ but since this is always the case we do not draw these labels.

In order to state our bounds, we will always denote such a graph by $(\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ is a set of directed edges for the vertex set $\mathcal{V}$. The distinguished "origin" is denoted by $0 \in \mathcal{V}$ and we use the notation $\mathcal{V}_0 = \mathcal{V} \setminus \{0\}$. We furthermore denote by $v_{\star,1}$ and $v_{\star,2}$ the two vertices that are connected to the origin by the distinguished edges and we set

$$\mathcal{V}_\star = \{0, v_{\star,1}, v_{\star,2}\}.$$

With all of these notations at hand, a labelled graph as above, together with the corresponding collection of kernels J_e and constants $I_e^{(k)}$ determines a number

$$\mathcal{I}_\lambda(J) \stackrel{\text{def}}{=} \int_{(\boldsymbol{R}^2)^{\mathcal{V}_0}} \prod_{e \in \mathcal{E}} \hat{J}_e(x_{e_-}, x_{e_+})\, dx, \tag{5.5}$$

where we also implicitly set $x_0 = 0$.

REMARK 5.1. At this stage, the careful reader may wonder what (5.5) actually means: some of the factors appearing there are distributions, so that this is not at all clear a priori. To clarify this, given any homogeneity $m \in \boldsymbol{R}$ and some $d > 0$, we define the following (semi)norm on the space of compactly supported functions that are smooth everywhere, except at the origin:

$$\|J\|_{m,d} = \sup_{|k|_{\mathfrak{s}} < d} \sup_{0 < |x|_{\mathfrak{s}} \le 1} |x|_{\mathfrak{s}}^{m+|k|_{\mathfrak{s}}} |D^k J(x)|.$$

Recall again that here x denotes a space-time point and $|x|_{\mathfrak{s}}$ denotes its parabolic norm. If we then replace each of the J_e by a smooth function $J_e^{(n)}$ such that $\|J_e - J_e^{(n)}\|_{m_e,d} \le 1/n$ for every e and define $\hat{J}_e^{(n)}$ from $J_e^{(n)}$ as above, then $\mathcal{I}_\lambda(J^{(n)})$ is well-defined. The quantity $\mathcal{I}_\lambda(J)$ is then defined as the limit of this quantity as $n \to 0$, provided that this limit exists and is independent of the approximating sequence for d sufficiently large. If the limit doesn't exist or depends on the approximating sequence, then we simply set $\mathcal{I}_\lambda(J) = \infty$.

There is a natural homogeneity associated to $\mathcal{I}_\lambda$ as follows. To each integration variable, we associate a homogeneity 3, which is the scaling dimension of parabolic space-time. To each factor $\hat{J}_e$ corresponding to a kernel of singularity m_e, we associate a homogeneity $-m_e$, except for the factors φ_λ which have homogeneity -3, and not 0. In other words, the total homogeneity of this expression is given by

$$\alpha = 3|\mathcal{V}_0| - 6 - \sum_{e \in \mathcal{E}} m_e = 3|\mathcal{V} \setminus \mathcal{V}_\star| - \sum_{e \in \mathcal{E}} m_e.$$

It is natural to guess that one then has $|\mathcal{I}_\lambda| \sim \lambda^\alpha$ for small values of λ. This is *not* the case in general! For example, it might happen that the integral in (5.5) does not even converge. Alternatively, it might happen that it converges, but the resulting expression has the "wrong" homogeneity.

In order to formulate the additional assumptions we will place on our labelled graph, we define, for any $\bar{\mathcal{V}} \subset \mathcal{V}$, the following subsets of $\mathcal{E}$:

$$\begin{aligned}
\mathcal{E}^\uparrow(\bar{\mathcal{V}}) &= \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} = e_- \ \& \ r_e > 0\},\\
\mathcal{E}^\downarrow(\bar{\mathcal{V}}) &= \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} = e_+ \ \& \ r_e > 0\},\\
\mathcal{E}_0(\bar{\mathcal{V}}) &= \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} = e\},\\
\mathcal{E}(\bar{\mathcal{V}}) &= \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} \neq \emptyset\}.
\end{aligned}$$

Here, we use the notation $e = (e_-, e_+)$ for a directed edge. Consider now the following assumption, where we use the shorthands $r_e^+ = (r_e \vee 0)$ and $r_e^- = -(r_e \wedge 0)$.

ASSUMPTION 5.2. *The labelled graph $(\mathcal{V}, \mathcal{E})$ satisfies the following properties.*

1. *For every edge $e \in \mathcal{E}$, one has $m_e - r_e^- < 3$.*
2. *For every subset $\bar{\mathcal{V}} \subset \mathcal{V}_0$ of cardinality at least 3, one has*

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{V}})} m_e < 3(|\bar{\mathcal{V}}| - 1). \tag{5.6}$$

3. *For every subset $\bar{\mathcal{V}} \subset \mathcal{V}$ containing 0 and of cardinality at least 2, one has*

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{V}})} m_e + \sum_{e \in \mathcal{E}^\uparrow(\bar{\mathcal{V}})} (m_e + r_e - 1) - \sum_{e \in \mathcal{E}^\downarrow(\bar{\mathcal{V}})} r_e < 3(|\bar{\mathcal{V}}| - 1). \tag{5.7}$$

4. *For every non-empty subset $\bar{\mathcal{V}} \subset \mathcal{V} \setminus \mathcal{V}_\star$, one has the bounds*

$$\sum_{e \in \mathcal{E}(\bar{\mathcal{V}}) \setminus \mathcal{E}^\downarrow(\bar{\mathcal{V}})} m_e + \sum_{e \in \mathcal{E}^\uparrow(\bar{\mathcal{V}})} r_e - \sum_{e \in \mathcal{E}^\downarrow(\bar{\mathcal{V}})} (r_e - 1) > 3|\bar{\mathcal{V}}|. \tag{5.8}$$

It turns out that Assumption 5.2 is sufficient to guarantee that the quantity $\mathcal{I}_\lambda$ does indeed have the correct scaling behaviour for small values of λ. This is the content of the following theorem, the proof of which can be found in [**HQ15**].

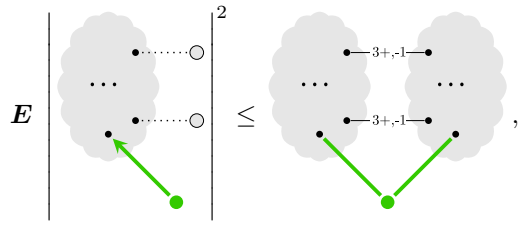
THEOREM 5.3. *Provided that Assumption 5.2 holds, there exists $d > 0$ and a constant C depending only on the number of vertices in $\mathcal{V}$ and on the values of the constants I_e, such that*

$$|\mathcal{I}_\lambda(J)| \le C\lambda^\alpha \prod_{e \in \mathcal{E}} \|J_e\|_{m_e, d}, \qquad \lambda \in (0, 1],$$

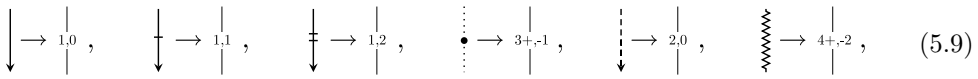
where $\alpha = 3|\mathcal{V} \setminus \mathcal{V}_\star| - \sum_{e\in\mathcal{E}} m_e$.

5.2. Construction of labelled graphs.

In our case, we will always consider the situation where the labelled graph $(\mathcal{V}, \mathcal{E})$ is built from a "half graph" in the following way:



where one furthermore performs the substitutions



(5.9)

and the bold green edges denote the distinguished edges $(v_{\star,i}, 0)$. Here, ----→ denotes the spatial derivative of the kernel K and ∿∿∿ denotes the kernel $\mathscr{R}Q_\varepsilon$ appearing in (5.14) below. The rationale for this is the fact that, for any $d > 0$, one has

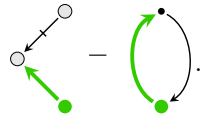
$$\|K\|_{1,d} + \|K'\|_{2,d} < \infty, \quad \sup_{\varepsilon\in(0,1]} \|\varrho_\varepsilon^{(2)}\|_{3,d} < \infty, \quad \sup_{\varepsilon\in(0,1]} \varepsilon^{-\kappa}\|\varrho_\varepsilon^{(2)}\|_{3+\kappa,d} < \infty, \tag{5.10}$$

for every $\kappa \in (0, 1)$.

5.2.1. Estimating the first term.

Let us now return to our proof of convergence for $\hat{\Pi}_0^{(\varepsilon)}$  . It is straightforward to verify by inspection that the graph appearing in (5.4) does indeed satisfy the assumptions of Theorem 5.3, so that one has a bound of the type $\boldsymbol{E}\big|\big(\hat{\Pi}_0^{(\varepsilon)}$ $\big)(\varphi_\lambda)\big|^2 \lesssim \lambda^{-2}$, uniformly over $\varepsilon, \lambda \in (0, 1]$.

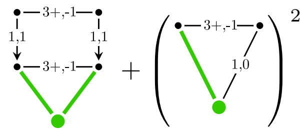
We now claim that once a bound of this type has been obtained, one automatically obtains convergence to a limiting random variable $\hat{\Pi}_0\tau$. Indeed, simply define $\hat{\Pi}_0\tau$ in the same way as $\hat{\Pi}_0^{(\varepsilon)}\tau$, but with all edges representing ϱ_ε removed. In our case, this yields the identity

$$\big(\hat{\Pi}_0$$ $$\big)(\varphi_\lambda) \overset{\text{def}}{=}$$  . (5.11)

(As a matter of fact, this definition only yields a random variable $\big(\hat{\Pi}_0$ $\big)(\varphi_\lambda)$ for every test function φ_λ. The fact that there exists a model-valued random variable $\hat{\Pi}_0$ such that

$(\hat\Pi_0\tau)(\varphi_\lambda)$ agrees with this almost surely will follow from [**Hai14**, Theorem 10.7], once all the relevant bounds have been obtained.) We then note that $(\hat\Pi_0^{(\varepsilon)} - \hat\Pi_0)(\varphi_\lambda)$ can be decomposed as a sum of terms, each of them looking like (5.2), but with some of the edges representing ϱ_ε now representing δ and exactly one of these edges representing $\varrho_\varepsilon - \delta$. As a consequence of the last bound in (5.10), we then obtain immediately the bound

$$\varepsilon^{-\kappa}\boldsymbol{E}\big|(\hat\Pi_0^{(\varepsilon)} - \hat\Pi_0)(\varphi_\lambda)\big|^2 \lesssim$$



$$\lesssim \lambda^{-2-2\kappa}, \tag{5.12}$$

where we wrote 3+ as a shorthand for $3+\kappa$. Noting that $2|\,\cdot\,| = -2-4\kappa$, this does indeed imply the bound required in (4.13) for the particular case $\tau =$. The remainder of this section is devoted to the proof of this bound for the remaining symbols τ with $|\tau| < 0$, see the list (4.1).

REMARK 5.4. Note that in general one does *not* have the identity

$$(\hat\Pi_0\,)(\varphi_\lambda) = (\hat\Pi_0\, \diamond \hat\Pi_0\,)(\varphi_\lambda), \tag{5.13}$$

with $\diamond$ the Wick product in the sense of white noise analysis. The discrepancy between the two expressions is the second term in (5.11). If however the support of the test function is located in the future, then (5.13) and therefore (4.12) (with $\tau =$) *does* hold thanks to the fact that the second term in (5.11) vanishes in that case. This is an immediate consequence of the non-anticipative nature of the kernel K.

5.3. Proof of Theorem 4.5.

We now proceed to give the proof of Theorem 4.5 for any symbol τ with $|\tau| < 0$. The proof is always essentially the same, so we only give the main steps.

5.3.1. Convergence for the symbols , , and .

In the preceding subsection, we have shown in detail that there exists a random distribution $\hat\Pi_0$ such that

$$\boldsymbol{E}\big|(\hat\Pi_0\,)(\varphi^\lambda)\big|^2 \lesssim \lambda^{2|\,\cdot\,|+\kappa}, \qquad \boldsymbol{E}\big|(\hat\Pi_0 - \hat\Pi_0^{(\varepsilon)})(\varphi^\lambda)\big|^2 \lesssim \varepsilon^\kappa\lambda^{2|\,\cdot\,|+\kappa}.$$

We have furthermore shown that the identity (5.13) holds, which indeed yields (4.12) for $\tau =$. Since $(\hat\Pi_0^{(\varepsilon)}\,)(\varphi) = (\hat\Pi_0^{(\varepsilon)}\,)(\tilde\varphi)$ with $\tilde\varphi(t,x) = x\varphi(t,x)$, the required properties and bounds for follow immediately from those for .

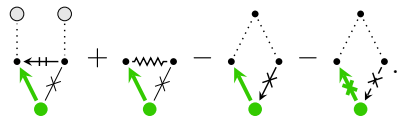
At this stage, we introduce the kernel Q_ε given by

$$Q_\varepsilon(z) = K(z)\,\varrho_\varepsilon^{(2)}(z),$$

and we use the notation ⟿ for the renormalised kernel

$$\mathscr{R}Q_\varepsilon(z) = K(z)\,\varrho_\varepsilon^{(2)}(z) - C_\varepsilon\delta_0(z). \tag{5.14}$$

Note that C_ε is precisely the integral of Q_ε and Q_ε is an even function of the spatial variable, so that $\mathscr{R}Q_\varepsilon$ annihilates every polynomial of parabolic degree strictly less than 2. Regarding  , we use these notations to obtain the identity

$\big(\hat{\Pi}_0^{(\varepsilon)}$  $\big)(\varphi_\lambda) =$ 

(5.15)

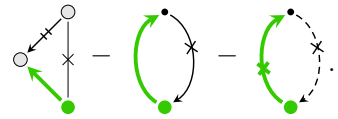
Here, we used notation —×— for the kernel $(t,x) \mapsto x$ (which is of homogeneity $+1$), →×⟶ for the kernel $(t,x) \mapsto xK(t,x)$, and similarly for - -×- ->. Note that while the function $(t,x) \mapsto x$ is of course not of compact support, we can replace it by a compactly supported function independent of ε without changing the values of these integrals. We are therefore back in the context of Theorem 5.3 and it is indeed possible to verify that each of these terms satisfies Assumption 5.2.

REMARK 5.5. Here and below, Assumption 5.2 can be verified "by hand", but this soon becomes rather tedious. The interested reader will find a small computer program at the URL http://www.hairer.org/paper/Trees.zip which verifies that Assumption 5.2 does indeed hold for all the graphs for which we make such a claim in this article.

Since one has

$$\|\mathscr{R}Q_\varepsilon\|_{-4-\kappa} \lesssim \varepsilon^\kappa, \tag{5.16}$$

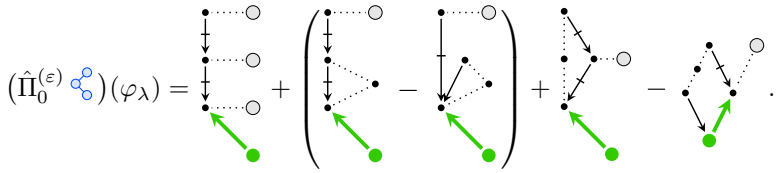
the term including this kernel vanishes in the limit. It follows that one has the identity

$\big(\hat{\Pi}_0$ $\big)(\varphi_\lambda) =$ 

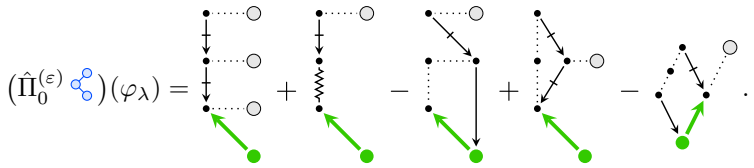
As before, the last two terms vanish if the test function is supported in the future. The first term on the other hand is easily seen to be equal to $\big(\hat{\Pi}_0$  $\big)(\varphi_\lambda) \diamond \big(\hat{\Pi}_0$  $\big)(\varphi_\lambda)$, which yields (4.12) in this case.

5.3.2. Convergence for the symbol  .

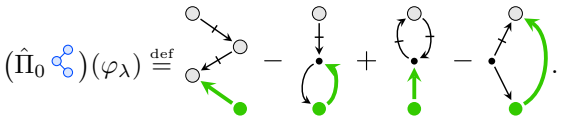
Regarding the symbol  , we combine (3.11c) with the expression (4.6) for Δ  and the expression (4.7) for Δ^M to obtain similarly to before



With the above notation for $\mathscr{R}Q_\varepsilon$, one then has



As before, one can verify that each of these terms separately satisfies Assumption 5.2 (after associating to them a labelled graph via the procedure outlined in Section 5.2), so that they satisfy the bounds (4.13). Since furthermore $\mathscr{R}Q_\varepsilon \to 0$ and $\varrho_\varepsilon \to \delta$ in the distributional sense as $\varepsilon \to 0$, one obtains in the limit



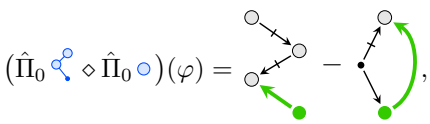
(5.17)

One might think that the penultimate term in this expression vanishes since the kernel K is non-anticipative. This would indeed be the case if this term were equal to  . In our case however, it does not vanish in general, unless the test function is supported in the future.

It remains to show (4.12), namely that if the test function φ has support located in the future, then

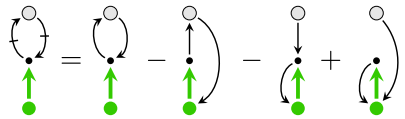
$$\big(\hat{\Pi}_0 \, \big)(\varphi) = \big(\hat{\Pi}_0 \, \diamond \, \hat{\Pi}_0 \, \big)(\varphi).$$

For this, we note that, as a consequence of (3.11), $\big(\hat{\Pi}_0$  $\big)(\varphi)$ is obtained from $\big(\hat{\Pi}_0$ $\big)(\varphi)$ by simply replacing  by  . Taking the Wick product with $\hat{\Pi}_0$ (which is nothing but the underlying white noise ξ) then has the effect of simply further replacing  by . In other words, we obtain



so that the difference between $\big(\hat{\Pi}_0$ $\big)(\varphi_\lambda)$ and $\big(\hat{\Pi}_0$ $\diamond \hat{\Pi}_0$ $\big)(\varphi_\lambda)$ is given by the second

and third terms in (5.17). As before, the second term vanishes if the test function φ_λ is supported in the future. The reason why this is also true for the third term is as follows. By definition, one has the identity

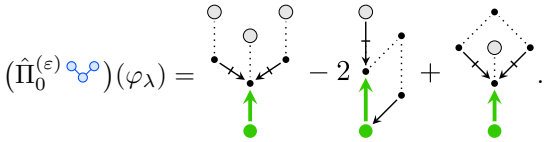


(5.18)

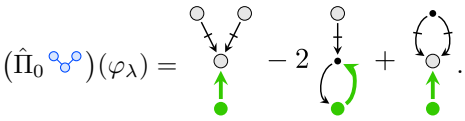
We now see that each of these terms contains a closed loop with all arrows pointing in the same direction. Since each of these arrows depicts either K or φ, both of which are supported in the future, this implies that the corresponding integrands vanish identically.

5.3.3. Convergence for the symbol .

We now turn to  . In this case, we obtain from (3.11c) and (4.6) the identity

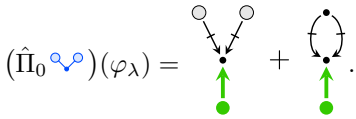


Performing the substitutions (5.9), it is straightforward to verify that the labelled graphs arising from these three expressions from the procedure outlined in Section 5.2 all satisfy Assumption 5.2, so that the required bound (4.13) holds. It follows that the limit as $\varepsilon \to 0$ is given by



(5.19)

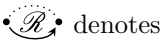
Similarly, one sees that

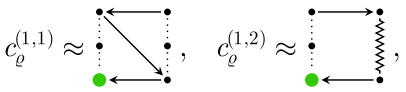


It follows that the difference between $\big(\hat{\Pi}_0$ $\big)(\varphi_\lambda)$ and $\big(\hat{\Pi}_0$ $\diamond\, \hat{\Pi}_0$ $\big)(\varphi_\lambda)$ is given by the second term in (5.19), which again vanishes if φ_λ is supported in the future.

5.3.4. Convergence for the symbol  .

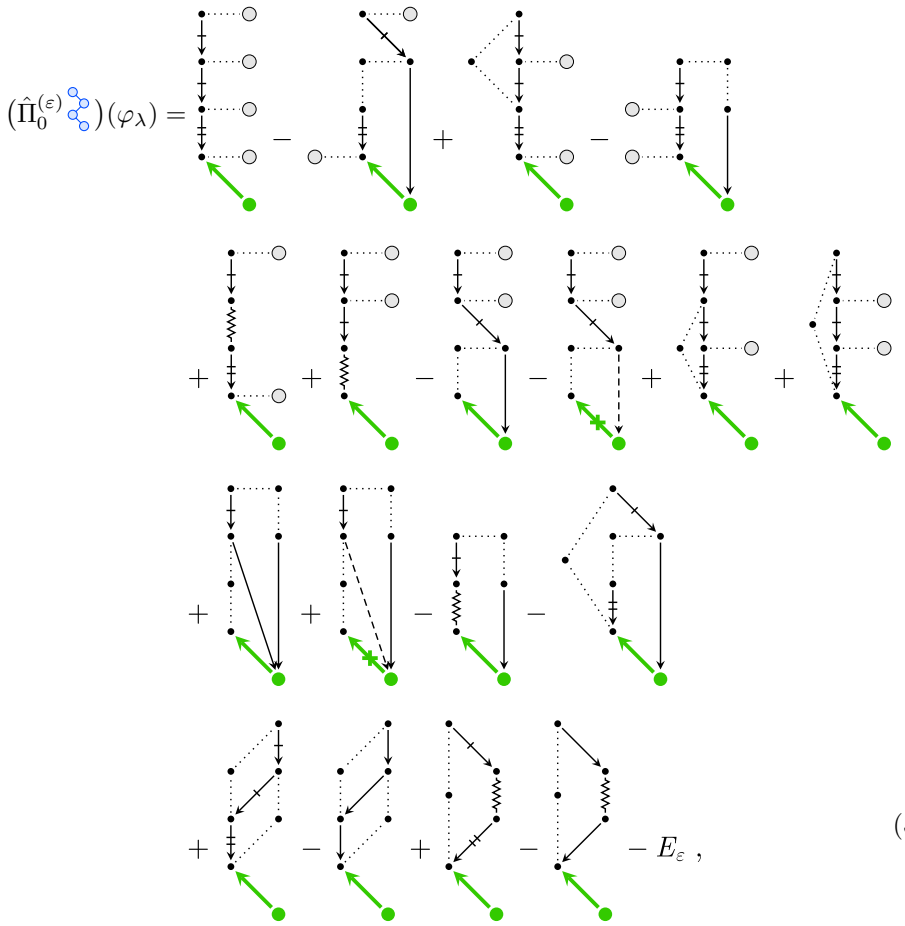
The renormalisation of this term involves the two constants $c_\varrho^{(1,1)}$ and $c_\varrho^{(1,2)}$. We note here that the notations used in (2.2) are slightly inconsistent from the ones employed here: dotted lines there denote the convolution of ϱ with itself instead of ϱ_ε, arrows

denote the heat kernel P instead of K, and  denotes the kernel $\mathscr{R}Q_1$ (i.e. $\mathscr{R}Q_\varepsilon$ with $\varepsilon = 1$). It is however straightforward to verify that this integral is invariant under rescaling of the variables by a factor ε so that, with current pictorial notations, one has



which would actually be identities if the arrows denoted the heat kernel without truncation. It is then a consequence of the convergence shown in Section 2.1 that the error E_ε implicit in the $\approx$ signs appearing in these expressions converges to 0 as $\varepsilon \to 0$.

Assume from now on without loss of generality that $\int \varphi^\lambda(z)\,dz = 1$. Combining this with the recursive definition of $\hat{\Pi}_0^{(\varepsilon)}$  , we then obtain

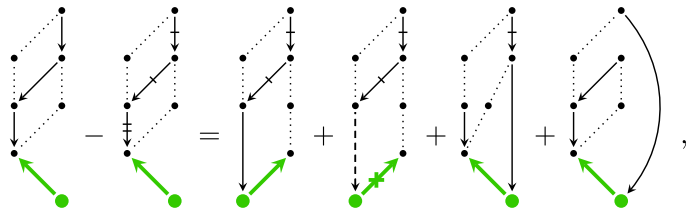


(5.20)

where we used the shorthand ----> for the kernel $K' = \partial_x K$ and  for the test function $(t,x) \mapsto x\varphi^\lambda(t,x)$. Note that if we set $\tilde{\varphi}(t,x) = x\varphi(t,x)$, then $\tilde{\varphi}$ is again an

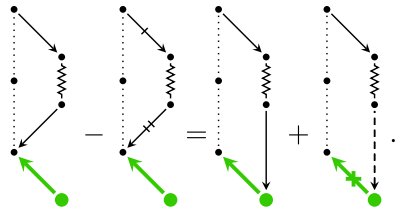
admissible test function and one has $x\varphi^\lambda(t,x) = \lambda\tilde\varphi^\lambda(t,x)$. As a consequence, when applying Theorem 5.3 to a graph with test function $\tilde\varphi$, one gains an additional power of λ. This however is exactly compensated by the fact that in this case one instance of the kernel K is replaced by K', thus lowering the total homogeneity of the graph by one.

It is a lengthy but straightforward task to verify that each of the terms appearing on the first three lines verify Assumption 5.2 when performing the "doubling" procedure of Section 5.2 and the substitutions (5.9) to turn them into labelled graphs, so that the required bounds hold for them. In order to obtain analogous bounds for the terms on the last line, one needs to exploit the fact that they create cancellations. More precisely, we rewrite these terms as



(5.21)

as well as



(5.22)

It is then possible to verify that each of these terms appearing in the right hand side of (5.21) as well as the first term appearing in the right hand side of (5.22) all separately give rise to graphs satisfying Assumption 5.2. The second term appearing on the right hand side of (5.22) fails Condition 3, but it can easily be dealt with "by hand": it simply consists of a convolution of (renormalized) kernels of respective homogeneities -3, -4 and -2, tested against the test function $\lambda\tilde\varphi^\lambda$. These can easily be bounded by repeatedly applying [**Hai14**, Lemma 3.14–3.16].

We now verify that one has

$$\big(\hat\Pi_0 \;\big)(\varphi) = \big(\hat\Pi_0 \;\diamond\, \hat\Pi_0 \;\big)(\varphi),$$

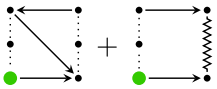
for φ with support in $\boldsymbol{R}_+ \times S^1$. Similarly to before, the right hand side is obtained from $\big(\hat\Pi_0 \;\big)(\varphi)$ by replacing ⟵ with ⟶○⟵ in its pictorial representation. It then follows immediately from (5.17) that this yields precisely the first four terms in (5.20). The first three terms on the second line of (5.20) contain a factor ⟿ and satisfy the assumptions of Theorem 5.3, so they vanish as $\varepsilon \to 0$ for any test function as a consequence of (5.16). It remains to show that all remaining terms vanish as $\varepsilon \to 0$ when

the test function φ is supported in the future. For the remaining terms on the second and third lines, this can be verified by a systematic use of the argument following (5.18). Regarding the two differences appearing on the last line, they can again be treated by the same arguments. Note however that it is crucial here to make use of the cancellations appearing there: the individual terms on the last line do in general *not* converge to 0!

5.3.5. Convergence for the symbol .

In this case, we note that similarly to the previous case one can write

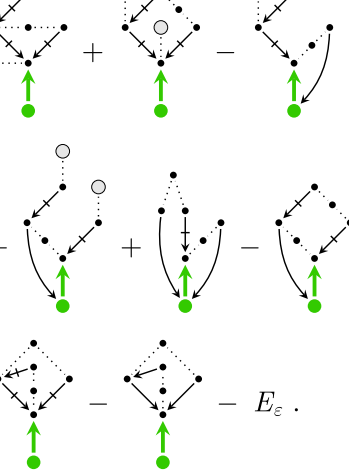
$$c_\varrho^{(2,1)} + c_\varrho^{(2,2)} =$$



$$+ E_\varepsilon \,,$$

for some error E_ε with $\lim_{\varepsilon\to 0} E_\varepsilon = 0$. With this identity at hand, we then obtain

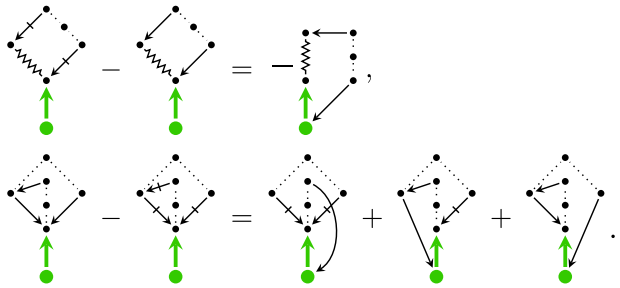
$$\big(\hat\Pi_0^{(\varepsilon)}$$



$$\big)(\varphi_\lambda) =$$

$$- E_\varepsilon \,. \tag{5.23}$$

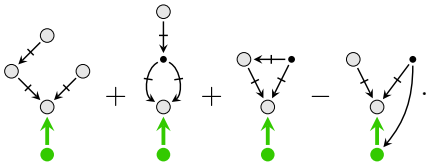
Again, each term appearing on the first two lines in this expression gives rise to a labelled graph satisfying Assumption 5.2. The terms appearing on the last line require us again to make use of cancellations similarly to what we did for the last two terms appearing in the expression for $\big(\hat\Pi_0^{(\varepsilon)}$ $\big)(\varphi_\lambda)$. This time, we use the identities



(5.24)

Similarly to before, the term appearing on the first line does not satisfy the assumptions of Theorem 5.3 but can again be dealt with by repeatedly invoking [**Hai14**, Lemma 3.14–3.16]. All the terms appearing on the right hand side of the second line on the other hand do give rise to labelled graphs satisfying Assumption 5.2.

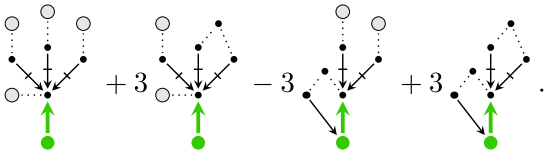
If the test function φ is supported in the future, then the limit of the right hand side of (5.23) as $\varepsilon \to 0$ is given by

$(\hat{\Pi}_0$  $)(\varphi_\lambda) =$ 

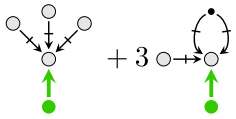
As before, it is now a straightforward task to verify that this is indeed equal to $(\hat{\Pi}_0$  $\diamond$ $\hat{\Pi}_0$ $)(\varphi_\lambda)$ as required.

5.3.6. Convergence for the symbol  .

This time, one obtains the identity

$(\hat{\Pi}_0^{(\varepsilon)}$  $)(\varphi_\lambda) =$ 

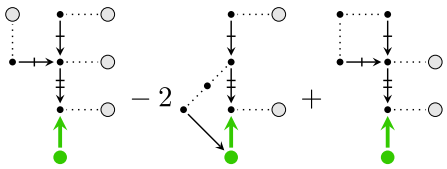
This time, each term generates a graph satisfying Assumption 5.2, so that the required bounds and convergence hold. When testing against a test function that is supported in the future, the limit as $\varepsilon \to 0$ of this expression becomes

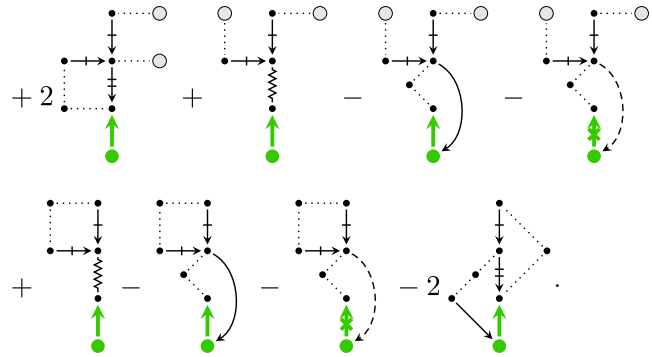
$(\hat{\Pi}_0^{(\varepsilon)}$ $)(\varphi) =$ 

As before, one can easily verify that this is equal to $(\hat{\Pi}_0$  $\diamond \hat{\Pi}_0$  $)(\varphi_\lambda)$, so that (4.12) is verified.

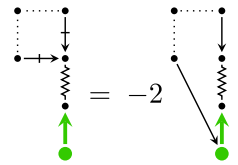
5.3.7. Convergence for the symbol  .

For this last term, we have

$(\hat{\Pi}_0^{(\varepsilon)}$  $)(\varphi_\lambda) =$ 



It is again a lengthy but ultimately straightforward task to verify that each term yields a graph satisfying Assumption 5.2, so that the required bounds and convergence hold, except for the first term appearing on the last line. This term however can be rewritten as



thus showing that it is identical to the term appearing on the first line of (5.24) and has already been dealt with.

An argument very similar to the ones given previously shows that, if we test $\hat{\Pi}_0^{(\varepsilon)}$  agains a test function φ supported in the future and take the limit $\varepsilon \to 0$, we do again obtain the identity (4.12). This finally concludes the proof of Theorem 4.5.

6. Identification of the limit.

We now show that, when using the Itô model constructed in Theorem 4.5, the corresponding solutions constructed in Theorem 3.10 coincide with the classical Itô solutions to the nonlinear stochastic heat equation. Let us first state the following fact.

LEMMA 6.1. *Let $\eta \in \mathcal{C}^\alpha(\boldsymbol{R} \times S^1)$ for $\alpha \in (-2, 0)$ and let $[s,t] \subset \boldsymbol{R}$. Then, there exists a unique distribution $\mathbf{1}_{[s,t]}\eta \in \mathcal{C}^\alpha$ such that*

$$\big(\mathbf{1}_{[s,t]}\eta\big)(\varphi) = \eta(\varphi),$$

for smooth test functions φ such that $\operatorname{supp}\varphi \subset [s,t] \times S^1$, and such that furthermore $\big(\mathbf{1}_{[s,t]}\eta\big)(\varphi) = 0$ if $\operatorname{supp}\varphi \cap [s,t] \times S^1 = \emptyset$. The map $\eta \mapsto \mathbf{1}_{[s,t]}\eta$ is continuous in $\mathcal{C}^\alpha$.

PROOF. In the Euclidean case this is known, see for example [**RS96**, Chapter 4.6.3]. In the more general parabolic case considered here, it is an easy corollary of [**Hai14**, Proposition 6.9]. (Noting that the "effective codimension" of the hyperplane

$\{t=0\}$ is 2 in the parabolic case and 1 in the Euclidean case.) □

With the help of this lemma, we are able to formulate the following result, which is the main ingredient in identifying our solution with the classical Itô solution. Here we use again the notation $\mathcal{T}_{\mathcal{U}}$ introduced in (3.4).

THEOREM 6.2. *Let $(\hat{\Pi}, \hat{F})$ be the Itô model built in Theorem 4.5 and let V be a $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$-valued random variable for some $\gamma > (3/2) + \kappa$ and $\eta \geq 0$, and let $T > 0$ be a bounded stopping time. Assume that the stochastic process $V\colon (t,x) \mapsto V(t,x) \in \mathcal{T}_{\mathcal{U}}$ is adapted to the filtration $\{\mathcal{F}_s\}_{s\in\boldsymbol{R}}$ generated by the underlying space-time white noise and that there exists some $p > 2$ such that $\boldsymbol{E}\|V\|^p_{\gamma,\eta;\mathfrak{K}} < \infty$ for the compact set $\mathfrak{K} = [0,T]\times S^1$.*

Denote by $\mathcal{R}$ the reconstruction operator associated to $(\hat{\Pi}, \hat{F})$. Then, for any smooth function $\psi\colon \boldsymbol{R}_+ \times S^1 \to \boldsymbol{R}$ with $\operatorname{supp}\psi \subset (0,\infty)\times S^1$, one has

$$\big(\mathbf{1}_{[0,T]}\mathcal{R}(V\Xi)\big)(\psi) = \int_0^T \big\langle v(t,\cdot)\psi(t,\cdot), dW(t)\big\rangle,$$

where $v(t,x) = \big\langle \mathbf{1}, V(t,x)\big\rangle$ denotes the component of V in the subspace spanned by $\mathbf{1}$. Here, $V\Xi$ denotes the modelled distribution given by $(V\Xi)(t,x) = V(t,x)\Xi$.

REMARK 6.3. Note that the space $\mathcal{D}^{\gamma,\eta}_{\mathcal{U}}$ itself is also random in this statement! The statement is somewhat surprising since it shows that in this situation $\mathcal{R}_T(V\Xi)$ only depends on v and not on any of the higher-order coefficients, which is certainly not the case in general.

REMARK 6.4. In practice, we will apply this theorem to the case $V = \hat{G}(U)$, where U is the solution to the fixed point equation (3.22).

PROOF OF THEOREM 6.2. Note first that, as a consequence of the canonical inclusion $\mathcal{D}^{\gamma,\eta} \subset \mathcal{D}^{\gamma',\eta}$ for $\gamma' < \gamma$ (obtained by setting to 0 all components in $\mathcal{T}_\alpha$ with $\alpha \in [\gamma', \gamma)$), we can assume without loss of generality that $\gamma < 7/4$.

Every centered square integrable $\mathcal{F}$-measurable random variable can be represented as a stochastic integral against W, as is shown in Lemma 1.1 from [**CW77**]. Hence it suffices to show that for any smooth and compactly supported test function ψ and for any adapted and uniformly Lipschitz continuous process Ψ with compact support, one has the identity

$$\boldsymbol{E}\left(\big(\mathbf{1}_{[0,T]}\mathcal{R}V\Xi\big)(\psi)\int_0^\infty \big\langle \Psi(t), dW(t)\big\rangle\right) = \boldsymbol{E}\big\langle v\psi, \Psi\big\rangle_T, \tag{6.1}$$

where $\langle\cdot,\cdot\rangle_T$ denotes the scalar product in $L^2([0,T]\times S^1)$.

We rely on the fact that, by [**Hai14**, Theorem 3.23] and Lemma 6.1, for any modelled distribution $f \in \mathcal{D}^{\bar\gamma,\bar\eta}_\Xi$ (the space of elements in $\mathcal{D}^{\bar\gamma,\bar\eta}$ with values in $\mathcal{T}_\Xi$) with $\bar\gamma > 0$, $\bar\eta > -2$, and any smooth test function ψ supported in a compact set $\mathfrak{K}$ not intersecting $\{(t,x) : t = 0\}$, one has the bound

$$\big|\big(\mathbf{1}_{[0,T]}\mathcal{R}f\big)(\psi) - \big\langle \mathbf{1}_{[0,T]}\mathcal{R}^n f, \psi\big\rangle_T\big| \lesssim C(\psi)2^{-\alpha n}\|f\|_{\bar\gamma,\bar\eta;\mathfrak{K}}\big(|\!|\!|\hat\Pi|\!|\!|_{\bar\gamma} + |\!|\!|\hat\Gamma|\!|\!|_{\bar\gamma}\big), \tag{6.2}$$

where α is any exponent with $\alpha \in (0,\bar\gamma)$, $|\!|\!|\hat\Gamma|\!|\!|_{\bar\gamma}$ was defined in (3.24),

$$|\!|\!|\hat\Pi|\!|\!|_{\bar\gamma} = \sup_{z,\lambda,\varphi}\sup_{|\tau|<\bar\gamma} \lambda^{-|\tau|}|(\Pi_z\tau)(\varphi_z^\lambda)|,$$

(with the supremum over z, φ and λ as in (3.15)) and where the sequence of functions $\mathbf{1}_{[0,T]}\mathcal{R}^n f$ is given by

$$\big(\mathbf{1}_{[0,T]}\mathcal{R}^n f\big)(z) = \mathbf{1}_{[0,T]}(t)\sum_{\bar z\in\Lambda_{\mathfrak{s}}^n(T)} \big(\hat\Pi_{\bar z} f(\bar z)\big)(\varphi_{\bar z}^n)\,\varphi_{\bar z}^n(z), \tag{6.3}$$

where t is the time coordinate of z. Here, $\Lambda_{\mathfrak{s}}^n(T)$ denotes the dyadic grid on $[0,T]\times S^1$, φ is the scale function of some sufficiently smooth multiresolution analysis and, writing $z=(t,x)$, $\bar z=(\bar t,\bar x)$, one sets

$$\varphi_{\bar z}^n(z) = 2^{3n/2}\varphi(2^n(x-\bar x))\varphi(2^{2n}(t-\bar t)).$$

(See also [**Hai14**, Section 3.3] for more details.) Our argument relies on the fact that the choice of multiresolution analysis in this construction is arbitrary. In particular, we can ensure that the support of φ is contained in the interval $[0,K]$ for some $K>0$ and we make such a choice from now on.

Using (6.3) with $f = V\Xi$ and then applying (4.12), we now have the identity

$$\begin{aligned}\big\langle \mathbf{1}_{[0,T]}\mathcal{R}^n V\Xi, \psi\big\rangle &= \sum_{z\in\Lambda_{\mathfrak{s}}^n}\big(\hat\Pi_z\Xi V(\bar z)\big)(\varphi_z^n)\,\big\langle\varphi_z^n,\psi\big\rangle_T \\ &= \sum_{z\in\Lambda_{\mathfrak{s}}^n}\big\langle\varphi_z^n,\psi\big\rangle_T\int_t^\infty\big\langle\big(\hat\Pi_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), dW(s)\big\rangle,\end{aligned}$$

where we wrote $\Lambda_{\mathfrak{s}}^n$ for the dyadic grid on all of $\boldsymbol{R}_+\times S^1$. Note now that if either $t\le T-K2^{-2n}$ or $t\ge T$, we have the identity

$$\begin{aligned}&\big\langle\varphi_z^n,\psi\big\rangle_T\int_t^\infty\big\langle\big(\hat\Pi_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), dW(s)\big\rangle \\ &\quad= \big\langle\varphi_z^n,\psi\big\rangle_T\int_t^T\big\langle\big(\hat\Pi_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), dW(s)\big\rangle.\end{aligned}$$

As a consequence, we can write

$$\big\langle \mathbf{1}_{[0,T]}\mathcal{R}^n V\Xi, \psi\big\rangle = \sum_{z\in\Lambda_{\mathfrak{s}}^n}\big\langle\varphi_z^n,\psi\big\rangle_T\int_t^T\big\langle\big(\hat\Pi_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), dW(s)\big\rangle + R_n,$$

where R_n is a sum of the order of 2^n terms, each of which is bounded (in squared expectation) by a multiple of $2^{-3n/2}$, so that $\boldsymbol{E}|R_n|^2 \lesssim 2^{-n}$.

Combining this with (6.2), we have now shown that there exists $\alpha > 0$ such that

$$\big(\mathbf{1}_{[0,T]}\mathcal{R}V\Xi\big)(\psi) = \sum_{z\in\Lambda_{\mathfrak{s}}^n} \big\langle \varphi_z^n, \psi\big\rangle_T \int_t^T \big\langle \big(\hat{\Pi}_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), dW(s)\big\rangle + \tilde{R}_n,$$

where $\tilde{R}_n$ satisfies $\boldsymbol{E}|\tilde{R}_n|^2 \lesssim 2^{-\alpha n}$. Consequently, using Itô's isometry, the left hand side of (6.1) equals the limit, as $n \to \infty$, of

$$\begin{aligned}\boldsymbol{E} \sum_{z\in\Lambda_{\mathfrak{s}}^n} \big\langle \varphi_z^n, \psi\big\rangle_T \int_t^T \big\langle \big(\hat{\Pi}_z V(z)\big)(s,\cdot)\varphi_z^n(s,\cdot), \Psi(s,\cdot)\big\rangle\, ds \\ = \boldsymbol{E} \sum_{z\in\Lambda_{\mathfrak{s}}^n} \big\langle \varphi_z^n, \psi\big\rangle_T \big\langle \big(\hat{\Pi}_z V(z)\big)(\cdot)\varphi_z^n, \Psi\big\rangle_T. \end{aligned} \tag{6.4}$$

At this stage we use the fact that, as a consequence of [**Hai14**, Proposition 3.28], there exists an exponent $\alpha > 0$ such that

$$\big|\big(\Pi_z V(z)\big)(\hat{z}) - v(z)\big| \le C|z-\hat{z}|_{\mathfrak{s}}^{\alpha} \|V\|_{\gamma;\mathfrak{K}} |\!|\!|\Gamma|\!|\!|_{\gamma},$$

where $\mathfrak{K}$ is some compact set containing z and $\hat{z}$. Inserting this into (6.4) and exploiting the fact that $\{\varphi_z^n\}_{z\in\Lambda_{\mathfrak{s}}^n}$ is an approximate resolution of the identity, it is now straightforward to verify that the limit of (6.4) as $n \to \infty$ equals the right hand side of (6.1) as required. □

As a corollary of this result, it is now straightforward to obtain the announced result, namely that the solution to our abstract fixed point problem driven by the Itô model is nothing but the classical weak solution to (1.2) interpreted in the Itô sense.

COROLLARY 6.5. *Let $(\hat{\Pi}, \hat{F})$ be the Itô model built in Theorem* 4.5, *let $\eta \in (0, (1/2) - \kappa)$, let $M > 0$, and let $U \in \mathcal{D}_{\mathcal{U}}^{\gamma,0}$ be the solution to the fixed point problem* (3.22) *given by Theorem* 3.10 *with respect to $(\hat{\Pi}, \hat{F})$, up to the first* (*stopping*) *time τ where $u = \mathcal{R}U$ satisfies $\|u(\tau,\cdot)\|_{\mathcal{C}^\eta} \ge M$.*

Then, the function u on $[0,\tau]$ coincides almost surely with the unique weak local solution to (1.2).

PROOF. Applying the reconstruction operator to both sides of (3.22) and using the equivalence between weak and mild solutions for the heat equation with a distributional inhomogeneity (see the proof of Theorem 3.2 in [**Wal86**]), we deduce that u satisfies the distributional identity

$$\partial_t u = \partial_x^2 u + H(u) + \mathcal{R}\big(\hat{G}(U)\Xi\big),$$

on $(0,\tau) \times S^1$. Since we know that $u \in \mathcal{C}^\eta$, each term appearing in this identity belongs

to $\mathcal{C}^\alpha$ for some $\alpha > -2$, so we can use again Lemma 6.1 to multiply this identity with $\mathbf{1}_{[0,t]}$ for any stopping time $t \le \tau$. Testing against a smooth test function ψ of x and applying Theorem 6.2, we thus obtain the identity

$$\langle u(t,\cdot),\psi\rangle = \langle u_0,\psi\rangle + \int_0^t \langle u(s,\cdot),\partial_x^2\psi\rangle\,ds + \int_0^t \langle H(u(s,\cdot)),\psi\rangle\,ds + \int_0^t \langle G(u(s,\cdot))\psi, dW(s)\rangle.$$

This is indeed the definition of a weak solution to (1.2), the uniqueness of which was shown in [**Wal86**], see Exercise 3.4. □

References

[BMS95] V. Bally, A. Millet and M. Sanz-Solé, Approximation and support theorem in Hölder norm for parabolic stochastic partial differential equations, Ann. Probab., **23** (1995), 178–222.

[CW77] R. Cairoli and J. Walsh, Martingale representations and holomorphic processes, Ann. Probab., **5** (1977), 511–521.

[DPZ92] G. Da Prato and J. Zabczyk, Stochastic Equations in Infinite Dimensions, Encyclopedia of Mathematics and its Applications, **44**, Cambridge University Press, 1992.

[FH14] P. Friz and M. Hairer, A Course on Rough Paths, Universitext. Springer-Verlag, Berlin, 2014, An introduction with applications.

[GT10] M. Gubinelli and S. Tindel, Rough evolution equations, Ann. Probab., **38** (2010), 1–75.

[Hai13] M. Hairer, Solving the KPZ equation, Ann. of Math. (2), **178** (2013), 559–664.

[Hai14] M. Hairer, A theory of regularity structures, Invent. Math., **198** (2014), 269–504.

[Hai15] M. Hairer, Introduction to regularity structures, Braz. J. Probab. Stat., **29** (2015), 175–210.

[HPP13] M. Hairer, É. Pardoux and A. Piatnitski, Random homogenization of a highly oscillatory singular potential, SPDEs: Anal. Comp., **1** (2013), 571–605.

[HQ15] M. Hairer and J. Quastel, A class of growth models rescaling to KPZ, 2015, preprint.

[HW13] M. Hairer and H. Weber, Rough Burgers-like equations with multiplicative noise, Probab. Theory Related Fields, **155** (2013), 71–126.

[KSXZ13] D. Khoshnevisan, J. Swanson, Y. Xiao and L. Zhang, Weak existence of a solution to a differential equation driven by a very rough fBm, ArXiv e-prints (2013), arXiv:1309.3613.

[LCL07] T. J. Lyons, M. Caruana and T. Lévy, Differential equations driven by rough paths, Lecture Notes in Math., **1908**, Springer, Berlin, 2007, 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.

[Nua06] D. Nualart, The Malliavin calculus and related topics, Probability and its Applications (New York), Springer-Verlag, Berlin, second ed., 2006.

[PP12] É. Pardoux and A. Piatnitski, Homogenization of a singular random one-dimensional PDE with time-varying coefficients, Ann. Probab., **40** (2012), 1316–1356.

[RS96] T. Runst and W. Sickel, Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations, de Gruyter Series in Nonlinear Analysis and Applications, **3**, Walter de Gruyter & Co., Berlin, 1996.

[SV72] D. Stroock and S. R. S. Varadhan, On the support of diffusion processes with applications to the strong maximum principle, Proc. Sixth Berkeley Symp. on Math. Statist. and Prob., **3** (1972), 333–359.

[Wal86] J. B. Walsh, An introduction to stochastic partial differential equations, In: École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., **1180**, 265–439. Springer, Berlin, 1986.

[WZ65a] E. Wong and M. Zakai, On the convergence of ordinary integrals to stochastic integrals,

Ann. Math. Statist., **36** (1965), 1560–1564.

[WZ65b] E. Wong and M. Zakai, On the relation between ordinary and stochastic differential equations, Internat. J. Engrg. Sci., **3** (1965), 213–229.

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

Étienne Pardoux
Aix-Marseille Université
CNRS, Centrale Marseille, I2M
UMR 7373 13453 Marseille
France
E-mail: etienne.pardoux@univ-amu.fr