Moment bounds for SPDEs with non-Gaussian fields and application to the Wong-Zakai problem

Ajay Chandra[*]   Hao Shen[†]

Abstract

Upon its inception the theory of regularity structures [7] allowed for the treatment for many semilinear perturbations of the stochastic heat equation driven by space-time white noise. When the driving noise is non-Gaussian the machinery of the theory can still be used but must be combined with an infinite number of stochastic estimates in order to compensate for the loss of hypercontractivity, as was done in [12]. In this paper we obtain a more streamlined and automatic set of criteria implying these estimates which facilitates the treatment of some other problems including non-Gaussian noise such as some general phase coexistence models [13], [16] - as an example we prove here a generalization of the Wong-Zakai Theorem found in [10].

Contents

[*]University of Warwick, United Kingdom.
  E-mail: `ajay.chandra@gmail.com`
[†]Columbia University, United States of America.
  E-mail: `pkushenhao@gmail.com`

Moment bounds for SPDEs with non-Gaussian fields

1 Introduction

In the paper [10] the main focus was the convergence of smooth approximations u_ϵ to the solution of the SPDE

$$\partial_t u = \partial_x^2 u + H(u) + G(u)\xi. \tag{1.1}$$

Here $u(t,x)$ is a function from $\mathbf{R}_+ \times S^1$ to $\mathbf{R}$, $H, G : \mathbf{R} \to \mathbf{R}$ are respectively twice and five-times continuously differentiable and ξ denotes space-time white noise. One can immediately obtain a solution u to (1.1) by viewing it as an infinite dimensional Itô integral equation in time.

The fundamental obstacle to interpreting (1.1) without stochastic calculus is the irregularity of ξ. The smooth approximations u_ϵ satisfy the above equation with ξ replaced by $\xi_\varepsilon := \xi * \varrho_\varepsilon$ where ϱ_ε is a mollifier converging to a space-time delta function as $\varepsilon \downarrow 0$. More concretely the authors of [10] set

$$\varrho_\varepsilon(t,x) := \varepsilon^{-3}\varrho(\varepsilon^{-2}t, \varepsilon^{-1}x),$$

where $\varrho : \mathbf{R}^2 \mapsto \mathbf{R}$ is an even, smooth, compactly supported function which integrates to 1. Let u_ε denote the classical solution to the equation driven by ξ_ε.

Unsurprisingly, the u_ε do not converge to u in general. One already sees this in finite dimensions where the Wong-Zakai Theorem [19, 20] (for more recent progress c.f. [14] and references therein) states that smooth approximations to an SDE converge to the *Stratonovich* solution to the SDE which in general differs from the Itô solution. Of course this discrepancy can be cured by "renormalizing" the SDE by inserting a Stratonovich-Itô correction term into the mollified SDEs. The main result of [10] is the corresponding result for the SPDE setting.

Theorem 1.1. *(Hairer-Pardoux [10]) Assume that H and G are of classes $\mathcal{C}^2$ and $\mathcal{C}^5$ respectively, both with bounded first derivatives. Let ϱ and ξ_ε be as above. Let u denote the Itô solution to (1.1) with some initial condition $u(0,\cdot) \in \mathcal{C}(S^1)$. Then there exist finite ε independent constants $C_\varrho^{(1)}, c_\varrho^{(1)}$, and $c_\varrho^{(2)}$ such that the following holds: If u_ε is the classical solution to the random PDE*

$$\partial_t u_\varepsilon = \partial_x^2 u_\varepsilon + \bar{H}(u_\varepsilon) - \frac{C_\varrho^{(1)}}{\varepsilon} G'(u_\varepsilon)G(u_\varepsilon) + G(u_\varepsilon)\,\xi_\varepsilon\,, \tag{1.2}$$

where

$$\bar{H}(u) \stackrel{def}{=} H(u) - c_\varrho^{(1)} G'(u)^3 G(u) - c_\varrho^{(2)} G''(u)G'(u)G^2(u)\,. \tag{1.3}$$

with initial condition $u_\varepsilon(0,\cdot) = u(0,\cdot)$ then 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. For any $\alpha \in (0, \frac{1}{2})$ and $t > 0$, the restriction of u_ε to $[t,T] \times S^1$ converges to u in probability for the topology of $\mathcal{C}^{\alpha/2,\alpha}([t,T] \times S^1)$ as $\varepsilon \downarrow 0$.

We defer exact formulae for the renormalization constants; they can be explicitly written as integrals involving the heat kernel and ϱ. The $\frac{1}{\varepsilon}$ term in (1.2) is exactly the Itô-Stratonovich correction term which diverges as expected - there is no notion of infinite dimensional analog of Stratonovich integration. The definition of $\bar{H}(u)$ also involves two finite renormalizations which are chosen so that it is precisely the Itô solution to which the u_ε are converging. In fact, along the proof [10] obtains a natural notion of solution to (1.1) which is pathwise - an analogous situation as in [15] (on rough paths) and [5] (on evolution equations).

In [10, Remark 1.7] Hairer and Pardoux ask if an analogous statement can be proven if one replaces the mollified space-time white noise $\xi_\varepsilon(t,x)$ with $\varepsilon^{-3/2}\zeta(\varepsilon^{-2}t,\varepsilon^{-1}x)$ where ζ is a *non-Gaussian* random field which is supported on smooth functions and satisfies a central limit theorem. They conjectured that in addition to the renormalization seen in the Gaussian case one would see additional terms of order $\varepsilon^{-\frac{1}{2}}$.

The question of [10, Remark 1.7] is our point of departure. Let ζ be stationary, centered, generically non-Gaussian random field on $\mathbf{R}^2$ which is almost surely continuous and for which all cumulants[1] exist and are exponentially decaying[2]. We also assume that ζ is normalized $\int \zeta(0)\zeta(z)\,dz = 1$. Let $\zeta^{(\varepsilon)}$ be a random field on $\mathbf{R} \times [-\frac{1}{2\varepsilon}, \frac{1}{2\varepsilon}]$ which is a periodization of ζ (see Remark 1.3). We then set

$$\zeta_\varepsilon(t,x) := \varepsilon^{-3/2}\zeta^{(\varepsilon)}(\varepsilon^{-2}t, \varepsilon^{-1}x). \tag{1.4}$$

Our result is then the following, which is proved using the theory of regularity structures developed in [7].

Theorem 1.2. *Let ζ be as above and H and G be as in Theorem 1.1 and as before u be the Itô solution to (1.1) started from some initial condition $u(0,\cdot) \in \mathcal{C}(S^1)$. Then there exist constants $\{C_\zeta^{(i)}\}_{i=1}^3$ and $\{c_\zeta^{(j)}\}_{j=1}^4$, all of which are independent of the parameter ε, such that the following holds: Suppose that u_ε is the classical solution to the random PDE*

$$
\begin{aligned}
\partial_t u_\varepsilon = \partial_x^2 u_\varepsilon + &\bar{H}(u_\varepsilon) - \frac{C_\zeta^{(1)}}{\varepsilon}G'(u_\varepsilon)G(u_\varepsilon) \\
&- \frac{C_\zeta^{(2)}}{\varepsilon^{1/2}}G'(u_\varepsilon)^2 G(u_\varepsilon) - \frac{C_\zeta^{(3)}}{\varepsilon^{1/2}}G''(u_\varepsilon)G^2(u_\varepsilon) + G(u_\varepsilon)\,\zeta_\varepsilon\,,
\end{aligned}
\tag{1.5}
$$

where

$$
\begin{aligned}
\bar{H}(u) \stackrel{\text{def}}{=} H(u) &- \frac{1}{6}c_\zeta^{(1)}G'''(u)G^3(u) - c_\zeta^{(2)}G'(u)^3 G(u) \\
&- \Big(\frac{1}{2}c_\zeta^{(3)} + c_\zeta^{(4)}\Big)G''(u)G'(u)G^2(u)
\end{aligned}
\tag{1.6}
$$

and u_ε is started with the initial condition $u(0,\cdot) \in \mathcal{C}(S^1)$.

Then for every $T > 0$, the family of random functions u_ε converges in law to the Itô solution u as $\varepsilon \downarrow 0$, with initial data u_0 in the space $\mathcal{C}^{\alpha/2,\alpha}([0,T] \times S^1)$, for any $\alpha \in (0, \frac{1}{2})$.

We now explictly specify the renormalization constants of appearing in the above theorem. We write

$$
\begin{aligned}
C_\zeta^{(1)} &= C^{\text{⚬⚬}} \qquad C_\zeta^{(2)} = C^{\text{⚬⚬}} \qquad C_\zeta^{(3)} = C^{\text{⚬⚬⚬}} \\
c_\zeta^{(1)} &= C^{\text{⚬⚬⚬⚬}} \qquad c_\zeta^{(2)} = C_1^{\text{⚬⚬⚬}} + C_2^{\text{⚬⚬⚬}} + C_3^{\text{⚬⚬⚬}} \\
c_\zeta^{(3)} &= C^{\text{⚬⚬⚬}} \qquad c_\zeta^{(4)} = C_1^{\text{⚬⚬⚬}} + C_2^{\text{⚬⚬⚬}} + C_3^{\text{⚬⚬⚬}}
\end{aligned}
\tag{1.7}
$$

---

[1] See Appendix A for definition of cumulants.
[2] See Definition A.1.

where the new constants appearing above are defined by the diagrammatic formulae

These diagrams represent integrals of various kernels. The black vertices • represent integrated variables in $\mathbf{R}^2$ while the single green vertex ● represents $0 \in \mathbf{R}^2$. A blue arrow ⟶ corresponds to the heat kernel P evaluated at the difference of its terminal vertex and initial vertex. The lightly blue shaded regions (e.g.) should be thought of as "hyperedges", they represent a cumulant of ζ evaluated at the positions of the vertices within the region. (See Appendix A for definition of the n-th cumulants, which will be written as $\mathfrak{C}_n$.) For example we have

$$C^{\includegraphics{}} = \int_{\mathbf{R}^2} \int_{\mathbf{R}^2} \mathrm{d}z_1 \, \mathrm{d}z_2 \; P(z_2 - z_1) P(0 - z_2) \mathfrak{C}_3(z_1, z_2, 0).$$

Finally, the notation ⩗⩗⩗⩗⩗ stands for a renormalized kernel. If the variables corresponding to its endpoints are z_1 and z_2 then it represents the kernel

$$P(z_1 - z_2)\mathfrak{C}_2(z_1 - z_2) - C^{\includegraphics{}}\delta(z_1 - z_2).$$

Remark 1.3. In (1.4) the field $\zeta^{(\varepsilon)}$ on $\mathbf{R} \times [-\frac{1}{2\varepsilon}, \frac{1}{2\varepsilon}]$ where $[-\frac{1}{2\varepsilon}, \frac{1}{2\varepsilon}]$ is identified with the circle of length $\frac{1}{\varepsilon}$ is said to be a periodization of ζ, in the same sense as in [12, Assumption 2.1], i.e. for every sufficiently small $\varepsilon > 0$, there exists a coupling of $\zeta^{(\varepsilon)}$ and ζ such that for every $T > 0$ and every $\delta > 0$,

$$\sup_{|t| \leq T\varepsilon^{-2}} \sup_{|x| \leq (1-\delta)/(2\varepsilon)} \lim_{\varepsilon \to 0} \varepsilon^{-3} \mathbf{E}|\zeta(t, x) - \zeta^{(\varepsilon)}(t, x)|^2 = 0 \,. \tag{1.8}$$

As an example, let μ be a Poisson point process on $\mathbf{R}^2 \times [0, 1]$ with uniform intensity measure, let $\varphi(t, x, a)$ be a continuous function bounded by $e^{-|t|-|x|}$, and set

$$\zeta(t, x) = \int_{\mathbf{R}^2 \times [0,1]} \varphi(t - s, x - y, a)\mu(ds, dy, da) - \mu(\varphi) \,. \tag{1.9}$$

Let $\mu^{(\varepsilon)}$ be the periodic extension to $\mathbf{R}^2 \times [0,1]$ of a Poisson point process on $\mathbf{R} \times [-1/(2\varepsilon), 1/(2\varepsilon)] \times [0,1]$ with uniform intensity measure, and

$$\varphi^{(\varepsilon)}(t,x,a) \stackrel{\text{def}}{=} \varphi(t,x,a)\tilde{\varphi}_{\varepsilon^{-1/4}}(t,x)$$

where $\tilde{\varphi}_{\varepsilon^{-1/4}}(t,x)$ is a continuous function that is equal to 1 when $|x| < \varepsilon^{-1/4}$ and 0 when $|x| > 2\varepsilon^{-1/4}$, and let $\zeta^{(\varepsilon)}$ be as in (1.9), with μ replaced by $\mu^{(\varepsilon)}$ and φ replaced by $\varphi^{(\varepsilon)}$. Then one can verify that for ε small enough, (1.8) is satisfied, where the natural coupling between ζ and $\zeta^{(\varepsilon)}$ is such that $\zeta = \zeta^{(\varepsilon)}$ on $\mathbf{R} \times [-\frac{1}{2}\varepsilon^{-1} + \varepsilon^{-\frac{1}{4}}, \frac{1}{2}\varepsilon^{-1} - \varepsilon^{-\frac{1}{4}}]$. Note that the cumulants of $\zeta^{(\varepsilon)}$ are allowed to have infinite range (but exponential decaying) in t.

Remark 1.4. The same results as Corollaries 1.10, 1.11, 1.13 in [10] (on local continuity of the solution with respect to the initial condition in pathwise sense, sharp regularity result for the solution etc.) can also be proved. Since the proofs follow along the same lines, we refrain from redoing them here.

1.1 Moment estimates for SPDE with non-Gaussian fields

Much of our paper is spent developing a set of criterion for estimating moments of certain non-Gaussian random variables which will be written as $(\Pi_0 \tau)(\varphi_0^\lambda)$ and defined in Section 2. These random variables can be represented as rooted trees. Each of these random variables is a multilinear functional of the driving noise ζ_ε and the tree describes how to write this functional as an integral - the edges correspond to kernels, at each leaf we have an occurrence of the driving noise, and all variables other than the root correspond to integrated space-time variables.

Roughly speaking[3], the p-th moments of this random variable can be written as a graphical sum where one takes p copies of this tree, identifies their roots, and then sums over all the possible ways of grouping the leaves into cumulants. For Gaussian noise only second cumulants appear but in general we have higher order cumulants which we view as hyperedges (edges incident to more than two vertices). We refer to this graphical sum as a cumulant expansion.

Estimates on moments are needed to establish regularity/homogeneity of certain random space-time processes via a generalization of the Kolmogorov continuity criterion, the random variable we are estimating is the analog of an "increment" of the process with the size of the increment given by a parameter λ. Our primary goal is, for each p, a bound of the type

$$\left|\mathbf{E}\big[(\Pi_0\tau)(\varphi_0^\lambda)^p\big]\right| \lesssim \lambda^{|\tau|p} \tag{1.10}$$

uniform in $\lambda \in (0,1]$ and ε small where $|\tau|$ is determined by the structure of the tree. Here, for any continuous function $\varphi : \mathbf{R}^2 \mapsto \mathbf{R}$, $z = (t,x) \in \mathbf{R}^2$, and $\lambda > 0$,

$$\varphi_z^\lambda(\bar{t},\bar{x}) \stackrel{\text{def}}{=} \lambda^{-3}\varphi(\lambda^{-2}(\bar{t}-t), \lambda^{-1}(\bar{x}-x)).$$

For the bound (1.10) we need to estimate, for each of the larger hypergraphs appearing in the cumulant expansion of the given moment, a complicated convolution of kernels and cumulants.

The paper [11] developed criteria for graphs which guarantee that they satisfy the desired upper bound. Moreover by hypercontractivity (e.g. [7, Lemma 10.5]) one has

$$\mathbf{E}\big[(\Pi_0\tau)(\varphi_0^\lambda)^p\big] \leq C_p\big|\mathbf{E}\big[(\Pi_0\tau)(\varphi_0^\lambda)^2\big]\big|^{p/2}.$$

Thus in the Gaussian case one needs (1.10) only for $p = 2$ which is obtained by checking a small number of graphs.

---

[3]See [9, Section 5] for a more in depth discussion.

When hypercontractivity is not available establishing (1.10) for every p involves estimating infinitely many hypergraphs[4]. This problem was tackled in [12] where a criterion for individual trees was given which implies that any larger graph built by merging p copies of this tree satisfies the necessary criterion in [11]. However the criteria of [12] requires one to do some manipulation on the trees and the larger graphs one got after the merging: (i) cumulants had to be replaced by collections of edges and good factors had to be distributed according to "epsilon allocation rules" and (ii) so called "positively - renormalized" edges had to be estimated by hand on a case by case basis leading to more trees to check.

In this paper we provide streamlined criteria for these trees, proving the sufficiency of these criterion will also be easier. By working with hyperedges issue (i) is avoided - this makes proving very general results like [16] much easier. Handling the positively-renormalized edges automatically deals with issue (ii) which makes the treatment of the Wong-Zakai problem much easier.

In Section 2 we fix our regularity structures and formulate the abstract fixed point problem for the Wong-Zakai equation in a space of modeled distributions. In Section 3 we prove Theorem 3.19 which states that certain criteria on the graphs yield the desired moment bounds. In Section 4 we apply the results obtained in Section 3 to the Wong-Zakai problem and prove Theorem 1.2.

2 Regularity Structures

The moment estimates we prove will be used as input for the theory of regularity structures developed in [7] (see also [6]). This machinery allows us to go from these estimates to the construction of an actual solution to the SPDE in question along with convergence of regularized and renormalized solutions to this limiting solution.

We refer readers looking for a detailed exposition to [8], [3][Ch. 15], and [1]; our description of the theory will be quite brief. The most basic object in the theory is a *regularity structure* which consists of a triple $(A, \mathcal{T}, \mathcal{G})$. The set $A \subset \mathbf{R}$ is a list of the possible homogeneities we allow in our expansions; it is assumed to be locally finite and bounded below. $\mathcal{T}$ is a graded vector space $\mathcal{T} = \oplus_{\alpha \in A} \mathcal{T}_\alpha$ where each $\mathcal{T}_\alpha$ is a Banach space with a distinguished basis. $\mathcal{G}$ is a group of continuous linear transformations on $\mathcal{T}$ with the property that for all $\alpha \in A$, $\tau \in \mathcal{T}_\alpha$, and $\Gamma \in G$ one has $(\Gamma \tau - \tau) \in \oplus_{\beta < \alpha} \mathcal{T}_\beta$. A regularity structure is used to describe "jets of abstract Taylor expansions"; the vector space $\mathcal{T}$ is the target space for the jets and the *structure group* $\mathcal{G}$ includes transformations on the target space corresponding to change of base-point operations.

2.1 The Wong-Zakai regularity structure

The specific regularity structure we use for our Wong-Zakai type model is exactly the same as the one used in [10]; in particular $\mathcal{T}$ is spanned by a set of indeterminants $\tau \in \mathcal{W}$, each carrying a homogeneity $|\tau|$. We first define a a larger class of indeterminants and then take $\mathcal{W}$ as an appropriate subset.

We start with the indeterminants $\mathbf{1}$, X_0, and X_1 which are the abstract counterparts of 1, t, and x - since our scaling is parabolic we set $|\mathbf{1}| = 0$, $|X_0| = 2$, and $|X_1| = 1$. Given a multi-index $k = (k_0, k_1) \in \mathbf{N}^2$ we write X^k as a shorthand for $X_0^{k_0} X_1^{k_1}$. For such k we set $|k|_{\mathfrak{s}} \stackrel{\text{def}}{=} 2k_0 + k_1$ so $|X^k| = |k|_{\mathfrak{s}}$. We also write $\bar{\mathcal{T}}$ for the commutative algebra generated by $\mathbf{1}$, X_0, and X_1, the set of *abstract polynomials*.

This set also carries a commutative, multiplicative structure (for which $\mathbf{1}$ will act as a unit). Given two indeterminants $\tau, \bar{\tau}$ we enforce that $|\tau \bar{\tau}| = |\tau| + |\bar{\tau}|$.

---

[4]One actually only needs (1.10) bound for p large enough to facilitate a Borel-Cantelli argument - but if $p > 10$ this is already daunting.

We also introduce the indeterminant Ξ which represents the driving noise ξ, we set $|\Xi| = -3/2 - \kappa$ where κ is a fixed, small positive parameter [5]. We also introduce an operator $\mathcal{I}(\cdot)$ on indeterminants, enforcing that $\mathcal{I}(X^k)$ vanish for any multi-index k; if $\mathcal{I}(\tau) \neq 0$ we enforce $|\mathcal{I}(\tau)| = |\tau| + 2$.

Define $\mathcal{U}$ to be the smallest collection of indeterminants which contains $\mathbf{1}$, X_0, and X_1 and that satisfies the conditions (i) $\tau \in \mathcal{U} \Rightarrow \mathcal{I}(\tau) \in \mathcal{U}$, (ii) $\tau \in \mathcal{U} \Rightarrow \mathcal{I}(\Xi\tau) \in \mathcal{U}$, and (iii) $\tau, \bar{\tau} \in \mathcal{U} \rightarrow \tau\bar{\tau} \in \mathcal{U}$. Finally we set

$$\mathcal{W} := \left\{ \tau \in \mathcal{U} \cup \left\{ \tau\Xi : \ \tau \in \mathcal{U} \right\} : \ |\tau| \leq \frac{5}{2} \right\}. \tag{2.1}$$

A is given by the set of homogeneities that appear in $\mathcal{W}$ which by [7, Lemma 8.10] is bounded below and locally finite, also for each $\alpha \in A$ the vector space $\mathcal{T}_\alpha$ spanned by indeterminants of homogeneity α in $\mathcal{W}$ is finite dimensional.

Often we write the elements of $\mathcal{W}$ as blue symbolic trees with $\Xi = \circ$. Each occurrence of the abstract integration map $\mathcal{I}$ is then denoted by a downward straight line. The product τ and $\bar{\tau}$ is represented by attaching the trees for τ and $\bar{\tau}$ at the root. For example, we have $\Xi \mathcal{I}^2(\Xi) = \<tree>$. Note that we never see an expression of the form Ξ^2 in $\mathcal{W}$. We also use the shorthand $\Xi X_1 = \otimes$. The elements in $\mathcal{W}$ with negative homogeneities are:

$$
|\circ| = -\frac{3}{2} - \kappa, \quad |\<tree>| = -1 - 2\kappa, \quad |\<tree>| = |\<tree>| = -\frac{1}{2} - 3\kappa,
$$
$$
|\otimes| = -\frac{1}{2} - \kappa, \quad |\<tree>| = |\<tree>| = |\<tree>| = |\<tree>| = -4\kappa, \quad |\<tree>| = |\<tree>| = -2\kappa. \tag{2.2}
$$

Having defined the $\mathcal{T}$ of the Wong-Zakai regularity structure, now we turn to defining a structure group $\mathcal{G}$. To do this we introduce another set of indeterminants denoted $\mathcal{W}_+$ and denote by $\mathcal{T}_+$ the commutative algebra they generate. The construction of the structure group can be summarized as follows: there will be a *single* "abstract" matrix of indeterminants from $\mathcal{T}_+$ which acts on $\mathcal{T}$ - all the individual elements of $\mathcal{G}$ arise by specifying an appropriate map $f \in \mathcal{T}_+^\star$ where $\mathcal{T}_+^\star$ is the set of algebra homomorphisms from $\mathcal{T}_+$ to $\mathbf{R}$.

Following [10] we set

$$\mathcal{W}_+ \stackrel{\text{def}}{=} \{X_0, X_1\} \cup \left\{ \mathcal{J}_k(\tau) : \ \tau \in \mathcal{W} \setminus \bar{\mathcal{T}}, \ k \in \mathbf{N}^2 \text{ with } |k|_{\mathfrak{s}} < |\tau| + 2 \right\}.$$

Here the operators $\mathcal{J}_k(\cdot)$ on $\mathcal{W}_+$ are analogous to $\mathcal{I}[\cdot]$ on $\mathcal{W}$. We use the convention that $\mathcal{J}_k(\tau) \stackrel{\text{def}}{=} 0$ if $|\tau| \leq -2 + |k|_{\mathfrak{s}}$.

The abstract matrix described earlier will be map $\Delta : \mathcal{T} \mapsto \mathcal{T} \otimes \mathcal{T}_+$ which we now define recursively on $\mathcal{T}$. The base cases are given by

$$\Delta \mathbf{1} \stackrel{\text{def}}{=} \mathbf{1} \otimes \mathbf{1}, \quad \Delta \Xi \stackrel{\text{def}}{=} \Xi \otimes \mathbf{1}, \quad \text{and} \quad \Delta X_i \stackrel{\text{def}}{=} X_i \otimes \mathbf{1} + \mathbf{1} \otimes X_i.$$

We then recursively set

$$\Delta(\tau\bar{\tau}) \stackrel{\text{def}}{=} \Delta(\tau) \cdot \Delta(\tau) \quad \text{and} \quad \Delta\mathcal{I}(\tau) \stackrel{\text{def}}{=} (\mathcal{I} \otimes \text{Id})\Delta\tau + \sum_{l,k \in \mathbf{N}^2} \frac{X^l}{l!} \otimes \frac{X^k}{k!} \mathcal{J}_{l+k}(\tau).$$

The product on the RHS of the first definition is component-wise. Also note that the sum in the second definition only has finitely many non-vanishing terms.

Given any $f \in \mathcal{T}_+^*$ we define a linear transformation $\Gamma_f : \mathcal{T} \mapsto \mathcal{T}$ by setting

$$\Gamma_f \tau \stackrel{\text{def}}{=} (\text{Id} \otimes f)\Delta\tau. \tag{2.3}$$

[5] fixing $\kappa \in (0, 1/10)$ suffices

We then define $\mathcal{G}$ to be the set of all linear transformations of the above form. The only non-trivial thing is to check is that $\mathcal{G}$ forms a group. This is done by equipping $\mathcal{T}_+$ with a Hopf-algebra structure for which Δ serves as a comodule coproduct - we refer the curious reader to [7, Section 8].

2.2 Admissible models and renormalized models

It is convenient to replace the heat kernel P with a truncation $K : \mathbf{R}^2 \setminus \{0\} \mapsto \mathbf{R}$ with the following properties: (i) $K(z)$ vanishes for $|z| > 1$, (ii) for $|z| < 1/2$, $K(z) = P(z)$, (iii) K is smooth on $\mathbf{R}^2 \setminus \{0\}$, and (iv) $\int_{\mathbf{R}^2} \mathrm{d}z\, K(z) z^k = 0$ for any multi-index k with $|k|_{\mathfrak{s}} < 3$.

The existence of such a K is not hard to show, see [7, Section 5], and we consider it fixed for the rest of the paper. We now have everything in place to define the set of *admissible* models $\mathscr{M}$ on the Wong-Zakai regularity structure.

A *model* is a pair of maps (Π, Γ) with $\Pi : \mathbf{R}^2 \mapsto \mathcal{L}(\mathcal{T}, \mathcal{S}'(\mathbf{R}^d))$ which we write $z \mapsto \Pi_z$ and $\Gamma : \mathbf{R}^2 \times \mathbf{R}^2 \mapsto \mathcal{G}$ which we write $(z, \bar{z}) \mapsto \Gamma_{z\bar{z}}$. Here $\mathcal{L}(\mathcal{T}, \mathcal{S}'(\mathbf{R}^2))$ is the space of linear maps from $\mathcal{T}$ into the space of tempered distributions $\mathcal{S}'(\mathbf{R}^2)$. These maps are required to satisfy the algebraic conditions

$$\Pi_z \Gamma_{z\bar{z}} = \Pi_{\bar{z}} \quad \text{and} \quad \Gamma_{z\bar{z}} \Gamma_{\bar{z}z'} = \Gamma_{zz'} \text{ for any } z, \bar{z}, z' \in \mathbf{R}^2. \tag{2.4}$$

Let $\mathcal{B}$ be the set of all $\varphi \in \mathcal{S}'(\mathbf{R}^2)$ supported on the ball of radius 1 and with $|D^k \varphi| \leq 1$ where D denoted differentiation and $k \in \mathbf{N}^2$ with $|k|_{\mathfrak{s}} \leq 2$. We also require that models satisfy the analytic bounds

$$\sup_{z \in \mathfrak{K}} \left| (\Pi_z \tau)(\varphi_z^\lambda) \right| \lesssim \lambda^{|\tau|} \quad \text{and} \quad \sup_{\substack{z, \bar{z} \in \mathfrak{K} \\ z \neq \bar{z}}} \| \Gamma_{z\bar{z}} \tau \|_\alpha \lesssim |z - \bar{z}|^{|\tau| - \alpha} \tag{2.5}$$

for each compact set $\mathfrak{K} \subset \mathbf{R}^2$, uniformly over $\varphi \in \mathcal{B}$, $\lambda \in (0, 1]$, $\tau \in \mathcal{W}$ and $\alpha \in A$. In the second bound $\| \tau \|_\alpha$ denotes the $\mathcal{T}_\alpha$-norm of the $\mathcal{T}_\alpha$ component of τ.

The set of models can be equipped with a family of pseudometrics indexed by compact sets $\mathfrak{K} \subseteq \mathbf{R}^2$ - for two models $Z = (\Pi, \Gamma)$ and $\bar{Z} = (\bar{\Pi}, \bar{\Gamma})$ one sets $\| Z; \bar{Z} \|_{\mathfrak{K}}$ to the maximum of two optimal constants for each of the the bounds of (2.5) where in the first and second bound the individual objects are replaced by differences $(\Pi_z - \bar{\Pi}_z)$ and $(\Gamma_{z\bar{z}} - \bar{\Gamma}_{z\bar{z}})$, respectively. Together the pseudometrics $\| \cdot; \cdot \|_{\mathfrak{K}}$ make the space of models a *non-linear metric-space* [6].

To formulate the condition of admissibility it is convenient to switch to parameterizing models by pairs of maps (Π, f) where Π is as before and $f : \mathbf{R}^2 \mapsto \mathcal{T}_+^*$ which we write $z \mapsto f_z$. The correspondance between f's and Γ's is given by $\Gamma_{z\bar{z}} = \Gamma_{f_z}^{-1} \Gamma_{f_{\bar{z}}}$, here we use the notation of (2.3). It is straightforward to verify that (Π, Γ) constructed in such a way automatically satisfies condition (2.4).

The notion of *admissibility* can then be stated as follows.

Definition 2.1. *A pair of maps (Π, f) as above is said to be* admissible *on $(\mathcal{T}, \mathcal{G})$ if the following conditions hold for all $z, \bar{z} \in \mathbf{R}^d$, and for any multi-index k,*

$$(\Pi_z \mathbf{1})(\bar{z}) = 1, \quad (\Pi_z X^k \tau)(\bar{z}) = (\bar{z} - z)^k (\Pi_z \tau)(\bar{z}), \quad f_z(X^k) = (-z)^k$$

and for every $\tau \in \mathcal{W}$ with $\mathcal{I}(\tau) \in \mathcal{W}$,

$$f_z(\mathcal{J}_k \tau) = -\mathbf{1}_{\{|k|_{\mathfrak{s}} < |\tau| + 2\}} \times \int_{\mathbf{R}^2} D^k K(z - \bar{z})(\Pi_z \tau)(\mathrm{d}\bar{z}),$$

$$(\Pi_z \mathcal{I} \tau)(\bar{z}) = \int_{\mathbf{R}^2} K(\bar{z} - z')(\Pi_z \tau)(\mathrm{d}z') + \sum_k \frac{(\bar{z} - z)^k}{k!} f_z(\mathcal{J}_k \tau).$$

---

[6] the non-linearity coming from the constraint (2.4)

We remark that if the pair (Π, f) is admissible and Π satisfies the first bound of (2.5) then the Γ's built from f satisfy the second and (Π, f) determines a model. We denote by $\mathscr{M}$ the complete metric space of admissible models.

We now describe one way to lift a continuous space-time function ψ to a corresponding admissible model $Z_\psi = (\Pi, f)$. The algebraic constraints placed on admissible models are quite strong - if we define

$$(\Pi_z \Xi)(\bar{z}) \stackrel{\text{def}}{=} \psi(\bar{z}) \quad \text{and} \quad (\Pi_z \tau \bar{\tau})(\bar{z}) \stackrel{\text{def}}{=} (\Pi_z \tau)(\bar{z}) \cdot (\Pi_z \tau)(\bar{z}) \tag{2.6}$$

then one can use the identities of Definition 2.1 to define the rest of the action of (Π, f). We call the model Z_ψ built this way the canonical model built from ψ - we use the shorthand $Z_\varepsilon = (\Pi^\varepsilon, f^\varepsilon) \stackrel{\text{def}}{=} Z_{\zeta_\varepsilon}$ where ζ_ε is our rescaled random field.

A defect of the family of models Z_ε is that they *do not* converge to a limiting model in $\mathscr{M}$ as $\varepsilon \downarrow 0$, the key difficulties coming from symbols τ which correspond to products of insufficiently regular space-time processes. We will have to modify this family to get a new collection of renormalized models $\hat{Z}_\varepsilon = (\hat{\Pi}^\varepsilon, \hat{f}^\varepsilon)$ - in general these new models will not satisfy the second identity of (2.6) - as an example we will have $\Pi_z^\varepsilon(\mathcal{I}[\Xi]\Xi)(\bar{z}) = \Pi_z^\varepsilon(\mathcal{I}[\Xi])(\bar{z})\Pi_z^\varepsilon(\Xi)(\bar{z}) - C^{\stackrel{\circ}{\circ}}\varepsilon^{-1}$, without this subtraction the RHS would not converge to a meaningful object.

It is a fairly non-trivial task to determine how to deform the product property of a canonical model and still be left with an admissible model. In the theory of regularity structures this type of deformation of the product property is encoded via the action of a linear map $M : \mathcal{T}_0 \mapsto \mathcal{T}_0$ for an appropriate subset $\mathcal{T}_0 \subset \mathcal{T}$. One then has the following theorem, which is combination of Prop. 8.36, Def. 8.41, Theorem 8.44 in [7] and Theorem B.1 of [11].

Theorem 2.2. *Let $\mathcal{T}_0 \subset \mathcal{T}$ satisfy the properties that $\oplus_{\alpha \leq 0} \mathcal{T}_\alpha \subset \mathcal{T}_0$ and*

$$\Delta \mathcal{T}_0 \subset \mathcal{T}_0 \otimes \mathrm{Alg}(\mathcal{T}_0)$$

where $\mathrm{Alg}(\mathcal{T}_0)$ is the subalgebra of $\mathcal{T}_+$ spanned by terms of the form $X^k \prod_i \mathcal{J}_{l_i}(\tau_i)$ with $\tau_i \in \mathcal{T}_0$.

Let $M : \mathcal{T}_0 \mapsto \mathcal{T}_0$ be a linear map that commutes with both the application of $\mathcal{I}[\cdot]$ and multiplication by X^k. Then there exist a unique linear, multiplicative map $\hat{M} : \mathcal{T}_+ \mapsto \mathcal{T}_+$ fixing abstract polynomials and a unique linear map $\Delta^M : \mathcal{T} \mapsto \mathcal{T} \otimes \mathcal{T}_+$ such that one has

$$\hat{M}\mathcal{J}_k = \mathcal{M}(\mathcal{J}_k \otimes 1)\Delta^M \quad \text{and} \quad (1 \otimes \mathcal{M})(\Delta \otimes 1)\Delta^M = (M \otimes \hat{M})\Delta.$$

Suppose furthermore that the map Δ^M is upper triangular, *that is for every $\alpha \in A$ and $\bar{\tau} \in \mathcal{T}_\alpha$ one has $\Delta^M \tau \in (\oplus_{\beta \geq \alpha} \mathcal{T}_\alpha) \otimes \mathcal{T}_+$. Then if (Π, f) is an admissible model then so is the renormalized model (Π^M, f^M) given by*

$$\Pi_z^M \stackrel{\text{def}}{=} (\Pi_z \otimes f_z)\Delta^M \quad \text{and} \quad f_z^M \stackrel{\text{def}}{=} f_z \circ \hat{M}.$$

Furthemore, the family of M satisfying the above properties form a group $\mathfrak{R}$ under composition.

Later, we will prescribe renormalization maps M^ε, sketch how one checks the upper-triangle condition for Δ^{M^ε}, and set $\hat{\Pi}^\varepsilon \stackrel{\text{def}}{=} (\Pi_z^\varepsilon \otimes f_z^\varepsilon)\Delta^{M^\varepsilon}$ and $\hat{f}^\varepsilon = f_z \circ \hat{M}^\varepsilon$. Returning to our previous example, one will have $\Delta^{M^\varepsilon}{}^{\stackrel{\circ}{\circ}} = M^\varepsilon{}^{\stackrel{\circ}{\circ}} \otimes 1$ and $M^\varepsilon{}^{\stackrel{\circ}{\circ}} = {}^{\stackrel{\circ}{\circ}} - C^{\stackrel{\circ}{\circ}}\varepsilon^{-1}$.

2.3 Modeled distributions and abstract fixed point problem

Given an admissible model Z, one can then start formulating abstract fixed point problems in spaces of modelled distributions $\mathcal{D}^{\gamma, \eta}$.

Definition 2.3. *Given an admissible model $Z \in \mathcal{M}$ and $\gamma, \eta \in \mathbf{R}$ we define the space of modelled distributions $\mathcal{D}^{\gamma,\eta}$ to be the set of all functions $U : \mathbf{R}^2 \mapsto \oplus_{\alpha < \gamma} \mathcal{T}_\alpha$ such that for every compact set $\mathfrak{K} \subset \mathbf{R}^2$ one has*

$$\|U\|_{\gamma,\eta} \stackrel{def}{=} \sup_{z \in \mathfrak{K}} \sup_{\alpha < \gamma} \frac{\|U(z)\|_\alpha}{|t|^{(\eta - \alpha)/2}} + \sup_{(z,\bar{z}) \in \mathfrak{K}^{(2)}} \sup_{\alpha < \gamma} \frac{\|U(z) - \Gamma_{z\bar{z}} U(\bar{z})\|_\alpha}{(|t| \wedge |\bar{t}|)^{(\eta - \gamma)/2} |z - \bar{z}|^{\gamma - \alpha}} < \infty. \tag{2.7}$$

Here $\mathfrak{K}^{(2)} \stackrel{def}{=} \{(z, \bar{z}) \in \mathfrak{K}^2 : |z - \bar{z}| \leq 1 \wedge \frac{1}{2}\sqrt{|t| \wedge |\bar{t}|}\}$.

One of the main theorems in [7] says that there exists a reconstruction operator $\mathcal{R}$ mapping the elements of $\mathcal{D}^{\gamma,\eta}$ to actual functions or distributions. In the space $\mathcal{D}^{\gamma,\eta}$ one can define the notions of multiplication and composition with smooth functions. It is also possible to construct a linear operator $\mathcal{P}$ on this space which represents the space-time convolution by the heat kernel, namely one has $\mathcal{R}\mathcal{P}U = P * \mathcal{R}U$ where $*$ is space-time convolution. These constructions allow us to formulate and solve abstract fixed point problems in the space $\mathcal{D}^{\gamma,\eta}$, and then apply the operator $\mathcal{R}$ on the abstract solution, which yields an actual function or distribution. For instance, regarding the equation (1.1), the abstract fixed point problem in the space $\mathcal{D}^{\gamma,\eta}$ is formulated as follows:

$$U = \mathcal{P}\big((\hat{H}(U) + \hat{G}(U)\Xi)\mathbf{1}_{t>0}\big) + Pu_0, \tag{2.8}$$

where Pu_0 is understood as naturally lifted to the abstract polynomials $\bar{\mathcal{T}}$. We actually consider this fixed point problem in a subspace $\mathcal{D}_{\mathcal{U}}^{\gamma,\eta} \subset \mathcal{D}^{\gamma,\eta}$ consisting of functions that take values in the span of $\mathcal{U}$ rather than $\mathcal{W}$ (see (2.1)).

The functions $\hat{H}, \hat{G}$ appearing in (2.8) are understood as follows. Given $U \in \mathcal{D}_{\mathcal{U}}^{\gamma;\eta}$ with $U(z) = u(z)\mathbf{1} + \tilde{U}(z)$ where $\tilde{U}(z)$ takes values in $\bigoplus_{\alpha > 0} \mathcal{T}_\alpha$ and a smooth function $G \colon \mathbf{R} \to \mathbf{R}$, we write

$$\big(\hat{G}(U)\big)(z) = G(u(z))\mathbf{1} + \sum_{k \geq 1} \frac{D^k G(u(z))}{k!} \tilde{U}(z)^k, \tag{2.9}$$

with the understanding that the product between any number of terms such that their homogeneity adds up to γ or more vanishes. Another property we will use is that $\mathcal{P}U - \mathcal{I}U \in \bar{\mathcal{T}}$, so any solution U to (2.8) satisfies

$$U(z) - \mathcal{I}\big(\hat{H}(U(z)) + \hat{G}(U(z))\Xi\big) \in \bar{\mathcal{T}}, \tag{2.10}$$

for all points $z = (t, x)$ with $t \in (0, T)$.

It follows from (2.10) and (2.9) that if we consider it as an element of $\mathcal{D}^{\gamma,\eta}$ with γ greater than, but sufficiently close to, $\frac{3}{2}$, then U is of the form

$$U = u\,\mathbf{1} + G(u)\,\text{⬩} + G'(u)G(u)\,\text{⬩} + u' X_1 \tag{2.11}$$
$$+ G'(u)^2 G(u)\,\text{⬩} + \frac{1}{2}G''(u)G^2(u)\,\text{⬩} + G'(u)u'\,\text{⬩},$$

for some functions u and u'. The symbols appearing here are introduced in Subsection 2.1. In $\mathcal{D}^{\gamma'}$ for $\gamma' > 0$ sufficiently close to 0 we have the identity

$$\hat{G}(U)\Xi = G(u)\,\circ + G'(u)G(u)\,\text{⬩} + G'(u)^2 G(u)\,\text{⬩} + G'(u)u'\,\otimes$$
$$+ \frac{1}{2}G''(u)G^2(u)\,\text{⬩} + \frac{1}{6}G'''(u)G^3(u)\,\text{⬩} + G'(u)^3 G(u)\,\text{⬩}$$
$$+ \frac{1}{2}G''(u)G'(u)G^2(u)\,\text{⬩} + G''(u)G'(u)G^2(u)\,\text{⬩}$$
$$+ G'(u)^2 u'\,\text{⬩} + G''(u)G(u)u'\,\text{⬩}. \tag{2.12}$$

This expansion is needed to derive the renormalized equations (1.5) and (1.6).

3 Graphical Moment Bounds

As discussed earlier the moments we need to bound can be represented by sums of graphs with hyperedges representing higher cumulants of the non-Gaussian noise. In this section we prove Theorem 3.19 which states that Assumption 3.17 for a given tree implies the desired bounds for every graph that appears in the aforementioned sum for the moment of that tree. We do not specialize to (1.1) and instead work in the general setting of d-dimensional space $\mathbf{R}^d$ with fixed scaling $\mathfrak{s} \in \mathbf{N}^d$.

3.1 Assumptions on kernels associated to hyper-edges

We start by recalling the notion of labeled coalescence trees, which will be useful for both the definition of norms on the kernels that are functions of more than two variables, and the proof to Theorem 3.7. A labeled coalescence tree (T, ℓ) is a rooted binary tree with every inner node v associated with a natural number ℓ_v which respects the partial order of the nodes, namely, $\ell_v \geq \ell_w$ whenever $v \geq w$ (i.e. w belongs to the shortest path connecting v to the root).

We denote by $\mathcal{T}_n$ the set of labeled coalescence trees with precisely n leaves. Given $(T, \ell) \in \mathcal{T}_n$ we define $\mathcal{D}(T, \ell) \subset (\mathbf{R}^d)^n$ to be the set of all tuples $(x_1, \ldots, x_n)$ such that for any $1 \leq i < j \leq n$ one has $|x_i - x_j| \sim 2^{-\ell_{v_i \wedge v_j}}$ where $v_1, \ldots, v_n$ denote the leaves of T.

Definition 3.1. *For any $\alpha \geq 0$ we define the following quantities and any function κ_n of $n > 2$ space-time points, we define*

$$\|\kappa_n\|_\alpha \stackrel{def}{=} \sup_{(T,\ell) \in \mathcal{T}_n} \left[2^{-\alpha \ell_\varrho} \sup_{(x_1,\ldots,x_n) \in \mathcal{D}(T,\ell)} |\kappa_n(x_1, \ldots, x_n)| \right].$$

Here ϱ denotes the root of the tree (T, ℓ).

Definition 3.2. *Given any $\alpha \geq 0$ and $p \in \mathbf{N}$, and a function $\kappa_2 : \mathbf{R}^d \setminus \{0\} \to \mathbf{R}$ set*

$$\|\kappa_2\|_\alpha \stackrel{def}{=} \left(\int_{\mathbf{R}^d} dx \, |\kappa_2(x)| \right) \vee \sup_{\substack{(T,\ell) \in \mathcal{T}_2 \\ k \in \mathbf{N}^d, \, |k|_\mathfrak{s} \leq p}} \left[2^{-(\alpha + |k|_\mathfrak{s}) \ell_\varrho} \sup_{(x_1, x_2] \in \mathcal{D}(T,\ell)} |\kappa_2(x_1 - x_2)| \right].$$

Remark 3.3. Note that the definition of $\| \cdot \|_{\alpha,p}$ given by Definition 3.2 dominates that given [7, Definition 10.12] for kernels of homogeneity α. The extra integral included in Definition 3.1 is needed in order to handle renormalizations that arise when $\alpha \geq |\mathfrak{s}|$.

Lemma 3.4. *Suppose the family of cumulants $\{\mathfrak{C}_n\}_{n \in \mathbf{N}}$ are exponentially decaying as in Definition A.1. Then for each $n \in \mathbf{N}$ one has*

$$\|\mathfrak{C}_n^{(\varepsilon)}\|_\alpha \lesssim \infty \tag{3.1}$$

uniform in $\varepsilon \in (0, 1]$ when $\alpha = n/2 \times |\mathfrak{s}|$, where the $\{\mathfrak{C}_n^{(\varepsilon)}\}_{n \in \mathbf{N}}$ are rescaled cumulants.

Proof. Without loss of generality suppose (A.1) with $\theta = e^{-1}$. Fix n, and suppose we are given $\vec{z} = (z_1, \ldots, z_n) \in (\mathbf{R}^d)^n$ - we can assume that $\mathrm{diam}(z_1, \ldots, z_n) = |z_1 - z_2|$. It follows that we have

$$\left| \mathfrak{C}_n^{(\varepsilon)}(z_1, \ldots, z_n) \right| \lesssim \varepsilon^{-n|\mathfrak{s}|/2} e^{-\varepsilon^{-1}|z_1 - z_2|}.$$

We want to show that if $\vec{z} \in \mathcal{D}(T, \ell)$ then the RHS above is bounded by some constant times $2^{n\ell(\varrho)|\mathfrak{s}|/2}$. If $2^{-\ell(\varrho)} < \varepsilon$ then this is immediate (just bound the exponential factor by one), so suppose instead that $2^{-\ell(\varrho)} \geq \varepsilon$. Then since $|z_1 - z_2| \sim 2^{-\ell(\varrho)}$ we have

$$\left| \mathfrak{C}_n^{(\varepsilon)}(z_1, \ldots, z_n) \right| \lesssim \varepsilon^{-n|\mathfrak{s}|/2} e^{-\varepsilon^{-1} 2^{-\ell(\varrho)}} \lesssim \varepsilon^{-n|\mathfrak{s}|/2} \times (\varepsilon^{-1} 2^{-\ell(\varrho)})^{-n|\mathfrak{s}|/2}$$

where we used the inequality $e^{-t} \lesssim t^{-n|\mathfrak{s}|/2}$ for $t \geq 1$. Thus the claim follows. $\square$

3.2 The entire graph with hyper-edges

We now state a modified version of the bound on generalized convolutions found in [11]. The difference here is that we allow for the presence of the hyperedges described above. The proof of our version of the bound essentially follows in the same way as that found in [11], so instead of giving a full proof here we only list the ways in which the proof needs to be modified.

The basic setting for these proofs is encoding the key properties of our generalized convolution as a (decorated) finite graph $G = (\mathcal{V}, \mathcal{E})$. $\mathcal{V}$ as before is the vertex set, which includes a subset of distinguished vertices $\mathcal{V}_\star$, one of which we call 0. The set $\mathcal{E}$ can be decomposed as $\mathcal{E}_2 \sqcup \mathcal{E}_h$ ($\sqcup$ meaning disjoint union) where $\mathcal{E}_2$ is the set of normal directed edges (denoted by ordered pairs (e_-, e_+) with $e_-, e_+ \in \mathcal{V}$) and $\mathcal{E}_h$ is the set of hyper-edges (subsets $e \subset \mathcal{V}$ with $|e| \geq 3$) [7].

We make further assumptions on the set $\mathcal{E}$ which we list below.

- For any distinct $e_1, e_2 \in \mathcal{E}_h$ one has $e_1 \cap e_2 = \emptyset$.
- For any $e \in \mathcal{E}_2$ one has $\left| e \cap \left(\bigcup_{\bar{e} \in \mathcal{E}_h} \bar{e} \right) \right| \leq 1$.
- For all $e \in \mathcal{E}_h$ one has $e \cap \mathcal{V}_\star = \emptyset$.

The edges $e \in \mathcal{E}$ are also decorated with labels a_e, r_e where $a_e \in \mathbf{R}$ and $r_e \in \mathbf{Z}$. We now list assumptions we make on these labels.

(a) For every $e \in \mathcal{E}_h$, $a_e = |e| |\mathfrak{s}|/2$ and $r_e = 0$.

(b) If $e \in \mathcal{E}_2$, $\bar{e} \in \mathcal{E}_h$, and $e \cap \bar{e} \neq \emptyset$, then $r_e \leq 0$ and $e \cap \bar{e} = \{e_-\}$. [8]

The edges $e \in \mathcal{E}_2$ are associated with kernels K_e which are smooth functions on $\mathbf{R}^d \times \mathbf{R}^d \setminus \{0\}$ and satisfy $\|K_e\|_{a_e, p} < \infty$ for any $p > 0$. The edges $e \in \mathcal{E}_h$ are associated with functions κ_e on $(\mathbf{R}^d)^{|e|}$ with $\|\kappa_e\|_{a_e} < \infty$ where $\|\kappa_e\|_{a_e}$ is defined in Definition 3.1. We will write $\kappa_e(e) \stackrel{\text{def}}{=} \kappa_e(x_1, \dots, x_{|e|})$ if $e = \{x_1, \dots, x_{|e|}\}$. For edges $e \in \mathcal{E}_2$ one also has renormalized kernels $\hat{K}_e$ as follows. If $r_e < 0$ then define the distribution

$$(\mathscr{R}K_e)(\varphi) = \int K_e(x) \left(\varphi(x) - \sum_{|k|_\mathfrak{s} < |r_e|} \frac{x^k}{k!} D^k \varphi(0) \right) dx + \sum_{|k|_\mathfrak{s} < |r_e|} \frac{I_{e,k}}{k!} D^k \varphi(0) \qquad (3.2)$$

where $\{I_{e,k}\}_{|k|_\mathfrak{s} < |r_e|}$ is a collection of real numbers, and the distributional "kernel" $\hat{K}_e$ acts on smooth φ on $\mathbf{R}^d \times \mathbf{R}^d$ by $\hat{K}_e(\varphi) \stackrel{\text{def}}{=} \frac{1}{2} \int \mathscr{R}K_e(\varphi_z) \, dz$ where $\varphi_z(\bar{z}) \stackrel{\text{def}}{=} \varphi((z + \bar{z})/2, (z - \bar{z})/2)$. For $r_e \geq 0$, we define

$$\hat{K}_e(x_{e_-}, x_{e_+}) = K_e(x_{e_+} - x_{e_-}) - \sum_{|j|_\mathfrak{s} < r_e} \frac{x_{e_+}^j}{j!} D^j K_e(-x_{e_-}) \,. \qquad (3.3)$$

To lighten notation we assume that the κ_e are always symmetric functions of their arguments. With these notations the key quantity of interest is as follows.

$$\mathcal{I}^G(\varphi_\lambda, K, \kappa) \stackrel{\text{def}}{=} \int_{(\mathbf{R}^d)^{\mathcal{V}_0}} \prod_{e \in \mathcal{E}_2} \hat{K}_e(x_{e_-}, x_{e_+}) \prod_{e \in \mathcal{E}_h} \kappa_e(e) \prod_{v \in \mathcal{V}_\star \setminus \{0\}} \varphi_\lambda(x_v) \, dx \,, \qquad (3.4)$$

where $x_v \in \mathbf{R}^d$ is the point corresponding to the vertex $v \in \mathcal{V}$.

---

[7]We sometimes use the term edge to refer to any element of $\mathcal{E}$, not just the elements of $\mathcal{E}_2$.

[8]This requirement allows us to ensure that we never need estimates on derivatives of the kernels associated to our hyperedges.

For any $\bar{\mathcal{V}} \subset \mathcal{V}$, the subsets $\mathcal{E}^{\uparrow}(\bar{\mathcal{V}})$ and $\mathcal{E}^{\downarrow}(\bar{\mathcal{V}})$ of $\mathcal{E}$ are defined in the same way as in [11], namely, $\mathcal{E}^{\uparrow}(\bar{\mathcal{V}}) = \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} = e_- \ \& \ r_e > 0\}$ and $\mathcal{E}^{\downarrow}(\bar{\mathcal{V}}) = \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} = e_+ \ \& \ r_e > 0\}$, in particular $\mathcal{E}^{\uparrow}(\bar{\mathcal{V}}), \mathcal{E}^{\downarrow}(\bar{\mathcal{V}}) \subset \mathcal{E}_2$ by our assumption on the label r_e. We also set

$$\mathcal{E}_0(\bar{\mathcal{V}}) \stackrel{\text{def}}{=} \{e \in \mathcal{E} : e \subset \bar{\mathcal{V}}\} , \text{ and } \mathcal{E}(\bar{\mathcal{V}}) \stackrel{\text{def}}{=} \{e \in \mathcal{E} : e \cap \bar{\mathcal{V}} \neq \emptyset\} . \tag{3.5}$$

We use the shorthands $r_e^+ = (r_e \vee 0)$ and $r_e^- = -(r_e \wedge 0)$. We now state our main assumptions on the labels $(a_e, r_e)_{e \in \mathcal{E}}$.

Assumption 3.5. *The labelled graph* $(\mathcal{V}, \mathcal{E})$ *satisfies the following properties.*

1. *For every edge* $e \in \mathcal{E}_2$, *one has* $a_e - r_e^- < |\mathfrak{s}|$.

2. *For every subset* $\bar{\mathcal{V}} \subset \mathcal{V}_0$ *of cardinality at least* 3, *one has*

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{V}})} a_e < |\mathfrak{s}| (|\bar{\mathcal{V}}| - 1) . \tag{A.2}$$

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

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

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

$$\sum_{e \in \mathcal{E}(\bar{\mathcal{V}}) \setminus \mathcal{E}^{\downarrow}(\bar{\mathcal{V}})} a_e + \sum_{e \in \mathcal{E}^{\uparrow}(\bar{\mathcal{V}})} r_e - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{\mathcal{V}})} (r_e - 1) > |\mathfrak{s}| |\bar{\mathcal{V}}| . \tag{A.4}$$

Remark 3.6. The second assumption above is automatic for $\bar{\mathcal{V}} = e \in \mathcal{E}_h$ since condition (A.2) then asks that $|\mathfrak{s}| |\bar{\mathcal{V}}|/2 < |\mathfrak{s}| (|\bar{\mathcal{V}}| - 1)$ which follows from $|e| = |\bar{\mathcal{V}}| \geq 3$.

The main result in this subsection is the following.

Theorem 3.7. *Under Assumption 3.5 we have the bound*

$$|\mathcal{I}^G(\varphi_\lambda, K, \kappa)| \lesssim \lambda^\alpha \prod_{e \in \mathcal{E}_2} \|K_e\|_{a_e;p} \prod_{e \in \mathcal{E}_h} \|\kappa_e\|_{a_e} \tag{3.6}$$

where $\alpha = |\mathfrak{s}||\mathcal{V} \setminus \mathcal{V}_\star| - \sum_{e \in \mathcal{E}} a_e$, $\lambda \in (0, 1]$, $p = \max\{|r_e| : e \in \mathcal{E}\} + 1$, *and the proportionality constant only depends on the structure of* G *and the labels* r_e.

Multiscale expansion

The proof of Theorem 3.7 is by a multiscale analysis implemented by a scale decomposition of all kernels. For $e \in \mathcal{E}_2$ the kernels $\hat{K}_e$ are decomposed into an infinite collection of kernels $\hat{K}_e^{(\mathbf{n})}$ with $\mathbf{n} \in \mathbf{N}^3$ just as in [11, Lemma A.4,A.5]. We remark that as in [11], for an edge $e \in \mathcal{E}_2$ without any kernel K_e associated, the kernel $\hat{K}_e^{(\mathbf{n})}$ is still defined and we set $(a_e, r_e) = (0, 0)$.

We also implement a multiscale decomposition for the κ_e as follows.

Definition 3.8. *Given* $e = \{v_1, \ldots, v_{|e|}\} \in \mathcal{E}_h$, *let me*

$$\mathbf{n}_e = \{\mathbf{n}_{i,j}\}_{1 \leq i < j \leq |e|} \in (\mathbf{N}^3)^{\frac{1}{2}|e|(|e|-1)}$$

where $\mathbf{n}_{i,j} = \mathbf{n}_{\{v_i, v_j\}} \in \mathbf{N}^3$. *We define* $\kappa_e^{(\mathbf{n}_e)}$ *as follows. We set* $\kappa_e^{(\mathbf{n}_e)} = 0$ *unless* $\mathbf{n}_{i,j} = (m_{i,j}, 0, 0)$ *for every* $i < j$; *in the latter case, we set*

$$\kappa_e^{(\mathbf{n}_e)}(x_{v_1}, \ldots, x_{v_N}) \stackrel{\text{def}}{=} \kappa_e(x_{v_1}, \ldots, x_{v_N}) \prod_{i \neq j} \Psi^{(m_{i,j})}(x_{v_i} - x_{v_j}) , \tag{3.7}$$

where $N = |e|$, $\Psi^{(n)}$ *is the cutoff function supported in the annulus of radius* $\sim 2^{-n}$ *with* k-*th derivative bounded by* $2^{|k|_\mathfrak{s} n}$, *and* $\sum_n \Psi^{(n)} = 1$.

We can then decompose our generalized convolution as

$$
\mathcal{I}^G(\varphi_\lambda, K, \kappa) = \sum_{\mathbf{n}} \int_{(\mathbf{R}^d)^{\mathcal{V}_0}} \prod_{e \in \mathcal{E}_2} \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}) \prod_{e \in \mathcal{E}_h} \kappa_e^{(\mathbf{n}_e)}(e) \prod_{v \in \mathcal{V}_\star \setminus \{0\}} \varphi_\lambda(x_v)\, dx.
$$

As in [11] for $\lambda \in (0, 1]$, let $\mathcal{N}_\lambda$ be the set of $\mathbf{n}$ such that $2^{-|\mathbf{n}_e|} \leq \lambda$ for every $e = (0, v)$ with $v \in \mathcal{V}_\star \setminus \{0\}$. Since φ_λ can be viewed as a kernel with $a = 0$ and norm being $\lambda^{-|\mathfrak{s}|}$, it suffices for the proof of Theorem 3.7 to show

$$
|\mathcal{I}_\lambda^G(K, \kappa)| \lesssim \lambda^\alpha \prod_{e \in \mathcal{E}_2} \|K_e\|_{a_e; p} \prod_{e \in \mathcal{E}_h} \|\kappa_e\|_{a_e} \tag{3.8}
$$

where $\alpha = |\mathfrak{s}||\mathcal{V}_0| - \sum_{e \in \mathcal{E}} a_e$ and $\mathcal{I}_\lambda^G(K, \kappa) = \sum_{\mathbf{n} \in \mathcal{N}_\lambda} \mathcal{I}_\lambda^{\mathbf{n}, G}(K, \kappa)$ with

$$
\mathcal{I}_\lambda^{\mathbf{n}, G}(K, \kappa) \overset{\text{def}}{=} \int_{(\mathbf{R}^d)^{\mathcal{V}_0}} \prod_{e \in \mathcal{E}_2} \hat{K}_e^{(\mathbf{n}_e)} \prod_{e \in \mathcal{E}_h} \kappa_e^{(\mathbf{n}_e)}\, dx. \tag{3.9}
$$

Multiscale clustering and coalescence trees

As in [11] one can associate a coalescence tree $(T, \ell) \in \mathcal{T}(\mathcal{V})$ to any collection of vertex positions $\{x_v\}_{v \in \mathcal{V}_0}$ with $x_v \in \mathbf{R}^d$ via a coalescing process. For any two nodes u, v of T, $u \wedge v$ denotes the common ancestor of u, v. For any edge $e \in \mathcal{E}$ of the graph (may be hyperedge), $e_\uparrow$ is the common ancestor of all the leaves for the points in e, and $e_\Uparrow$ is the immediate ancestor of $e_\uparrow$.

Given a labelled tree $(T, \ell) \in \mathcal{T}(\mathcal{V})$ and a constant $c > 0$, we define the set $\mathcal{N}(T, \ell)$ of functions $\mathbf{n} : \mathcal{V}^2 \to \mathbf{N}^3$ as in [11] with the additional constraint that for every $e \in \mathcal{E}_h$ and every $\{v, w\} \subseteq e$ we enforce $\mathbf{n}_{(v, w)} = (m, 0, 0)$ with $|m - \ell_{v \wedge w}| \leq c$. If $\{v, w\} \notin \mathcal{E}_2$ and $\{v, w\} \not\subset e$ for all $e \in \mathcal{E}_h$, then the set $\mathcal{N}(T, \ell)$ imposes no requirement on $\mathbf{n}_{(v, w)}$.

Lemma 3.9. *There exists $c > 0$ such that the following holds: for every $\mathbf{n}$ with the property that integral in (3.9) is non-vanishing, there exists an element $(T, \ell) \in \mathcal{T}(\mathcal{V})$ with $\mathbf{n} \in \mathcal{N}(T, \ell)$.*

Proof. The only difference in our setting versus that of [11, Lemma A.9] are the additional constraints imposed by the requiring the support of $\kappa^{(\mathbf{n}_e)}$ to be non-empty for every $e \in \mathcal{E}_h$, however the argument remains exactly the same. $\qquad\square$

Lemma 3.9 allows us to bound $\mathcal{I}_\lambda^G$ by a sum over labelled trees:

$$
|\mathcal{I}_\lambda^G(K, \kappa)| \lesssim \sum_{(T, \ell) \in \mathcal{T}_\lambda(\mathcal{V})} \sum_{\mathbf{n} \in \mathcal{N}(T, \ell)} |\mathcal{I}_\lambda^{\mathbf{n}, G}(K, \kappa)|, \tag{3.10}
$$

where $\mathcal{T}_\lambda(\mathcal{V}) \subset \mathcal{T}(\mathcal{V})$ is the set of coalescence trees such that $2^{-\ell_{v \wedge w}} \leq \lambda$ for any $v, w \in \mathcal{V}_\star$. As in [11], when one wants to implement negative renormalizations to get a better (convergent) bound on the contribution from certain problematic labeled trees $(T, \ell) \in \mathcal{T}_\lambda(\mathcal{V})$ the procedure is to replace the integrand appearing in (3.9) with a cleverly chosen function $\tilde{\mathcal{K}}^{\mathbf{n}}(x)$ (see below) which satisfies $\operatorname{supp} \tilde{\mathcal{K}}^{\mathbf{n}} \subset \mathcal{D}(T, \ell)$ and integrates to the same value. We then write

$$
|\mathcal{I}_\lambda^G(K, \kappa)| \lesssim \sum_{(T, \ell) \in \mathcal{T}_\lambda(\mathcal{V})} \left(\prod_{v \in T^\circ} 2^{-\ell_v |\mathfrak{s}|} \right) \sup_{\mathbf{n} \in \mathcal{N}(T, \ell)} \sup_x |\tilde{\mathcal{K}}^{\mathbf{n}, T}(x)|,
$$

where T° is the set of inner nodes of T.

In [11] the key criterion used to get the bound (3.8) is the following lemma. The distinguished node v_* in this lemma will correspond to the internal node of T which is first common ancestor of the leaves $\mathcal{V}_*$.

Lemma 3.10. *([11, Lemma A.10]) Given a coalescence tree T with a fixed distinguished node $v_\star$, and the set of labelings denoted by $\mathcal{N}_\lambda(T)$ satisfying $2^{-\ell_{v_\star}} < \lambda$, together with a function $\eta : T^\circ \to \mathbf{R}$, we write*

$$\mathcal{I}_\lambda(\eta) = \sum_{n \in \mathcal{N}_\lambda(T)} \prod_{v \in T^\circ} 2^{-n_v \eta_v} .$$

Furthermore suppose that the two following conditions on η hold:

1. *For every $u \in T^\circ$, one has $\sum_{v \geq u} \eta(v) > 0$.*
2. *For every $u \in T^\circ$ such that $u \leq v_*$, one has $\sum_{v \not\geq u} \eta(v) < 0$.*

Then it follows that one has $\mathcal{I}_\lambda(\eta) \lesssim \lambda^{|\eta|}$ uniform for $\lambda \in (0, 1]$ where $|\eta| \stackrel{def}{=} \sum_{v \in T^\circ} \eta(v)$.

Since Lemma 3.10 is a result only about coalescence trees, and has nothing to do with the graph $(\mathcal{V}, \mathcal{E})$, we do not need to re-prove this lemma.

The goal is to show that Assumption 3.5 implies that for any labeled tree $(T, \ell) \in \mathcal{T}_\lambda(\mathcal{V})$ we can find an $\eta : T^\circ \mapsto \mathbf{R}$ such that: (i) the bound

$$|\mathcal{I}_\lambda^{\mathbf{n}, G}(K, \kappa)| \lesssim \prod_{v \in T^\circ} 2^{-n_v \eta(v)}$$

holds uniform in $\mathbf{n} \in \mathcal{N}(T, \ell)$ and (ii) the above function η satisfies the two conditions in Lemma 3.10.

Definition of η and proof of the theorem

As in [11] let $A^- \subset \mathcal{E}$ be the subset of edges e with $r_e < 0$ such that $e_\uparrow$ only has two descendants e_- and e_+ in the tree T. Given any edge $e = (e_-, e_+)$ and any $r > 0$, we define an operator $\mathscr{Y}_e^r$ acting on sufficiently smooth functions $V : \mathbf{R}^\mathcal{V} \to \mathbf{R}$ by

$$\left(\mathscr{Y}_e^r V\right)(x) = V(x) - \sum_{|k|_\mathfrak{s} < r} \frac{(x_{e_+} - x_{e_-})^k}{k!} \left(D_{e_+}^k V\right)(P_e(x)) ,$$

where D_{e_+} is differentiation with respect to the coordinate x_{e_+} and $\left(P_e(x)\right)_v = x_v$ if $v \neq e_+$ and $\left(P_e(x)\right)_v = x_{e_-}$ if $v = e_+$ (in other words it turns x_{e_+} to x_{e_-}).

We replace the integrand in (3.10) by

$$\tilde{\mathcal{K}}^{(\mathbf{n})}(x) = \left(\mathscr{Y}_{e^{(k)}}^{r_{e^{(k)}}} \cdots \mathscr{Y}_{e^{(1)}}^{r_{e^{(1)}}} \Big(\prod_{e \in \mathcal{E}_2 \setminus A^-} K_e^{(\mathbf{n}_e)} \prod_{e \in \mathcal{E}_h} \kappa_e^{(\mathbf{n}_e)} \Big)\right)(x) \prod_{e \in A^-} K_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+})$$

where $A^- = \{e^{(1)}, \ldots, e^{(k)}\}$. By our assumption (see assumption (b) in the beginning of this subsection), if $e \in A^-$ intersects a hyper-edge, then the intersection is the single vertex e_-. Therefore the operator $\mathscr{Y}_e^r$ leaves $\kappa_e^{(\mathbf{n}_e)}$ unchanged, which is very important in the following proofs.

Define

$$\tilde{\eta}(v) = |\mathfrak{s}| + \sum_{e \in \mathcal{E}} \tilde{\eta}_e(v) , \tag{3.11}$$

$$\tilde{\eta}_e(v) = \eta_e(v) + |r_e| \mathbf{1}_{e \in A^-} \left(\mathbf{1}_{e_\uparrow}(v) - \mathbf{1}_{e_\Uparrow}(v)\right)$$

where

$$\eta_e(v) = -a_e \mathbf{1}_{e_\uparrow}(v) + r_e \left(\mathbf{1}_{e_+ \wedge 0}(v) - \mathbf{1}_{e_\uparrow}(v)\right) \mathbf{1}_{r_e > 0, e_+ \wedge 0 > e_\uparrow}$$
$$+ (1 - r_e - a_e) \left(\mathbf{1}_{e_- \wedge 0}(v) - \mathbf{1}_{e_\uparrow}(v)\right) \mathbf{1}_{r_e > 0, e_- \wedge 0 > e_\uparrow} .$$

Although the definition looks the same as in [11], the $e \in \mathcal{E}$ here may be a hyper-edge, and in the case $e \in \mathcal{E}_h$ we have $\tilde{\eta}_e(v) = -\frac{1}{2}|\mathfrak{s}||e|\mathbf{1}_{e_\uparrow}(v)$. In other words for a hyper-edge $e = \{v_1, \ldots, v_n\}$, we add a weight of $\alpha = |\mathfrak{s}|n/2$ to the first common ancestor of $v_1, \ldots, v_n$.

Lemma 3.11. *The functions $\tilde{\mathcal{K}}^{(\mathbf{n})}$ satisfy the bound*

$$\Big(\prod_{v \in T^\circ} 2^{-\ell_v |\mathfrak{s}|} \Big) \sup_{\mathbf{n} \in \mathcal{N}(T, \ell)} \sup_{x} |\tilde{\mathcal{K}}^{(\mathbf{n})}(x)| \lesssim \prod_{v \in T^\circ} 2^{-\ell_v \tilde{\eta}(v)} \tag{3.12}$$

uniform in $\mathbf{n} \in \mathcal{N}(T, \ell)$.

Proof. Since the operator $\mathscr{Y}_e^r$ leaves $\kappa_e^{(\mathbf{n}_e)}$ unchanged, the functions $\tilde{\mathcal{K}}^{(\mathbf{n})}$ can be factored as

$$\tilde{\mathcal{K}}^{(\mathbf{n})} = \Big(\prod_{e \in A^-} K_e^{(\mathbf{n}_e)} \prod_{e \in \mathcal{E}_h} \kappa_e^{(\mathbf{n}_e)} \Big) \cdot \tilde{\mathcal{K}}_1^{(\mathbf{n})} \tag{3.13}$$

and the last factor here is

$$\tilde{\mathcal{K}}_1^{(\mathbf{n})}(x) \overset{\text{def}}{=} \sum_{k \in \partial A^-} \int_{(\mathbf{R}^d)^{A^-}} \Big(\prod_{e \in A^-} D_{x_{e_+}}^{k_e} \Big) \Big(\prod_{e \notin A^-} K_e^{(\mathbf{n}_e)}(x|y) \Big) \prod_{e \in A^-} \mathcal{Q}_x^{k,e}(dy_e) \tag{3.14}$$

We refer to [11, Lemma A.16] for the precise definition of the notations ∂A^-, $x|y$ and $\mathcal{Q}_x^{k,e}(dy_e)$ appearing above, but only remark that $\tilde{\mathcal{K}}_1^{(\mathbf{n})}$ and the first product on the right hand side of (3.13) can both be bounded as in [11] by the right hand side of (3.12) if $\tilde{\eta}(v)$ were defined as in (3.11) with $\sum_{e \in \mathcal{E}}$ replaced by $\sum_{e \in \mathcal{E}_2}$. By the multiplicative structure of the second factor, it remains to show that for each $e \in \mathcal{E}_h$

$$\sup_{x \in \mathbf{R}^d} |\kappa_e^{(\mathbf{n}_e)}(x)| \lesssim \prod_{v \in T^\circ} 2^{-\ell_v \tilde{\eta}_e(v)} = 2^{-\ell_{e_\uparrow} \tilde{\eta}_e(e_\uparrow)} = \big(2^{\ell_{e_\uparrow}} \big)^{\frac{1}{2}|\mathfrak{s}||e|} .$$

This follows immediately from Definitions 3.1 and 3.8. Note that the cut-off functions $\Psi^{(m_{i,j})}$ in (3.7) impose that the tree $(\bar{T}, \bar{\ell})$ over the $|e|$ vertices of e induced from the tree (T, ℓ) has precisely $e_\uparrow$ as its root. $\qquad\square$

The proof of Theorem 3.7 is finished with the following lemma.

Lemma 3.12. *For each $T \in \mathcal{T}(\mathcal{V})$ the map $\tilde{\eta} : T^\circ \mapsto \mathbf{R}$ as given in (3.11) satisfies the two conditions of Lemma 3.10 and one also has $|\tilde{\eta}| = |\mathfrak{s}| \times (|\mathcal{V}| - 1) - \sum_{e \in \mathcal{E}(\mathcal{V})} a_e$.*

Proof. The proof of [11][Lemma A.19] applies to our situation verbatim so we just give a sketch here. Fix $\nu \in T^\circ$, we write $L_\nu \subset \mathcal{V}$ for the set of leaves which are descendants of ν in T.

When one calculates $\sum_{v \geq \nu} \tilde{\eta}(v)$ the result takes three different forms. If $L_\nu = e$ for some e with $r_e < 0$ then it takes the value $|\mathfrak{s}| - a_e + r_e$. Otherwise, the value depends on whether $0 \notin L_\nu$ or $0 \in L_\nu$ - in the former case the value of the sum is given by the difference of the righthand and lefthand sides of (A.2) of Assumption 3.5 while in the latter it is given by the difference of the righthand and lefthand sides of (A.3) of Assumption 3.5 - in both cases one takes $\bar{\mathcal{V}} = L_\nu$.

On the other hand, if $\nu \leq \nu_*$ then $\sum_{v \not\geq \nu} \tilde{\eta}(v)$ is given by the difference of the righthand and lefthand sides of (A.4) with $\bar{\mathcal{V}} = \mathcal{V} \setminus L_\nu$. $\qquad\square$

3.3 The elementary graphs

In this subsection we will show that Assumption 3.17 imposed on an "elementary graph" will imply that Assumption 3.5 holds for any (much larger) graph constructed from Wick contracting the elementary graph. Basically the graphs showing in the pictures in Section 4 are all examples of elementary graphs.

Definition 3.13. *An elementary graph H is a connected graph labelled with (a_e, r_e) for each edge e as in Subsection 3.2 above, with only one special non-zero vertex $v_\star$, namely $H_\star = \{0, v_\star\}$. Additionally there is a unique distinguished edge $e_\star = \{0, v_\star\}$ attached*

to 0 - other edges may also connect to 0 but they are not called distinguished edges. The set $H_0 = H \setminus \{0\}$ can be decomposed as a union of two disjoint subsets of vertices $H_0 = H_{ex} \sqcup H_{in}$ which we call external and internal vertices, respectively.

We also enforce that $\deg(v) = 1$ for every $v \in H_{ex}$ and $\deg(v) \geq 2$ for every $v \in H_{in}$. Any edge e with $e \cap H_{ex} \neq \emptyset$ will be called an "external edge". Edges $e \neq e_\star$ which are not external edges are called "internal edges". The unique internal vertex connected to an external vertex v will be denoted $i(v)$. We require that $v_\star \in H_{in}$, and that for every external edge e one has $a_e = |\mathfrak{s}|$. [9] We also enforce that for all edges e with $|e| = 2$ one has $a_e < 2|\mathfrak{s}|$.

We can construct graphs $\mathcal{V}$ by *Wick contracting* several copies of H, similarly as in [12], except that we now build hyper-edges over external vertices instead of identifying them. For a set D we denote by $\mathcal{P}(D)$ the set of partitions of D.

Definition 3.14. *Given a set A and an integer $p > 1$, let $\{A^{(i)}\}_{i=1}^{p}$ be p copies of A and let D be their p-fold disjoint union - that is $D = \sqcup_{i=1}^{p} A^{(i)}$. For $\pi \in \mathcal{P}(D)$ we say that*

$$\pi \in \mathcal{P}_w(D; A, p) \subset \mathcal{P}(D)$$

if for every $B \in \pi$, one cannot find $1 \leq i \leq p$ such that $B \subset A^{(i)}$. In other words, we enforce that every block of the partition π must contain elements from at least two different copies of A. In particular, one must have $|B| > 1$.

Definition 3.15. *Suppose that we are given an elementary graph H and an integer $p > 1$. For $1 \leq i \leq p$ let $H^{(i)}$ be a copy of the graph H. Suppose that we are also given a partition $\pi \in \mathcal{P}_w(\sqcup_{i=1}^{p} H_{ex}^{(i)}; H_{ex}, p)$. From H, p, and π we will construct a labeled graph $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ which will be called a p-fold Wick contraction of H.*

To define the vertex set $\mathcal{U}$ we first start with $\sqcup_{i=1}^{p} H^{(i)}$ and then identify all the p copies of the distinguished vertex 0. The edge set is given by

$$\mathcal{E}(\mathcal{U}) \stackrel{def}{=} \left(\sqcup_{i=1}^{p} \mathcal{E}_0(H^{(i)}) \right) \sqcup \mathcal{E}_c(\mathcal{U}) , \qquad \text{where } \mathcal{E}_c(\mathcal{U}) \stackrel{def}{=} \pi.$$

As in the last section we have the decomposition $\mathcal{E}(\mathcal{U}) = \mathcal{E}_2(\mathcal{U}) \sqcup \mathcal{E}_h(\mathcal{U})$.

Each edge $e \in \mathcal{E}(\mathcal{U})$ is naturally associated with a label (a_e, r_e), which is $(|\mathfrak{s}|, -1)$ if $e \in \mathcal{E}_c(\mathcal{U}) \cap \mathcal{E}_2(\mathcal{U})$, or $(|e||\mathfrak{s}|/2, 0)$ if $e \in \mathcal{E}_c(\mathcal{U}) \cap \mathcal{E}_h(\mathcal{U})$, or otherwise inherits the label (a_e, r_e) from H. We also set $\mathcal{U}_0 \stackrel{def}{=} \mathcal{U} \setminus \{0\}$ and $\mathcal{U}_\star \subset \mathcal{U}$ to be given by the set $\sqcup_{i=1}^{p} H_\star^{(i)}$ with all the copies of 0 identified.

While we have defined enough structure to formulate Assumption 3.5 for $\mathcal{G}$, it turns out that this labeled graph does *not* quite satisfy this assumption in general; the second inequality will be violated whenever one has a block $B = \{u, v\}$ in π of cardinality 2. Pictorially one then has

$$
\begin{array}{ccc}
u & \!\!\!\!\!\text{---}\!\!\!\!\! & v \\
\diagup & & \diagdown \\
i(u) & & i(v)
\end{array}
\tag{3.15}
$$

where as before, for an external vertex $w \in H_{ex}^{(i)}$ we denote by $i(w)$ the unique element of $H_{in}^{(i)}$ which is connected to w by an edge of $\mathcal{E}_0(H^{(i)})$. In this scenario a subset $V \subset \{u, v, i(u), i(v)\}$ with $|V| > 2$ is called a *bad chain* for $\mathcal{G}$.

The outer two edges of (3.15) each carry a label $(|\mathfrak{s}|, -1)$. While they are divergent by power counting we expect them to be integrable since they represent approximate identities. The solution is to perform the integration of vertices u and v *before* we

---

[9]In practice one can always attach a new edge with label $|\mathfrak{s}|$ representing the Dirac function to an external vertex to ensure this assumption holds.

perform our multiscale analysis. Pictorially we replace (3.15) by

$$i(u) \quad\rule{2cm}{0.4pt}\quad i(v)$$

which carries a label $(|\mathfrak{s}|, -1)$.

Definition 3.16. *Given a a partition $\pi \in \mathcal{P}_w(\sqcup_{i=1}^p H_{ex}^{(i)}; H_{ex}, p)$ we define a labeled graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}'(\mathcal{V}))$ which we call a reduced p-fold Wick contraction of H. Let $\mathcal{G} = (\mathcal{U}, \mathcal{E}(\mathcal{U}))$ denote the corresponding non-reduced p-fold Wick contraction as in Definition 3.15. $\mathcal{G}'$ represents the reduced graph obtained after we have integrated out the following variables* [10]*

$$\mathcal{U}_{rem} = \bigsqcup_{B \in \pi \cap \mathcal{E}_2(\mathcal{U})} B.$$

The vertex set is given by $\mathcal{V} = \mathcal{U} \setminus \mathcal{U}_{rem}$, while the edge set is given by

$$\mathcal{E}'(\mathcal{V}) \stackrel{\text{def}}{=} \left\{ e \in \sqcup_{i=1}^p \mathcal{E}_0(H^{(i)}) : e \cap \mathcal{U}_{rem} = \emptyset \right\} \sqcup \mathcal{E}'_c(\mathcal{V}),$$

$$\text{where} \quad \mathcal{E}'_c(\mathcal{V}) \stackrel{\text{def}}{=} \left(\pi \setminus \mathcal{E}_2(\mathcal{U}) \right) \sqcup \{\{i(u), i(v)\} : \{u, v\} \in \pi \cap \mathcal{E}_2(\mathcal{U})\}.$$

As before, we write $\mathcal{E}'(\mathcal{V}) = \mathcal{E}'_2(\mathcal{V}) \sqcup \mathcal{E}'_h(\mathcal{V})$. The edges of $\mathcal{E}'_c(\mathcal{V}) \cap \mathcal{E}'_2(\mathcal{V})$ are given the label $(|s|, -1)$. Since $\mathcal{E}'(\mathcal{V}) \setminus [\mathcal{E}'_c(\mathcal{V}) \cap \mathcal{E}'_2(\mathcal{V})] \subset \mathcal{E}(\mathcal{U})$ we can just let all the other edges inherit their labeling from $\mathcal{G}$.

Finally we let $\mathcal{V}_0 = \mathcal{V} \setminus \{0\}$ and $\mathcal{V}_\star = \mathcal{U}_\star$ and also define a retraction map $\mathfrak{r} : \mathcal{U} \mapsto \mathcal{V}$ associated to $\mathcal{G}$ given by

$$\mathfrak{r}(a) \stackrel{\text{def}}{=} \begin{cases} i(a) & \text{if } a \in \mathcal{U}_{rem}, \\ a & \text{otherwise} . \end{cases}$$

We now state our counterpart of Assumption 3.5 for the elementary graphs H. For various sets of edges we recall the notation (3.5).

Assumption 3.17. *The labelled graph $(H, \mathcal{E})$ satisfies the following properties.*

1. *For every edge $e \in \mathcal{E}_2$, one has $a_e - r_e^- < |\mathfrak{s}|$ where $r_e^- = -(r_e \wedge 0)$.*

2. *For every subset $\bar{H} \subset H_0$ of cardinality at least 3, one has*

$$\sum_{e \in \mathcal{E}_0(\bar{H})} a_e < |\mathfrak{s}| \left(|\bar{H}_{in}| + \frac{1}{2}(|\bar{H}_{ex}| - 1 - \mathbf{1}_{\bar{H}_{ex}=\emptyset}) \right). \tag{H.2}$$

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

$$\sum_{e \in \mathcal{E}_0(\bar{H})} a_e + \sum_{e \in \mathcal{E}^\uparrow(\bar{H})} (a_e + r_e - 1) - \sum_{e \in \mathcal{E}^\downarrow(\bar{H})} r_e < |\mathfrak{s}| \left(|\bar{H}_{in}| + \frac{1}{2}|\bar{H}_{ex}| \right). \tag{H.3}$$

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

$$\sum_{e \in \mathcal{E}(\bar{H}) \setminus \mathcal{E}^\downarrow(\bar{H})} a_e + \sum_{e \in \mathcal{E}^\uparrow(\bar{H})} r_e - \sum_{e \in \mathcal{E}^\downarrow(\bar{H})} (r_e - 1) > |\mathfrak{s}| \left(|\bar{H}_{in}| + \frac{1}{2}|\bar{H}_{ex}| \right). \tag{H.4}$$

Our goal is to show that Assumption 3.17 on an elementary graph H will imply assumptions 3.5 for any reduced p-fold Wick contraction built from H. It is more convenient to instead prove a weaker version of Assumption 3.5 for the corresponding non-reduced Wick contraction. We have the following lemma which is a straightforward consequence of our definitions.

---

[10] In plain words, $\mathcal{U}_{rem}$ is the set of vertices like u and v in (3.15) for which we want to remove from the vertex set.

Lemma 3.18. *Let H be an elementary graph, $\mathcal{G} = (\mathcal{U}, \mathcal{E}(\mathcal{U}))$ be a p-fold Wick contraction of H, and $\mathcal{G}' = (\mathcal{V}, \mathcal{E}'(\mathcal{V}))$ be the corresponding reduced p-fold Wick contraction. Suppose that $\mathcal{G}$ satisfies items 1,3, and 4 of Assumption 3.5 and that Eq. (A.2) holds for every subset $\bar{\mathcal{U}} \subset \mathcal{U}_0$ of cardinality at least 3 which is not a bad chain. Then the graph $\mathcal{G}'$ satisfies Assumption 3.5.*

Proof. The fact that $\mathcal{G}'$ will satisfy item 1 is quite clear so we focus on the other items. The first key point is that for any of the conditions 2,3, and 4 of Assumption 3.5, given an appropriate $\bar{\mathcal{V}} \subset \mathcal{V}$ the difference between the LHS and RHS's of the needed inequality remains the same if one replaces $\bar{\mathcal{V}}, \mathcal{E}'(\bar{\mathcal{V}}), \mathcal{E}'_0(\bar{\mathcal{V}})$, and $\mathcal{E}'^{\downarrow}(\bar{\mathcal{V}})$ by $\bar{\mathcal{U}} \stackrel{\text{def}}{=} \mathfrak{r}^{-1}(\bar{\mathcal{V}}), \mathcal{E}(\bar{\mathcal{U}}), \mathcal{E}_0(\bar{\mathcal{U}})$, and $\mathcal{E}^{\downarrow}(\bar{\mathcal{U}})$, respectively - one must have $|\bar{\mathcal{U}}| = 2n$ in which case making this switch increases both the LHS and RHS by $2n|\mathfrak{s}|$. The second point is that for $|\bar{\mathcal{V}}| > 2$ the set $\mathfrak{r}^{-1}(\bar{\mathcal{V}})$ will not be a bad chain. $\square$

We will denote the p copies of H by $H^{(1)}, \dots, H^{(p)}$ and write $H_{in}^{(j)}$ and $H_{ex}^{(j)}$ for the sets of internal vertices and external vertices of these copies.

Theorem 3.19. *Let H be an elementary graph satisfying Assumption 3.17. Let $p \geq 2$, and let $(\mathcal{V}, \mathcal{E})$ be a reduced p-fold Wick contraction of H constructed as in Definition 3.16. Then $(\mathcal{V}, \mathcal{E})$ satisfies Assumption 3.5. In particular, the conclusion (3.6) of Theorem 3.7 holds.*

Proof. Let $(\mathcal{U}, \mathcal{E}(\mathcal{U}))$ be as in Definition 3.15, namely, the p-fold Wick contraction from which $(\mathcal{V}, \mathcal{E}'(\mathcal{V}))$ is constructed. It suffices to prove that $(\mathcal{U}, \mathcal{E}(\mathcal{U}))$ satisfies the assumptions of Lemma 3.18.

Item 1 of the assumption on $(H, \mathcal{E})$, together with the definition of the labels for edges in $\mathcal{E}_c(\mathcal{U})$, obviously implies item 1 of the assumption on $(\mathcal{U}, \mathcal{E})$.

To prove item 2 let $\bar{\mathcal{U}} \subset \mathcal{U}_0$ be vertex set of cardinality at least 3 which is not a bad chain. We aim to show (A.2). We first show that it is sufficient to treat the case where $\bar{\mathcal{U}}$ is *connected*. First we claim it suffices to treat the case where no connected component of $\bar{\mathcal{U}}$ is a bad chain - if there is such at least one such connected component it suffices to prove (A.2) for the union of all the other connected components of $\bar{\mathcal{U}}$.

If $\bar{\mathcal{U}}$ has connected components of size 1 then it suffices to check (A.2) for the smaller vertex set where one drops these components; and the same holds for components of size 2 (here one uses the assumption $|a_e| < 2|\mathfrak{s}|$ in Definition 3.13). If all the components of $\bar{\mathcal{U}}$ have cardinality at least 3, then if (A.2) holds for each of these components then summing up these bounds yields an even stronger bound for $\bar{\mathcal{U}}$. This covers all the disconnected cases so we assume that $\bar{\mathcal{U}}$ is *connected*.

Let $\bar{H}^{(i)} \stackrel{\text{def}}{=} \bar{\mathcal{U}} \cap H^{(i)}$. Let $J_1, J_2, J_{\geq 3}$ be three disjoint subsets of $\{1, \dots, p\}$, such that $|\bar{H}^{(i)}| = 1$ for each $i \in J_1$, and $|\bar{H}^{(i)}| = 2$ for each $i \in J_2$, and $|\bar{H}^{(i)}| \geq 3$ for each $i \in J_{\geq 3}$. The LHS of (A.2) is then

$$
\begin{aligned}
\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e &= \sum_{i \in J_1} \sum_{e \in \mathcal{E}_0(\bar{H}^{(i)})} a_e + \sum_{i \in J_2} \sum_{e \in \mathcal{E}_0(\bar{H}^{(i)})} a_e + \sum_{i \in J_{\geq 3}} \sum_{e \in \mathcal{E}_0(\bar{H}^{(i)})} a_e + \sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}}) \cap \mathcal{E}_c(\mathcal{U})} a_e \\
&\leq |J_2||\mathfrak{s}| + \sum_{i \in J_{\geq 3}} |\mathfrak{s}| \left[|\bar{H}_{in}^{(i)}| + \frac{1}{2}(|\bar{H}_{ex}^{(i)}| - 1 - \mathbf{1}_{\bar{H}_{ex}^{(i)} = \varnothing}) \right] + \frac{|\mathfrak{s}|}{2} \sum_{i \in J_1 \cup J_2 \cup J_{\geq 3}} |\bar{H}_{ex}^{(i)}|
\end{aligned}
\tag{3.16}
$$

For going to the bottom lines we use the following reasoning for each of the terms on the first line: (i) the sum over J_1 is obviously zero, (ii) since $|\bar{\mathcal{U}}| \geq 3$ and is assumed to be connected, the edges involved in the J_2-summation are all external with labels $|\mathfrak{s}|$, (iii) one can apply (H.2) for the sum over $J_{\geq 3}$, and (iv) one has

$$
\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}}) \cap \mathcal{E}_c(\mathcal{U})} a_e = |\mathfrak{s}|/2 \times \sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}}) \cap \mathcal{E}_c(\mathcal{U})} |e| \leq |\mathfrak{s}|/2 \times \sum_{i \in J_1 \cup J_2 \cup J_{\geq 3}} |\bar{H}_{ex}^{(i)}|.
\tag{3.17}
$$

Observe that if $J_{\geq 3} \neq \emptyset$ then by (H.2) the inequality in (3.16) is actually strict.

We now deal with the case where $J_1 = J_2 = \emptyset$ and $|J_{\geq 3}| = n \geq 1$. If $J_{\geq 3} = \{i\}$ then the last term on the first line of (3.16) must vanish so

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e < |\mathfrak{s}| \left(|\bar{H}_{in}^{(i)}| + \frac{1}{2}(|\bar{H}_{ex}^{(i)}| - 1 - \mathbf{1}_{\bar{H}_{ex}^{(i)} = \emptyset}) \right) \tag{3.18}$$

with the RHS being bounded above by $|\mathfrak{s}|(|\bar{\mathcal{U}}| - 1)$, as desired. If $n > 1$ we must have $\bar{H}_{ex}^{(i)} \neq \emptyset$ for all $i \in J_{\geq 3}$ and our claim follows from the fact that

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e < |\mathfrak{s}| \times \sum_{i \in J_{\geq 3}} \left[|\bar{H}_{in}^{(i)}| + (1/2 + 1/2)|\bar{H}_{ex}^{(i)}| - \frac{1}{2} \right] = |\mathfrak{s}|(|\bar{\mathcal{U}}| - n/2). \tag{3.19}$$

Note that in all remaining cases one must have $\bar{H}_{ex}^{(i)} \neq \emptyset$ for $i \in J_1 \sqcup J_2 \sqcup J_{\geq 3}$. The case when $|J_{\geq 3}| \geq 1$ and $J_1 \sqcup J_2 \neq \emptyset$ follows similarly to the two cases treated above: the upper bounds in (3.18) and (3.19) will apply if we increase them by $(\frac{1}{2}|J_1| + \frac{3}{2}|J_2|) \times |\mathfrak{s}|$ but the quantity $|\mathfrak{s}|(|\bar{\mathcal{U}}| - 1)$ we compare them against goes up by $(|J_1| + 2|J_2|) \times |\mathfrak{s}|$.

Henceforth we assume $J_{\geq 3} = \emptyset$. Suppose that $(|J_1|, |J_2|) = (1, 1)$ or $(0, 2)$. Since $\bar{\mathcal{U}}$ is not a bad chain it must be the case that $\mathcal{E}_0(\bar{\mathcal{U}}) \cap \mathcal{E}_c(\mathcal{U}) = \emptyset$. Then by (3.16) one has $\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e \leq |J_2| \times |\mathfrak{s}|$ which is strictly smaller than $|\mathfrak{s}| \times (|J_1| + 2|J_2| - 1)$.

In the remaining scenarios $J_{\geq 3} = \emptyset$ and $(|J_1|, |J_2|) \neq (1, 1)$ or $(0, 2)$ - (A.2) then follows by observing that

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e \leq |\mathfrak{s}| \times \left(\frac{3}{2} \times |J_2| + \frac{1}{2}|J_1| \right) < (2|J_2| + |J_1| - 1) \times |\mathfrak{s}| = |\mathfrak{s}|(|\bar{\mathcal{U}}| - 1).$$

We now turn to proving that (H.3) implies (A.3). Let $\bar{\mathcal{U}} \subset \mathcal{V}$ be of cardinality at least two with $\bar{\mathcal{U}} \ni 0$. Let J be the set of $j \in \{1, \ldots, p\}$ with the property that $|\bar{\mathcal{U}} \cap H^{(j)}| \geq 2$. We immediately have

$$\sum_{e \in \mathcal{E}^{\uparrow}(\bar{\mathcal{U}})} (a_e + r_e - 1) - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{\mathcal{U}})} r_e = \sum_{i \in J} \left(\sum_{e \in \mathcal{E}^{\uparrow}(\bar{H}^{(i)})} (a_e + r_e - 1) - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{H}^{(i)})} r_e \right). \tag{3.20}$$

Applying item 3 of Assumption 3.17 and the bound (3.17) with the summation set replaced by J gives

$$\sum_{e \in \mathcal{E}_0(\bar{\mathcal{U}})} a_e + \sum_{e \in \mathcal{E}^{\uparrow}(\bar{\mathcal{U}})} (a_e + r_e - 1) - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{\mathcal{U}})} r_e$$

$$\leq \sum_{i \in J} \left(\sum_{e \in \mathcal{E}_0(\bar{H}^{(i)})} a_e + \sum_{e \in \mathcal{E}^{\uparrow}(\bar{H}^{(i)})} (a_e + r_e - 1) - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{H}^{(i)})} r_e + \frac{|\mathfrak{s}|}{2}|\bar{H}_{ex}^{(i)}| \right)$$

$$< |\mathfrak{s}| \times \sum_{i \in J} \left(|\bar{H}_{in}^{(i)}| + |\bar{H}_{ex}^{(i)}| \right) = |\mathfrak{s}|(|\bar{\mathcal{U}}| - 1),$$

where for the last equality we note that 0 is neither internal nor external.

We now show (H.4) implies (A.4). Suppose $\bar{\mathcal{U}} \subset \mathcal{U} \setminus \mathcal{U}_\star$. Using the bound

$$\sum_{e \in \mathcal{E}(\bar{\mathcal{U}}) \cap \mathcal{E}_c(\mathcal{U})} a_e \geq \sum_{i=1}^{p} |\bar{H}_{ex}^{(i)}|,$$

a decomposition similar to (3.20), and applying (H.4) to each non-empty $\bar{H}^{(i)}$ gives

$$\sum_{e \in \mathcal{E}(\bar{\mathcal{U}})} a_e + \sum_{e \in \mathcal{E}^{\uparrow}(\bar{\mathcal{U}})} r_e - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{\mathcal{U}})} (r_e - 1)$$

$$\geq \sum_{i=1}^{p} \left[\sum_{e \in \mathcal{E}(\bar{H}^{(i)}) \setminus \mathcal{E}^{\downarrow}(\bar{H}^{(i)})} a_e + \sum_{e \in \mathcal{E}^{\uparrow}(\bar{H}^{(i)})} r_e - \sum_{e \in \mathcal{E}^{\downarrow}(\bar{H}^{(i)})} (r_e - 1) + \frac{|\mathfrak{s}|}{2} |\bar{H}_{ex}^{(i)}| \right]$$

$$> |\mathfrak{s}| \times \sum_{i=1}^{p} \left(|\bar{H}_{in}^{(i)}| + |\bar{H}_{ex}^{(i)}| \right) \;=\; |\mathfrak{s}| \times |\bar{\mathcal{U}}|.$$

$\square$

We now state a lemma which is a partial converse to the above theorem in the case of symmetric pairings of elementary graphs. A symmetric pairing of H is a special type of Wick contraction - one has $p = 2$ and $\pi = \{\{v^{(1)}, v^{(2)}\} : \; v \in H_{ex}\}$ where for $v \in H$ we write $v^{(i)}$ for v's instantiation in $H^{(i)}$.

Lemma 3.20. *For a given elementary graph H, let $(\mathcal{V}, \mathcal{E}')$ be the reduced two-fold Wick contraction corresponding to the symmetric pairing of H. Suppose that $(\mathcal{V}, \mathcal{E}')$ satisfies Assumption 3.5, then H satisfies Assumption 3.17.*

Proof. Clearly H satisfies the first item of Assumption 3.17 as a consequence of $(\mathcal{V}, \mathcal{E}')$ satisfying the first item of Assumption 3.5.

For the other items we first observe that $\mathcal{V} = H_{in}^{(1)} \sqcup H_{in}^{(2)} \sqcup \{0\}$ and

$$\mathcal{E}' \setminus \left[\mathcal{E}_0(H_{in}^{(1)}) \sqcup \mathcal{E}_0(H_{in}^{(2)}) \right] = \cup_{v \in H_{ex}} \{i^{(1)}(v), i^{(2)}(v)\} \tag{3.21}$$

where $i^{(j)}(v)$ is the instantiation of $i(v)$ in $H_{ex}^{(j)}$.

For item 2 fix some $\bar{H} \subseteq H_0$. If $\bar{H} \cap H_{ex} = \emptyset$ then (H.2) follows by applying (A.2) to $\bar{\mathcal{V}} = \bar{H}^{(1)}$ - the right hand side of (H.2) is equal to $|\mathfrak{s}| (|\bar{H}| - 1)$.

We claim that given appropriate $\bar{H} \subset H$, the needed criteria for the other case of item 2 (where $\bar{H} \cap H_{ex} \neq \emptyset$) or items 3 or 4 are equivalent to the corresponding items of Assumption 3.5 for the set $\bar{\mathcal{V}} \subset \mathcal{V}$, where $\bar{\mathcal{V}} \overset{\text{def}}{=} \bar{H}_{in}^{(1)} \sqcup \bar{H}_{in}^{(2)}$ if $\bar{H} \not\ni 0$ and $\bar{\mathcal{V}} \overset{\text{def}}{=} \bar{H}_{in}^{(1)} \sqcup \bar{H}_{in}^{(2)} \sqcup \{0\}$ if $\bar{H} \ni 0$.

Since the arguments are so similar we only do the other case of item 2 here. Suppose $\bar{H} \subset H_0$ with $\bar{H} \cap H_{ex} \neq \emptyset$. Defining $\bar{\mathcal{V}}$ as described and then applying (A.2) together with (3.21) to $\bar{\mathcal{V}}$ gives

$$2 \sum_{e \in \mathcal{E}_0(\bar{H})} a_e - |\mathfrak{s}| \times |\bar{H}_{ex}| = \sum_{e \in \mathcal{E}_0(\bar{\mathcal{V}})} a_e < |\mathfrak{s}|(|\bar{\mathcal{V}}| - 1) = |\mathfrak{s}|(2|\bar{H}_{in}| - 1),$$

which is exactly the desired condition for $\bar{H}$. $\square$

Remark 3.21. The above lemma establishes a type of "hypercontractive" or "equivalence of moments" bound in the non-Gaussian setting.

We state two lemmas and a remark before proceeding to the stochastic estimates.

Lemma 3.22. *Suppose that the elementary graph $H = (\mathcal{V}, \mathcal{E})$ satisfies Assumption 3.17, and that $H' = (\mathcal{V}, \mathcal{E}')$ is another elementary graph which has the same vertex set $\mathcal{V}$. Assume that there exist internal edges $e_1, e_2 \in \mathcal{E}$ with $a_{e_j} = |e_j||\mathfrak{s}|/2$ and $r_{e_j} \leq 0$ for $j = 1, 2$, and that $\mathcal{E}' = (\mathcal{E} \setminus \{e_1, e_2\}) \cup e$ with $e = e_1 \cup e_2$ and $a_e = |e||\mathfrak{s}|/2$. In plain words H' is formed by merging e_1, e_2 into one hyper-edge e. Then, H' also satisfies Assumption 3.17.*

Proof. For items 1-3 of Assumption 3.17 and any allowable subset $\bar{H} \subset \mathcal{V}$, the LHS of the bounds for $\bar{H}$ as a subgraph of H' is always smaller or equal to the LHS for $\bar{H}$ as a subgraph of H, while the RHS remains the same.

For item 4 and any allowable subset $\bar{H}$, the LHS of the bound for $\bar{H}$ as a subgraph of H' is always larger or equal to the LHS for $\bar{H}$ as a subgraph of H, while the RHS again remains the same.

Therefore if H satisfies Assumption 3.17, then after merging e_1, e_2 into one hyperedge the new graph also satisfies Assumption 3.17. □

Lemma 3.23. *For graphs H such that $H_{ex} = \emptyset$, Assumption 3.17 for H is equivalent with Assumption 3.5 for $\mathcal{V} = H$.*

Proof. Immediate upon comparing the two assumptions with the condition $H_{ex} = \emptyset$. □

Remark 3.24. When we do our stochastic estimates there are symbols τ and self-contractions π on τ such that the elementary graph $H_{\tau,\pi}$ will have a bad chain, failing to abide Assumption 3.17. However, it is easy to see that the finite set of bad chains of $H_{\tau,\pi}$ can be eliminated by integrating a noise vertex in each one yielding an abiding $\tilde{H}_{\tau,\pi}$. Below this is done implicitly whenever the scenario arises.

4 Application: Wong-Zakai theorem for non-Gaussian noise

We now apply the machinery of the preceding sections to prove Theorem 1.2.

4.1 Renormalization

We fix $\mathcal{T}_0 \subset \mathcal{T}$ to be the span of $\mathcal{W}_0$ where

$$\mathcal{W}_0 \stackrel{\text{def}}{=} \left\{ \circ, \, \mathord{\vcenter{\hbox{τ_1}}}, \, \mathord{\vcenter{\hbox{τ_2}}}, \, \mathord{\vcenter{\hbox{τ_3}}}, \, \otimes, \, \mathord{\vcenter{\hbox{τ_4}}}, \, \mathord{\vcenter{\hbox{τ_5}}}, \, \mathord{\vcenter{\hbox{τ_6}}}, \, \mathord{\vcenter{\hbox{τ_7}}}, \, \mathord{\vcenter{\hbox{τ_8}}}, \, \mathord{\vcenter{\hbox{τ_9}}}, \, \mathbf{1}, \, \mathord{\vcenter{\hbox{τ_{10}}}}, \, \mathord{\vcenter{\hbox{τ_{11}}}}, \, \mathord{\vcenter{\hbox{τ_{12}}}}, \, \mathord{\vcenter{\hbox{τ_{13}}}} \right\}.$$

Our maps $M : \mathcal{T}_0 \mapsto \mathcal{T}_0$ will be of the form

$$M = \exp\left(-\sum_{i=1}^{7} \ell_i L^{(i)}\right) \tag{4.1}$$

where $\{\ell_i\}_{i=1}^7 \subset \mathbf{R}$ and the $\{L^{(i)}\}_{i=1}^7$ are nilpotent linear operators on $\mathcal{T}_0$ given as follows. $L^{(1)}$ is defined in the same way as the map called in L in [10]: it iterates over all occurrences of $\mathord{\vcenter{\hbox{$\bullet$}}}$ as a "subsymbol" of τ and "erases" it in the graphical notation:

$$L^{(1)}\mathord{\vcenter{\hbox{a}}} = \mathbf{1}, \qquad L^{(1)}\mathord{\vcenter{\hbox{b}}} = \mathord{\vcenter{\hbox{c}}}, \qquad L^{(1)}\mathord{\vcenter{\hbox{d}}} = 2\,\mathord{\vcenter{\hbox{e}}},$$
$$L^{(1)}\mathord{\vcenter{\hbox{f}}} = 3\,\mathord{\vcenter{\hbox{g}}}, \qquad L^{(1)}\mathord{\vcenter{\hbox{h}}} = \mathord{\vcenter{\hbox{i}}} + \mathord{\vcenter{\hbox{j}}}, \qquad L^{(1)}\mathord{\vcenter{\hbox{k}}} = \mathord{\vcenter{\hbox{l}}} + \mathord{\vcenter{\hbox{m}}}, \tag{4.2}$$
$$L^{(1)}\mathord{\vcenter{\hbox{n}}} = \mathord{\vcenter{\hbox{o}}} + 2\,\mathord{\vcenter{\hbox{p}}}, \qquad L^{(1)}\mathord{\vcenter{\hbox{q}}} = X_1, \qquad L^{(1)}\mathord{\vcenter{\hbox{r}}} = X_1.$$

Note that there is no term $\mathord{\vcenter{\hbox{$\bullet$}}}$ appearing in $L^{(1)}\mathord{\vcenter{\hbox{s}}}$ because $\mathcal{I}(\mathbf{1}) = 0$, and similarly for the other terms. We furthermore have $L^{(1)}\tau = 0$ for every $\tau \in \mathcal{W}_0$ which is not one of the above nine elements. Regarding the other maps, one has

$$L^{(2)}\mathord{\vcenter{\hbox{a}}} = \mathbf{1}, \quad L^{(2)}\mathord{\vcenter{\hbox{b}}} = \mathord{\vcenter{\hbox{c}}}, \quad L^{(2)}\mathord{\vcenter{\hbox{d}}} = \mathord{\vcenter{\hbox{e}}}$$

and $L^{(2)}\tau = 0$ for every $\tau \in \mathcal{W}_0 \setminus \{\mathord{\vcenter{\hbox{a}}}, \mathord{\vcenter{\hbox{b}}}, \mathord{\vcenter{\hbox{d}}}\}$, and

$$L^{(3)}\mathord{\vcenter{\hbox{a}}} = \mathbf{1}, \quad L^{(3)}\mathord{\vcenter{\hbox{b}}} = 3\,\mathord{\vcenter{\hbox{c}}}, \quad L^{(3)}\mathord{\vcenter{\hbox{d}}} = \mathord{\vcenter{\hbox{e}}}$$

and $L^{(3)}\tau = 0$ for every $\tau \in \mathcal{W}_0 \setminus \{\mathord{\vcenter{\hbox{a}}}, \mathord{\vcenter{\hbox{b}}}, \mathord{\vcenter{\hbox{d}}}\}$, and

$$L^{(4)}\mathord{\vcenter{\hbox{a}}} = \mathbf{1}, \quad L^{(5)}\mathord{\vcenter{\hbox{b}}} = \mathbf{1}, \quad L^{(6)}\mathord{\vcenter{\hbox{c}}} = \mathbf{1}, \quad L^{(7)}\mathord{\vcenter{\hbox{d}}} = \mathbf{1}.$$

and $L^{(i)}\tau = 0$ for every $\tau \in \mathcal{W}_0 \setminus \{\tau\}$ with $(i, \tau) \in \{(4, \mathord{\vcenter{\hbox{a}}}), (5, \mathord{\vcenter{\hbox{b}}}), (6, \mathord{\vcenter{\hbox{c}}}), (7, \mathord{\vcenter{\hbox{d}}})\}$.

Recall that $\mathfrak{R}$ is the renormalization group introduced in Theorem 2.2.

Theorem 4.1. *For any choice of the constants ℓ_j, and $M \in \mathfrak{R}$.*

Proof. One can check that $L^{(i)} L^{(j)} \tau = 0$ for for all $\tau \in \mathcal{W}_0$ and any i, j - thus the operators $L^{(i)}$ all commute and one actually has $M = I - \sum_{i=1}^{7} \ell_i L^{(i)}$. Furthermore since $\mathfrak{R}$ is a group it suffices to check, for $1 \leq j \leq 7$, the upper-triangularity of Δ^{M_j} where $M_j \overset{\text{def}}{=} e^{-\ell_j L_j}$. This can be checked by computation.

The $j = 1$ case requires the most work but the computations for this case are exactly the same as those found in [10][Sec 4.2] for the operator called L.

For $j \geq 2$ we observe that one has, for all $\tau \in \mathcal{T}_0$, $\hat{M}_j \mathcal{J}_k(\tau) = \mathcal{J}_k(M_j \tau)$ and $\Delta^{M_j} = M_j \otimes \mathrm{Id}$. Clearly Δ^{M_j} is upper triangular. $\qquad\square$

We then define, for each $\varepsilon \in (0, 1]$, a map $M_\varepsilon \in \mathfrak{R}$ by specifying the constants

$$\ell_1 = \varepsilon^{-1} C^{\stackrel{\circ}{\circ}} , \quad \ell_2 = \varepsilon^{-\frac{1}{2}} C^{\stackrel{\circ}{\circ}} , \quad \ell_3 = \varepsilon^{-\frac{1}{2}} C^{\stackrel{\circ\circ}{}} , $$
$$\ell_4 = c_\zeta^{(2)} , \quad \ell_5 = c_\zeta^{(4)} , \quad \ell_6 = c_\zeta^{(1)} , \quad \ell_7 = c_\zeta^{(3)} .$$

(4.3)

Recall that the constants on the right hand side are defined in (1.7).

4.2 Moment bounds

In order to identify the limit of our sequence of models $(\hat{\Pi}^{(\varepsilon)}, \hat{\Gamma}^{(\varepsilon)})$ as the Itô model we follow the approach in [12] and introduce another level of regularization $\zeta_{\varepsilon,\bar{\varepsilon}} = \varrho_{\bar{\varepsilon}} * \zeta_\varepsilon$ with $\bar{\varepsilon} > 0$ and $\varrho_{\bar{\varepsilon}} = \bar{\varepsilon}^{-3} \varrho(\bar{\varepsilon}^{-2}t, \bar{\varepsilon}^{-1}x)$, where ϱ is smooth, compactly supported, even in the space variable and integrating to one.

We then construct the renormalized model $(\hat{\Pi}^{(\varepsilon,\bar{\varepsilon})}, \hat{\Gamma}^{(\varepsilon,\bar{\varepsilon})})$ from $\zeta_{\varepsilon,\bar{\varepsilon}}$ together with the renormalization maps $M_{\varepsilon,\bar{\varepsilon}}$ with constants specified in the same way as in (4.3), except that we replace every P by our truncation K, every $\mathfrak{C}_n$ by $\mathfrak{C}_n^{(\varepsilon)}$, and finally drop the factors ε^{-1} and $\varepsilon^{-\frac{1}{2}}$ in (4.3). For example, the constant $\varepsilon^{-1} C^{\stackrel{\circ}{\circ}}$ in (4.3) is replaced by
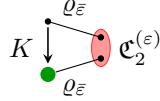
One may find that for a fixed $\varepsilon > 0$, sending $\bar{\varepsilon}$ to zero does not exactly recover the renormalization constants for $(\hat{\Pi}^{(\varepsilon)}, \hat{\Gamma}^{(\varepsilon)})$; for instance in the latter model the renormalization constants are defined via P. This does not matter for two reasons: (i) we will only consider the situation that ε is much less than $\bar{\varepsilon}$; (ii) when we bound the moments for $(\hat{\Pi}_x^{(\varepsilon)} \tau)(\varphi_x^\lambda)$ below, we will actually replace the constants for $\hat{\Pi}^{(\varepsilon)}$ by the ones defined via K and $\mathfrak{C}_n^{(\varepsilon)}$ here, with an error that goes to zero as $\varepsilon \to 0$. More precisely, for the constants $\varepsilon^{-1} C^{\stackrel{\circ}{\circ}}, \varepsilon^{-\frac{1}{2}} C^{\stackrel{\circ}{\circ}}, \varepsilon^{-\frac{1}{2}} C^{\stackrel{\circ\circ}{}}$, by exponential decay of $\mathfrak{C}_n$ and the fact that $P(z) = K(z)$ for $|z| < 1$, one can easily see by a scaling argument that this error is bounded by $\theta^{\frac{1}{\varepsilon}}$ with $\theta \in (0, 1)$; thus even though these errors are multiplied by "graphs" which may diverge as $\varepsilon \to 0$ they still vanish in that limit. For the other constants one can also argue as in [10] that the error of such replacement vanishes as $\varepsilon \to 0$.

Proposition 4.2. *Let $(\hat{\Pi}^{(\varepsilon)}, \hat{\Gamma}^{(\varepsilon)})$ and $(\hat{\Pi}^{(\varepsilon,\bar{\varepsilon})}, \hat{\Gamma}^{(\varepsilon,\bar{\varepsilon})})$ be defined as above. There exist $\kappa, \eta > 0$ such that for every $\tau \in \{\circ, \stackrel{\circ}{\circ}, \stackrel{\circ}{\bullet}, \stackrel{\bullet}{\bullet}, \stackrel{\circ}{\circ}, \stackrel{\circ\circ}{}, \stackrel{\circ\circ}{\circ}, \stackrel{\circ}{\vee}, \stackrel{\circ\circ}{\circ}\}$ and every $p > 0$, one has*

$$\mathbf{E}|(\hat{\Pi}_x^{(\varepsilon)} \tau - \hat{\Pi}_x^{(\varepsilon,\bar{\varepsilon})} \tau)(\varphi_x^\lambda)|^p \lesssim \bar{\varepsilon}^\kappa \lambda^{p(|\tau|_{\mathfrak{s}} + \eta)}$$

(4.4)

uniformly in all $\bar{\varepsilon} \in (0, 1]$ and all $\varepsilon \in (0, 1]$ sufficiently small depending on $\bar{\varepsilon}$, all $\lambda \in (0, 1]$, all test functions $\varphi \in \mathcal{B}$, and all $x \in \mathbf{R}^2$.

Remark 4.3. We will actually mainly focus on a weaker bound

$$\mathbf{E}|(\hat{\Pi}_x^{(\varepsilon)}\tau)(\varphi_x^\lambda)|^p \lesssim \lambda^{p(|\tau|+\eta)} , \tag{4.5}$$

and (4.4) will turn out to follow in the same way.

We introduce more graphical notation. We denote by ──▶ a generic test function φ_λ rescaled to scale λ and label it by $(0,0)$. An arrow ───▶ labeled by $(1,0)$ represents the kernel K. A barred arrow ──┼──▶ labeled by $(1,1)$ represents a function $K(t-s,y-x) - K(-s,-x)$, where (s,x) and (t,y) are the coordinates of the starting and end point respectively. A double barred arrow ──┼┼──▶ labeled by $(1,2)$ represents a function $K(t-s,y-x) - K(-s,-x) - y\,K'(-s,-x)$. The dotted lines ·········· represent the Dirac distributions and are labeled by $(3,-1)$. A polygon with n points inside ⬛ ($n=4$ here) represents the cumulant $\mathfrak{C}_n^{(\varepsilon)}(z_1,\cdots,z_n)$ and is labeled by $(-3n/2,0)$.

We also draw an arrow ----▶ with label $(2,0)$ for the kernel $K'=\partial_x K$ and ──✕──▶ for the test function $(t,x)\mapsto x\varphi^\lambda(t,x)$. Note that $\tilde\varphi(t,x)=x\varphi(t,x)$ is again an admissible test function and one has $x\varphi^\lambda(t,x)=\lambda\tilde\varphi^\lambda(t,x)$. As a consequence, when a test function φ is replaced with test function $\tilde\varphi$, one gains an additional power of λ.

4.2.1 The symbols ⚬, ⚬, and ⚬

We start with the simplest object ⚬ after the noise. By translation invariance we take the point x in (4.5) to be 0. Using Definition A.3 which allows us to represent a product of the noises as Wick products, one has

$$\left(\hat{\Pi}_0^{(\varepsilon)}⚬\right)(\varphi_0^\lambda) = \quad + \quad - \varepsilon^{-1}C^{⚬} \quad = \quad - \quad + E_\varepsilon \tag{4.6}$$

where E_ε here and below is an error which arises from the difference between $\varepsilon^{-1}C^{⚬}$ and $C_\varepsilon^{⚬}$ and goes to zero.

Now if we compare the above graphs with the corresponding ones in [10, Eq. (5.2)] in the Gaussian noise case, we realize that they are essentially the same graphs. Since in [10] it is checked that the symmetric pairing of the first graph, as well as the last graph above, satisfy Assumption 3.5, our Lemma 3.20 and Lemma 3.23 immediately yield Assumption 3.17 for the two graphs on the RHS of (4.6), and therefore by Theorem 3.19 one concludes the desired bounds for $\left(\hat{\Pi}_0^{(\varepsilon)}⚬\right)(\varphi_0^\lambda)$.

The other two symbols ⚬ and ⚬ are bounded in the same way. In these cases, graphs appearing in the expansions are the same with those in the Gaussian case. However, in general, for other symbols below, there will be new terms due to the nontrivial higher cumulants of our non-Gaussian noise (we sometimes refer to them as just "new terms"). The discussion here shows that we only need to treat with these new terms.

4.2.2 The symbol ⚬

Besides the terms appearing in [10], we have the following new terms in the expression for $\left(\hat{\Pi}_0^{(\varepsilon)}⚬\right)(\varphi_\lambda)$:

$$\quad - \varepsilon^{-\frac{1}{2}}C^{⚬} \quad = \quad - \left(\quad - \quad \right) + E_\varepsilon$$

It is straightforward to check that Assumption 3.17 is satisfied for the two graphs on the RHS, which yields the desired bounds $\mathbf{E}|(\hat{\Pi}_x^{(\varepsilon)} \,\text{⚬⚬})(\varphi_x^\lambda)|^p \lesssim \lambda^{p(-\frac{1}{2}+\eta)}$.

4.2.3 The symbol ⚬⚬⚬

We now turn to ⚬⚬⚬. In this case, besides the terms shown in [10], we have the following new terms in the expression for $\big(\hat{\Pi}_0^{(\varepsilon)} \,\text{⚬⚬⚬}\big)(\varphi_\lambda)$:

One can also check that Assumption 3.17 holds for these two graphs.

4.2.4 The symbol ⚬⚬⚬⚬

The new terms for $\big(\hat{\Pi}_0^{(\varepsilon)} \,\text{⚬⚬⚬⚬}\big)(\varphi_\lambda)$ are:

The sum of the last two terms can be written as a sum of 11 graphs (expanding the terms represented by the barred arrows yield 12 terms, and the renormalization constant cancels one of them); each of them has a fourth order hyper-edge which can be split into two edges as $\Rightarrow$. After this split all the graphs satisfy Assumption 3.5 as checked in [10] so they satisfy Assumption 3.17 by Lemmas 3.23 and 3.22.

After cancellations with renormalization constant we only need to check the assumption for the following graphs
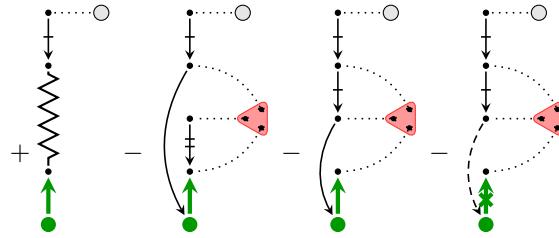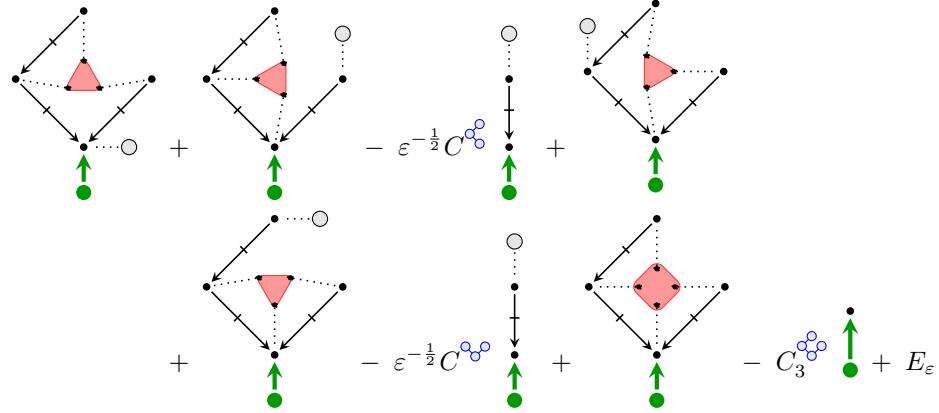
where represents the kernel
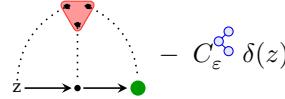
understood as a function of z, with label $(3\frac{1}{2}, -1)$. It can be checked that these eight graphs all satisfy the conditions in Assumption 3.17.
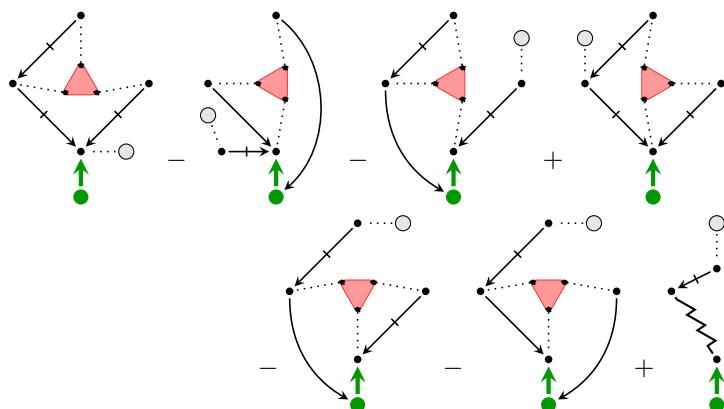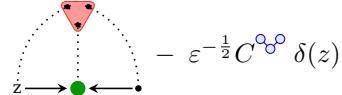
4.2.5 The symbol

The new terms are

For the last two terms, after cancellation with $C_{3,\varepsilon}^{\,\cdots}$, there are seven terms, each having a hyper-edge that can be split into two edges representing a cumulant of the top and bottom noises and a cumulant of the left and right noises. These terms all satisfy Assumption 3.5 as checked in [10].
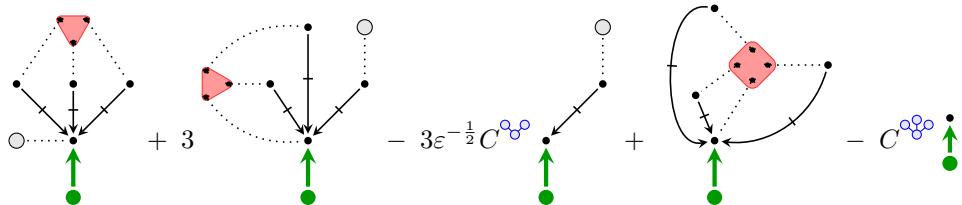
So we are left with

where ⋰⋰⋰⋰ represents the kernel

$$
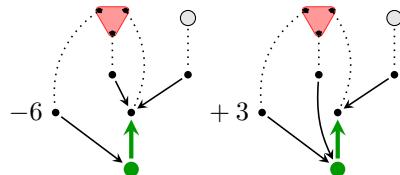\text{(graph)} \;-\; \varepsilon^{-\frac{1}{2}} C^{\,\text{(symbol)}} \,\delta(z)
$$

labeled by $(3\tfrac{1}{2}, -1)$. It can be checked that these seven graphs all satisfy the conditions in Assumption 3.17.
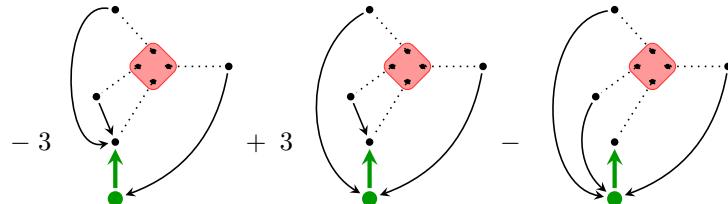
4.2.6 The symbol (symbol)

The new terms are

$$
\text{(graph)} \;+\; 3\,\text{(graph)} \;-\; 3\varepsilon^{-\frac{1}{2}} C^{\,\text{(symbol)}}\,\text{(graph)} \;+\; \text{(graph)} \;-\; C^{\,\text{(symbol)}}\,\text{(graph)}
$$

The sum of the 2nd and the 3rd terms is

$$
-6\,\text{(graph)} \;+\; 3\,\text{(graph)}
$$

The sum of the last two terms is

$$
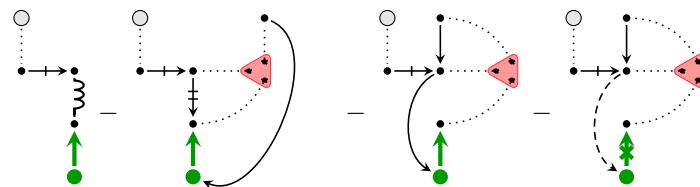-3\,\text{(graph)} \;+\; 3\,\text{(graph)} \;-\; \text{(graph)}
$$

Therefore, for this symbol, six graphs remain to be checked. Indeed, they all satisfy Assumption 3.17.

4.2.7 The symbol (symbol)

The new terms are

$$
2\,\text{(graph)} \;-\; \frac{2C^{\,\text{(symbol)}}}{\sqrt{\varepsilon}}\,\text{(graph)} \;+\; \text{(graph)} \;+\; \text{(graph)} \;+\; \text{(graph)} \;-\; C^{\,\text{(symbol)}}\,\text{(graph)}
$$
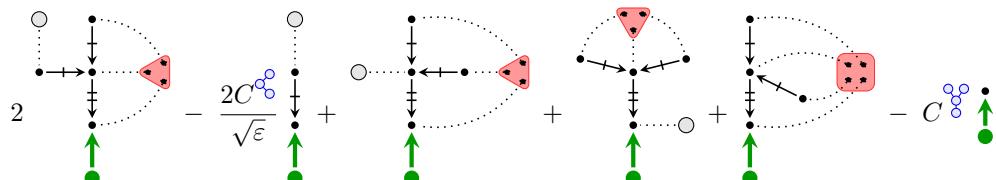
The sum of the first two terms is 2 times

$$
\text{(graph)} \;-\; \text{(graph)} \;-\; \text{(graph)} \;-\; \text{(graph)}
$$

where is

labeled by $(3\tfrac{1}{2}, -1)$.

The 4th term is equal to (noting that the heat kernels in annihilate constants)

Finally the sum of the last two terms is

The first two graphs above have a fourth order hyper-edge, by splitting it into two edges and using Lemma 3.22 they become the graphs checked in [10].

Therefore, for this symbol, there are still nine graphs to be checked. It is straightforward (though a bit tedious) to check that they all satisfy Assumption 3.17.

of Proposition 4.2. Collecting all the results of this section, we obtain the weaker bound (4.5). The bound (4.4), follows in essentially the same way as the first bound, by the argument in [12] (see also the verification of the second bound in [7, Theorem 10.7]). Indeed, as we consider the difference between $\hat{\Pi}_0^{(\varepsilon)}\tau$ and $\hat{\Pi}_0^{(\varepsilon,\bar\varepsilon)}\tau$ for any $\tau \neq \Xi$, we obtain the same graphs as above, and in each graph some of the instances of δ are replaced by $\varrho_{\bar\varepsilon}$ and exactly one instance is replaced by $\delta - \varrho_{\bar\varepsilon}$. By the bound $\|\delta - \varrho_{\bar\varepsilon}\|_{-3-\kappa} \lesssim \bar\varepsilon^\kappa$ one obtains the same bound as (4.5) with an extra factor $\bar\varepsilon^\kappa$, which is exactly required by the bound (4.4). The case $\tau = \Xi$ can be also proved as in [12]. $\qquad\square$

4.3 Proof of the main theorem

It was shown in [10] that if we replace ζ_ε by ξ_ε with $\xi_\varepsilon = \varrho_\varepsilon * \xi$ where ξ is the Gaussian space-time white noise, the renormalised models built from ξ_ε converge to the limit $\hat{Z} = (\hat\Pi, \hat\Gamma)$ called the Itô model, which satisfies the following property. For every $\tau \in \mathcal{U}$ and every (t,x), the process $s \mapsto \left(\hat\Pi_{(t,x)}\tau\right)(s,\cdot)$ 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

$$\left(\hat\Pi_{(t,x)}\Xi\tau\right)(\varphi) = \int_t^\infty \left\langle \left(\hat\Pi_{(t,x)}\tau\right)(s,\cdot)\varphi(s,\cdot), dW(s)\right\rangle, \tag{4.7}$$

where the integral on the right is the Itô integral.

The goal of this section is to show that our renormalised models $(\hat\Pi^{(\varepsilon)}, \hat\Gamma^{(\varepsilon)})$ built from ζ_ε defined above converge to the same limit. We prove this by applying a "diagonal argument" as in [12].

Theorem 4.4. *Let $\hat{Z}_\varepsilon = (\hat\Pi^{(\varepsilon)}, \hat\Gamma^{(\varepsilon)})$ be the renormalised model built from ζ_ε defined in the previous sections (with the choice of renormalisation constants given by (1.7)). Let $\hat{Z} = (\hat\Pi, \hat\Gamma)$ be the Itô random model. Then, as $\varepsilon \to 0$, one has $\hat{Z}_\varepsilon \to \hat{Z}$ in distribution in the space $\mathscr{M}$ of admissible models for $\mathscr{T}$.*

Proof. Since the topology on $\mathscr{M}$ is generated by the pseudometrics defined in Section 2.2, it is sufficient to show that

$$\lim_{\varepsilon \to 0} \mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}\| = 0 \,, \tag{4.8}$$

where the compact domain that the pseudometric depends on is omitted in the notation since it does not matter much in the sequel. We have

$$\mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}\| \leq \mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}_{\varepsilon,\bar\varepsilon}\| + \mathbf{E} \|\hat{Z}_{\varepsilon,\bar\varepsilon}, \hat{Z}_{0,\bar\varepsilon}\| + \mathbf{E} \|\hat{Z}_{0,\bar\varepsilon}, \hat{Z}\| \,,$$

which is valid for any fixed $\bar\varepsilon > 0$ and for any coupling between ξ and ζ_ε, where $\hat{Z}_{\varepsilon,\bar\varepsilon}$ is the previously defined "intermediate" model, and the model $\hat{Z}_{0,\bar\varepsilon}$ is obtained again in the same way as $\hat{Z}_{\varepsilon,\bar\varepsilon}$ but built from the noise $\zeta_{0,\bar\varepsilon} \stackrel{\text{def}}{=} \xi * \varrho_{\bar\varepsilon}$, with renormalization constants defined in the same way as those for $\hat{Z}_{\varepsilon,\bar\varepsilon}$ except that every $\varrho_{\bar\varepsilon}$ is replaced by a delta function. Therefore,

$$\lim_{\varepsilon \to 0} \mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}\| \leq \lim_{\bar\varepsilon \to 0} \lim_{\varepsilon \to 0} \big(\mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}_{\varepsilon,\bar\varepsilon}\| + \mathbf{E} \|\hat{Z}_{\varepsilon,\bar\varepsilon}, \hat{Z}_{0,\bar\varepsilon}\| + \mathbf{E} \|\hat{Z}_{0,\bar\varepsilon}, \hat{Z}\| \big) \,. \tag{4.9}$$

For the last term, $\lim_{\bar\varepsilon \to 0} \mathbf{E} \|\hat{Z}_{0,\bar\varepsilon}, \hat{Z}\|^p = 0$ for every p was already shown in [10]. Note that since ξ is Gaussian, all the renormalization constants associated with the model $\hat{Z}_{0,\bar\varepsilon}$ containing cumulants of order higher than two vanish; the other constants slightly differ from the corresponding ones in [10] because our constants for $\hat{Z}_{0,\bar\varepsilon}$ are defined via the kernel K, but this error vanishes as $\bar\varepsilon \to 0$.

Regarding the first term, by [7, Thm 10.7] and the second bound in Theorem 4.2, we obtain the bound $\mathbf{E} \|\hat{Z}_\varepsilon, \hat{Z}_{\varepsilon,\bar\varepsilon}\| \lesssim \bar\varepsilon^\kappa$ uniformly over ε sufficiently small, so that this term also vanishes.

It therefore remains to bound the second term. This follows from the same argument as in [12]. Firstly, one has a central limit theorem for the noise ζ_ε, namely for every $\alpha < -\frac{3}{2}$ the field ζ_ε converges in law to space-time white noise ξ in $\mathcal{C}^\alpha(\mathbf{R} \times S^1)$, see [12, Prop. 6.1]. Therefore

$$\lim_{\varepsilon \to 0} \mathbf{E} \|\zeta_{0,\bar\varepsilon} - \zeta_{\varepsilon,\bar\varepsilon}\|^p_{\mathcal{C}^1;\mathfrak{K}} = 0 \,. \tag{4.10}$$

for any bounded domain $\mathfrak{K}$. Also, the map from the space of stationary and almost surely periodic noise equipped with L^p of $\mathcal{C}^1$ norm to the space of random admissible models is continuous. Therefore one has

$$\lim_{\varepsilon \to 0} \mathbf{E} \|\hat{Z}_{\varepsilon,\bar\varepsilon}, \hat{Z}_{0,\bar\varepsilon}\| = 0 \,,$$

for every fixed (sufficiently small) $\bar\varepsilon > 0$, so that the second term in (4.9) also vanishes, thus concluding the proof. $\square$

We finally collect all our results and prove the main theorem.

of Theorem 1.2. By [10, Thm 3.9] or [7, Sec. 7] one has for every $\varepsilon > 0$ a unique maximal solution U^ε to the fixed point problem (2.8) in $\mathcal{D}^{\gamma,0}$ with respect to the renormalized model $\hat{\Pi}^{(\varepsilon)}$, and solution U respect to the Itô model $\hat{\Pi}$, and by [10, Thm 4.7] we have $\lim_{\varepsilon \to 0} \mathbf{P}(\|U; U^\varepsilon\|_{\gamma,0} > \varepsilon^\theta) = 0$ for every $\theta < \kappa/2$.

Using the identity

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

with M_ε defined above (4.3), together with (1.7), (4.2), we see that this expression is non-zero for the symbols $\mathbf{1}$, ○, ⅗, ⅗, ⅗, ⅗, ⅗ and ⅗, where one has

$$\big(\hat{\Pi}^{(\varepsilon)}_z \mathbf{1}\big)(z) = 1 \,, \quad \big(\hat{\Pi}^{(\varepsilon)}_z ○\big)(z) = \xi_\varepsilon(z) \,, \quad \big(\hat{\Pi}^{(\varepsilon)}_z ⅗\big)(z) = -C^{(1)}_\zeta/\varepsilon \,,$$

$$\big(\hat{\Pi}^{(\varepsilon)}_z ⅗\big)(z) = -C^{(2)}_\zeta/\sqrt{\varepsilon} \,, \quad \big(\hat{\Pi}^{(\varepsilon)}_z ⅗\big)(z) = -C^{(3)}_\zeta/\sqrt{\varepsilon} \,,$$

$$\big(\hat{\Pi}_z^{(\varepsilon)}\, \text{⬡}\big)(z) = -c_\zeta^{(2)}\,, \quad \big(\hat{\Pi}_z^{(\varepsilon)}\, \text{⬡}\big)(z) = -c_\zeta^{(1)}\,,$$
$$\big(\hat{\Pi}_z^{(\varepsilon)}\, \text{⬡}\big)(z) = -c_\zeta^{(3)}\,, \quad \big(\hat{\Pi}_z^{(\varepsilon)}\, \text{⬡}\big)(z) = -c_\zeta^{(4)}\,.$$

Write $\mathcal{R}_\varepsilon$ as the reconstruction map with respect to the renormalized model $\hat{\Pi}^{(\varepsilon)}$. Combining the above identities with (2.12) it follows that one has the identity

$$\mathcal{R}_\varepsilon\big(\hat{G}(U)\Xi\big)(z) = G(u(z))\xi_\varepsilon(z) - \frac{C_\zeta^{(1)}}{\varepsilon}G'(u(z))G(u(z))$$
$$- \frac{C_\zeta^{(2)}}{\varepsilon^{1/2}}G'(u(z))^2 G(u(z)) - \frac{C_\zeta^{(3)}}{\varepsilon^{1/2}}G''(u(z))G^2(u(z))$$
$$- \frac{1}{6}c_\zeta^{(1)}G'''(u(z))G^3(u(z)) - c_\zeta^{(2)}G'(u(z))^3 G(u(z))$$
$$- \Big(\frac{1}{2}c_\zeta^{(3)} + c_\zeta^{(4)}\Big)G''(u(z))G'(u(z))G^2(u(z))$$

By Theorem 4.4, the limiting model is the Itô model; it is proved in [10, Corollary 6.3] that $u = \mathcal{R}U$ constructed with respect to the Itô model is indeed the classical weak solution interpreted in the Itô sense ([18, 2]) which concludes the proof. $\qquad\square$

A Cumulants and Wick products

We define joint cumulant functions $\mathfrak{C}_n(z_1, \ldots, z_n)$ of a space-time centered random field ζ. Given any subset A of space-time points we will define a joint cumulant function $\mathfrak{C}(\bar{A})$ of ζ. The definition operates recursively in $|A|$, one sets

$$\mathfrak{C}(A) \overset{\text{def}}{=} \mathbf{E}\Big(\prod_{z\in A}\zeta(z)\Big) - \sum_{\substack{\pi\in\mathcal{P}(A)\\\pi\neq\{A\}}}\prod_{B\in\pi}\mathfrak{C}(B) \tag{A.1}$$

where $\mathcal{P}(A)$ denotes the set of all partitions of A. The key cumulant identity comes from moving all the cumulants of (A.1) to the LHS.

We write $\mathfrak{C}_n$ for the n-th joint cumulant function of the field ζ at n space-time points $A = \{z_1, \ldots, z_n\}$:

$$\mathfrak{C}_n(z_1, \ldots, z_n) = \mathfrak{C}(A)\,. \tag{A.2}$$

Note that $\mathfrak{C}_1 = 0$ since ζ is assumed to be centered. We write $\mathfrak{C}_n^{(\varepsilon)}$ for the n-th cumulant function of ζ_ε

Definition A.1. *We say that ζ has exponentially decaying cumulants if there exists $\theta \in (0,1)$ such that for each $n \in \mathbf{N}$ one has the bound*

$$|\mathfrak{C}_n(z_1, \ldots, z_n)| \lesssim \theta^{\mathrm{diam}(z_1, \ldots, z_n)} \tag{A.3}$$

uniform in $z_1, \ldots, z_n \in \mathbf{R}^d$ where we set $\mathrm{diam}(z_1, \ldots, z_n) \overset{\text{def}}{=} \sup_{1\leq i,j\leq n}|z_i - z_j|$.

We give an example of a random field ζ satisfying the above property.

Example A.2. Let μ be a finite positive measure on $\mathcal{C}(\mathbf{R}^d)$ such that for every $\varphi \in \mathrm{supp}(\mu)$ and $p > 1$ one has

$$\int \Big(\sup_{z\in\mathbf{R}^d} e^{2|z|_\mathfrak{s}}|\varphi(z)|\Big)^p \mu(d\varphi) < \infty$$

Denote by $\hat{\mu}$ a realisation of the Poisson point process on $\mathcal{C}(\mathbf{R}^d) \times \mathbf{R}^d$ with intensity $\mu \otimes dz$ and set

$$\zeta(z) = \int \varphi(z - z')\,\hat{\mu}(d\varphi \otimes dz') - \int\int \varphi(z)\,dz\mu(d\varphi)\,.$$

Then ζ satisfies Definition A.1.

We now give the definition of the Wick product of a generically non-Gaussian random field evaluated at a collection of points.

Definition A.3. *Given a collection of space-time points A, the Wick product* $:\prod_{z \in A} \zeta(z):$ *is defined recursively in $|A|$ by setting* $:1: = 1$ *and then setting*

$$
\begin{aligned}
:\prod_{z \in A} \zeta(z): &\overset{\text{def}}{=} \prod_{z \in A} \zeta(z) - \sum_{B \subsetneq A} :\prod_{z \in B} \zeta(z): \mathbf{E}\Big(\prod_{z \in A \setminus B} \zeta(z) \Big) \qquad\qquad (\text{A.4}) \\
&= \prod_{z \in A} \zeta(z) - \sum_{B \subsetneq A} :\prod_{z \in B} \zeta(z): \sum_{\pi \in \mathcal{P}(A \setminus B)} \prod_{\bar{B} \in \pi} \mathfrak{C}(\bar{B}) \,.
\end{aligned}
$$

Note that $:\prod_{z \in B} \zeta(z):$ is understood as an operation on $\{\zeta(z)\}_{z \in B}$, not the product. The second line of (A.4) follows from the first by applying (A.1). Note that any non-empty Wick product is always mean zero.

The key Wick product identity comes from moving all the subtracted terms of on the RHS of the second line of (A.4) to the first line of the RHS.

We close with a lemma which is sometimes called "diagram formula" in the literature [11] and generalizes (A.1). It states that moments of Wick products can be calculated by summing over partitions without "self-contractions".

Lemma A.4. *For every $p \geq 1$ and $m \geq 1$, one has*

$$
\mathbf{E}\Big(\prod_{k=1}^{p} :\zeta_\varepsilon(x_k^{(1)}) \cdots \zeta_\varepsilon(x_k^{(m)}): \Big) = \sum_{\pi \in \mathcal{P}_M(M \times P)} \prod_{B \in \pi} \mathfrak{C}_\varepsilon(B) \qquad\qquad (\text{A.5})
$$

where $M \overset{\text{def}}{=} \{i : 1 \leq i \leq m\}$, $P \overset{\text{def}}{=} \{k : 1 \leq k \leq p\}$, and $\mathcal{P}_M(M \times P)$ denotes the set of all partitions $\pi \in \mathcal{P}(M \times P)$ with the property that for every $B \in \pi$ there exist $(i,k),(i',k') \in B$ such that $k \neq k'$ (in particular $|B| > 1$).

References

[1] Ajay Chandra and Hendrik Weber, *Stochastic PDEs, regularity structures, and interacting particle systems*, arXiv preprint arXiv:1508.03616 (2015).

[2] Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, second ed., Encyclopedia of Mathematics and its Applications, vol. 152, Cambridge University Press, Cambridge, 2014. MR-3236753

[3] Peter K. Friz and Martin Hairer, *A course on rough paths: with an introduction to regularity structures*, Universitext, Springer, Cham, 2014. MR-3289027

[4] L. Giraitis and D. Surgailis, *Multivariate Appell polynomials and the central limit theorem*, Dependence in probability and statistics (Oberwolfach, 1985), Progr. Probab. Statist., vol. 11, Birkhäuser Boston, Boston, MA, 1986, pp. 21–71. MR-0899984

[5] Massimiliano Gubinelli and Samy Tindel, *Rough evolution equations*, Ann. Probab. **38** (2010), no. 1, 1–75. MR-2599193

[6] Martin Hairer, *Solving the KPZ equation*, Ann. of Math. (2) **178** (2013), no. 2, 559–664. MR-3071506

[7] Martin Hairer, *A theory of regularity structures*, Invent. Math. **198** (2014), no. 2, 269–504. MR-3274562

[8] Martin Hairer, *Introduction to regularity structures*, Braz. J. Probab. Stat. **29** (2015), no. 2, 175–210. MR-3336866

[9] Martin Hairer, *Regularity structures and the dynamical φ_3^4 model*, arXiv preprint arXiv:1508.05261 (2015). MR-3468250

---

[11] See [17, 4] or [12].

[10] Martin Hairer and Étienne Pardoux, *A Wong-Zakai theorem for stochastic PDEs*, J. Math. Soc. Japan **67** (2015), no. 4, 1551–1604. MR-3417505

[11] Martin Hairer and Jeremy Quastel, *A class of growth models rescaling to KPZ*, ArXiv e-prints (2015).

[12] Martin Hairer and Hao Shen, *A central limit theorem for the KPZ equation*, arXiv preprint arXiv:1507.01237 (2015).

[13] Martin Hairer and Weijun Xu, *Large scale behaviour of 3D phase coexistence models*, ArXiv e-prints (2016).

[14] Ioannis Karatzas and Johannes Ruf, *Pathwise solvability of stochastic integral equations with generalized drift and non-smooth dispersion functions*, Ann. Inst. Henri Poincaré Probab. Stat. **52** (2016), no. 2, 915–938. MR-3498016

[15] Terry J. Lyons, Michael Caruana, and Thierry Lévy, *Differential equations driven by rough paths*, Lecture Notes in Mathematics, vol. 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. MR-2314753

[16] Hao Shen and Weijun Xu, *Weak universality of dynamical φ_3^4: non-Gaussian noise*, ArXiv e-prints (2016).

[17] D. Surgailis, *On poisson multiple stochastic integrals and associated equilibrium markov processes*, Theory and Application of Random Fields, Lecture Notes in Control and Information Sciences, vol. 49, Springer, 1983, pp. 233–248. MR-0799947

[18] John B. Walsh, *An introduction to stochastic partial differential equations*, École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., vol. 1180, Springer, Berlin, 1986, pp. 265–439. MR-0876085

[19] E. Wong and M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist. **36** (1965), 1560–1564. MR-0195142

[20] E. Wong and M. Zakai, *On the relation between ordinary and stochastic differential equations*, Internat. J. Engrg. Sci. **3** (1965), 213–229. MR-0183023