Electron. J. Probab. **18** (2013), no. 3, 1–34.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2004

The Stratonovich heat equation: a continuity result and weak approximations

Aurélien Deya* Maria Jolis† Lluís Quer-Sardanyons‡

Abstract

We consider a Stratonovich heat equation in $(0,1)$ with a nonlinear multiplicative noise driven by a trace-class Wiener process. First, the equation is shown to have a unique mild solution. Secondly, convolutional rough paths techniques are used to provide an almost sure continuity result for the solution with respect to the solution of the 'smooth' equation obtained by replacing the noise with an absolutely continuous process. This continuity result is then exploited to prove weak convergence results based on Donsker and Kac-Stroock type approximations of the noise.

Keywords: convergence in law; stochastic heat equation; Stratonovich integral; convolutional rough paths theory.
AMS MSC 2010: 60H10, 60H05, 60H07.
Submitted to EJP on May 3, 2012, final version accepted on November 23, 2012.

1 Introduction and main results

The main motivation of the paper comes from [3], where the authors consider, for some fixed $T>0$, the stochastic heat equation

$$\frac{\partial Y^n}{\partial t}(t,x)-\frac{\partial^2 Y^n}{\partial x^2}(t,x)=\dot{\theta}^n(t,x),\quad (t,x)\in[0,T]\times[0,1], \tag{1.1}$$

with some initial data and Dirichlet boundary conditions, where the random fields $(\dot{\theta}^n)_{n\geq 1}$ verify that the family of processes $\theta^n(t,x):=\int_0^t\int_0^x\dot{\theta}^n(s,y)\,dyds$ converge *in law*, in the space $\mathcal{C}([0,T]\times[0,1])$ of continuous functions, to the Brownian sheet. Then, sufficient conditions on θ^n are provided such that Y^n converges *in law*, as $n\to\infty$, to the mild solution Y of

$$\frac{\partial Y}{\partial t}(t,x)-\frac{\partial^2 Y}{\partial x^2}(t,x)=\dot{W}(t,x),\quad (t,x)\in[0,T]\times[0,1],$$

*Institut Élie Cartan Nancy, France.
E-mail: Aurelien.Deya@iecn.u-nancy.fr
†Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, Spain.
E-mail: mjolis@mat.uab.cat
‡Departament de Matemàtiques, Facultat de Ciències, Universitat Autònoma de Barcelona, Spain.
E-mail: quer@mat.uab.cat

where $\dot{W}(t,x)$ stands for the space-time white noise. Applications of this result include the case of a Donsker type approximation, as well as a Kac-Stroock type approximation in the plane.

Such *diffusion approximation* issues for stochastic PDEs have been extensively studied in the literature. Let us quote here Walsh [32], Manthey [20, 21], Tindel [30], Carmona and Fouque [8], Florit and Nualart [12], just to mention but a few.

Now, following the line of [3], a natural question to be dealt with is to try to get the same type of weak convergence in a non-additive situation, that is when the term $\dot{\theta}^n(t,x)$ in (1.1) is replaced with $f(Y^n(t,x))\dot{\theta}^n(t,x)$, for some sufficiently smooth function $f : \mathbb{R} \to \mathbb{R}$. In this case, one expects that the limit equation is of Stratonovich type, as it was the case in [8] and [12] (see also [2, 29] for examples of a similar behaviour). This phenomenon has been recently illustrated by Bal in [1] as well, for the weak approximation of a linear parabolic equation in $\mathbb{R}^d$ with random potential given by $Y^n(t,x)\dot{\theta}^n(x)$.

Going back to our setting, and focusing first on what we expect to be our limit equation, we should consider the Stratonovich heat equation

$$\frac{\partial Y}{\partial t}(t,x) - \frac{\partial^2 Y}{\partial x^2}(t,x) = f(Y(t,x)) \circ \dot{W}(t,x), \quad (t,x) \in [0,T] \times [0,1]. \tag{1.2}$$

Unfortunately, a well-known drawback in this situation is that the solution admits only very low regularity (see [31]), a major obstacle for our treatment of the non-linearity of the problem. For this reason, we have chosen to restrict our attention to the case of a trace-class noise. To be more specific, we will assume that $\dot{W}$ is the formal derivative of a $L^2(0,1)$-valued Wiener process $\{W_t,\, t \in [0,T]\}$ with covariance operator Q satisfying the following property:

Hypothesis 1.1. *Let $(e_k)_{k\geq 1}$ be the basis of eigenfunctions for the Dirichlet Laplacian Δ in $L^2(0,1)$ given by $e_k(x) := \sqrt{2}\sin(k\pi x)$, $x \in [0,1]$. We assume that there exists a sequence of non-negative real numbers $(\lambda_k)_{k\geq 1}$ and a parameter $\eta > 0$ such that $Qe_k = \lambda_k e_k$ for every $k \geq 1$ and $\sum_{k\geq 1}(\lambda_k k^{4\eta}) < \infty$. Without loss of generality, we assume that $\eta \in (0, \frac{1}{8})$.*

In particular, for any fixed $t \geq 0$, the process W_t can be expanded in $L^2(\Omega; L^2(0,1))$ as

$$W_t = \sum_{k\geq 1} \sqrt{\lambda_k}\beta_t^k\, e_k, \tag{1.3}$$

where $(\beta^k)_{k\geq 1}$ is a family of independent Brownian motions. Note that the condition $\sum_{k\geq 1}(\lambda_k \cdot k^{4\eta}) < \infty$ is only slightly stronger than the usual trace-class hypothesis $\sum_{k\geq 1}\lambda_k < \infty$, insofar as η can be chosen as small as one wishes. For instance, it covers the case where $Q = (\mathrm{Id} - \Delta)^{-r}$ with $r > \frac{1}{2}$.

Another change with respect to [3] lies in our formulation of the study: compared to the random field approach in [3], here it has turned out to be more convenient to use the Hilbert-space-valued setting of Da Prato and Zabczyk [9]. In particular, we are interested in the mild form of equation (1.2), which is given by

$$Y_t = S_t\psi + \int_0^t S_{t-u}(f(Y_u) \circ dW_u), \qquad t \in [0,T], \tag{1.4}$$

where from now on, we will use the notation $Y_t(\cdot) := Y(t,\cdot)$, ψ is some initial condition and $f : \mathbb{R} \to \mathbb{R}$ is a smooth enough mapping. As usual, $(S_t)_{t>0}$ denotes the strongly continuous semigroup of operators generated by $-\Delta$.

A first part of the paper (Section 2) will be devoted to the interpretation of (1.4) as a *Stratonovich* equation, and it will allow us to exhibit an existence and uniqueness result for the solution. We should mention here that the stochastic heat equation in the Stratonovich framework has already been studied in various settings, most of them in the case of a linear multiplicative noise (see e.g. [6, 7, 17]). Once we have given a full sense to (1.4), our strategy to study weak approximations of the solution could be stated in the following loose form:

(a) We will first establish an *almost sure* continuity result (in some suitable space-time topology) for the solution of (1.4) with respect to the solution of the 'smooth' heat equation obtained by replacing W with an absolutely continuous process $\widetilde{W}$ (see Theorem 1.2).
(b) Then, for two particular families of absolutely continuous processes approximating W, we will rely on our continuity result to show convergence towards the solution in some possibly different probability space, leading us to the expected *weak* convergence (see Theorem 1.3).

Our strategy to compare the solution Y of (1.4) with 'smooth' solutions is based on a genuine *rough-paths* type expansion of the equation, which follows the ideas of [16, 11, 10]. Rough-paths techniques have indeed proved to be very efficient as far as approximation of non-linear systems in finite dimension is concerned (see [13, Chapter 17]), and it is therefore natural to address the same question in this infinite-dimensional background. Note that the model given by (1.4) differs from those studied in [11, 10], where only finite-dimensional noises are considered, forcing us to revise most of the technical details behind this procedure (see Section 3).

In order to state the above-mentioned results with more details, we need to introduce the spaces in which our random variables will take their values. First, as far as the spatial regularity is concerned, the fractional Sobolev spaces must come into the picture. Thus, for every $\alpha \in \mathbb{R}$ and $p \geq 2$, we will denote by $\mathcal{B}_{\alpha,p}$ the fractional Sobolev space of order α based on $L^p(0,1)$, that is

$$\varphi \in \mathcal{B}_{\alpha,p} \Longleftrightarrow (-\Delta)^{\alpha}\varphi \in L^p(0,1),$$

where Δ stands for the Dirichlet Laplacian in $L^2(0,1)$ (see e.g. [26] for a thorough study of these spaces). For the sake of conciseness, we will write $\mathcal{B}_\alpha$ for $\mathcal{B}_{\alpha,2}$ and $\mathcal{B}$ for $\mathcal{B}_0 = L^2(0,1)$ throughout the paper. We will also denote by $\mathcal{B}_\infty$ the set of continuous functions on $[0,1]$, endowed with the supremum norm.

Of course, we will also have to deal with the time regularity of our processes. So, for any subinterval $I \subset [0,T]$ and any Banach space V, we define $\mathcal{C}^0(I;V)$ as the space of continuous functions $y : I \to V$ and set

$$\mathcal{N}[y;\mathcal{C}^0(I;V)] := \sup_{t\in I}\|y_t\|_V.$$

Moreover, for any $\lambda > 0$, we introduce the space $\mathcal{C}^\lambda(I;V)$ of λ-Hölder continuous V-valued functions endowed with the seminorm

$$\mathcal{N}[y;\mathcal{C}^\lambda(I;V)] := \sup_{s<t\in I}\frac{\|y_t - y_s\|_V}{|t-s|^\lambda}. \tag{1.5}$$

Note that in the case where $I = [0,T]$, we will often write $\mathcal{C}^\lambda(V)$ for $\mathcal{C}^\lambda([0,T];V)$.

Now, consider any process $\widetilde{W}$ defined on the same probability space as W and with absolutely continuous paths in $\mathcal{B}_{\eta,2p}$, for every integer $p \geq 1$ (recall that η has been defined in Hypothesis 1.1). Then, let $\{\widetilde{Y}_t,\, t \in [0,T]\}$ be the unique solution of the Riemann-Lebesgue equation (considered in a pathwise sense):

$$\widetilde{Y}_t = S_t\widetilde{\psi} + \int_0^t S_{t-u}(f(\widetilde{Y}_u)\cdot d\widetilde{W}_u), \qquad t \in [0,T], \tag{1.6}$$

where $\tilde{\psi} \in \mathcal{B}$. As evoked earlier, our first main result will consist in comparing such a solution $\widetilde{Y}$ with the solution Y of (1.4). This result can be stated as follows.

Theorem 1.2. *Assume that Hypothesis 1.1 holds true for W and some parameter $\eta > 0$, and let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^3$, bounded with bounded derivatives. In addition, pick $\gamma \in (\frac{1}{2}, \frac{1}{2}+\eta)$ and assume that the initial condition ψ (resp. $\widetilde{\psi}$) in (1.4) (resp. in (1.6)) belongs to $\mathcal{B}_\gamma$. Then there exist $\varepsilon > 0$ and $p \geq 1$ such that*

$$\mathcal{N}[Y-\widetilde{Y};\mathcal{C}^0(\mathcal{B}_\gamma)] \leq F_{\varepsilon,p}\Big(\|\psi\|_{\mathcal{B}_\gamma}, \|\widetilde{\psi}\|_{\mathcal{B}_\gamma}, \mathcal{N}[W;\mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})], \mathcal{N}[\widetilde{W};\mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})]\Big) \Big\{\|\psi-\widetilde{\psi}\|_{\mathcal{B}_\gamma} + \mathcal{N}[W-\widetilde{W};\mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})]\Big\}, \tag{1.7}$$

for some deterministic function $F_{\varepsilon,p} : (\mathbb{R}^+)^4 \to \mathbb{R}^+$ bounded on bounded sets.

The topologies involved in this statement are directly inherited from our rough-paths analysis of the equation, and their relevance should therefore become clear through the lines of Section 3 (see in particular the proof of the central Proposition 3.10). Note that this bound certainly remains valid with respect to some Hölder norm (in time) for the left-hand side of (1.7), as our arguments will suggest it. However, due to the technicality of the rough-paths procedure, we have preferred to focus on the behaviour of the supremum norm (see also Remark 3.13).

Our next step will consist in applying the above Theorem 1.2 - on some possibly larger probability space - to two particular families of absolutely continuous processes that approximate W, so as to retrieve weak convergence results for the solution. To define these approximation processes, we will make use of the following additional notation. Namely, on a probability space $(\Omega,\mathcal{F},P)$, given a sequence $(X^k)_{k\geq 1}$ of centered i.i.d processes admitting moments of any order, we set, for every $t \geq 0$,

$$\mathbf{W}(X^\cdot)_t := \sum_{k\geq 1}\sqrt{\lambda_k}X_t^k e_k.$$

Thanks to our forthcoming Proposition 4.3, we know that $\mathbf{W}(X^\cdot)$ is indeed well-defined as a process on $(\Omega,\mathcal{F},P)$ with values in $\mathcal{B}_{\eta,2p}$, for all $p \geq 1$. Let us also specify that, given a sequence $(\beta^n)_{n\geq 1}$ of real-valued processes, we will henceforth denote by $(\beta^{n,\cdot})_{n\geq 1} = (\beta^{n,k})_{n,k\geq 1}$ a generic sequence of independent copies of $(\beta^n)_{n\geq 1}$ (defined on a possibly larger probability space).

The two families of approximations at the core of our study can now be introduced as follows (we fix $T = 1$ for the sake of clarity):

(i) The *Donsker approximation* $W^n := \mathbf{W}(S^{n,\cdot})$, where S^n is a sequence of appropriately rescaled random walks. To be more specific, let $(Z_j)_{j\geq 1}$ be a family of i.i.d random variables with mean zero, unit variance and admitting moments of any order. Then, for each $n \in \mathbb{N}$, set

$$S_t^n := n^{-1/2}\Big\{\sum_{j=1}^{i-1} Z_j + \frac{t-(i-1)/n}{1/n}\, Z_i\Big\} \quad \text{if } t \in \Big[\frac{i-1}{n}, \frac{i}{n}\Big], \quad \text{with } i \in \{1,\dots,n\}. \tag{1.8}$$

Recall that, by Donsker Invariance Principle (see e.g. [19, Thm. 4.20]), S^n is known to converge in law to the standard Brownian motion in $\mathcal{C}^0([0,1];\mathbb{R})$, as $n \to \infty$.

(ii) The *Kac-Stroock approximation* $W^n := \mathbf{W}(\theta^{n,\cdot})$, where θ^n stands for the classical Kac-Stroock approximation of the (one-dimensional) Brownian motion. Precisely, introduce a standard Poisson process N and a Bernoulli variable ζ independent of N, with $P(\zeta = 1) = 1/2$. Then, set

$$\theta^n_t := \sqrt{n} \int_0^t (-1)^{\zeta + N(ns)} ds. \tag{1.9}$$

Here again, the sequence $(\theta^n)_{n\geq 1}$ thus defined converges in law in $\mathcal{C}^0([0,1];\mathbb{R})$, as $n \to \infty$, to a standard Brownian motion (see e.g [18, 24]).

Of course, the one-dimensional weak convergence of S^n (resp. θ^n) towards the Brownian motion is a priori not sufficient for us to apply Theorem 1.2. Our aim is to turn this one-dimensional weak convergence into an almost sure convergence result for $\mathbf{W}(S^{n,\cdot})$ (resp. $\mathbf{W}(\theta^{n,\cdot})$) with respect to the topology involved in (1.7), and this will appeal in particular to Skorokhod embedding arguments (see Section 4). Together with Theorem 1.2, the strategy ends up with the following statement.

Theorem 1.3. *Under the hypotheses of Theorem 1.2, fix an initial condition $\psi = \tilde{\psi} \in \mathcal{B}_\gamma$, and denote by Y^n the (Riemann-Lebesgue) solution of (1.6) associated with either the Donsker approximation $W^n = \mathbf{W}(S^{n,\cdot})$ or the Kac-Stroock approximation $W^n = \mathbf{W}(\theta^{n,\cdot})$. Then, as $n \to \infty$, Y^n converges in law to Y in the space $\mathcal{C}^0(\mathcal{B}_\gamma)$.*

The paper is organized as follows. Section 2 is devoted to a few preliminaries on the theoretical study of the Stratonovich heat equation (1.2). The rough-paths type analysis of this equation is performed in Section 3, and it will lead us to the proof of our continuity result Theorem 1.2. In Section 4, we will tackle the approximation issue for the above-defined Donsker and Kac-Stroock processes by exhibiting a general convergence criterion (see Proposition 4.5), which will entail Theorem 1.3. Eventually, we have added an appendix with material on fractional Sobolev spaces and the proof of a technical result needed in Section 3.4.

Remark 1.4. *At first sight, the reader familiar with rough-paths type continuity results may be surprised at the absence of some 'Lévy-area' term in our bound (1.7). Otherwise stated, the convergence of an approximation W^n towards W (with respect to some appropriate topology) is sufficient to guarantee the convergence of the associated solution. In fact, on this particular point, the situation is very similar to the case of a one-dimensional SDE with so-called* commuting *vector fields, i.e.,*

$$dY_t = b(Y_t)\, dt + \sum\nolimits_{i=1}^n \sigma_i(Y_t) \circ dB^i_t \quad \textit{with} \quad \sigma_i'(x)\sigma_j(x) = \sigma_j'(x)\sigma_i(x) \quad \textit{for all } i,j = 1,\dots,n. \tag{1.10}$$

It is a well-known fact (see for instance [28]) that under this commuting assumption, the solution Y of (1.10) appears as a continuous functional of the sole noise B (that is, no need for any Lévy-area component). In a certain way, Equation (1.2) fits the above pattern. Indeed, for fixed $x \in (0,1)$, the noisy perturbation can be written as $\sum_{i=0}^\infty [\sqrt{\lambda_i}\, e_i(x) f(Y_t(x))] \circ d\beta^i_t$, i.e., we (morally) deal with $n = \infty$ and $\sigma_i(\cdot) = \sqrt{\lambda_i} e_i(x) f(\cdot)$ in (1.10). So, at least at this heuristic level, our continuity result (1.7) becomes quite natural. In a more specific way, we will see that due to the commuting property, the Lévy-area term arising from the rough-paths analysis of (1.2) can be easily reduced to some continuous functional of W (Lemma 3.9).

Remark 1.5. *As we shall see it in Section 3, our proof of Theorem 1.2 heavily relies on the properties of the fractional Sobolev spaces $\mathcal{B}_{\alpha,p}$, which we have recalled in the appendix. Unfortunately, many of these properties become much more restrictive as soon as the underlying space dimension is larger than 2, as illustrated by the classical Sobolev embeddings. This accounts for our choice to stick to a one-dimension heat equation. Note however that our considerations on the theoretical study of (1.4) (Section 2) could be easily extended to a multidimensional setting.*

Remark 1.6. *The results in this paper remain actually valid for any operator A of the form $A = -\partial_x(a \cdot \partial_x) + c$, where $c \geq 0$ and $a : [0,1] \to \mathbb{R}$ is a continuously differentiable function. Indeed, as explained in [10, Section 2.1], such an operator A also generates an analytic semigroup of contractions and one can identify the domains $\mathcal{D}(A_p^\alpha)$ of its fractional powers with the spaces $\mathcal{B}_{\alpha,p}$, which is sufficient to follow the lines of our reasoning.*

Unless otherwise stated, any constant c or C appearing in our computations below is understood as a generic constant which might change from line to line without further mention.

2 The Stratonovich integral

Recall that we are interested in the following mild equation:

$$Y_t = S_t\psi + \int_0^t S_{t-u}(f(Y_u) \circ dW_u), \quad t \in [0,T],\ \psi \in \mathcal{B}, \tag{2.1}$$

where $(S_t)_{t\geq 0}$ denotes the strongly continuous semigroup of operators generated by $-\Delta$ with Dirichlet boundary conditions, and W is assumed to satisfy Hypothesis 1.1.

The integral appearing in (2.1) is thus understood in some *Stratonovich sense*, an interpretation to be clarified in a convolutional setting, which is the main purpose of this first section. Once endowed with this interpretation, it turns out that (2.1) reduces to a common mild Itô equation with an additional drift term, and accordingly the existence and uniqueness of Y can be derived from well-known results (see Section 2.2).

Note that the following regularity assumption on f will prevail throughout the section.

Hypothesis 2.1. *The function $f : \mathbb{R} \to \mathbb{R}$ is bounded, of class $\mathcal{C}^2$ and with bounded derivatives.*

2.1 The Stratonovich integral

In order to interpret $\int_0^t S_{t-u}(f(Y_u) \circ dW_u)$, we restrict our attention to a particular class of processes Y. Namely, we assume that, on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, $\{Y_t,\, t \in [0,T]\}$ is the unique $\mathcal{B}$-valued mild solution of the following equation:

$$dY_t - \Delta Y_t dt = V_t^1 dt + V_t^2 dW_t, \quad Y_0 = \psi \in \mathcal{B}, \tag{2.2}$$

for some $\mathcal{F}_t$-adapted random fields $\{V_t^i,\, t \in [0,T]\}$, $i = 1, 2$, with continuous paths in $\mathcal{B}$ (recall that $\mathcal{B} := L^2(0,1)$). Moreover, we assume that

$$\sup_{t\leq T} E[\|V_t^2\|_{\mathcal{B}}^2] < +\infty. \tag{2.3}$$

In fact, such a process Y is explicitly given (see e.g. [9]) by

$$Y_t = S_t\psi + \int_0^t S_{t-s}V_s^1 ds + \int_0^t S_{t-s}\left(V_s^2 \cdot dW_s\right). \tag{2.4}$$

Now, a natural idea to define the integral in (2.1) in some Stratonovich sense would be the following: introduce the kernel $G_{t-s}(x,y)$ of S_{t-s} and, with the representation (1.3) of W in mind, set

$$\left[\int_0^t S_{t-s}\left(f(Y_s)\circ dW_s\right)\right](x) \text{ " = " } \sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(\int_0^t \langle G_{t-s}(x,*)f(Y_s),e_j\rangle_{\mathcal{B}}\circ d\beta_s^j\right),$$

where the symbol $*$ denotes the space variable and each integral

$$\int_0^t \langle G_{t-s}(x,*)\cdot f(Y_s),e_j\rangle_{\mathcal{B}}\circ d\beta_s^j$$

is interpreted in the (classical) Stratonovich sense. Nevertheless, it is a well-known fact that the process Y defined by (2.4) is not always a $\mathcal{B}$-valued semimartingale (in other words, Y is not always a strong solution of (2.4), see e.g. [9, Sec. 5.6]), making the definition of these integrals quite obscure at first sight.

To overcome this difficulty, we consider a standard semimartingale approximation of Y: for every $\varepsilon>0$, let Y^ε be the unique (strong) solution of

$$dY_t^\varepsilon - \Delta_\varepsilon Y_t^\varepsilon dt = V_t^1 dt + V_t^2 dW_t,\quad Y_0^\varepsilon=\psi,$$

where

$$-\Delta_\varepsilon := \frac{1}{\varepsilon}(\mathrm{Id}-(\mathrm{Id}-\varepsilon\Delta)^{-1})$$

stands for the Yosida approximation of $-\Delta$. In particular, $-\Delta_\varepsilon$ defines a monotone and bounded operator which converges pointwise to $-\Delta$ (see e.g. [4]). Then, for every fixed $\varepsilon>0$, Y^ε is a semimartingale, and we have (see e.g. [9, Proposition 7.5])

$$\lim_{\varepsilon\to 0}\sup_{t\le T} E\left[\|Y_t^\varepsilon - Y_t\|_{\mathcal{B}}^2\right]=0. \tag{2.5}$$

This extrinsic procedure will lead us to the following interpretation:

Proposition 2.2. *With the above notations, the family of Stratonovich integrals defined for all $t\in[0,T], x\in(0,1)$ by*

$$\left[\int_0^t S_{t-s}\left(f(Y_s^\varepsilon)\circ dW_s\right)\right](x) := \lim_{u\nearrow t}\sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(\int_0^u \langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}}\circ d\beta_s^j\right), \tag{2.6}$$

where the latter limit is considered in $L^2(\Omega)$, converges in $L^1(\Omega;\mathcal{C}^0([0,T];\mathcal{B}))$ as ε tends to 0. Its limit, that we denote by $\int_0^\cdot S_{\cdot-s}\left(f(Y_s)\circ dW_s\right)$, satisfies the relation

$$\int_0^t S_{t-s}\left(f(Y_s)\circ dW_s\right) = \int_0^t S_{t-s}\left(f(Y_s)\cdot dW_s\right) + \int_0^t S_{t-s}\left(V_s^2\cdot f'(Y_s)\cdot P\right)ds, \tag{2.7}$$

where $P(\xi):=\frac{1}{2}\sum_{k=1}^\infty \lambda_k e_k(\xi)^2$ and the notation $\int_0^t S_{t-s}\left(f(Y_s)\cdot dW_s\right)$ refers to the (usual) Itô integral.

Thus, the Stratonovich integral in (2.1) will henceforth be understood as in the latter proposition, in the class of processes Y satisfying an equation of the form (2.2). Note that the relation (2.7) provides us with a familiar decomposition for the Stratonovich integral as the sum of an Itô integral and a trace term, and it must be compared with the decomposition for the (standard) Stratonovich integral.

As a first step in the proof of Proposition 2.2, observe that the two terms in the right-hand side of (2.7) are indeed well-defined processes in $L^2(\Omega;\mathcal{B})$. This is a straightforward consequence of the boundedness of f, f', the trace-class assumption on W, and the fact that P defines a uniformly bounded function.

We point out that, in the definition (2.6), we first restrict the integral to $(0,u)$ with $u<t$ in order to avoid the singularity in the derivative of the kernel G. This will be clarified in the proof of the next lemma.

Lemma 2.3. *With the above notations, we have that, for all $u\in(0,t)$,*

$$\sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(\int_0^u \langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}}\circ d\beta_s^j\right) \tag{2.8}$$
$$=\left[\int_0^u S_{t-s}\left(f(Y_s^\varepsilon)\cdot dW_s\right)\right](x)+\int_0^u [S_{t-s}(f'(Y_s^\varepsilon)\cdot V_s^2\cdot P)](x)\,ds.$$

Proof. For any fixed $(u,x)\in(0,t)\times(0,1)$ and $j\in\mathbb{N}$, the process $s\mapsto\langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}}$, $s\in[0,u]$, defines a (real-valued) semimartingale. Hence, we can use Itô's formula to assert that

$$\begin{aligned}\langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}} &= \langle G_t(x,*)\cdot f(\psi),e_j\rangle_{\mathcal{B}}+\int_0^s\langle\partial_t G_{t-r}(x,*)\cdot f(Y_r^\varepsilon),e_j\rangle_{\mathcal{B}}\,dr\\ &\quad+\int_0^s\langle G_{t-r}(x,*)\big\{(\Delta_\varepsilon Y_r^\varepsilon+V_r^1)\cdot f'(Y_r^\varepsilon)+\frac{1}{2}V_r^2\cdot f''(Y_r^\varepsilon)\big\},e_j\rangle_{\mathcal{B}}\,dr\\ &\quad+\sum_{k=1}^{\infty}\sqrt{\lambda_k}\int_0^s\langle G_{t-r}(x,*)\cdot f'(Y_r^\varepsilon)\cdot V_r^2\cdot e_k,e_j\rangle_{\mathcal{B}}\,d\beta_r^k.\end{aligned} \tag{2.9}$$

The hypotheses on f and V^i, and the fact that $s\le u<t$, guarantee that all terms on the right-hand side above are well-defined. More precisely, using the spectral decomposition of G given by

$$G_{t-r}(x,y)=\sum_{k=1}^{\infty}e^{-k^2\pi^2(t-r)}e_k(x)e_k(y),$$

one proves that, P-a.s.,

$$\int_0^s\langle\partial_t G_{t-r}(x,*)\cdot f(Y_r^\varepsilon),e_j\rangle_{\mathcal{B}}\,dr\le C\sum_{k=1}^{\infty}\Big(e^{-k^2\pi^2(t-s)}-e^{-k^2\pi^2 t}\Big),$$

and the latter is finite since $s\in(0,t)$. As far as the second pathwise integral on the right-hand side of (2.9) is concerned, we have, for instance,

$$\begin{aligned}E\left[\left|\int_0^s\langle G_{t-r}(x,*)\cdot\Delta_\varepsilon Y_r^\varepsilon\cdot f'(Y_r^\varepsilon),e_j\rangle_{\mathcal{B}}\,dr\right|^2\right] &\le C\,E\left[\int_0^s\int_0^1 G_{t-r}(x,y)|[\Delta_\varepsilon Y_r^\varepsilon](y)|^2\,dydr\right]\\ &\le C\left(\sup_{r\le T}E[\|Y_r^\varepsilon\|_{\mathcal{B}}^2]\right)\int_0^s\frac{1}{\sqrt{t-r}}dr<+\infty.\end{aligned}$$

Here, we have used the fact that $0\le G_{t-r}(x,y)\le(2\pi(t-r))^{-1/2}\,e^{-\frac{(x-y)^2}{2(t-r)}}$. Similarly, one easily proves that the last term in (2.9) is a well-defined square-integrable random variable.

Plugging the expression (2.9) in (2.8) and using the definition of the (standard)

Stratonovich integral (see e.g. formula (3.9) in [19, p. 156]), we end up with

$$\begin{aligned}
&\sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(\int_0^u \langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}}\circ d\beta_s^j\right)\\
&\quad=\sum_{j=1}^{\infty}\sqrt{\lambda_j}\left(\int_0^u \langle G_{t-s}(x,*)\cdot f(Y_s^\varepsilon),e_j\rangle_{\mathcal{B}}\, d\beta_s^j\right)\\
&\qquad\quad+\frac{1}{2}\sum_{j=1}^{\infty}\lambda_j\int_0^u \langle G_{t-s}(x,*)\cdot f'(Y_s^\varepsilon)\cdot V_s^2\cdot e_j,e_j\rangle_{\mathcal{B}}\, ds\\
&\quad=\left[\int_0^u S_{t-s}\left(f(Y_s^\varepsilon)\cdot dW_s\right)\right](x)+\int_0^u\left[S_{t-s}(f'(Y_s^\varepsilon)\cdot V_s^2\cdot P)\right](x)\, ds,
\end{aligned}$$

which concludes the proof. □

We can now go back to our main statement.

Proof of Proposition 2.2. First, owing to the previous lemma, we have that the limit on the right-hand side of (2.6) equals to

$$\left[\int_0^t S_{t-s}\left(f(Y_s^\varepsilon)\cdot dW_s\right)\right](x)+\int_0^t [S_{t-s}(f'(Y_s^\varepsilon)\cdot V_s^2\cdot P)](x)\, ds.$$

This can be proved using the bounded convergence theorem. Hence, the proof reduces to the two assertions:

$$\lim_{\varepsilon\to 0}\sup_{t\le T}E\left[\left\|\int_0^t S_{t-s}\left(f(Y_s^\varepsilon)\cdot dW_s\right)-\int_0^t S_{t-s}\left(f(Y_s)\cdot dW_s\right)\right\|_{\mathcal{B}}\right]=0. \tag{2.10}$$

and

$$\lim_{\varepsilon\to 0}\sup_{t\le T}E\left[\left\|\int_0^t S_{t-s}(f'(Y_s^\varepsilon)\cdot V_s^2\cdot P)ds-\int_0^t S_{t-s}(f'(Y_s)\cdot V_s^2\cdot P)ds\right\|_{\mathcal{B}}\right]=0. \tag{2.11}$$

Let us first deal with (2.10). By the isometry property of the stochastic integral, the boundedness of S_{t-s} and the assumptions on f, we have:

$$\begin{aligned}
&E\left[\left\|\int_0^t S_{t-s}\left(f(Y_s^\varepsilon)\cdot dW_s\right)-\int_0^t S_{t-s}\left(f(Y_s)\cdot dW_s\right)\right\|_{\mathcal{B}}^2\right]\\
&\quad=E\left[\int_0^t\sum_{k=1}^{\infty}\lambda_k\|S_{t-s}([f(Y_s^\varepsilon)-f(Y_s)]e_k)\|_{\mathcal{B}}^2\, ds\right]\\
&\quad\le\sum_{k=1}^{\infty}\lambda_k\, E\left[\int_0^T\int_0^1|f(Y^\varepsilon(s,y))-f(Y(s,y))|^2|e_k(y)|^2\, dyds\right]\\
&\quad\le C\sum_{k=1}^{\infty}\lambda_k\sup_{t\le T}E\big[\|Y_t^\varepsilon-Y_t\|_{\mathcal{B}}^2\big]\le C\sup_{t\le T}E\big[\|Y_t^\varepsilon-Y_t\|_{\mathcal{B}}^2\big],
\end{aligned}$$

upon recalling that $\sum_{k=1}^{\infty}\lambda_k<\infty$.

In order to prove (2.11), we use the Sobolev embedding $L^1(0,1)\subset\mathcal{B}_{-\frac{1}{4}-\varepsilon}$ and the

assumptions on f and V^2. In fact, we have

$$\begin{aligned} & E\Big[\Big\| \int_0^t S_{t-s}\big([f'(Y_s^\varepsilon) - f'(Y_s)] \cdot V_s^2 \cdot P\big)\, ds\Big\|_{\mathcal{B}}\Big] \qquad (2.12)\\ &\leq \; C\,\|P\|_{\mathcal{B}_\infty} \int_0^t |t-s|^{-\frac{1}{4}-\varepsilon}\, E\big[\|[f'(Y_s^\varepsilon) - f'(Y_s)] \cdot V_s^2\|_{L^1}\big]\, ds \\ &\leq \; C\,\|P\|_{\mathcal{B}_\infty} \int_0^t |t-s|^{-\frac{1}{4}-\varepsilon}\, E\big[\|[f'(Y_s^\varepsilon) - f'(Y_s)]\|_{\mathcal{B}} \|V_s^2\|_{\mathcal{B}}\big]\, ds \\ &\leq \; C\,\Big(\sup_{t\leq T} E[\|Y_t^\varepsilon - Y_t\|_{\mathcal{B}}^2]^{1/2}\Big)\Big(\sup_{t\leq T} E[\|V_t^2\|_{\mathcal{B}}^2]^{1/2}\Big) \end{aligned}$$

(recall that $\|\cdot\|_{\mathcal{B}_\infty}$ refers to the supremum norm on $[0,1]$). Therefore, by the assumptions on V^2, the convergence (2.5) guarantees that (2.10) and (2.11) hold, and this lets us conclude the proof. □

2.2 Existence and uniqueness of solution

With the notations of the previous section, consider the following mild (Itô) equation:

$$Y_t = S_t\psi + \int_0^t S_{t-s}\,(f(Y_s)\cdot dW_s) + \int_0^t S_{t-s}(f'(Y_s)\cdot f(Y_s)\cdot P)\, ds, \qquad (2.13)$$

where we recall that $P(\xi) = \sum_{k=1}^\infty \lambda_k e_k(\xi)^2$. Hypotheses 1.1 and 2.1 allow us to apply standard methods and guarantee that this equation admits a unique $L^2(\Omega;\mathcal{B})$-valued solution Y (see [9]). In particular, we observe that Y solves an equation of the form (2.2) with $V_t^1 = f'(Y_t)\cdot f(Y_t)\cdot P$ and $V_t^2 = f(Y_t)$ and these random fields fulfill the assumptions specified in the previous section. Thus, for all $t\in[0,T]$, we can define the Stratonovich integral $\int_0^t S_{t-s}\,(f(Y_s)\circ dW_s)$ through Proposition 2.2 and we know that

$$\int_0^t S_{t-s}\,(f(Y_s)\circ dW_s) = \int_0^t S_{t-s}\,(f(Y_s)\cdot dW_s) + \int_0^t S_{t-s}(f'(Y_s)\cdot f(Y_s)\cdot P)ds,$$

which yields that Y is also a solution of (2.1).

Conversely, due to (2.7), it is readily checked that any solution of (2.1) in the class of processes satisfying an equation of the form (2.2) is also a solution of (2.13) (use the uniqueness of V^1, V^2 in (2.4)). This provides us with the following existence and uniqueness result.

Theorem 2.4. *Assume that Hypotheses 1.1 and 2.1 are both satisfied and that $\psi\in\mathcal{B}$. Then, there exists a unique $\mathcal{B}$-valued process $\{Y_t,\, t\in[0,T]\}$ which solves*

$$Y_t = S_t\psi + \int_0^t S_{t-u}(f(Y_u)\circ dW_u), \quad t\in[0,T].$$

Moreover, Y has a version with continuous paths and it holds that $\sup_{t\leq T} E[\|Y_t\|_{\mathcal{B}}^2] < \infty$.

3 A rough-paths type analysis of the equation

Let us now turn to the proof of Theorem 1.2. As announced in the Introduction, our strategy is based on a rough-paths type expansion of the equation. Accordingly, a few ingredients taken from the so-called *convolutional rough paths theory*, that is rough paths theory adapted to mild evolution equation, must be introduced in the first place.

3.1 Tools from (convolutional) rough paths theory

We gather here some preliminary material borrowed from [16] (see also [11, 10]). As underlined in the latter references, a key point towards a fruitful pathwise analysis of (1.4) lies in the following elementary observation: due to the semigroup property $S_{t+t'} = S_t \cdot S_{t'}$, it holds that, for any $s < t$,

$$\begin{aligned} Y_t - Y_s &= \int_s^t S_{t-u}(f(Y_u) \circ dW_u) + \int_0^s \big[S_{t-u} - S_{s-u}\big](f(Y_u) \circ dW_u) \\ &= \int_s^t S_{t-u}\left(f(Y_u) \circ dW_u\right) + a_{ts} Y_s, \quad \text{where} \quad a_{ts} := S_{t-s} - \mathrm{Id}\,. \end{aligned}$$

Otherwise stated, by setting $(\hat{\delta}Y)_{ts} := (Y_t - Y_s) - a_{ts}Y_s$, the equation (1.4) can be equivalently written in the convenient form:

$$Y_0 = \psi, \qquad (\hat{\delta}Y)_{ts} = \int_s^t S_{t-u}\left(f(Y_u) \circ dW_u\right), \quad 0 \le s \le t \le 1. \tag{3.1}$$

This should be compared with the behaviour of solutions to standard (stochastic) differential equations: if $X_t = a + \int_0^t \sigma(X_u)\, dB_u$, then $(\delta X)_{ts} := X_t - X_s = \int_s^t \sigma(X_u)\, dB_u$. Then, in a rough-paths setting, we are naturally led to extend the definition of $\hat{\delta}$ to processes with 2 variables, as follows:

Notation 3.1. *For all processes $y : [0,T] \to \mathcal{B}$ and $z : \mathcal{S}_2 \to \mathcal{B}$, where $\mathcal{S}_2 := \{(s,t) \in [0,T]^2 :\ s \le t\}$ denotes the two-dimensional simplex, we set, for $s \le u \le t \in [0,T]$:*

$$(\delta y)_{ts} := y_t - y_s, \qquad (\hat{\delta} y)_{ts} := (\delta y)_{ts} - a_{ts} y_s = y_t - S_{t-s} y_s, \tag{3.2}$$

$$(\hat{\delta} z)_{tus} := z_{ts} - z_{tu} - S_{t-u} z_{us}. \tag{3.3}$$

To make the notations (3.2)-(3.3) even more legitimate in this convolutional context, let us point out the following algebraic properties:

Proposition 3.2. *For any $y : [0,T] \to \mathcal{B}$, it holds:*

(i) Telescopic sum: $\hat{\delta}(\hat{\delta} y)_{tus} = 0$ and $(\hat{\delta} y)_{ts} = \sum_{i=0}^{n-1} S_{t-t_{i+1}} (\hat{\delta} y)_{t_{i+1} t_i}$ for any partition $\{s = t_0 < t_1 < \ldots < t_n = t\}$ of an interval $[s,t]$ of $[0,T]$.

(ii) Chasles relation: if $\mathcal{J}_{ts} := \int_s^t S_{t-u}\,(y_u \cdot dW_u)$, then $\hat{\delta}\mathcal{J} = 0$.

Both points (i) and (ii) are straighforward consequences of the semigroup property. Now, in accordance with the new expression (3.1) for the equation, a $\hat{\delta}$-version of the classical Hölder norm must come into the picture. To this end, fix a subinterval $I \subset [0,T]$ and a Banach space V. Then, if $y : I \to V$ and $\lambda > 0$, set

$$\mathcal{N}[y; \hat{\mathcal{C}}^\lambda(I;V)] := \sup_{s<t\in I} \frac{\|(\hat{\delta} y)_{ts}\|_V}{|t-s|^\lambda}, \tag{3.4}$$

and define $\hat{\mathcal{C}}^\lambda(I;V)$ as the set of processes $y : I \to V$ such that $\mathcal{N}[y; \hat{\mathcal{C}}^\lambda(I;V)] < \infty$.

As we will see it in the sequel, a proper control for the expansion of $\int_s^t S_{t-u}(f(Y_u) \circ dW_u)$ also requires the extension of both definitions (1.5) and (3.4) to processes with 2 or 3 variables. Precisely, if $z : \mathcal{S}_2 \to V$ and $h : \mathcal{S}_3 \to V$, where $\mathcal{S}_3 := \{(t,u,s) \in [0,T]^3 : s \le u \le t\}$, we set

$$\mathcal{N}[z; \mathcal{C}_2^\lambda(I;V)] := \sup_{s<t\in I} \frac{\|z_{ts}\|_V}{|t-s|^\lambda}, \qquad \mathcal{N}[h; \mathcal{C}_3^\lambda(I;V)] := \sup_{s<u<t\in I} \frac{\|h_{tus}\|_V}{|t-s|^\lambda}, \tag{3.5}$$

and we define $\mathcal{C}_2^\lambda(I;V)$ (resp. $\mathcal{C}_3^\lambda(I;V)$) along the same lines as $\hat{\mathcal{C}}^\lambda(I;V)$. Observe for instance that if $y \in \mathcal{C}_2^\lambda(I;\mathcal{L}(V,W))$ and $z \in \mathcal{C}_2^\beta(I;V)$, then the process h defined as $h_{tus} = y_{tu}z_{us}$ ($s \le u \le t \in I$) belongs to $\mathcal{C}_3^{\lambda+\beta}(I;W)$.

Note that when $I = [0,T]$, we will more simply write $\mathcal{C}_k^\lambda(V) := \mathcal{C}_k^\lambda(I;V)$ for $k \in \{1,2,3\}$. Besides, from now on, we use the following convenient notation for products of processes.

Notation 3.3. *If $g : \mathcal{S}_n \to \mathcal{L}(V,W)$ and $h : \mathcal{S}_m \to W$ (with $n,m \in \{1,2,3\}$), we define the product $gh : \mathcal{S}_{n+m-1} \to W$ by the formula*

$$(gh)_{t_1\dots t_{m+n-1}} := g_{t_1\dots t_n} h_{t_n\dots t_{n+m-1}}. \tag{3.6}$$

With this convention, it is readily checked that if $g : \mathcal{S}_2 \to \mathcal{L}(\mathcal{B}_\kappa, \mathcal{B}_\alpha)$ and $h : [0,T] \to \mathcal{B}_\kappa$, then $\hat\delta(gh) : \mathcal{S}_3 \to \mathcal{B}_\alpha$ obeys the rule:

$$\hat\delta(gh) = (\hat\delta g)h - g(\delta h). \tag{3.7}$$

To end up with this toolbox, let us report what may be seen as the cornerstone result of the convolutional rough paths theory, namely the existence of (some kind of) an inverse operator for $\hat\delta$, denoted by $\hat\Lambda$, and which will play a prominent role in our forthcoming decomposition (3.9). In brief, this operator allows us to get both a nice expression and a sharp estimate for the *regular* terms, i.e., the terms with Hölder regularity strictly larger than 1, that arise from the expansion of $\int_s^t S_{t-u}(f(Y_u)\circ dW_u)$ (see in particular the proof of Lemma 3.12).

Theorem 3.4. *Fix an interval $I \subset [0,T]$, a parameter $\kappa \ge 0$ and let $\mu > 1$. For any $h \in \mathcal{C}_3^\mu(I;\mathcal{B}) \cap \operatorname{Im}\hat\delta$, there exists a unique element*

$$\hat\Lambda h \in \cap_{\alpha\in[0,\mu)} \mathcal{C}_2^{\mu-\alpha}(I;\mathcal{B}_\alpha)$$

such that $\hat\delta(\hat\Lambda h) = h$. Moreover, $\hat\Lambda h$ satisfies the following contraction property: for all $\alpha \in [0,\mu)$,

$$\mathcal{N}[\hat\Lambda h; \mathcal{C}_2^{\mu-\alpha}(I;\mathcal{B}_\alpha)] \le c_{\alpha,\mu}\, \mathcal{N}[h; \mathcal{C}_3^\mu(I;\mathcal{B})]. \tag{3.8}$$

The proof of this result can be found in [16, Theorem 3.5].

3.2 A rough-paths type expansion of the solution

We are now ready to settle our reasoning, which applies to a smooth enough vector field f:

Hypothesis 3.5. *The function $f : \mathbb{R} \to \mathbb{R}$ in (3.1) is of class $\mathcal{C}^3$, bounded and with bounded derivatives.*

Our main task will actually consist in establishing the following pathwise decomposition for the solution Y to (3.1):

Theorem 3.6. *Assume that both Hypotheses 1.1 and 3.5 hold true. Fix $\gamma \in (\frac12, \frac12+\eta)$ and assume that $\psi \in \mathcal{B}_\gamma$. Then the $\hat\delta$-variations of the solution Y to (3.1) can be expanded as*

$$(\hat\delta Y)_{ts} = \int_s^t S_{t-u}(f(Y_u)\circ dW_u) = L_{ts}^W(f(Y_s)) + L_{ts}^{WW}(f(Y_s)\cdot f'(Y_s)) + \hat\Lambda_{ts}\big(R^Y\big), \tag{3.9}$$

where we have set, for all $s<u<t$,

$$L^W_{ts}(\varphi) := \int_s^t S_{t-u}(\varphi \cdot dW_u), \tag{3.10}$$

$$L^{WW}_{ts}(\varphi) := \int_s^t S_{t-u}\left(\varphi \cdot (\delta W)_{us} \cdot dW_u\right) + \int_s^t S_{t-u}\left(\varphi \cdot P\right) du, \tag{3.11}$$

and

$$R^Y_{tus} := -\hat{\delta}\big(L^W f(Y) + L^{WW}\big(f(Y)\cdot f'(Y)\big)\big)_{tus}. \tag{3.12}$$

The theorem must be read as follows: in the expansion of $\int_s^t S_{t-u}(f(Y_u)\circ dW_u)$, we can exhibit a *main* term, namely

$$L^W_{ts}(f(Y_s)) + L^{WW}_{ts}(f(Y_s)\cdot f'(Y_s)),$$

and a *residual* term $\hat{\Lambda}_{ts}\big(R^Y\big)$ with Hölder regularity strictly larger than 1, in the sense of Theorem 3.4 (take $\alpha = 0$ in (3.8)). Besides, from the decomposition (3.9), we can somehow conclude that the whole dynamics induced by W is "encoded" through the two (stochastic) operator-valued processes L^W and L^{WW}. So, before we turn to the proof of (3.9), let us elaborate on the properties of these two processes.

3.3 The couple (L^W, L^{WW})

At this point, we consider L^W_{ts} and L^{WW}_{ts} as stochastic linear operators acting on the space of smooth functions φ. The following (straightforward) relation accounts for the algebraic behaviour of the couple (L^W, L^{WW}): it is the convolutional analog of the classical Chen's relation between a process and its Lévy area (see [15]).

Proposition 3.7. *The processes L^W and L^{WW} obey the following algebraic rules: For all $s<u<t$ and all smooth function φ,*

$$(\hat{\delta}L^W)_{tus}(\varphi) = 0 \quad , \quad (\hat{\delta}L^{WW})_{tus}(\varphi) = L^W_{tu}(\varphi\cdot(\delta W)_{us}). \tag{3.13}$$

Now, it matters to identify the regularity properties of L^W and L^{WW} as 2-variables processes. A first clue in this direction is given by the following (a.s.) regularity result for the noise W itself.

Lemma 3.8. *Under Hypothesis 1.1, one has (a.s.) $W \in \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})$ for every integer $p\geq 1$ and every small $\varepsilon>0$.*

Proof. By using our forthcoming Proposition 4.3, we deduce that

$$E\big[\|(\delta W)_{ts}\|^{2pq}_{\mathcal{B}_{\eta,2p}}\big] \leq C_{p,q} E\big[|(\delta\beta)_{ts}|^{2pq}\big] \leq C_{p,q}\,|t-s|^{pq},$$

for all $q\geq 1$, and the result is now a straightforward consequence of the Garsia-Rodemich-Rumsey Lemma 6.1 (take $\delta^*=\delta$ and $R=\delta W$ in the latter statement). □

Our second ingredient towards the regularity properties of (L^W, L^{WW}) relies on two successive observations. First, due to their relative simplicity, the two expressions (3.10)-(3.11) can be integrated by parts. Then, owing to the some obvious commuting properties, we can turn L^{WW} into an easy-to-handle functional of δW. This is what we propose to detail in the proof of the following Lemma.

Lemma 3.9. *For every smooth function φ and all $s<t\in[0,T]$, the following formulas hold true (a.s.):*

$$L_{ts}^{W}(\varphi)=S_{t-s}(\varphi\cdot(\delta W)_{ts})-\int_{s}^{t}\Delta S_{t-u}(\varphi\cdot(\delta W)_{tu})\,du, \tag{3.14}$$

$$L_{ts}^{WW}(\varphi)=\frac{1}{2}\left\{S_{t-s}(\varphi\cdot(\delta W)_{ts}^{2})-\int_{s}^{t}\Delta S_{t-u}\left(\varphi\cdot\left[(\delta W)_{tu}^{2}+2(\delta W)_{tu}\cdot(\delta W)_{us}\right]\right)\,du\right\}. \tag{3.15}$$

Proof. With the expansion (1.3) in mind, it is easily checked, by setting

$$W_{t}^{N}:=\sum_{i=1}^{N}\sqrt{\lambda_{i}}\beta_{t}^{i}e_{i},$$

that $L_{ts}^{W}(\varphi)=\lim_{N\to\infty}\sum_{i=1}^{N}\int_{s}^{t}S_{t-u}(\varphi\cdot dW_{u}^{N})$ and

$$\int_{s}^{t}S_{t-u}(\varphi\cdot(\delta W)_{us}\cdot dW_{u})=\lim_{N\to\infty}\int_{s}^{t}S_{t-u}(\varphi\cdot(\delta W^{N})_{us}\cdot dW_{u}^{N}),$$

where the limits are taken in $L^{2}(\Omega,\mathcal{B})$. The proof then reduces to applications of Itô's formula and we only elaborate on (3.15). For fixed $i,j\in\{1,\dots,N\}$, apply Itô's formula to the (random) function $F_{s,t}^{i,j}:[s,t]\times\mathbb{R}\times\mathbb{R}\to\mathcal{B}$ defined by

$$F_{s,t}^{i,j}(u,x,y):=S_{t-u}(\varphi\cdot e_{i}\cdot e_{j})\big[(x-\beta_{s}^{i})(y-\beta_{s}^{j})-(\delta\beta^{i})_{ts}(\delta\beta^{j})_{ts}\big]$$

so as to deduce

$$0=F_{s,t}^{i,j}(t,\beta_{t}^{i},\beta_{t}^{j})=F_{s,t}^{i,j}(s,\beta_{s}^{i},\beta_{s}^{j})-\int_{s}^{t}\Delta S_{t-u}(\varphi\cdot e_{i}\cdot e_{j})\big[(\delta\beta^{i})_{us}(\delta\beta^{j})_{us}-(\delta\beta^{i})_{ts}(\delta\beta^{j})_{ts}\big]\,du$$
$$+\int_{s}^{t}S_{t-u}(\varphi\cdot e_{i}\cdot e_{j})(\delta\beta^{i})_{us}\,d\beta_{u}^{j}+\int_{s}^{t}S_{t-u}(\varphi\cdot e_{i}\cdot e_{j})(\delta\beta^{j})_{us}\,d\beta_{u}^{i}+\mathbf{1}_{\{i=j\}}\int_{s}^{t}S_{t-u}(\varphi\cdot e_{i}\cdot e_{j})\,du.$$

By taking the sum over i,j, we deduce the formula

$$S_{t-s}(\varphi\cdot(\delta W^{N})_{ts}^{2})-\int_{s}^{t}\Delta S_{t-u}\big(\varphi\cdot[(\delta W^{N})_{ts}^{2}-(\delta W^{N})_{us}^{2}]\big)\,du$$
$$=2\int_{s}^{t}S_{t-u}(\varphi\cdot(\delta W^{N})_{us}\cdot dW_{u}^{N})+\int_{s}^{t}S_{t-u}\big(\varphi\cdot\big(\sum_{i=1}^{N}\lambda_{i}e_{i}^{2}\big)\big)$$

and by passing to the limit (in $L^{2}(\Omega,\mathcal{B})$), we get

$$L_{ts}^{WW}(\varphi)=\frac{1}{2}\Big\{S_{t-s}(\varphi\cdot(\delta W)_{ts}^{2})+\int_{s}^{t}\Delta S_{t-u}\big(\varphi\cdot[(\delta W)_{ts}^{2}-(\delta W)_{us}^{2}]\big)\,du\Big\}.$$

Formula (3.15) immediately follows.

□

We are now in a position to extend both L_{ts}^{W} and L_{ts}^{WW} to larger classes of functions φ and retrieve the following (a.s.) bounds, which will be at the core of our identification procedure:

Proposition 3.10. *Under the hypotheses of Theorem 3.6, for any small $\varepsilon > 0$, there exists $\tilde{\varepsilon} > 0$ and $p \geq 1$ such that (almost surely)*

$$\mathcal{N}[L^W; \mathcal{C}_2^{\frac{1}{2}-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))] + \mathcal{N}[L^W; \mathcal{C}_2^{(\frac{1}{2}-\gamma+\eta)-\varepsilon}(\mathcal{L}(\mathcal{B}_{1/2},\mathcal{B}_\gamma))] + \mathcal{N}[L^W; \mathcal{C}_2^{\frac{1}{4}-\varepsilon}(\mathcal{L}(\mathcal{B}_\infty,\mathcal{B}_\infty))] \leq c_{\varepsilon,\tilde{\varepsilon},p}\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})], \quad (3.16)$$

$$\mathcal{N}[L^{WW}; \mathcal{C}_2^{1-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))] + \mathcal{N}[L^{WW}; \mathcal{C}_2^{(1-\gamma+\eta)-\varepsilon}(\mathcal{L}(\mathcal{B}_{1/2},\mathcal{B}_\gamma))] + \mathcal{N}[L^{WW}; \mathcal{C}_2^{\frac{3}{4}-\varepsilon}(\mathcal{L}(\mathcal{B}_\infty,\mathcal{B}_\infty))] \leq c_{\varepsilon,\tilde{\varepsilon},p}\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})]^2, \quad (3.17)$$

for some constant $c_{\varepsilon,\tilde{\varepsilon},p}$.

Note here how important the assumption $\gamma < \frac{1}{2} + \eta$ in Theorem 3.6 to ensure that $(\frac{1}{2} - \gamma + \eta) - \varepsilon > 0$ for any small enough $\varepsilon > 0$.

Proof. In fact, thanks to the representation formulas (3.14)-(3.15) and the pathwise regularity of W (Lemma 3.8), all of these bounds can be derived from the classical properties of the fractional Sobolev spaces (see Appendix A). For instance, owing to (5.3), one has, for any $p \geq 1$ and $\alpha \geq \frac{1}{4p}$,

$$\|\varphi \cdot (\delta W)_{ts}\|_{\mathcal{B}_{-\alpha}} \leq c_{\alpha,p}\|(\delta W)_{ts}\|_{L^{2p}(0,1)}\|\varphi\|_{\mathcal{B}}, \quad (3.18)$$

so that for any $\tilde{\varepsilon}$ small enough,

$$\begin{aligned} \|S_{t-s}(\varphi \cdot (\delta W)_{ts})\|_{\mathcal{B}} &\leq c_{\alpha,p}\,|t-s|^{-\alpha}\,\|(\delta W)_{ts}\|_{L^{2p}(0,1)}\|\varphi\|_{\mathcal{B}} \qquad \text{(use (5.1))} \\ &\leq c_{\alpha,p,\tilde{\varepsilon}}\,|t-s|^{\frac{1}{2}-\tilde{\varepsilon}-\alpha}\,\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})]\,\|\varphi\|_{\mathcal{B}}. \end{aligned}$$

In the same way,

$$\begin{aligned} \|\Delta S_{t-u}(\varphi \cdot (\delta W)_{tu}\|_{\mathcal{B}} &\leq c_{\alpha,p}\,|t-s|^{-1-\alpha}\,\|(\delta W)_{ts}\|_{L^{2p}(0,1)}\|\varphi\|_{\mathcal{B}} \qquad \text{(use (5.1))} \\ &\leq c_{\alpha,p,\varepsilon}\,|t-s|^{(\frac{1}{2}-\tilde{\varepsilon}-\alpha)-1}\,\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})]\,\|\varphi\|_{\mathcal{B}}. \end{aligned}$$

By taking α small enough, i.e., p large enough, we get the expected bound, namely

$$\mathcal{N}[L^W; \mathcal{C}_2^{\frac{1}{2}-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))] \leq c_{\varepsilon,\tilde{\varepsilon},p}\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})].$$

The other estimates for L^W can be proved along the same lines. As far as L^{WW} is concerned, observe for instance that if $\varepsilon > 0$ is small enough, then one has

$$\begin{aligned} &\|\Delta S_{t-u}(\varphi \cdot (\delta W)_{tu} \cdot (\delta W)_{us})\|_{\mathcal{B}_\gamma} \\ &\quad\leq c\,|t-u|^{(\eta-\gamma)-1}\,\|\varphi \cdot (\delta W)_{tu} \cdot (\delta W)_{us}\|_{\mathcal{B}_\eta} \\ &\quad\leq c\,|t-u|^{(\eta-\gamma)-1}\,\|(\delta W)_{tu} \cdot (\delta W)_{us}\|_{\mathcal{B}_\eta}\|\varphi\|_{\mathcal{B}_{1/2}} \qquad \text{(use (5.4))} \\ &\quad\leq c\,|t-u|^{(\eta-\gamma)-1}\,\|(\delta W)_{tu}\|_{\mathcal{B}_{\eta,4}}\|(\delta W)_{us}\|_{\mathcal{B}_{\eta,4}}\|\varphi\|_{\mathcal{B}_{1/2}} \qquad \text{(use (5.5))} \\ &\quad\leq c_\varepsilon\,|t-u|^{(\frac{1}{2}+\eta-\gamma-\tilde{\varepsilon})-1}\,|u-s|^{\frac{1}{2}-\tilde{\varepsilon}}\,\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,4})]^2\|\varphi\|_{\mathcal{B}_{1/2}}, \end{aligned}$$

which entails that

$$\mathcal{N}[L^{WW}; \mathcal{C}_2^{(1-\gamma+\eta)-\varepsilon}(\mathcal{L}(\mathcal{B}_{1/2},\mathcal{B}_\gamma))] \leq c_{\varepsilon,\tilde{\varepsilon},p}\mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,4})]^2.$$

The (analogous) proofs for the other bounds are left to the reader. □

3.4 Proof of Theorem 3.6

First, we need to justify that the right-hand side of the decomposition (3.9) is well-defined. This will rely (among others) on the following a priori controls for the solution Y. For the sake of clarity, we have postponed the proof of this statement to Appendix B.

Lemma 3.11. *Under the hypotheses of Theorem 3.6, one has (almost surely)*

$$Y \in \hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty) \cap \mathcal{C}^0(\mathcal{B}_\gamma), \tag{3.19}$$

$$K^Y := \hat{\delta} Y - L^W f(Y) \in \mathcal{C}_2^{\frac{1}{2}+\eta}(\mathcal{B}). \tag{3.20}$$

Recall that according to our convention (3.6), the definition of K^Y in (3.20) must be understood as $K^Y_{ts} := (\hat{\delta} Y)_{ts} - L^W_{ts}(f(Y_s))$ for every $s < t \in [0,T]$.

Lemma 3.12. *Under the hypotheses of Theorem 3.6, let Z be the process given by $Z_0 = \psi$ and*

$$(\hat{\delta} Z)_{ts} = L^W_{ts}(f(Y_s)) + L^{WW}_{ts}(f(Y_s) \cdot f'(Y_s)) + \hat{\Lambda}_{ts}\big(R^Y\big). \tag{3.21}$$

Then, almost surely, Z is well-defined as an element of $\hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty) \cap \mathcal{C}^0(\mathcal{B}_\gamma)$, and there exists a constant $\lambda > 0$ such that for any subinterval $I = [\ell_1, \ell_2] \subset [0,T]$, one has

$$\mathcal{N}[Z; \hat{\mathcal{C}}^{2\eta}(I;\mathcal{B}_\infty)] + \mathcal{N}[Z; \mathcal{C}^0(I;\mathcal{B}_\gamma)] \le \|Z_{\ell_1}\|_{\mathcal{B}_\gamma} + c_{W,f}|I|^\lambda \mathcal{N}[Y; \mathcal{Q}(I)], \tag{3.22}$$

where we have set

$$\mathcal{N}[Y; \mathcal{Q}(I)] := \mathcal{N}[Y; \hat{\mathcal{C}}^{2\eta}(I;\mathcal{B}_\infty)] + \mathcal{N}[Y; \mathcal{C}^0(I;\mathcal{B}_\gamma)] + \mathcal{N}[K^Y; \mathcal{C}_2^{\frac{1}{2}+\eta}(I;\mathcal{B})]$$

Proof. First, according to Theorem 3.4, we need to justify that $R^Y \in \mathcal{C}_3^\mu(\mathcal{B})$ for some $\mu > 1$. To this end, expand R using the algebraic rules (3.7) and (3.19), which gives

$$R^Y_{tus} = L^W_{tu} N_{us} + L^{WW}_{tu} \delta(f(Y) \cdot f'(Y))_{us}, \tag{3.23}$$

with $N_{us} := \delta(f(Y))_{us} - (\delta W)_{us} \cdot f(Y_s) \cdot f'(Y_s)$. Thanks to (3.17) and (3.19), it is readily checked that $L^{WW}\delta(f(Y)\cdot f'(Y)) \in \mathcal{C}_3^{1+2\eta-\varepsilon}(\mathcal{B})$ for any small $\varepsilon > 0$, since

$$\begin{aligned} \|\delta(f(Y)\cdot f'(Y))_{us}\|_{\mathcal{B}} &\le c\,\|(\delta Y)_{us}\|_{\mathcal{B}} \le c\,\{\|(\hat{\delta} Y)_{us}\|_{\mathcal{B}} + \|a_{us} Y_s\|_{\mathcal{B}}\} \\ &\le c\,\{|u-s|^{2\eta}\mathcal{N}[Y;\mathcal{C}^{2\eta}(\mathcal{B}_\infty)] + |u-s|^\gamma \mathcal{N}[Y;\mathcal{C}^0(\mathcal{B}_\gamma)]\}, \end{aligned}$$

where we have used (5.2) to get the last inequality (recall that $a_{ts} := S_{t-s} - \mathrm{Id}$).
Then, as far as $L^W N$ is concerned, let us expand N using standard differential calculus, which provides us with the expression

$$\begin{aligned} N_{us} = \int_0^1 dr\, f'(Y_s + r(\delta Y)_{us}) \cdot \big\{K^Y_{us} + a_{us} Y_s + L^{aW}_{us}(f(Y_s))\big\} \\ + \int_0^1 dr\, [f'(Y_s + r(\delta Y)_{us}) - f'(Y_s)] \cdot (\delta W)_{us} \cdot f(Y_s), \end{aligned} \tag{3.24}$$

where the additional operator-valued process L^{aW} is defined by

$$L^{aW}_{ts}(\varphi) := \int_s^t a_{tu}(\varphi \cdot dW_u) = a_{ts}(\varphi \cdot (\delta W)_{ts}) - \int_s^t \Delta S_{t-u}(\varphi \cdot (\delta W)_{tu}) du.$$

Now, since $L^W \in \mathcal{C}_2^{\frac{1}{2}-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))$, it is sufficient to prove that $N \in \mathcal{C}_2^{\frac{1}{2}+\varepsilon}(\mathcal{B})$ for some small $\varepsilon > 0$. But, with the expansion (3.24) in hand, this becomes an easy consequence of the a priori controls given by Lemma 3.11, together with the regularity property:

$$\mathcal{N}[L^{aW}; \mathcal{C}_2^{(\frac{1}{2}+\eta)-\varepsilon}(\mathcal{L}(\mathcal{B}_{1/2},\mathcal{B}))] \le c_{\varepsilon,\tilde{\varepsilon},p} \mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})],$$

derived from (5.4). Note in particular how important the assumption $\gamma > 1/2$, insofar as, by (5.2),

$$\|f'(Y_s + r(\delta Y)_{us}) \cdot (a_{us} Y_s)\|_{\mathcal{B}} \leq C_f \|a_{us} Y_s\|_{\mathcal{B}} \leq C|u-s|^{\gamma} \mathcal{N}[Y; \mathcal{C}^0(\mathcal{B}_\gamma)].$$

We are thus in a position to apply $\hat{\Lambda}$ to R^Y, and so Z is properly defined through (3.21). The regularity of Z and the bound (3.22) are immediate consequences of (3.16)-(3.17) and the contraction property (3.8) of $\hat{\Lambda}$. The details are left to the reader.

□

Remark 3.13. *Although not optimal, the two regularity results (3.19) and (3.20) are thus sufficient for us to prove that the right-hand side of the decomposition (3.9) is indeed well-defined. We also retrieve an important stability phenomenon here: Y and Z both belong to the same space $\hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty) \cap \mathcal{C}^0(\mathcal{B}_\gamma)$. A posteriori, this accounts for our choice in favor of this particular topology.*

We can eventually proceed to prove Theorem 3.6.

Proof of Theorem 3.6. We need to identify the increments of Y with those of the process Z defined in Lemma 3.12. To do so, we naturally rely on some expansion of the right-hand side of (2.13). Precisely, we have that

$$\begin{aligned}\int_s^t S_{t-u}(f(Y_u) \cdot dW_u) + \int_s^t S_{t-u}(P \cdot f(Y_u) \cdot f'(Y_u))\, du \\ = \quad L^W_{ts}(f(Y_s)) + L^{WW}_{ts}(f(Y_s) \cdot f'(Y_s)) + J^Y_{ts},\end{aligned}$$

with

$$J^Y_{ts} := \int_s^t S_{t-u}(P \cdot \delta(f(Y) \cdot f'(Y))_{us})\, du + \int_s^t S_{t-u}(N^Y_{us} \cdot dW_u), \tag{3.25}$$

where the process $N^Y_{ts} = \delta(f(Y))_{ts} - (\delta W)_{ts} \cdot f(Y_s) \cdot f'(Y_s)$ has already been considered in the proof of Lemma 3.12. Therefore, with this notation, it holds that

$$\hat{\delta}(Z - Y) = \hat{\Lambda}_{ts}(R^Y) - J^Y_{ts}.$$

Now, by the contraction property (3.8), we know that $\hat{\Lambda}(R^Y) \in \mathcal{C}_2^{\mu_1}(\mathcal{B})$ for some $\mu_1 > 1$. Besides, with the same ingredients as in the proof of Lemma 3.11 (Burkholder-Davis-Gundy inequality plus Lemma 6.1, see Appendix B), we can easily lean on the expansion (3.24) of N to prove that $J^Y \in \mathcal{C}_2^{\mu_2}(\mathcal{B})$ for some $\mu_2 > 1$ (note that $\hat{\delta} J^Y = R^Y$). Consequently, $\hat{\delta}(Z - Y) \in \mathcal{C}_2^{\mu}(\mathcal{B})$ with $\mu = \inf(\mu_1, \mu_2) > 1$, and this entails that $\hat{\delta}(Z - Y) = 0$. Indeed, for any partition $\mathcal{P}_{[s,t]} = \{s = t_1 < \ldots < t_n = t\}$ of $[s,t]$, one has, due to the telescopic sum property reported in Proposition 3.2,

$$\|\hat{\delta}(Z-Y)_{ts}\|_{\mathcal{B}} \leq \sum_i \|\hat{\delta}(Z-Y)_{t_{i+1}t_i}\|_{\mathcal{B}} \leq c \sum_i |t_{i+1} - t_i|^{\mu} \leq |\mathcal{P}_{[s,t]}|^{\mu-1} |t-s|,$$

and we conclude by letting the mesh $|\mathcal{P}_{[s,t]}| := \max_i |t_{i+1} - t_i|$ tend to 0. □

As a straightforward consequence of the decomposition (3.9), we can exhibit an almost sure bound for Y in terms of W. Indeed, by plugging the estimate (3.22) back into the equation, we deduce that for any subinterval $I = [\ell_1, \ell_2] \subset [0,T]$,

$$\mathcal{N}[Y; \hat{\mathcal{C}}^{2\eta}(I; \mathcal{B}_\infty)] + \mathcal{N}[Y; \mathcal{C}^0(I; \mathcal{B}_\gamma)] \leq \|Y_{\ell_1}\|_{\mathcal{B}_\gamma} + C_W |I|^{\lambda} \mathcal{N}[Y; \mathcal{Q}(I)]$$

for some constant $\lambda > 0$, and similar estimates for $K^Y = L^{WW}(f(Y) \cdot f'(Y)) + \hat{\Lambda}(R^Y)$ finally show that

$$\mathcal{N}[Y; \mathcal{Q}(I)] \leq \|Y_{\ell_1}\|_{\mathcal{B}_\gamma} + C_W |I|^{\lambda} \mathcal{N}[Y; \mathcal{Q}(I)].$$

At this point, a basic patching argument easily leads us to the following statement:

Corollary 3.14. *Under the hypotheses of Theorem 3.6, there exist $\varepsilon > 0$ and $p \geq 1$ such that*

$$\mathcal{N}[Y; \mathcal{Q}([0,T])] \leq G_{\varepsilon,p}\big(\|\psi\|_{\mathcal{B}_\gamma}, \mathcal{N}[W; \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})]\big) \tag{3.26}$$

for some deterministic function $G_{\varepsilon,p} : (\mathbb{R}^+)^2 \to \mathbb{R}^+$ bounded on bounded sets.

3.5 Comparison with smooth solutions

The previous considerations will allow us to prove our continuity result (Theorem 1.2) and for this purpose, we first go back to the case where the driving noise is an absolutely continuous process $\widetilde{W}$ (with values in $\mathcal{B}_{\eta,2p}$), *assumingly defined on the same probability space as W*. In this situation, our mild equation is naturally understood in a pathwise sense as a classical (Riemann-Lebesgue) mild equation, i.e.,

$$\widetilde{Y}_t = S_t\widetilde{\psi} + \int_0^t S_{t-u}(f(\widetilde{Y}_u)\cdot d\widetilde{W}_u) = S_t\widetilde{\psi} + \int_0^t S_{t-u}(f(Y_u)\cdot \widetilde{W}'_u)\, du, \tag{3.27}$$

and the (pathwise) existence and uniqueness of the solution $\widetilde{Y}$ follows from standard PDE results. The key step towards a comparison between Y and $\widetilde{Y}$ lies in the following result, which points out the similarity between the couple (L^W, L^{WW}) at the core of the previous considerations and the couple $(L^{\widetilde{W}}, L^{\widetilde{W}\widetilde{W}})$ constructed from $\widetilde{W}$:

Lemma 3.15. *Define the operator-valued processes $L^{\widetilde{W}}$ and $L^{\widetilde{W}\widetilde{W}}$ in the classical Riemann-Lebesgue sense as*

$$L^{\widetilde{W}}_{ts}(\varphi) := \int_s^t S_{t-u}(\varphi\cdot d\widetilde{W}_u) \quad , \quad L^{\widetilde{W}\widetilde{W}}_{ts}(\varphi) := \int_s^t S_{t-u}\left(\varphi\cdot(\delta\widetilde{W})_{us}\cdot d\widetilde{W}_u\right), \tag{3.28}$$

for every smooth function φ. Then both formulas (3.14) and (3.15) remain valid when substituting $\widetilde{W}$ for W, and accordingly the bounds (3.16) and (3.17) hold true for $\widetilde{W}$ as well.

Proof. It suffices to replace the use of Itô's formula in the proof of Lemma 3.9 with standard integration by parts. Indeed, as an absolutely continuous process, $\widetilde{W}$ obeys the rules of standard differential calculus and one has for instance

$$\widetilde{\mathbf{W}}^{\mathbf{2}}_{ts} := \int_s^t (\delta\widetilde{W})_{us}\cdot d\widetilde{W}_u = \frac{1}{2}(\delta\widetilde{W})^2_{ts}.$$

Consequently, it holds that

$$\begin{aligned}
L^{\widetilde{W}\widetilde{W}}_{ts}\varphi &= \int_s^t S_{t-u}(\varphi\cdot d_u(\widetilde{\mathbf{W}}^{\mathbf{2}}_{us})) \\
&= \int_s^t S_{t-u}(\varphi\cdot d_u(\widetilde{\mathbf{W}}^{\mathbf{2}}_{us} - \widetilde{\mathbf{W}}^{\mathbf{2}}_{ts})) \\
&= \frac{1}{2}S_{t-u}(\varphi\cdot(\delta\widetilde{W})^2_{ts}) - \frac{1}{2}\int_s^t \Delta S_{t-u}(\varphi\cdot[(\delta\widetilde{W})^2_{ts} - (\delta\widetilde{W})^2_{us}])du \\
&= \frac{1}{2}S_{t-u}(\varphi\cdot(\delta\widetilde{W})^2_{ts}) - \frac{1}{2}\int_s^t \Delta S_{t-u}(\varphi\cdot\left[(\delta\widetilde{W})^2_{us} + 2(\delta\widetilde{W})_{tu}\cdot(\delta\widetilde{W})_{us}\right])\, du,
\end{aligned}$$

which precisely fits the pattern of (3.15). □

Another consequence of the similarity between (L^W, L^{WW}) and $(L^{\widetilde{W}}, L^{\widetilde{W}\widetilde{W}})$ through the two formulas (3.14) and (3.15) is a set of (readily-checked) Lipschitz-type bounds: with the notations of Proposition 3.10, one has, for some polynomial expression $c_{W,\widetilde{W}}$,

$$\mathcal{N}[L^W - L^{\widetilde{W}}; \mathcal{C}_2^{\frac{1}{2}-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))] \le c_{W,\widetilde{W}}\, \mathcal{N}[W - \widetilde{W}; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})], \tag{3.29}$$

$$\mathcal{N}[L^{WW} - L^{\widetilde{W}\widetilde{W}}; \mathcal{C}_2^{1-\varepsilon}(\mathcal{L}(\mathcal{B},\mathcal{B}))] \le c_{W,\widetilde{W}}\, \mathcal{N}[W - \widetilde{W}; \mathcal{C}^{\frac{1}{2}-\tilde{\varepsilon}}(\mathcal{B}_{\eta,2p})], \tag{3.30}$$

and this bound remains valid for all of the other topologies involved in Proposition 3.10. Then, as far as the solution $\widetilde{Y}$ is concerned, note that

$$(\hat{\delta}\widetilde{Y})_{ts} = L^{\widetilde{W}}_{ts} f(\widetilde{Y}_s) + L^{\widetilde{W}\widetilde{W}}_{ts}(f(\widetilde{Y}_s)\cdot f'(Y_s)) + J^{\widetilde{Y}}_{ts}$$

with $J^{\widetilde{Y}}_{ts} := \int_s^t S_{t-u}\big(\big[\delta f(\widetilde{Y})_{us} - (\delta\widetilde{W})_{us}\cdot f(Y_s)\cdot f'(Y_s)\big]\cdot d\widetilde{W}_u\big)$, and it is obvious in this (absolutely continuous) situation that $J^{\widetilde{Y}} \in \mathcal{C}_2^{\mu}(\mathcal{B})$ for some $\mu > 1$. Therefore, we can easily follow the lines of our previous identification procedure (see the proofs of Lemma 3.12 and Theorem 3.6) in order to exhibit a similar formula for the $\hat{\delta}$-variations of $\widetilde{Y}$:

Lemma 3.16. *Under the hypotheses of Theorem 1.2, assume that $\widetilde{\psi} \in \mathcal{B}_\gamma$. Then the $\hat{\delta}$-variations of the solution $\widetilde{Y}$ to (3.27) can be expanded as*

$$(\hat{\delta}\widetilde{Y})_{ts} = L^{\widetilde{W}}_{ts}(f(\widetilde{Y}_s)) + L^{\widetilde{W}\widetilde{W}}_{ts}(f(\widetilde{Y}_s)\cdot f'(\widetilde{Y}_s)) + \hat{\Lambda}_{ts}\big(R^{\widetilde{Y}}\big), \tag{3.31}$$

where $R^{\widetilde{Y}}_{tus} := -\hat{\delta}\big(L^{\widetilde{W}} f(\widetilde{Y}) + L^{\widetilde{W}\widetilde{W}}\big(f(\widetilde{Y})\cdot f'(\widetilde{Y})\big)\big)_{tus}$. In particular, the bound (3.26) remains valid for $\widetilde{Y}$ when replacing ψ (resp. W) with $\widetilde{\psi}$ (resp. $\widetilde{W}$).

With these identifications in hand, the proof of Theorem 1.2 becomes a matter of a standard rough-paths argument, and we only sketch out the main steps of the procedure (see e.g. the proof of [10, Lemma 5.2] for further details on the computations).

Proof of Theorem 1.2. In order to compare Y with $\widetilde{Y}$, we can now rely on their respective decompositions (3.9) and (3.31). By setting $g := ff'$, we get that

$$\begin{aligned}\hat{\delta}(Y-\widetilde{Y})_{ts} = {} & \big\{\big[L^W_{ts} - L^{\widetilde{W}}_{ts}\big] f(Y_s) + \big[L^{WW}_{ts} - L^{\widetilde{W}\widetilde{W}}_{ts}\big](g(Y_s)\big\} \\ & + \big\{L^{\widetilde{W}}_{ts}\big[f(Y_s) - f(\widetilde{Y}_s)\big] + L^{\widetilde{W}\widetilde{W}}_{ts}\big[g(Y_s) - g(\widetilde{Y}_s)\big]\big\} + \hat{\Lambda}_{ts}\big(R^Y - R^{\widetilde{Y}}\big),\end{aligned} \tag{3.32}$$

with a similar splitting for $R^Y - R^{\tilde{Y}}$ (based on the expansion (3.24)). Now, as in Lemma 3.12, we consider the following appropriate topology:

$$\mathcal{N}[Y-\tilde{Y}; \mathcal{Q}(I)] := \mathcal{N}[Y-\tilde{Y}; \hat{\mathcal{C}}^{2\eta}(I;\mathcal{B}_\infty)] + \mathcal{N}[Y-\tilde{Y}; \mathcal{C}^0(I;\mathcal{B}_\gamma)] + \mathcal{N}[K^Y - K^{\tilde{Y}}; \mathcal{C}_2^{\frac{1}{2}+\eta}(I;\mathcal{B})].$$

By using the decomposition (3.32) and the bounds (3.29)-(3.30), standard differential calculus shows that for any subinterval $I = [\ell_1, \ell_2]$ of $[0,T]$,

$$\begin{aligned}&\mathcal{N}[Y-\tilde{Y}; \mathcal{Q}(I)] \\ &\quad \le C_{W,\tilde{W},\psi,\tilde{\psi}}\left\{\|Y_{\ell_1} - \tilde{Y}_{\ell_1}\|_{\mathcal{B}_\gamma} + \mathcal{N}[W-\tilde{W}; \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})] + |I|^\lambda \mathcal{N}[Y-\tilde{Y}; \mathcal{Q}(I)]\right\},\end{aligned}$$

for some constant $\lambda > 0$. As in Corollary 3.14, we can then rely on an elementary patching argument to reach the global bound (1.7).

□

Remark 3.17. *The above strategy sheds new light on the classical Itô-Stratonovich correction phenomenon arising in the approximation of stochastic heat equations. Indeed, on the one hand, it emphasizes that the convergence of $\widetilde{Y}$ towards Y reduces to the convergence of $(L^{\widetilde{W}}, L^{\widetilde{W}\widetilde{W}})$ towards (L^W, L^{WW}), and on the other, continuous bounds such as (3.30) clearly highlight the relevance of the Stratonovich interpretation of L^{WW} in this context. In a way, the correction phenomenon is therefore more directly observed through the decomposition (3.11) of L^{WW} as the sum of an Itô integral and a trace term.*

4 Approximations in law

We now aim to prove our approximation result, that is Theorem 1.3. Thus, from now on, we assume that the hypotheses in Theorem 1.3 are all satisfied. Recall that the approximation processes involved in this statement, namely the Donsker and the Kac-Stroock approximations, have been specified in the Introduction (see (1.8) and (1.9)), as well as the notations $\mathbf{W}$ and $\beta^{n,\cdot}$. Besides, in this part of the paper we take $T = 1$ for the sake of simplicity.

4.1 Preliminary results

As a first step towards Theorem 1.3, we need to check that the processes we have constructed via $\mathbf{W}$ are indeed well-defined. To do so, we will make use of the following bound.

Lemma 4.1. *Fix $n \geq 1$. Let $X_1^{(n)}, \ldots, X_n^{(n)}$ be independent centered random variables with moments of any order and $f_n : \{1, \ldots, n\} \to \mathbb{R}$. Then for every $r \geq 1$, there exists a constant C_r which only depends on r such that*

$$E\Big[\Big|\sum_{i=1}^n f_n(i) X_i^{(n)}\Big|^{2r}\Big] \leq C_r \Big(\sum_{i=1}^n f_n(i)^2\Big)^r \cdot \Big(\sup_{1\leq i\leq n} E\big[|X_i^{(n)}|^{2r}\big]\Big).$$

This inequality can be easily deduced from the following result, which is clear for $r = 1$ and was proved by Rosenthal for $r > 1$ (see [25, Thm. 3]).

Theorem 4.2. *Let $Y_1, \ldots, Y_n$ be independent centered random variables satisfying $E\big[|Y_i|^{2r}\big] < \infty$, where $r \geq 1$. Then, there exists a constant C_r such that*

$$E\Big[\Big|\sum_{i=1}^n Y_i\Big|^{2r}\Big] \leq C_r \max\Big\{\sum_{i=1}^n E|Y_i|^{2r}, \Big(\sum_{i=1}^n E|Y_i|^2\Big)^r\Big\}.$$

The transition from real-valued to $\mathcal{B}_{\eta,2p}$-valued processes will be ensured by the following result.

Proposition 4.3. *Let $(X_k)_{k\geq 1}$ be a sequence of centered i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, P)$. Assume that each X_i has moments of any order, and consider a sequence $(\lambda_k)_{k\geq 1}$ of positive numbers such that $\sum_{k\geq 1} \lambda_k\, k^{4\eta} < \infty$ for some (fixed) $\eta > 0$. Then, for every $p, q \geq 1$, the random series of functions $\sum_k \sqrt{\lambda_k} X_k e_k$ converges in $L^{2pq}(\Omega, \mathcal{B}_{\eta,2p})$ to an element $\mathbf{X}$ which satisfies*

$$E\big[\|\mathbf{X}\|^{2pq}_{\mathcal{B}_{\eta,2p}}\big] \leq C_{p,q,\lambda,\eta} E\big[|X_1|^{2pq}\big], \tag{4.1}$$

for some constant $C_{p,q,\lambda,\eta}$ which only depends on p, q and $\sum_{k\geq 1} \lambda_k\, k^{4\eta}$.

Proof. Set $\mathbf{X}^n := \sum_{k=1}^n \sqrt{\lambda_k} X_k e_k$ and observe that $\|\mathbf{X}^m - \mathbf{X}^n\|_{\mathcal{B}_{\eta,2p}} = \|\mathbf{X}^{(m,n),\eta}\|_{L^{2p}(0,1)}$, where we have set $\mathbf{X}^{(m,n),\eta}(\xi) := \sum_{k=n+1}^m k^{2\eta}\sqrt{\lambda_k} X_k e_k(\xi)$. Then, by Jensen's inequality,

$$E\big[\|\mathbf{X}^{(m,n),\eta}\|^{2pq}_{L^{2p}(0,1)}\big] = E\bigg[\Big(\int_0^1 d\xi\, |\mathbf{X}^{(m,n),\eta}(\xi)|^{2p}\Big)^q\bigg] \le \int_0^1 d\xi\, E\big[|\mathbf{X}^{(m,n),\eta}(\xi)|^{2pq}\big],$$

and thanks to Lemma 4.1, we get

$$\begin{aligned} E\big[\|\mathbf{X}^{(m,n),\eta}\|^{2pq}_{L^{2p}(0,1)}\big] &\le C_{p,q} E\big[|X_1|^{2pq}\big] \int_0^1 d\xi \Big(\sum_{k=n+1}^m \lambda_k\, k^{4\eta}\, e_k(\xi)^2\Big)^{pq} \\ &\le C_{p,q} E\big[|X_1|^{2pq}\big] \Big(\sum_{k=n+1}^m \lambda_k\, k^{4\eta}\Big)^{pq} \end{aligned} \tag{4.2}$$

due to the uniform bound $\|e_k\|_{\mathcal{B}_\infty} \le \sqrt{2}$. In particular, $E\big[\|\mathbf{X}^m - \mathbf{X}^n\|^{2pq}_{\mathcal{B}_{\eta,2p}}\big]$ tends to zero as both m and n tend to infinity, so that $\mathbf{X}^n$ converges in $L^{2pq}(\Omega, \mathcal{B}_{\eta,2p})$. The bound (4.1) can of course be derived from (4.2). □

In particular, due to Hypothesis 1.1, we can conclude that $W^n = \mathbf{W}(S^{n,\cdot})$ and $W^n = \mathbf{W}(\theta^{n,\cdot})$ are indeed well-defined processes with values in $\mathcal{B}_{\eta,2p}$. Let us now get a little bit closer to the assumptions of Theorem 1.2 by checking that in both cases, W^n admits an absolutely continuous version.

Lemma 4.4. *For any fixed $n \ge 1$, both the Donsker approximation $W^n = \mathbf{W}(S^{n,\cdot})$ and the Kac-Stroock approximation $W^n = \mathbf{W}(\theta^{n,\cdot})$ have an absolutely continuous version with values in $\mathcal{B}_{\eta,2p}$, for all $p \ge 1$.*

Proof. Since the (deterministic) approximation grid for $S^{n,k}$ does not depend on k, it is easily seen that

$$\mathbf{W}(S^{n,\cdot})_t = \mathbf{W}(S^{n,\cdot})_{\frac{i}{n}} + n \cdot \Big(t - \frac{i}{n}\Big) \cdot \big\{\mathbf{W}(S^{n,\cdot})_{\frac{i+1}{n}} - \mathbf{W}(S^{n,\cdot})_{\frac{i}{n}}\big\} \qquad \text{if } t \in \Big[\frac{i}{n}, \frac{i+1}{n}\Big].$$

In particular, $\mathbf{W}(S^{n,\cdot})$ is a piecewise linear process (with values in $\mathcal{B}_{\eta,2p}$) and accordingly it is absolutely continuous.

As far as the Kac-Stroock approximation is concerned, first we can see that it has a continuous version with values in $\mathcal{B}_{\eta,2p}$. Indeed, applying Proposition 4.3,

$$\begin{aligned} E\Big[\big\|\delta(\mathbf{W}(\theta^{n,\cdot}))_{ts}\big\|^{2pq}_{\mathcal{B}_{\eta,2p}}\Big] &\le C\, E\Big[\big|\delta(\theta^n)_{ts}\big|^{2pq}\Big] \\ &= C\, E\Big[\Big|\int_s^t \sqrt{n}\,(-1)^{\zeta+N(nu)} du\Big|^{2pq}\Big] \le C n^{pq} |t-s|^{2pq}. \end{aligned}$$

For the sake of clarity, we will also denote by $\mathbf{W}(\theta^{n,\cdot})$ this continuous version. To prove the existence of an absolutely continuous version, we will see that with probability 1,

$$\mathbf{W}(\theta^{n,\cdot})_t = \int_0^t \mathbf{W}(\dot\theta^{n,\cdot})_s\, ds, \quad \text{for any } t \in [0,1], \tag{4.3}$$

where $\dot\theta^n_t := \sqrt{n}\cdot(-1)^{\zeta+N(nt)}$. Indeed, thanks to Proposition 4.3, $\mathbf{W}(\dot\theta^{n,\cdot})_t$ is well-defined for every $t \in [0,1]$ as an element of $L^{2p}(\Omega, \mathcal{B}_{\eta,2p})$ and

$$E\bigg[\int_0^1 \|\mathbf{W}(\dot\theta^{n,\cdot})_s\|_{\mathcal{B}_{\eta,2p}} ds\bigg] \le \int_0^1 \Big(E\big[\|\mathbf{W}(\dot\theta^{n,\cdot})_s\|^{2p}_{\mathcal{B}_{\eta,2p}}\big]\Big)^{\frac{1}{2p}} ds \le C_p \int_0^1 \Big(E\big[|\dot\theta^n_s|^{2p}\big]\Big)^{\frac{1}{2p}} ds < \infty.$$

As a consequence $\mathbf{W}(\dot\theta^{n,\cdot})$ is (a.s.) Bochner-integrable. Moreover, for each $t\in[0,1]$,

$$\mathbf{W}(\theta^{n,\cdot})_t=\lim_{N\to\infty}\sum_{k=1}^N\sqrt{\lambda_k}\,\theta_t^{n,k}e_k=\lim_{N\to\infty}\int_0^t\Big(\sum_{k=1}^N\sqrt{\lambda_k}\,\dot\theta_s^{n,k}e_k\Big)ds\qquad\text{in } L^{2p}(\Omega,\mathcal{B}_{\eta,2p})$$

and

$$E\Big[\Big\|\int_0^t\Big(\sum_{k=1}^N\sqrt{\lambda_k}\,\dot\theta_s^{n,k}e_k\Big)ds-\int_0^t\mathbf{W}(\dot\theta^{n,\cdot})_s\,ds\Big\|_{\mathcal{B}_{\eta,2p}}^{2p}\Big]\le\int_0^1E\Big[\Big\|\sum_{k=N+1}^\infty\sqrt{\lambda_k}\,\dot\theta_s^{n,k}e_k\Big\|_{\mathcal{B}_{\eta,2p}}^{2p}\Big]ds$$
$$\le C_{pq}\Big(\sup_{s\in[0,1]}E\big[|\dot\theta_s^n|^{2pq}\big]\Big)\Big(\sum_{k=N+1}^\infty\lambda_k\,k^{4\eta}\Big),$$

by similar arguments as in the proof of Proposition 4.3. Since the last expression tends to 0 as $N\to\infty$, we obtain that for each $t\in[0,1]$

$$\mathbf{W}(\theta^{n,\cdot})_t=\int_0^t\mathbf{W}(\dot\theta^{n,\cdot})_s\,ds\quad\text{a.s.}$$

Thus, since $\{\mathbf{W}(\theta^{n,\cdot})_t,\,t\in[0,1]\}$ and $\{\int_0^t\mathbf{W}(\dot\theta^{n,\cdot})_s\,ds,\,t\in[0,1]\}$ are both continuous processes, we can conclude that

$$P\Big\{\mathbf{W}(\theta^{n,\cdot})_t=\int_0^t\mathbf{W}(\dot\theta^{n,\cdot})_s\,ds,\ \forall t\in[0,1]\Big\}=1.$$

□

4.2 A general convergence criterion

One of our key ingredients to prove Theorem 1.3 via Theorem 1.2 lies in the following statement, which puts forward sufficient conditions for an approximation of the noise (defined on the same probability space) to converge with respect to the topology involved in (1.7).

Proposition 4.5. *Let $(\beta^n)_{n\ge1}$ be a sequence of centered processes and β a Brownian motion, all defined on a same probability space $(\Omega,\mathcal{F},P)$, and such that the following two conditions are satisfied:*

(i) For every integer $p\ge1$, there exists a constant C_p such that for all $s,t\in[0,1]$ and all $n\ge1$,

$$E\big[|\beta_t^n-\beta_s^n|^{2p}\big]\le C_p\,|t-s|^p\,.$$

(ii) For every integer $p\ge1$, there exists a constant C_p such that for all $n\ge1$,

$$\sup_{t\in[0,1]}E\big[|\beta_t^n-\beta_t|^{2p}\big]\le C_pn^{-\nu p},$$

for some fixed parameter $\nu>0$.

Then if we consider independent copies $(\beta^{n,k})_{k\ge1}$ (resp. $(\beta^k)_{k\ge1}$) of β^n (resp. β) on a same probability space, we have that, for any integer $p\ge1$ and any $\varepsilon>0$,

$$\mathcal{N}[\mathbf{W}(\beta^{n,\cdot})-\mathbf{W}(\beta^\cdot);\mathcal{C}^{\frac12-\varepsilon}(\mathcal{B}_{\eta,2p})]\underset{n\to\infty}{\longrightarrow}0\qquad\text{a.s.}$$

Let us first see how to combine the above conditions (i) and (ii) so as to exhibit convergent bounds in Hölder topology.

Lemma 4.6. *Under the hypotheses of Proposition 4.5, for all integers $n, p \geq 1$, all $\varepsilon \in (0,1)$ and $s < t \in [0,1]$, one has*

$$E\big[|\delta(\beta^n - \beta)_{ts}|^{2p}\big] \leq C_p \frac{|t-s|^{(1-\varepsilon)p}}{n^{p\varepsilon\nu}},$$

for some constant C_p which only depends on p.

Proof. If $|t-s| \leq n^{-\nu}$, then due to the condition (i), it holds that

$$E\big[|\delta(\beta^n - \beta)_{ts}|^{2p}\big] \leq C_p\big\{E\big[|\beta^n_t - \beta^n_s|^{2p}\big] + E\big[\beta_t - \beta_s|^{2p}\big]\big\} \leq C_p\,|t-s|^p \leq C_p \frac{|t-s|^{(1-\varepsilon)p}}{n^{\nu\varepsilon p}}.$$

On the other hand, if $|t-s| > n^{-\nu}$, one has, thanks to the condition (ii),

$$E\big[|\delta(\beta^n - \beta)_{ts}|^{2p}\big] \leq C_p \sup_{t\in[0,1]} E\big[|\beta^n_t - \beta_t|^{2p}\big] \leq C_p\, n^{-\nu p} \leq C_p \frac{|t-s|^{(1-\varepsilon)p}}{n^{\nu\varepsilon}}.$$

□

Proof of Proposition 4.5. By using successively Proposition 4.3 and Lemma 4.6, we get, for any $q \geq 1$,

$$E\Big[\|\delta\big(\mathbf{W}(\beta^{n,\cdot}) - \mathbf{W}(\beta^{\cdot})\big)_{ts}\|^{2pq}_{\mathcal{B}_{\eta,2p}}\Big] \leq C_{p,q,\eta}\, E\Big[|\delta(\beta^n - \beta)_{ts}|^{2pq}\Big] \leq C_{p,q,\eta} \frac{|t-s|^{(1-\varepsilon)pq}}{n^{pq\nu\varepsilon}}.$$

We are thus in a position to apply the Garsia-Rodemich-Rumsey Lemma 6.1 (with $\delta^* = \delta$) and assert that, for q large enough,

$$\begin{aligned} &E\Big[\mathcal{N}[\mathbf{W}(\beta^{n,\cdot}) - \mathbf{W}(\beta^{\cdot}); \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})]^{2pq}\Big] \\ &\leq \; C_{p,q,\eta} \iint_{[0,T]^2} \frac{E\Big[\|\delta\big(\mathbf{W}(\beta^{n,\cdot}) - \mathbf{W}(\beta^{\cdot})\big)_{ts}\|^{2pq}_{\mathcal{B}_{\eta,2p}}\Big]}{|t-s|^{2pq(\frac{1}{2}-\varepsilon)+2}}\, dsdt \\ &\leq \; C_{p,q,\eta}\, n^{-\varepsilon pq\nu} \iint_{[0,T]^2} |t-s|^{pq\varepsilon-2}\, dsdt \;\leq\; C_{p,q,\eta}\, n^{-\varepsilon pq\nu}. \end{aligned}$$

As a result, it holds that

$$P\big(\mathcal{N}[\mathbf{W}(\beta^{n,\cdot}) - \mathbf{W}(\beta^{\cdot}); \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})] > n^{-\varepsilon\nu/4}\big) \leq C_{p,q,\eta}\, n^{-\varepsilon pq\nu/2},$$

which, thanks to the Borell-Cantelli Lemma, leads us to the conclusion, that is

$$\mathcal{N}[\mathbf{W}(\beta^{n,\cdot}) - \mathbf{W}(\beta^{\cdot}); \mathcal{C}^{\frac{1}{2}-\varepsilon}(\mathcal{B}_{\eta,2p})] \to 0 \qquad \text{a.s.}$$

as n tends to infinity. □

Example: As an immediate illustration of Proposition 4.5, let us consider here the *Wong-Zakai* approximation of a given noise W satisfying Hypothesis 1.1. Precisely, set

$$W^n_t := W_{\frac{i}{n}} + n \cdot \Big(t - \frac{i}{n}\Big) \cdot \big\{W_{\frac{i+1}{n}} - W_{\frac{i}{n}}\big\} \quad \text{for } t \in \Big[\frac{i}{n}, \frac{i+1}{n}\Big],$$

and denote by Y^n the solution of the equation

$$Y^n_t = S_t\psi + \int_0^t S_{t-u}(f(Y^n_u) \cdot dW^n_u),$$

understood in the classical Riemann-Lebesgue sense. Note that W^n can be equivalently described as follows: with the expansion (1.3) of W in mind, i.e. $W = \mathbf{W}(\beta^{\cdot})$, we have that $W^n = \mathbf{W}(\beta^{n,\cdot})$, where, for each $k \geq 1$, $\beta^{n,k}$ stands for the linear interpolation of β^k with mesh $\frac{1}{n}$. Therefore, it suffices to check that the conditions (i) and (ii) in Proposition 4.5 are satisfied by $\beta^n := \beta^{n,1}$, which is a matter of elementary computations (it can be also seen as a particular case of the forthcoming Proposition 4.8).

Together with Theorem 1.2, we retrieve the following almost sure approximation result:

Proposition 4.7. *Under the hypotheses of Theorem 1.2, let Y^n be the Wong-Zakai approximation of (1.4) with mesh $\frac{1}{n}$ and initial condition ψ. Then, as $n \to \infty$, one has $\mathcal{N}[Y - Y^n; \mathcal{C}^0(\mathcal{B}_\gamma)] \to 0$ a.s.*

This almost sure result in a non-linear situation is closely related to those of [5] or [2], where Wong-Zakaï approximations for some parabolic type equations have been considered. We also note that convergence in law for this type of approximations in the framework of stochastic evolution equations has been studied in [29].

Now, let us turn to the proof of the weak approximation results of Theorem 1.3, and which successively involve the Donsker approximation $\beta^n = S^n$ and the Kac-Stroock approximation $\beta^n = \theta^n$. In both cases, we wish to exploit the criterion of Proposition 4.5, which naturally leads us to the following 2-step procedure:

Step 1: Show that Condition (i) is satisfied, i.e., $\sup_n E\big[|\beta^n_t - \beta^n_s|^{2p}\big] \leq C_p\,|t-s|^p$.

Step 2: Find a probability space $(\bar\Omega, \bar{\mathcal{F}}, \bar P)$, a sequence $\bar\beta^n$ and a Brownian motion $\bar\beta$, both defined on $(\bar\Omega, \bar{\mathcal{F}}, \bar P)$, such that $\bar\beta^n \sim \beta^n$ and $\sup_{t\in[0,1]} \bar E\big[|\bar\beta^n_t - \bar\beta_t|^{2p}\big] \leq C_p n^{-\nu p}$ for some fixed parameter $\nu > 0$.

Once these two conditions have been checked, the proof of the weak convergence $Y^n \to Y$ in $\mathcal{C}^0(\mathcal{B}_\gamma)$ becomes a straightforward consequence of Theorem 1.2 and Proposition 4.5, since $\mathbf{W}(\bar\beta^{n,\cdot}) \sim \mathbf{W}(\beta^{n,\cdot})$ and accordingly, if $\bar Y^n$ denotes the solution of (1.6) associated with $\bar W^n := \mathbf{W}(\bar\beta^{n,\cdot})$, it holds that $\bar Y^n \sim Y^n$.

Note that for both approximations S^n and θ^n, the result in **Step 2** will be derived from a Skorokhod embedding argument (see [27]). In the Donsker situation (Proposition 4.9), this relies on a classical strategy towards the celebrated invariance principles (see [22, Section 5.3]). In the Kac-Stroock situation (Proposition 4.11), we will take advantage of an identification result due to Griego, Heath and Ruiz-Moncayo (see [14]).

4.3 Donsker approximation

Here, we proceed to tackle the above 2-step procedure for the Donsker approximation S^n.

Step 1 (Donsker case):

Proposition 4.8. *For every $p \geq 1$, there exists a positive constant C_p such that, for all $0 \leq s < t \leq 1$,*

$$\sup_{n\in\mathbb{N}} E\big[|S^n_t - S^n_s|^{2p}\big] \leq C_p|t-s|^p. \tag{4.4}$$

Proof. First, note that S^n can also be expressed as

$$S^n_t = n^{1/2}\sum_{i=1}^n \Big(\int_0^t \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n}]}(u)\,du\Big)Z_i.$$

Then, by Lemma 4.1, we have

$$E\big[|S_t^n - S_s^n|^{2p}\big] = n^p E\Big[\Big|\sum_{i=1}^n \Big(\int_s^t \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n}]}(u)\,du\Big) Z_i\Big|^{2p}\Big]$$
$$\leq C_p\, n^p E\big[|Z_1|^{2p}\big]\Big(\sum_{i=1}^n \Big(\int_s^t \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n}]}(u)\,du\Big)^2\Big)^p$$
$$\leq C_p\, n^p \Big(\max_{i=1,\dots,n}\Big\{\int_s^t \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n}]}(u)\,du\Big\}\Big)^p \Big(\sum_{i=1}^n \int_s^t \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n}]}(u)\,du\Big)^p \leq C_p|t-s|^p.$$

□

Step 2 (Donsker case):

Proposition 4.9. *Let $(Z_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. centered random variables with unit variance. Then, there exists a probability space $(\bar\Omega, \bar{\mathcal{F}}, \bar P)$, a Brownian motion $\bar\beta$ defined on it and, for each $n \geq 1$, a family of independent random variables $(\bar Z_i^{(n)})_{i=1,\dots,n}$ with the same law as Z_i, such that the following is satisfied. Set*

$$\bar S_t^n := n^{-1/2}\Big\{\sum_{j=1}^{i-1} \bar Z_j^{(n)} + \frac{t-(i-1)/n}{1/n}\,\bar Z_i^{(n)}\Big\} \quad \text{if } t \in \Big[\frac{i-1}{n}, \frac{i}{n}\Big], \quad \text{with } i \in \{1,\dots,n\}.$$

Then, for every integer $p \geq 1$,

$$\sup_{t\in[0,1]} E\big[|\bar\beta(t) - \bar S_t^n|^{2p}\big] \leq C_p n^{-p/4}.$$

Proof. As mentioned earlier, it is based on a general Skorokhod embedding theorem (see [27, p. 163]), which, in our particular situation, can be stated as follows : there exists a probability space $(\bar\Omega, \bar{\mathcal{F}}, \bar P)$, a Brownian motion $\bar\beta$ defined on it and, for each $n \in \mathbb{N}$, a sequence $\{\tau_i^{(n)}\}_{i=1,\dots,n}$ of independent and positive random variables such that the random vector

$$\Big(\bar\beta\big(\tau_1^{(n)}\big), \bar\beta\big(\tau_1^{(n)} + \tau_2^{(n)}\big), \dots, \bar\beta\big(\tau_1^{(n)} + \cdots + \tau_n^{(n)}\big)\Big)$$

has the same law as

$$\Big(\frac{1}{\sqrt n} Z_1, \frac{1}{\sqrt n}\big(Z_1 + Z_2\big), \dots, \frac{1}{\sqrt n}\big(Z_1 + \cdots + Z_n\big)\Big).$$

Moreover, it holds that $E\big[\tau_i^{(n)}\big] = E\big[(Z_1/\sqrt n)^2\big] = \frac{1}{n}$ and

$$E\big[|\tau_i^{(n)}|^m\big] \leq C_m E\Big[\Big(\frac{1}{\sqrt n} Z_i\Big)^{2m}\Big] \leq \frac{C_m}{n^m}, \quad \text{for any } m \in \mathbb{N}.$$

Set $T_0^n := 0$ and $T_i^{(n)} := \sum_{j=1}^i \tau_j^{(n)}$ for $i \geq 1$. With this notation, we can infer that

$$\bar\beta\big(T_i^{(n)}\big) - \bar\beta\big(T_{i-1}^{(n)}\big) \sim \frac{1}{\sqrt n} Z_i.$$

We define now

$$\bar Z_i^{(n)} = \sqrt n\big\{\bar\beta\big(T_i^{(n)}\big) - \bar\beta\big(T_{i-1}^{(n)}\big)\big\} \sim Z_i$$

and

$$\bar S_t^n = n^{-1/2}\Big\{\sum_{j=1}^{i-1} \bar Z_j^{(n)} + \frac{t-(i-1)/n}{1/n}\,\bar Z_i^{(n)}\Big\} \quad \text{if } t \in \Big[\frac{i-1}{n}, \frac{i}{n}\Big].$$

Observe that, if $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$,

$$\bar{S}_t^n = \bar{\beta}\big(T_{i-1}^{(n)}\big) + \frac{t-(i-1)/n}{1/n}\Big\{\bar{\beta}\big(T_i^{(n)}\big) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\Big\},$$

and hence, for $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$, we have that

$$\begin{aligned} E\Big[\big|\bar{\beta}(t) - \bar{S}_t^n\big|^{2p}\Big] &\leq C_p E\Big[\big|\bar{\beta}(t) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big|^{2p}\Big] + C_p E\Big[\big|\frac{1}{\sqrt{n}} Z_i\big|^{2p}\Big] \\ &\leq C_p E\Big[\big|\bar{\beta}(t) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big|^{2p}\Big] + C_p n^{-p}. \end{aligned}$$

Thus, we only need to bound the first term in the latter expression, and to this end, we will use the following decomposition:

$$E\Big[\big|\bar{\beta}(t) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big|^{2p}\Big] = A_1^n + A_2^n,$$

with

$$A_1^n = E\Big[\big|\bar{\beta}(t) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big|^{2p} \mathbf{1}_{\{|t - T_{i-1}^{(n)}| \leq n^{-1/4}\}}\Big]$$

and

$$A_2^n = E\Big[\big|\bar{\beta}(t) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big|^{2p} \mathbf{1}_{\{|t - T_{i-1}^{(n)}| > n^{-1/4}\}}\Big].$$

On the one hand, the maximal inequality for Brownian motion yields

$$\begin{aligned} A_1^n &\leq E\Big[\max_{s \in [(t-n^{-1/4}) \vee 0,\, t]} |\bar{\beta}(s) - \bar{\beta}(t)|^{2p}\Big] + E\Big[\max_{s \in [t,\, (t+n^{-1/4}) \wedge 1]} |\bar{\beta}(s) - \bar{\beta}(t)|^{2p}\Big] \\ &\leq 2\, E\Big[\max_{h \in [0, n^{-1/4}]} |\bar{\beta}(h)|^{2p}\Big] \leq C_p E\big[|\bar{\beta}(n^{-1/4})|^{2p}\big] \leq C_p n^{-p/4}. \end{aligned} \tag{4.5}$$

On the other hand, by Cauchy-Schwarz inequality, we have

$$A_2^n \leq C_p \Big\{E\big[\bar{\beta}\big(T_{i-1}^{(n)}\big)^{4p}\big] + E\big[\bar{\beta}\big(t\big)^{4p}\big]\Big\}^{1/2} \Big\{P\big(\big|t - T_{i-1}^{(n)}\big| > n^{-1/4}\big)\Big\}^{1/2} \tag{4.6}$$

Note that, by Lemma 4.1,

$$E\big[\bar{\beta}\big(T_{i-1}^{(n)}\big)^{4p}\big] = E\Big[\Big(\frac{1}{\sqrt{n}}\big(Z_1 + \ldots + Z_i\big)\Big)^{4p}\Big] \leq C_p n^{-2p} i^{2p} \leq C_p.$$

Thus, in order to estimate the term A_2^n, we only need to study the probability appearing in (4.6). To do so, observe first that since $t \in [\frac{i-1}{n}, \frac{i}{n}]$, we have, for n such that $\frac{1}{2}\, n^{-1/4} > \frac{1}{n}$ (that is, for $n \geq 3$),

$$P\big(\big|t - T_{i-1}^{(n)}\big| > n^{-1/4}\big) \leq P\Big(\big|T_{i-1}^{(n)} - \frac{i-1}{n}\big| > \frac{1}{2} n^{-1/4}\Big). \tag{4.7}$$

Then, using again Lemma 4.1, we get

$$\begin{aligned} P\Big(\big|T_{i-1}^{(n)} - \frac{i-1}{n}\big| > \frac{1}{2} n^{-1/4}\Big) &\leq C_p\, n^{p/2} E\Big[\Big(\sum_{j=1}^{i-1} \Big\{\tau_j^{(n)} - \frac{1}{n}\Big\}\Big)^{2p}\Big] \\ &\leq C_p\, n^{p/2} i^p n^{-2p} \;\leq\; C_p n^{-p/2}. \end{aligned} \tag{4.8}$$

Therefore, $A_2^n \leq C_p\, n^{-p/4}$, which concludes the proof. □

4.4 Kac-Stroock approximation

Along the same lines as in the Donsker case, we proceed now to analyze the Kac-Stroock approximations based on θ^n.

Step 1 (Kac-Stroock case):

Proposition 4.10. *For every integer $p \geq 1$, there exists a positive constant C_p such that, for all $0 \leq s < t \leq 1$,*

$$\sup_{n\in\mathbb{N}} E\big[|\theta^n_t - \theta^n_s|^{2p}\big] \leq C_p|t-s|^p. \tag{4.9}$$

Proof. We have that

$$\begin{aligned}
E\big[|\theta^n_t - \theta^n_s|^{2p}\big] &= E\bigg[\Big(\sqrt{n}\int_s^t (-1)^{\zeta+N(nu)}du\Big)^{2p}\bigg] = n^p E\bigg[\Big(\int_s^t (-1)^{N(nu)}du\Big)^{2p}\bigg] \\
&= C_p n^p E\Big[\int_s^t\cdots\int_s^t (-1)^{N(nu_1)+N(nu_2)+\cdots+N(nu_{2p})}du_1\cdots du_{2p}\Big] \\
&= C_p n^p E\Big[\int_s^t\cdots\int_s^t \mathbf{1}_{\{u_1<u_2<\cdots<u_{2p}\}}(-1)^{N(nu_1)+N(nu_2)+\cdots+N(nu_{2p})}du_1\cdots du_{2p}\Big],
\end{aligned}$$

where in the latter equality we have used the symmetry of the integrand. Taking into account that the two possible values of random variable $(-1)^{N(nu_1)+\cdots+N(nu_{2p})}$ only depend on the fact that the exponent is even or odd, we can write the latter expression above as

$$C_p n^p\Big[\int_s^t\cdots\int_s^t \mathbf{1}_{\{u_1<u_2<\cdots<u_{2p}\}}E\Big((-1)^{\sum_{i=1}^p N(nu_{2i})-N(nu_{2i-1})}\Big)du_1\cdots du_{2p}\Big]. \tag{4.10}$$

Using that for $u_1 < u_2 < \cdots < u_{2p}$, the random variables $N(nu_{2i}) - N(nu_{2i-1})$ are independent with Poisson distribution of parameter $n(u_{2i} - u_{2i-1})$, we have that (4.10) is equal to

$$C_p n^p\Big[\int_s^t\cdots\int_s^t \mathbf{1}_{\{u_1<u_2<\cdots<u_{2p}\}}e^{-2n\left[\sum_{i=1}^p(u_{2i}-u_{2i-1})\right]}du_1\cdots du_{2p}\Big].$$

This term can be bounded by

$$\begin{aligned}
&C_p n^p\Big[\int_s^t\cdots\int_s^t \mathbf{1}_{\{u_1<u_2\}}\cdots\mathbf{1}_{\{u_{2q-1}<u_{2p}\}}e^{-2n\left[\sum_{i=1}^p(u_{2i}-u_{2i-1})\right]}du_1\cdots du_{2p}\Big]. \\
&\quad= C_p n^p\Big(\int_s^t\int_s^{u_2} e^{-2n(u_2-u_1)}du_1\,du_2\Big)^p = C_p n^p\Big(\int_s^t \frac{1}{2n}(1-e^{-2n(u_2-s)})du_2\Big)^p \\
&\quad\leq C_p(t-s)^p.
\end{aligned}$$

This concludes the proof. □

Step 2 (Kac-Stroock case):

Proposition 4.11. *There exists a probability space $(\bar\Omega, \bar{\mathcal{F}}, \bar P)$, a Brownian motion $\bar\beta$ defined on it and, for each $n \in \mathbb{N}$, a process $\bar\theta^n$ with the same law as θ^n in (1.9) such that, for any $\nu \in (0, \frac{1}{4})$ and any $p \in \mathbb{N}$,*

$$\sup_{t\in[0,1]} E\Big[|\bar\beta(t) - \bar\theta^n(t)|^{2p}\Big] \leq C_{p,\nu}\, n^{-p\nu}, \tag{4.11}$$

for some constant $C_{p,\nu}$.

Proof. First of all, it is clear that we can suppose $p(1/4-\nu) \geq 1$ (otherwise, we can use Jensen's inequality). Then, following the lines of [14, Section 2], we consider a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ with the following mutually independent objects defined on it:

(i) a Brownian motion $\bar{\beta}$,

(ii) for each $n \in \mathbb{N}$, a sequence of independent random variables $\{\xi_i^{(n)},\, i \in \mathbb{N}\}$ such that $\xi_i^{(n)}$ has an exponential distribution with parameter $2\sqrt{n}$,

(iii) a sequence $\{k_i,\, i \in \mathbb{N}\}$ of independent random variables such that $P\{k_i = 1\} = P\{k_i = -1\} = 1/2$.

The i.i.d. random variables $k_1\xi_1^{(n)}$, $k_2\xi_2^{(n)}, \ldots$, verify $E\big[k_i\xi_i^{(n)}\big] = 0$ and $E\big[\big(k_i\xi_i^{(n)}\big)^2\big] = \frac{1}{2n}$. Therefore, with the same result of Skorokhod as the one quoted in the proof of Lemma 4.9 (see [27, p. 163]), there exists a sequence of independent positive random variables $\tau_1^{(n)}$, $\tau_2^{(n)} \ldots$, such that $\bar{\beta}(\tau_1^{(n)})$, $\bar{\beta}(\tau_1^{(n)} + \tau_2^{(n)}) \ldots$ have the same law as $k_1\xi_1^{(n)}$, $k_1\xi_1^{(n)} + k_2\xi_2^{(n)} \ldots$, respectively. Moreover, it holds that

$$E\big[\tau_i^{(n)}\big] = E\big[\big(k_i\xi_i^{(n)}\big)^2\big] = \frac{1}{2n},$$

and, for each $m \in \mathbb{N}$,

$$E\big[|\tau_i^{(n)}|^m\big] \leq C_m E\big[|k_i\xi_i^{(n)}|^{2m}\big] \leq \frac{C_m}{n^m}.$$

Set $T_i^{(n)} := \sum_{j=1}^{i} \tau_j^{(n)}$ and define

$$\widetilde{\tau}_i^{(n)} := n^{-1/2}\big|\bar{\beta}\big(T_i^{(n)}\big) - \bar{\beta}\big(T_{i-1}^{(n)}\big)\big| \quad , \quad \widetilde{T}_i^{(n)} := \sum_{j=1}^{i} \widetilde{\tau}_j^{(n)}.$$

Then, let $\bar{\theta}^n = \{\bar{\theta}^n(t),\, t \geq 0\}$ be a piecewise linear process given on the grid $\widetilde{T}_1^{(n)}, \widetilde{T}_2^{(n)}, \ldots$ by

$$\bar{\theta}^n(\widetilde{T}_i^{(n)}) := \bar{\beta}(T_i^{(n)}) \sim \sum_{j=1}^{i} k_j\xi_j^{(n)},$$

and $\bar{\theta}^n(0) = 0$. The $\widetilde{\tau}_i^{(n)}$'s are independent random variables exponentially distributed with parameter $2n$, and it is proved in [14] that the process $\bar{\theta}^n$ thus defined has the same law as θ^n.

Now, to show (4.11), we decompose the term $E\Big[|\bar{\beta}(t) - \bar{\theta}^n(t)|^{2p}\Big]$ as the sum of the following two terms:

$$E_1^n := E\Big[|\bar{\beta}(t) - \bar{\theta}^n(t)|^{2p}\, \mathbf{1}_{\{t \in [0, \widetilde{T}_{8n}^{(n)}]\}}\Big]$$

and

$$E_2^n := E\Big[|\bar{\beta}(t) - \bar{\theta}^n(t)|^{2p}\, \mathbf{1}_{\{t > \widetilde{T}_{8n}^{(n)}\}}\Big].$$

Let us first study E_1^n. If t belongs to $A_\ell^n := \big[\widetilde{T}_{\ell-1}^{(n)}, \widetilde{T}_\ell^{(n)}\big)$ for some $\ell = 1, \ldots, 8n$, we have that

$$\bar{\theta}^n(t) - \bar{\beta}(t) = \bar{\beta}\big(T_{\ell-1}^{(n)}\big) - \bar{\beta}(t) + \frac{t - \widetilde{T}_{\ell-1}^{(n)}}{\widetilde{\tau}_\ell^{(n)}}\Big\{\bar{\beta}\big(T_\ell^{(n)}\big) - \bar{\beta}\big(T_{\ell-1}^{(n)}\big)\Big\}.$$

So

$$\begin{aligned} E_1^n &= \sum_{\ell=1}^{8n} E\big[|\bar\beta(t)-\bar\theta^n(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\big] \\ &\le C_p\sum_{\ell=1}^{8n} E\big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\big] + C_p\sum_{\ell=1}^{8n} E\big[|\bar\beta(T_{\ell}^{(n)})-\bar\beta(T_{\ell-1}^{(n)})|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\big] \\ &\le C_p\sum_{\ell=1}^{8n} E\big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\big] + C_p\, n^{1-p}, \end{aligned} \tag{4.12}$$

where, for the last inequality, we have used the fact that $\bar\beta(T_\ell^{(n)})-\bar\beta(T_{\ell-1}^{(n)}) \sim k_1\xi_1$. Now, for any fixed $\ell\in\{1,\dots,8n\}$, write

$$\begin{aligned} E\big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\big] &= E\Big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\mathbf{1}_{\{|t-T_{\ell-1}^{(n)}|\le n^{-1/4}\}}\Big] \\ &\quad + E\Big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\mathbf{1}_{\{|t-T_{\ell-1}^{(n)}|> n^{-1/4}\}}\Big]. \end{aligned} \tag{4.13}$$

The first term in (4.13) can be bounded with the same argument as in the proof of Proposition 4.9 (see (4.5)), which gives

$$E\Big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\mathbf{1}_{\{|t-T_{\ell-1}^{(n)}|\le n^{-1/4}\}}\Big] \le C_p\, n^{-p/4}.$$

As far as the second term in (4.13) is concerned, we have

$$\begin{aligned} &E\Big[|\bar\beta(T_{\ell-1}^{(n)})-\bar\beta(t)|^{2p}\mathbf{1}_{\{t\in A_\ell^n\}}\mathbf{1}_{\{|t-T_{\ell-1}^{(n)}|> n^{-1/4}\}}\Big] \\ &\qquad \le \Big\{E\big[|\bar\beta(t)-\bar\beta(T_{\ell-1}^{(n)})|^{4p}\big]\Big\}^{1/2}\Big\{\bar P\big(t\in A_\ell^n, |t-T_{\ell-1}^{(n)}|>n^{-1/4}\big)\Big\}^{1/2}, \end{aligned} \tag{4.14}$$

and since

$$E\Big[\big|\bar\beta\big(T_{\ell-1}^{(n)}\big)\big|^{4p}\Big] = E\Big[\big|\sum_{j=1}^{\ell-1}k_j\xi_j^{(n)}\big|^{4p}\Big] \le C_p n^{2p} E\Big[\big|k_1\xi_1^{(n)}\big|^{4p}\Big] \le C_p,$$

we only have to focus on the probability appearing in (4.14). To do so, let us notice that

$$\begin{aligned} &\bar P\big(t\in A_\ell^n, |t-T_{\ell-1}^{(n)}|>n^{-1/4}\big) \\ &\quad\le \bar P\big(|t-\widetilde T_{\ell-1}^{(n)}|\le\widetilde\tau_\ell^n, |t-T_{\ell-1}^{(n)}|>n^{-1/4}\big) \\ &\quad\le \bar P\big(|t-\widetilde T_{\ell-1}^{(n)}|\le\widetilde\tau_\ell^n, \widetilde\tau_\ell^{(n)}\le\frac{1}{2}n^{-1/4}, |t-T_{\ell-1}^{(n)}|>n^{-1/4}\big) + \bar P\big(\widetilde\tau_\ell^{(n)}>\frac{1}{2}n^{-1/4}\big) \\ &\quad\le \bar P\big(|t-\widetilde T_{\ell-1}^{(n)}|\le\frac{1}{2}n^{-1/4}, |t-T_{\ell-1}^{(n)}|>n^{-1/4}\big) + C_p\, n^{-3p/2} \\ &\quad\le \bar P\big(|T_{\ell-1}^{(n)}-\widetilde T_{\ell-1}^{(n)}|>\frac{1}{2}n^{-1/4}\big) + C_p\, n^{-3p/2}. \end{aligned}$$

Then

$$\bar P\big(|T_{\ell-1}^{(n)}-\widetilde T_{\ell-1}^{(n)}|>\frac{1}{2}n^{-1/4}\big) \le C_p\, n^{p/2}\Big\{E\Big[\big|T_{\ell-1}^{(n)}-\frac{l-1}{2n}\big|^{2p}\Big]+E\Big[\big|\widetilde T_{\ell-1}^{(n)}-\frac{l-1}{2n}\big|^{2p}\Big]\Big\} \le C_p\, n^{-p/2},$$

where we have used the same argument as in (4.8) to get the last bound. Going back to (4.12), we deduce that $E_1^n \le C_p\, n\, n^{-p/4} \le C_p\, n^{-\nu p}$, since p is assumed to satisfy $p(\frac{1}{4}-\nu)\ge 1$.

Eventually, we must deal with E_2^n. In fact, we have that

$$E_2^n \le \Big\{E\Big[\big|\bar\theta^n(t)-\bar\beta(t)\big|^{4p}\Big]\Big\}^{1/2}\Big\{P\big(t>\widetilde T_{8n}^{(n)}\big)\Big\}^{1/2} \le C_p\Big\{P\Big(t>\sum_{j=1}^{8n}\widetilde\tau_j^{(n)}\Big)\Big\}^{1/2},$$

where we have used Lemma 4.10. If we denote by N_n a Poisson process with intensity $2n$, we can write

$$P\Big(t>\sum_{j=1}^{8n}\widetilde\tau_j^{(n)}\Big) \le P\Big(1>\sum_{j=1}^{8n}\widetilde\tau_j^{(n)}\Big) \le P\big(N_n(1)\ge 8n\big),$$

because the $\widetilde\tau_j^{(n)}$'s are independent random variables exponentially distributed with parameter $2n$. The latter probability can be bounded by using Stirling's inequality, as follows:

$$\begin{aligned} P\big(N_n(1)\ge 8n\big) &= \sum_{k=8n}^{\infty} e^{-2n}\frac{(2n)^k}{k!} \le C\,e^{-2n}\sum_{k=8n}^{\infty}\frac{(2n)^k}{\sqrt{2\pi k}\Big(\frac{k}{e}\Big)^k} \\ &= C\,e^{-2n}\sum_{k=8n}^{\infty}\Big(\frac{2en}{k}\Big)^k\frac{1}{\sqrt{2\pi k}} \le C\,e^{-2n}\sum_{k=8n}^{\infty}\Big(\frac{e}{4}\Big)^k \le Ce^{-2n}. \end{aligned}$$

This lets us conclude the proof. □

5 Appendix A: fractional Sobolev spaces

We gather here some classical properties of the fractional Sobolev spaces $(\mathcal{B}_{\alpha,p})_{\alpha\in\mathbb{R},p\in\mathbb{N}}$, which are extensively used throughout the paper. We recall the notations $\mathcal{B}_\alpha$ for $\mathcal{B}_{\alpha,2}$ and $\mathcal{B}$ for $\mathcal{B}_0$. Let us first label the following well-known regularizing properties of the semigroup (see [23]).

Proposition 5.1. *Fix two parameters $\lambda<\alpha\in\mathbb{R}$. Then, for every $\varphi\in\mathcal{B}_\lambda$ and $t>0$,*

$$\|S_t\varphi\|_{\mathcal{B}_\alpha}\le c\,t^{-(\alpha-\lambda)}\|\varphi\|_{\mathcal{B}_\lambda}\quad,\quad \|\Delta S_t\varphi\|_{\mathcal{B}_\alpha}\le c\,t^{-1-(\alpha-\lambda)}\|\varphi\|_{\mathcal{B}_\lambda}. \tag{5.1}$$

and for every $\psi\in\mathcal{B}_\alpha$,

$$\|S_t\psi-\psi\|_{\mathcal{B}_\lambda}\le c\,t^{\alpha-\lambda}\|\psi\|_{\mathcal{B}_\alpha}\quad,\quad \|\Delta S_t\psi\|_{\mathcal{B}_\lambda}\le c\,t^{-1+(\alpha-\lambda)}\|\psi\|_{\mathcal{B}_\alpha}. \tag{5.2}$$

The next results are taken from the exhaustive book [26] on fractional Sobolev spaces. With the notations of the latter reference, our space $\mathcal{B}_{\alpha,p}$ ($\alpha\in\mathbb{R},p\in\mathbb{N}$) corresponds to $F^{2\alpha}_{p,2}$. Let us first report some properties regarding pointwise multiplication of functions. Due to the multiplicative perturbation in (1.4), it is indeed natural that these results should intervene at some point. In the statement, the notation $E\cdot F\subset G$ must be understood as $\|\varphi\cdot\psi\|_G\le c\,\|\varphi\|_E\|\psi\|_F$ for every $\varphi\in E,\psi\in F$.

Proposition 5.2. *The following properties hold true:*

1. *([26, Section 2.4.4]) One has*

$$L^r(0,1)\subset\mathcal{B}_{-\alpha}\quad\textit{if}\quad\alpha\ge\frac12-\frac14, \tag{5.3}$$

and in particular:

$$L^p(0,1)\cdot\mathcal{B}\subset\mathcal{B}_{-\alpha}\quad\textit{if}\quad\alpha\ge\frac{1}{2p}.$$

2. *([26, Section 4.6.1]) Let $\alpha_1 < \alpha_2$ be such that $\alpha_1 + \alpha_2 > 0$ and $\alpha_2 > \frac{1}{4}$. Then*
$$\mathcal{B}_{\alpha_1} \cdot \mathcal{B}_{\alpha_2} \subset \mathcal{B}_{\alpha_1}. \tag{5.4}$$
In particular, $\mathcal{B}_\alpha$ is an algebra as soon as $\alpha > \frac{1}{4}$.
3. *([26, Section 4.8.2]) Let $\alpha \geq 0$ and $p_1, p_2, p \geq 2$ be such that $2\alpha < \frac{1}{p_i}$ ($i \in \{1,2\}$) and $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$. Then*
$$\mathcal{B}_{\alpha,p_1} \cdot \mathcal{B}_{\alpha,p_2} \subset \mathcal{B}_{\alpha,p}. \tag{5.5}$$

Let us also label here the classical Sobolev embedding

$$\mathcal{B}_{\alpha,p} \subset \mathcal{B}_\infty \quad \text{if} \quad 2\alpha > \frac{1}{p}, \tag{5.6}$$

which yields in particular:

$$\mathcal{B}_\alpha \subset \mathcal{B}_\infty \quad \text{as soon as} \quad \alpha > \frac{1}{4}. \tag{5.7}$$

Finally, in order to handle the non-linearity in (1.4), we resort at some point to the following stability result for composition of functions (see [26, Section 5.3.6]): if $f : \mathbb{R} \to \mathbb{R}$ is differentiable with bounded derivative, then for every $\alpha \in [0, \frac{1}{2}]$ and $\varphi \in \mathcal{B}_\alpha$,

$$\|f(\varphi)\|_{\mathcal{B}_\alpha} \leq c_f \left\{1 + \|\varphi\|_{\mathcal{B}_\alpha}\right\}. \tag{5.8}$$

Here, f is also understood as its associated Nemytskii operator, i.e., $f(\varphi)(\xi) := f(\varphi(\xi))$.

6 Appendix B: A priori estimates on the solution

It only remains to prove the two a priori controls (3.19) and (3.20) for the solution Y of (1.4) (or equivalently (3.1)). To do so, we will rely on the following result, taken from [11, Lemma 6.5], and which extends the classical Garsia-Rodemich-Rumsey in two directions: 1) it covers the case of $\hat{\delta}$-variations and 2) it applies to more general processes defined on the 2-dimensional simplex $\mathcal{S}_2 = \{(t,s) \in [0,T]^2 : t \geq s\}$.

Lemma 6.1. *Let $\delta^* = \delta$ or $\hat{\delta}$. For every $\alpha, \beta \geq 0$ and $p, q \geq 1$, there exists a constant c such that, for any $R : \mathcal{S}_2 \to \mathcal{B}_{\alpha,p}$,*

$$\mathcal{N}[R; \mathcal{C}_2^\beta([0,T]; \mathcal{B}_{\alpha,p})] \leq c \left\{ U_{\beta+\frac{2}{q},q,\alpha,p}(R) + \mathcal{N}[\delta^* R; \mathcal{C}_3^\beta([0,T]; \mathcal{B}_{\alpha,p}] \right\},$$

where

$$U_{\beta,q,\alpha,p}(R) = \left[\int_{0 \leq u < v \leq T} \left(\frac{\|R_{vu}\|_{\mathcal{B}_{\alpha,p}}}{|v-u|^\beta} \right)^q du dv \right]^{1/q}.$$

Proof of Lemma 3.11. In both cases, we will resort to the previous Lemma, which essentially reduces the problem to moment estimates. Thus, the following Burkholder-Davis-Gundy type inequality (borrowed from [9, Lemma 7.7]) naturally comes into play: for every $\alpha \geq 0$, one has, by setting $U_0 := Q^{1/2}(\mathcal{B})$,

$$E\Big[\Big\| \int_s^t S_{t-u}(f(Y_u) \cdot dW_u) \Big\|_{\mathcal{B}_\alpha}^{2q} \Big] \leq c_q \Big(\int_s^t E\Big[\|S_{t-u}(f(Y_u) \cdot *)\|_{HS(U_0,\mathcal{B}_\alpha)}^{2q} \Big]^{\frac{1}{q}} du \Big)^q, \tag{6.1}$$

where the notation $HS(U_0, \mathcal{B}_\alpha)$ refers to the space of Hilbert-Schmidt operators defined on U_0 and taking values in $\mathcal{B}_\alpha$. Note also that the family $(\lambda_k e_k)$ defines an orthonormal basis of U_0 and accordingly

$$\|S_{t-u}(f(Y_u) \cdot *)\|_{HS(U_0,\mathcal{B}_\alpha)} = \Big(\sum_k \lambda_k \|S_{t-u}(f(Y_u) \cdot e_k)\|_{\mathcal{B}_\alpha}^2 \Big)^{1/2}. \tag{6.2}$$

Now, to show that $Y \in \hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty)$, observe first that for every $q \geq 1$ and any small $\varepsilon > 0$,

$$\begin{aligned} &E\big[\|(\hat{\delta}Y)_{ts}\|^{2q}_{\mathcal{B}_\infty}\big] \leq c_q\, E\Big[\|(\hat{\delta}Y)_{ts}\|^{2q}_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\Big] \\ &\leq c_q\Big\{E\Big[\Big\|\int_s^t S_{t-u}(f(Y_u)\cdot dW_u)\Big\|^{2q}_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\Big] + E\Big[\Big(\int_s^t \|S_{t-u}(P\cdot f(Y_u)\cdot f'(Y_u))\|_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}du\Big)^{2q}\Big]\Big\}. \end{aligned} \tag{6.3}$$

The second summand in (6.3) is trivially bounded by $c_p\,|t-s|^{2q(\frac{3}{4}-\varepsilon)}$ since

$$\|S_{t-u}(P\cdot f(Y_u)\cdot f'(Y_u))\|_{\mathcal{B}_{\frac{1}{4}+\varepsilon}} \leq c\,|t-u|^{-\frac{1}{4}-\varepsilon}\,\|P\cdot f(Y_u)\cdot f'(Y_u)\|_{\mathcal{B}} \leq c\,|t-u|^{-\frac{1}{4}-\varepsilon}.$$

As far as the first summand in (6.3) is concerned, observe that

$$\sum_k \lambda_k \|S_{t-u}(f(Y_u)\cdot e_k)\|^2_{\mathcal{B}_{\frac{1}{4}+\varepsilon}} \leq c\sum_k \lambda_k\,|t-u|^{-\frac{1}{2}-2\varepsilon}\,\|f(Y_u)\cdot e_k\|_{\mathcal{B}} \leq c\,|t-u|^{-\frac{1}{2}-2\varepsilon},$$

which, owing to (6.1) and (6.2), entails that

$$E\Big[\Big\|\int_s^t S_{t-u}(f(Y_u)\cdot dW_u)\Big\|^{2q}_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\Big] \leq c_q\,|t-s|^{(\frac{1}{2}-2\varepsilon)q}.$$

Going back to (6.3), we are in a position to apply Lemma 6.1 and assert that $Y \in \hat{\mathcal{C}}^{\frac{1}{4}-}(\mathcal{B}_\infty) \subset \hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty)$ (we recall that η is assumed to belong to $(0,\frac{1}{8})$). Note that since $\psi \in \mathcal{B}_\gamma$, these computations also prove that $\sup_{t\in[0,T]} E\big[\|Y_t\|^{2q}_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\big] < \infty$ for $\varepsilon > 0$ small enough, which will be used in the sequel.

In order to show that $Y \in \mathcal{C}^0(\mathcal{B}_\gamma)$, let us write, like in (6.3),

$$\begin{aligned} &E\Big[\|(\hat{\delta}Y)_{ts}\|^{2q}_{\mathcal{B}_\gamma}\Big] \\ &\leq\ c_q\Big\{E\Big[\Big\|\int_s^t S_{t-u}(f(Y_u)\cdot dW_u)\Big\|^{2q}_{\mathcal{B}_\gamma}\Big] + E\Big[\Big(\int_s^t \|S_{t-u}(P\cdot f(Y_u)\cdot f'(Y_u))\|_{\mathcal{B}_\gamma}du\Big)^{2q}\Big]\Big\} \\ &\leq\ c_q\Big\{E\Big[\Big\|\int_s^t S_{t-u}(f(Y_u)\cdot dW_u)\Big\|^{2q}_{\mathcal{B}_\gamma}\Big] + \Big(\int_s^t |t-u|^{-\gamma}\,du\Big)^{2q}\Big\}. \end{aligned} \tag{6.4}$$

Then one has successively

$$\begin{aligned} E\Big[\|S_{t-u}(f(Y_u)\cdot *)\|^{2q}_{HS(U_0,\mathcal{B}_\gamma)}\Big] \ &=\ E\Big[\Big(\sum_k \lambda_k\|S_{t-u}(f(Y_u)\cdot e_k)\|^2_{\mathcal{B}_\gamma}\Big)^q\Big] \\ &\leq\ c_q\,|t-u|^{-2q(\gamma-\eta)}\,E\Big[\Big(\sum_k \lambda_k\|f(Y_u)\cdot e_k\|^2_{\mathcal{B}_\eta}\Big)^q\Big] \\ &\leq\ c_q\,|t-u|^{-2q(\gamma-\eta)}\,E\Big[\Big(\sum_k \lambda_k\|f(Y_u)\|^2_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\|e_k\|^2_{\mathcal{B}_\eta}\Big)^q\Big] \qquad \text{(use (5.4))} \\ &\leq\ c_q\,|t-u|^{-2q(\gamma-\eta)}\,\Big(\sum_k(\lambda_k\cdot k^{4\eta})\Big)^q\big\{1+\sup_{t\in[0,T]}E\big[\|Y_t\|^{2q}_{\mathcal{B}_{\frac{1}{4}+\varepsilon}}\big]\big\} \qquad \text{(use (5.8))}, \end{aligned}$$

with $2(\gamma-\eta)<1$ and $\sum_k(\lambda_k\cdot k^{4\eta})<\infty$. Thanks to (6.1) and (6.2), we can go back to (6.4) and deduce that $E\big[\|(\hat{\delta}Y)_{ts}\|^{2q}_{\mathcal{B}_\gamma}\big] \leq c_q\,|t-s|^{2q\varepsilon}$ for some small $\varepsilon>0$. Since $\psi\in\mathcal{B}_\gamma$, this proves in particular that $Y\in\mathcal{C}^0(\mathcal{B}_\gamma)$.

Let us now turn to K^Y and notice first that $\hat{\delta}K^Y = L^W\delta f(Y)$, so

$$\begin{aligned} \|(\hat{\delta}K^Y)_{tus}\|_{\mathcal{B}} \ &\leq\ \|L^W_{tu}\|_{\mathcal{L}(\mathcal{B},\mathcal{B})}\|\delta(f(Y))_{us}\|_{\mathcal{B}} \\ &\leq\ c_{W,f}\,|t-u|^{\frac{1}{2}-\eta}\,\|(\delta Y)_{us}\|_{\mathcal{B}} \qquad \text{(use (3.16))} \\ &\leq\ c_{W,f}\,|t-u|^{\frac{1}{2}-\eta}\Big\{\|(\hat{\delta}Y)_{us}\|_{\mathcal{B}} + \|a_{us}Y_s\|_{\mathcal{B}}\Big\} \\ &\leq\ c_{W,f}\Big\{|t-s|^{\frac{1}{2}+\eta}\,\mathcal{N}[Y;\hat{\mathcal{C}}^{2\eta}(\mathcal{B}_\infty)] + |t-s|^{\frac{1}{2}-\eta+\gamma}\,\mathcal{N}[Y;\mathcal{C}^0(\mathcal{B}_\gamma)]\Big\}. \end{aligned}$$

Besides, since

$$K^Y_{ts} = \int_s^t S_{t-u}(\delta(f(Y))_{us} \cdot dW_u) + \int_s^t S_{t-u}(P \cdot f(Y_u) \cdot f'(Y_u))\, du,$$

it is easy to check that $E\left[\|K^Y_{ts}\|^{2q}_{\mathcal{B}}\right] \leq c_q\,|t-s|^{(1+4\eta)q}$ (use (6.1) and (6.2) as above). We are thus in a position to apply Lemma 6.1 and conclude that $K^Y \in \mathcal{C}_2^{\frac{1}{2}+\eta}(\mathcal{B})$.

□

References

[1] G. Bal, *Convergence to SPDEs in Stratonovich form*, Comm. Math. Phys. **292** (2009), no. 2, 457–477. MR-2544739

[2] V. Bally, A. Millet, and M. Sanz-Solé, *Approximation and support theorem in Hölder norm for parabolic stochastic partial differential equations*, Ann. Probab. **23** (1995), no. 1, 178–222. MR-1330767

[3] X. Bardina, M. Jolis, and L. Quer-Sardanyons, *Weak convergence for the stochastic heat equation driven by Gaussian white noise*, Electron. J. Probab. **15** (2010), no. 39, 1267–1295. MR-2678391

[4] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. MR-0348562

[5] Z. Brzeźniak and F. Flandoli, *Almost sure approximation of Wong-Zakai type for stochastic partial differential equations*, Stochastic Process. Appl. **55** (1995), no. 2, 329–358. MR-1313027

[6] R. Buckdahn and J. Ma, *Stochastic viscosity solutions for nonlinear stochastic partial differential equations. I*, Stochastic Process. Appl. **93** (2001), no. 2, 181–204. MR-1828772

[7] R. Buckdahn and J. Ma, *Stochastic viscosity solutions for nonlinear stochastic partial differential equations. II*, Stochastic Process. Appl. **93** (2001), no. 2, 205–228. MR-1831830

[8] R.A. Carmona and J-P. Fouque, *A diffusion approximation result for two parameter processes*, Probab. Theory Related Fields **98** (1994), no. 3, 277–298. MR-1262967

[9] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, Cambridge, 1992. MR-1207136

[10] A. Deya, *A discrete approach to Rough Parabolic Equations*, Electron. J. Probab. **16** (2011), 1489–1518. MR-2827468

[11] A. Deya, M. Gubinelli, and S. Tindel, *Non-linear rough heat equations*, Probab. Theory Related Fields **153** (2012), no. 1-2, 97–147. MR-2925571

[12] C. Florit and D. Nualart, *Diffusion approximation for hyperbolic stochastic differential equations*, Stochastic Process. Appl. **65** (1996), no. 1, 1–15. MR-1422876

[13] P. Friz and N. Victoir, *Multidimensional dimensional processes seen as rough paths*, Cambridge University Press, 2010. MR-2604669

[14] R.J. Griego, D. Heath, and A. Ruiz-Moncayo, *Almost sure convergence of uniform transport processes to Brownian motion*, Ann. Math. Statist. **42** (1971), 1129–1131. MR-0278389

[15] M. Gubinelli, *Controlling rough paths*, J. Funct. Anal. **216** (2004), no. 1, 86–140. MR-2091358

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

[17] Y. Hu and D. Nualart, *Stochastic heat equation driven by fractional noise and local time*, Probab. Theory Related Fields **143** (2009), no. 1-2, 285–328. MR-2449130

[18] M. Kac, *A stochastic model related to the telegrapher's equation*, Rocky Mountain J. Math. **4** (1974), 497–509, Reprinting of an article published in 1956, Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972). MR-0510166

[19] I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR-1121940

[20] R. Manthey, *Weak convergence of solutions of the heat equation with Gaussian noise*, Math. Nachr. **123** (1985), 157–168. MR-809342

[21] R. Manthey, *Weak approximation of a nonlinear stochastic partial differential equation*, Random partial differential equations (Oberwolfach, 1989), Internat. Ser. Numer. Math., vol. 102, Birkhäuser, Basel, 1991, pp. 139–148. MR-1185745

[22] P. Mörters and Y. Peres, *Brownian motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010. MR-2604525

[23] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, vol. 44, Springer-Verlag, New York, 1983. MR-710486

[24] M. Pinsky, *Differential equations with a small parameter and the central limit theorem for functions defined on a finite Markov chain*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **9** (1968), 101–111. MR-0228067

[25] H.P. Rosenthal, *On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables*, Israel J. Math. **8** (1970), 273–303. MR-0271721

[26] T. Runst and W. Sickel, *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*, de Gruyter Series in Nonlinear Analysis and Applications, vol. 3, Walter de Gruyter & Co., Berlin, 1996. MR-1419319

[27] A. V. Skorokhod, *Studies in the theory of random processes*, Izdat. Kiev. Univ., Kiev, 1961, in Russian. English Translation: Addison-Wesley, 1965. MR-0185620

[28] Héctor J. Sussmann, *On the gap between deterministic and stochastic ordinary differential equations*, Ann. Probability **6** (1978), no. 1, 19–41. MR-0461664

[29] G. Tessitore and J. Zabczyk, *Wong-Zakai approximations of stochastic evolution equations*, J. Evol. Equ. **6** (2006), no. 4, 621–655. MR-2267702

[30] S. Tindel, *Diffusion approximation for elliptic stochastic differential equations*, Stochastic analysis and related topics, V (Silivri, 1994), Progr. Probab., vol. 38, Birkhäuser Boston, Boston, MA, 1996, pp. 255–268. MR-1396335

[31] S. Tindel, *Stochastic parabolic equations with anticipative initial condition*, Stochastics Stochastics Rep. **62** (1997), no. 1-2, 1–20. MR-1489179

[32] J.B. Walsh, *A stochastic model of neural response*, Adv. in Appl. Probab. **13** (1981), no. 2, 231–281. MR-612203

Acknowledgments. We thank Samy Tindel for helpful discussions on the topic of the paper. We are also very grateful to two anonymous referees for their careful reading and suggestions, which have led to significant improvements in the presentation of our results. Maria Jolis and Lluís Quer-Sardanyons are supported by the grant MCI-FEDER Ref. MTM2009-08869.